E-Library of Medieval Chant Manuscript Transcriptions

Louis W. G. Barton
University of Oxford
St Anne’s College
Oxford OX2 6HS, U.K.
louis.barton.je.77@aya.yale.edu

John A. Caldwell
Faculty of Music, University of Oxford
St Aldate’s
Oxford OX1 1DB, U.K.
john.caldwell@music.ox.ac.uk

Peter G. Jeavons
Oxford Univ. Computing Laboratory
Wolfson Building, Parks Road
Oxford OX1 3QD, U.K.
peter.jeavons@comlab.ox.ac.uk

ABSTRACT
In this paper we present our rationale and design principles for a distributed e-library of medieval chant manuscript transcriptions. We describe the great variety in neumatic notations, in order to motivate a standardised data representation that is lossless and universal with respect to these musical artefacts. We present some details of the data representation and an XML Schema for describing and delivering transcriptions via the Web. We argue against proposed data formats that look simpler, on the grounds that they will inevitably lead to fragmentation of digital libraries. We plan to develop applications software that will allow users to take full advantage of the carefully designed representation we describe, while shielding users from its complexity. We argue that a distributed e-library of this kind will greatly facilitate scholarship, education, and public appreciation of these artefacts.

Categories and Subject Descriptors
H.3.7 [Information Storage and Retrieval]: Digital Libraries – Standards; J.5 [Applications]: Arts and Humanities – Performing arts (music); E.2 [Data]: Data Storage Representations

General Terms: Design, Standardization

Keywords
Digital libraries, data representation, musical notation, chant, medieval manuscripts, transcription, search, comparison, XML

1. INTRODUCTION
The earliest Western European writing of music dates from the ninth century in liturgical manuscripts: a kind of musical notation called neumes. Many thousands of these manuscripts survive from the Middle Ages, and they are among the great monuments of European culture. Figure 1 is an excerpt from one such manuscript. Normally, the chant text runs along the baseline, and its neumation comprises the symbols written above it.

For centuries in Western Europe, chants were memorised from oral transmission; early neumed manuscripts were not ‘sheet music’ from which one could sight-sing without already knowing the melody. Exactly what scribes intended to convey in neumatic symbols is a difficult area of study. The symbols probably convey subtleties of vocalisation that cannot be written in modern musical notation. An e-library of symbolic transcriptions would be of great value to this study, but for an e-library to have lasting value, some hard problems of data representation must be solved first.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL’05, June 7–11, 2005, Denver, Colorado, USA
Copyright 2005 ACM 1-58113-876-8/05/0006…$5.00.

Several distinct “families” of neumation have been classified, and various “species” within the families, resulting from local traditions and the experimental process of inventing musical notation. Each species comprises many symbols; their taxonomies sometimes overlap, but not entirely. Figure 2 shows a few of the written forms in which one abstract symbol, the clivis, appears.

Byzantine neume notations pose special problems not evidenced in Western notations, although they share many basic conceptual elements and some common chant texts. Paleo-Byzantine notations, for example, involve ‘complex signs’ that denote a group of notes or an entire melisma (see, Figure 3).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL’05, June 7–11, 2005, Denver, Colorado, USA
Copyright 2005 ACM 1-58113-876-8/05/0006…$5.00.

This research project is supported by Eduserv.

Figure 1. Latin manuscript Douce 222 (excerpt), Bodleian Library, University of Oxford.

Several distinct “families” of neumation have been classified, and various “species” within the families, resulting from local traditions and the experimental process of inventing musical notation. Each species comprises many symbols; their taxonomies sometimes overlap, but not entirely. Figure 2 shows a few of the written forms in which one abstract symbol, the clivis, appears.

Figure 2. Twelve examples of the clivis symbol (a downward vocalisation of two notes) written in various neumatic species.

Figure 3. Greek text superscribed with neumatic symbols in the Chartres species of Paleo-Byzantine notation.

Tens of thousands of manuscripts containing chant notation survive, comprising perhaps millions of chant “documents.” (A “document” is a single item of chant in a particular manuscript.) No one knows how many exist. We estimate that there are at least ten thousand Latin manuscript pages containing neume notation in the Oxford University holdings alone, and this figure does not include many others in Greek, Russian, and so on. Today, only a fraction of the extant neumed documents have been adequately catalogued, and far fewer have been studied in detail by scholars.
Early neumed manuscripts are important to scholars and musicians, even when the music they contain is not understood. In some cases, these manuscripts contain music that is not available from later manuscripts in square-neume notation — they are the only surviving records. Even when old melodies are preserved in later sources, early sources are still valuable: their notation shows vocalisation details that were impossible to show in later, simpler notation. Comparing early sources with each other and with later material may reveal how notations and melodies evolved, vectors of transmission for repertories, the age of manuscripts, and so on.

2. THE GOAL OF THIS PROJECT

The goal of our project is to create the infrastructure software needed for a stable e-library of chant manuscript transcriptions to grow as a distributed resource, where cultural assets of many institutions may be interconnected and interoperable in a scalable way. This task has many facets, including: data-entry programs; multi-lingual description protocols for catalogues or search query; graphical browsing; and transformations for data reuse in a variety of applications, such as word processing, analysis, and pedagogy. Currently, we are creating some initial content mainly as a test suite for software validation [9]. We expect that the e-library will be self-sustaining after the infrastructure is operational, such that chant scholars can freely add content simply by making their transcriptions accessible to the World-Wide Web.

Transcriptions can be hyperlinked to related materials on the Web. (The default ‘wrapper’ we use for transcriptions is an XML format that includes hyperlink ‘hooks’ for this purpose, see [9].) Figure 4 shows the principal kinds of content that are appropriate for interconnection: the manuscript icon at the top denotes a chant source; it is transcribed and indexed along the arrowed lines.

Figure 4. Manuscript content in an e-library: photographs, transcriptions, and indexes; audio; and secondary sources.

A source (or, one “document”) is a ‘primary source’ (in historical terms) when it is represented either by a photograph or by a symbolic transcription. It can be transcribed as text-only (perhaps with parallel translation provided), or as chant text with neumation. In the latter case, it must be possible to generate a diplomatic facsimile of the source in its native notational style. (A transcription in modern musical notation is an interpretation, and so we would not consider it as a ‘primary source’.) Sources may be indexed in special collections catalogues of the holding institutions, or in online directories that specialise in this kind of content; in both cases, description meta-data may be part of the index entries. General search engines can provide query-access to sources, and we expect query-access will become increasingly sophisticated in the future. An audio file (whether a recording of singing, a MIDI sequence, etc.) is an interpretation and so is not a ‘primary source’. There may also be guides to reading neumation, commentary and analysis, cultural context, bibliographies and discographies, and so on, associated with a source. Together, these provide a rich e-library resource, adequate for many kinds of study, from general appreciation to serious research. Except for ‘closed’ data standards (such as, Flash), all these materials can be saved to local disk for secondary processing by offline programs.

Figure 5 shows part of the diplomatic facsimile of a Middle Byzantine source, as produced by our beta-version visualisation software [9]. The lower line shows tonal movement (parentheses denote that the interval is approximate; depending on the source notation, tonal movement could be transcribed by exact intervals or as just up/down direction). The dot before “martyr” is clickable — it opens a popup window with information about this martyria sign. This visualisation is generated directly from transcription data and can be displayed in Web browsers that support XSLT (Extensible Stylesheet Language Transformations). It is fast enough to allow users to browse through transcriptions. This visualisation software is just one of many possible programs that could serve the same purpose. The point is that the transcription data contain enough information so that, a source can be displayed in symbolic form that looks similar to the original document. The data format is self-describing, and so transcriptions can easily be imported to databases or transformed to other data formats for analysis, typesetting, etc.

Figure 5. Visualization (beta version) generated in a Web browser from transcription data of a neumed Greek text.

For this e-library architecture to work, however, one facet is crucial: there is an urgent and compelling need for a standard representation for transcriptions. It must be transparent, portable, computationally efficient, and appropriate to all species of neume notations. Satisfying these requirements necessarily introduces a level of complexity in the data that will not be needed by all users. Simpler encoding, however, necessitates multiple formats. These would not be formally equivalent to one another: some details needed in one encoding would not be present in another.

3. THE ENCODING PROBLEM

3.1 Fragmentation of Digital Libraries

In a ‘qualitative’ field, where a proof-theoretic methodology is lacking, scientific progress will depend on recording source artefacts in a “lossless” way. By this, we mean that the ‘raw’ transcription data must record the content and features of artefacts with a minimum of interpretation. A data representation must not colour the empirical evidence; asserting that such-and-such a thing ‘must have been’ is speculation, not proof. Representation must also be ‘universal’ (meaning that, a single standard of coding covers all species of notation); this promotes data sharing and reuse, and encourages new knowledge to arise from comparative study. Also, the representation should be optimised as necessary so that the most complex class of common operations will be computationally feasible, that is to say, scalable and efficient.
Medieval chant musicology is a ‘narrow vertical market’ in software: the core group of users is small, but the requirements are complex. It is counterproductive if musicologists invent their own coding schemes and transcribe sources to them. This causes fragmentation of transcription libraries and duplication of effort in programming. Considering the large number of sources in need of transcription and the shortage of funds in this field, it is vital that chant scholars should understand the gravity of this. Many aspects of chant software are appropriate for ‘maverick’ or ‘go it alone’ solutions (such as online courseware and new types of analysis), but the data representation is not. Some species of neumation are readable by only a few scholars in the world; their transcriptions should be preserved as interoperable data — not as isolated or idiosyncratic codes. A major challenge for ‘universal’ data representation is to encompass the perspectives of these experts, and yet guarantee the comparability of all transcriptions.

3.2 The Seduction of Simplification

Figure 6 is an example of Western neumation after the four-line ‘staff’ had gained widespread use. The obvious difference to our previous examples is the use of a scale (literally, “ladder”) for more exact recording of pitch. The trend toward measurement was applied also to lengths of notes, but the use of a metrical beat never took hold in chant as it did in secular music, such as dance.

![Figure 6. Part of an Ambrosian chant of the Latin liturgy, written in square-neume notation on a four-line staff.](image)

By 1500, mechanical printing for music was feasible. Mostly due to its lower cost (a manuscript book could cost as much as a house), this ended a 700-year run of Latin chant calligraphy and confirmed a uniform style of neumation. (Chant calligraphy in the East, however, persists to some extent even today.) One can understand this by comparison to a folk story like “Snow White”: Though most of them are duplicates. Because the Liber Usualis is the standard reference, there is an almost overwhelming opinion that ‘computerisation’ of Gregorian chant just means encoding this one species. This species is visually akin to common practice (i.e., modern) musical notation. If neumes have the ‘rhythmic’

values given to them by the Liber Usualis, it would be plausible that neumation could be subsumed under common practice notation. (Some adjustments are required: e.g., to omit bar lines.) From the ‘pastoral’ perspective of parish churches, maybe all that is needed is a music-editing program that supports square-neume notation, especially if it includes MIDI output for the organ!

3.3 Separation from General Music

Undoubtedly, square-neumation was a precursor of modern notation, but the view that neumes are just an archaic type of common practice notation distorts their meaning and the place of chant among genres of music. The High Middle Ages were the apex of medieval tradition, not the start of modernity. Square-neume notation is tied to early sources both by technology and by content: i.e., the cantillation of Latin, Greek, Slavonic, Armenian, Hebrew, and Coptic liturgical texts. Square-neumed sources ought to be regarded as part of the corpus of chant manuscripts — they have almost no commonality with Mozart, Brahms, Stravinsky, Broadway musicals, or even the hymns sung in churches today.

Early species of neumation convey little or no information about the pitch or duration of notes. (Indeed, the concept of ‘note’ may itself be an anachronism.) As emergent systems of writing, early neumation cannot be subsumed under today’s ‘Cartesian’ or ‘deterministic’ view of musical notation. Further, the meanings of some early symbols are not fully understood by scholars. A data representation for neumes must be more concerned with satisfying the requirements of early neumation than with being compatible with modern music and the assumptions of modern notation.

It is impossible to know how chant sounded a thousand years ago. Perhaps the early sources show ‘sanitised’ styles of singing technique. Some scholars [15] argue that a source might reflect just one of many possible vocalisations — that ancient chant was ‘formulatic’ and involved some degree of improvisation. This is consistent with the melodic formula signs of Byzantine ekphonic notation (see, ‘comma’ signs of Figure 7), apparently pictographic shapes for melodies in some notations (see, Figure 9), and recent scholarship on cantillation in the ancient Hebrew liturgy. In the case of Latin chant, in the early twentieth century the Vatican (for pastoral reasons) declared that the reconstructed Gregorian chant of the monks of Solesmes is ‘authentic’. With this, a particular style of singing came to be presumed ‘authentic’, using discrete pitches and ‘rhythmic’ values. Nevertheless, we assert that audio rendition must be excluded from a primary representation.

3.4 Uncertainty in Reconstructed Semantics

The Solesmes monastery (France) has tried during the past 150 years to reconstruct an ‘original’ Gregorian chant by studying early sources [11][12][13]. An analytical approach is necessary, because the chain of oral transmission in singing and writing chant was broken centuries ago. It is widely presumed that the Solesmes interpretation of glyph semantics is definitive. Their reconstruction is, however, speculative in significant ways. This is summarised by a monk in the U.K., whom we quote here.²

---

² Even if the reconstructed notations were entirely correct, oral interpretation would still be open. Alternative vocalization has been explored especially by Dominique Vellard [6] and in [16]. Compare to renditions of Eastern chant by Marie Keyrouz [7].


---

With the re-founding of Solesmes, in the mid-19th century, the modern study of chant commenced through a sincere attempt to restore the ancient form of the melodies. By collecting and collating chant manuscripts the monks developed a plausible method of decoding the neumatic symbols of the earliest extant manuscripts. … A significant part of the Solesmes editorial work was the addition of rhythmic signs to the melodies indicating the presence or likelihood of rhythmic variation in what otherwise appeared to be a succession of equal length notes. Similar restoration work was undertaken in Germany with surprisingly different results. … We should not presume that the Solesmes method was left unchallenged. Significant voices were raised against the “equal notes” approach both on the Continent and in Britain. Dom Gregory Murray published a series of articles proposing a radically different interpretation of the early manuscripts and within the Solesmes family itself there were voices questioning the simplicity of the Gajard [rhythmical] approach.

A member of our team, Jacqueline Elemans is undertaking a reanalysis of the statistical methods used by Solesmes, particularly the work they did in preparation of the Graduel Romain, édition critique (GR:EC). In a recent paper [5], she argues that the group of sources they chose as a ‘statistical sample’ necessarily skewed the results of their analysis. We quote from her paper here.

The data of the first sondage [i.e., sample] point towards a significant split between Eastern and Western manuscripts, with just a handful of manuscripts in between. … For the next stage the monks selected 33 manuscripts not later than the twelfth century. … From this, they hoped to determine the relationships between the main groups found in the first sondage. … Here, the monks appear to have run into another problem that they did not recognise: the new body of variant places shows very different relationships. … Of course, we know now that the differences between the three subsets are well within the statistical margin of uncertainty for such small samples. … Summing up, the work of GR:EC was fraught with difficulties. Some of these the monks were well aware of but could not deal with in the absence of computing power. Other problems remained unrecognised until now. The main technical problems limiting the reliability of the result of GR:EC are data selection and sample size.

Despite their good faith and years of careful work, it is clear from the Solesmes reconstruction that better methods are needed to handle the scope and complexity of these artefacts. In our experience, it seems that the great majority of chant musicologists understand already that computers are needed. We welcome any initiatives that build computer software toward this end. We must re-emphasise, however, that data representation is the crucial part. Data representation is less exciting than creating analysis methods or assembling digital libraries of content. Musicologists are eager to have comprehensive content and powerful search tools as soon as can be done, because these are useful to their work. Without understanding the importance of data representation, they are willing to take shortcuts that can have disastrous consequences.

4. PRIOR DEVELOPMENTS

4.1 Types of Encoding

Nearly all attempts to digitally encode neume notations can be said to fall into one of four ‘generations’ of data-representation technology, by the following classification scheme:

(a) “first order” data representations that use single ASCII characters as proxies for glyphs, such that, a character is mapped to a glyph by a font (usually in word processors);
(b) “first order” data representations that use binary patterns as tokens for semantic entities (usually in music editors);
(c) “second order” (‘modal’) data representations that use ASCII strings as names of semantic entities (interpreted);
(d) “first order” data representations that use Unicode™-compatible code points as tokens for semantic entities.

4.2 Existing Software

There is a plethora of software for common-practice notated music, and also some software for square-neuation. None of this software is adequate for transcribing early sources. They prejudice scientific study, because the encoding itself forces unwarranted assumptions. Imagine that German texts were transcribed in 7-bit ASCII code: any diacritical marks (or, accents) would have to be dropped, because ASCII has no codes for them. The situation gets worse with Greek. The Unicode has a much larger codespace for this reason. German texts lacking diacritics may be understood in context, but information that is lacking from chant transcriptions cannot be reliably inferred. Innovative software is needed that can represent early neuation on its own terms and permit algorithmic procedures (like search and comparison of melodic patterns) in an acceptable computing-time, especially across notational species.

The standard reference about data representations for music is the book Beyond Midr [10]. Data representations and computer programs for music are listed on the Web by Gerd Castan [4], which is the best index for finding current information. We have studied many of these representations in detail, and considered practically all of them under our classification scheme of § 4.1.

Of the dozens of representations proposed in the past (too many to discuss here), nearly all that handle notation were designed for common practice notation. They organise data by ‘events’ in quantified time, where the metric of quantification is a beat (this does not necessary relate to clock time). This concept is deeply rooted in most of these data representations, so much so that it is impossible to transcribe chant without forcing symbols to occur at specific beats in the underlying time-quantification.

An exception to the beat framework is (α)-encoding, which allows transcribers to place font symbols at any locations on the page. Analytical procedures on symbols under (α)-encoding are, however, infeasible in the general case.

4.3 Two Particular Cases

We discuss here two particular coding standards that were designed for common-practice notation but that claim to satisfy the requirements of neume notation. The first is Guido (named after Guido d’Arezzo, who perfected a method of pitched-notation on staff lines). Guido is a (γ)-encoding that has four dimensions of information: note name; accidentals; octave; and duration. Of these, only the note name is required in all transcriptions. Note name can be written in one of three ways: pitch letters of the diatonic scale (‘a’ through ‘g’); or pitch letters of the chromatic scale; or degrees of solfège (doh, re, mi, …). Guido is claimed to be “flexible,” “simple,” “extensible” (meaning, presumably, that it...
can be extended to accommodate any musical notation, and superior to (δ)-encoding by virtue of being ‘human-readable’. We favour simplicity in data representation if it does not compromise semantic content. If Guido were formally equivalent to our data representation, then our only objection would be over the implications of (γ)-encoding for computational feasibility. In fact, Guido is not formally equivalent. To assume that all musical notations contain discrete pitches is a cultural-centric view.

The second standard we discuss is the Humdrum Toolkit. It is not actually a data representation, but it provides a file format that serves as a ‘wrapper’ for music transcriptions in a variety of data representations. A data representation that uses (γ)-encoding was written for Humdrum, into which its author transcribed square-neume noted music by the 12th-century composer Hildegard of Bingen. This encoding has been demonstrated as effective for using Humdrum’s analytical tools to produce new knowledge about Hildegard’s music. The “Hildegard” data representation can undoubtedly be extended for square-neume notation in general.

Nevertheless, such examples do not solve the problem we are considering here, even if the whole repertoire of the Liber Usualis were transcribed. Any study limited to one notational species (no matter how large the sample) is not representative of the diversity in the whole population. More importantly, the demonstration that a data representation works in one particular application does not address the critical question of scalability: the computational feasibility of pattern search on a large, distributed data set.

5. DESIGN DECISIONS

5.1 Dimensions of Information

We have identified the minimum “dimensions of information” in neumation that are necessary to a “lossless” encoding as three:

(i) glyph form (as a graphic or font-surrogate for a glyph), with optional X- and Y- spatial metric or ordering;
(ii) pitch (relative to a tonal mode or a scale), with optional note ‘size’ or duration; and
(iii) tonal movement as a vector of melody (as an approximate or specific interval, or as equal/up/down/unknown).

Of these, only (iii) is required for encoding all sources: it ensures that sources across all species are comparable by tonal movement.

5.2 A Language of Transcription

Setting aside any significance of colour (e.g., the red ‘T’ in Figure 6), and ignoring ‘background noise’ (such as in Figure 1, where texture of the sheepskin parchment appears as speckles in the image), sources have a binary aspect: the scribe’s pen was dragged on the writing surface, or the pen was lifted. In general, an area of continuous ink constitutes a ‘glyph’. (There are two exceptions: some glyphs have breaks in the ink, viz., consider the lowercase letters ‘i’ and ‘j’; and some glyphs cross one another, such as the staff lines and neumatic symbols of Figure 6 — they are distinct marks but they overlap.) A glyph may have been formed in more than one stroke. Identifying pen strokes is a problem of duc tus, which we ignore in the representation, because those who study handwriting want to examine the manuscript — symbolic transcriptions are not of primary value to their work.

A common feature of nearly all human writing is that glyphs are placed in an order that has some linearity about it, and this ordering contributes to meaning. In the cultures that invented neumatic writing, it was already established that text should be written in a line, from left to right. Invariably, neumes stand in relation to a unit of text (e.g., the syllable, the colon, etc.). The modern idea that music can be composed, and then ‘lyrics’ added, is an anachronism: cantillation signs were added to a written text. Speech determined the cadence and emphasis of cantillation, not a rhythmic beat. From the earliest-known species of neumation, called ekphonic notation (Figure 7), one can see clearly that the basic concept was that notation ‘marks-up’ the text.

Figure 7. Part of a Greek text, with ekphonic notation (the comma-shaped glyphs, below) and prosodic marks (above).

The prima facie chant text is easy to discern: liturgical texts both in Eastern and Western scribal traditions were copied with great fidelity over many centuries. If a calligraphic handwriting is difficult to read, then the text can be verified against a ‘control’ document containing the same text in a more easily-read hand.

The logically consistent way to represent neumation, then, is to use the text (not a framework of beats) as the fundamental data structure, and correlate sequences of neumes to units of text. The correlation, in principle, could be implemented using pointers, but we decided to interchange text and neumation a single sequence. One can intermix these and maintain computational efficiency by using code points in disjoint ranges of (δ)-encoding [cf. § 4.1]. The codespace of neumation is not overloaded, and so no “mode switching” is done during parsing of the data. The neumation data can be read without ambiguity at any location in a data stream. Our data structure is a data stream because we treat transcriptions as strings of symbols, that is, as a language. Since coding of chant texts in Unicode is trivial, the problem reduces to designing an effective data representation for neumation as strings of symbols.

5.3 Data Independence

At a meeting of the International Musicological Society, we presented a beta version of our data representation. At the end of our presentation, one influential musicologist asked, “What will this program do for me?” Undoubtedly, the question was well intentioned, but it reveals an all-too-common confounding of data representation and programs. The confusion is relatively harmless when it is in the minds of end-users, but it can have catastrophic effects when in the minds of those who design or create software.

Certainly, within programs it is beneficial to encapsulate data structures together with primitive operations on those structures inside object classes. That way, access to and mutation of data is centralised and controlled. Perhaps, historical emphasis on the ADT (abstract data type) model made some programmers overly eager to integrate ‘code and data’ behind an abstraction barrier.

\[ τάντα \ δέ γέραπται \]

\[ τάντα \ δέ γέραπται \]
Maybe, too, the prestigious place that algorithms hold in curricula has caused some neglect in the study of data except for data structures. By contrast, people who have acquired a sophisticated understanding of database systems, or who have thought deeply about semantic data modelling, are more likely to fully appreciate the importance of separating process from data representation. Some progress has been made in this regard with the advent of XML, especially XML Schema (which has stronger constraints on data types and data values than does DTD). Unfortunately, we have seen a rash of XML DTDs (Data Type Definitions) where the integration demons are up to their old trick again: making DTDs overly use-specific, essentially making the data dependent on processes. In our view, design ought to be data-driven. Whenever possible, procedures should be dynamically generated from the data definitions.

The path of chant software is ‘littered’ with the bit-rotting corpses of programs that used (β)-encoding and were unable to adapt to the evolving, virtual world of the Web, XML, Java™, Unicode, and other new technologies. It is okay to have ‘throw-away’ programs if their failure produces new insights and design documentation that can be adjusted and re-implemented (which most of these programs did not produce). The tragic part is that an uncounted, presumably huge, number of hours were spent by expert musicologists in making transcriptions that have become obsolete and unusable, and may end up being a total waste.

The falsely held panacea du jour is that bit-rot will be cured by ‘open source programs’: viz., that the source code of programs should be public property and that volunteers ought to do the design and implementation of program functionality. Aside from the problem of ‘who pays the rent’, making programs open source does not necessarily cure the defect that (β)-encoding programs have an untenable design. What is needed is not so much ‘open source’ but open standards, by which we mean the following.

(a) The way information is represented as data should be designed separately from any particular application.

(b) The data representation should be both generalised and comprehensive, to the greatest degree foresight permits.

(c) A standard should be centrally controlled – even if freely licensed – in order to avoid fragmentation.

(d) It should be fully documented and ‘transparent’.

Many different programs can be written to use the data representation, either in whole or in part, but data should be exchanged in the standard representation. When better programs are needed, then the old programs can be scrapped without invalidating transcriptions that exist already in digital libraries. A benefit of XML is that, one data format can be readily converted to another via CSS (Cascading StyleSheets) or XSLT. Two data representations are formally equivalent, however, only if there exists a one-to-one mapping between the semantic entities on both sides. Transformation typically involves ignoring parts of a data representation’s semantics. Publishing a data representation as an ‘open’ standard and writing it in XML so that it will be transformable, do not in themselves guarantee our criteria, above.

5.4 Frequency-of-Change Criterion

Dr Elemans’ concern over the size and quality of statistical samples [cf., § 3.4] points to the hardest part in designing a data representation for neumes: one can call it the ‘chicken or the egg’ paradox (viz., “which came first, ...”). Hypothetically, if a large number of chant manuscripts existed in digital transcription, then statistical methods could reliably determine which entities of information have significant variance. Theoretically, one might find that the Solesmes conclusions about rhythm turned out to be valid; this would imply that a data representation for Western neumation could have an underlying structure of beats without prejudice to transcriptions. Furthermore, if some entity were relatively invariant within individual sources, one could safely put this entity in a (γ)-encoding. This is because mode-switching (or, disambiguation) of overloaded codes would not add substantial overhead, because these operations would be infrequent. Conversely, if a feature changes frequently in sources, then it ought to be in (δ)-encoding, so that complex operations like pattern-matching searches across (potentially) many thousands of transcriptions can be done in acceptable computation-time.

The paradox is that, these statistical analyses need a large amount of transcription data, but one doesn’t know how to transcribe the sources until after the analyses are done. Analysis of a small sample selected by intuition can give good results only by luck.

5.5 A Definition of Semantics

We can define the semantics (or, meaning) of neume notation in terms of relations. These relations are not necessarily lists of attributes in a static data structure as is the case with relational databases (where relationship is implied by equal attribute values). Generally, relations may be explicit and named; a relation may be heterogeneous in the data types of its domain and range; and so forth. This more flexible view of relations is consistent with our intention of creating a language of transcription. Like natural language, it can be as complex as need be. Semantics, then, are flexible correspondences between parts of the language, without necessarily a strict mapping from domain to range. The relations conform to a formal grammar that decides whether ‘sentences’ are well-formed. The language, however, is not generated from the grammar; instead, it is defined as the set of ‘sentences’ that come into being by the transcription of actual sources.

If we had a witness (i.e., someone who knew how to interpret the sources), s/he might teach us the semantics of neume notations by verbal explanations or by chanting examples. Unfortunately, there exists no witness.5 Oral transmission of Western chant ceased as an indirect effect of the Plague in Europe: as choir apprenticeships declined, and endowed choir seats increasingly were filled by political appointments, a crisis arose for continuing the daily cycle chant. The solution was to let choirs sight-sing from simple, quantified musical notation instead of from memory. This hastened the evolution of neumation toward a ‘Cartesian’

5 Very little was written contemporaneously to explain neumation theoretically, and there has been no witness for Western chant in many centuries. Traditionally, noble families dedicated a young son to a monastery or cathedral for the lengthy apprenticeship of memorizing the chant liturgy. Today, it seems incredible that such prodigious amounts of information could have been memorized (perhaps more than 3,000 melodies and chant texts). Wright [17] has proven, however, that chanting was done from memory. He studied chapter records of the number of candles purchased over several years by Notre Dame Cathedral; it is impossible that chants were sight-sung from manuscripts during the hours of darkness (Vespers, Compline, and Matins), because there were not nearly enough candles to provide light for that. In isolated places of the world today, one can hear epic poems recited from memory. History says too that the Iliad was recited from memory in ancient times. (C.f., also [14] and [17].)
model. The details of cantillation were lost due, simultaneously, to the break in oral transmission and the simplification of notation (in what we call “flattened semantics”). The clivis symbol came to be written uniformly as \( \hat{\mathfrak{g}} \); Figure 2 shows some of the variety that was suppressed. For instance, Cardine [3] says that a clivis written with rounded top (cf., glyphs A and D of Figure 2) shows “lightness of the notes”—this subtlety was simplified out.

Lacking a witness, and lacking a priori statistical results that are reliable, the semantics of neumation in the data representation must be inferred from relations between transcriptions. These relations may become clear only gradually as the e-library is built and as mathematical methods can be applied to these data. Nevertheless, we need a starting point of classification of features.

5.6 Taxonomic Classification

Even if there were a witness who could communicate semantics by explanation or example, this would not provide us a data representation. Computers do not yet have enough anthropomorphic “intelligence” to transform verbal descriptions or audio examples into effective abstractions. Let’s say someone (the agent) is given several photos of faces of people s/he has never seen before. The agent can discern which photos show the same person in various moods, and which show different people. (Of course, the agent might be unsure sometimes, and the test could be designed to trick the agent, but people are generally successful at this task.) At present, computers are not capable of doing this. A human agent must pre-determine what the object classes are.

Consider our experimental NeumeOCR program, shown in Figure 8. It has a ‘knowledge base’ of glyph exemplars (lower left) that were pre-determined by a human agent; and an input area (lower right) where click-and-drag motion of the user’s mouse is traced. When the user releases the mouse button, the program displays the name of the neume form (in this case, virga) that it ‘thinks’ is most similar to what the user has drawn. This program uses neural network computation, but it could have used a genetic algorithm or other method and still have the same logical problem: a human must establish the classes (in this experiment they are punctum, virga, pes, clivis), and provide glyph exemplars, a fitness function, or some other criteria of classification. One might think that a program could automatically construct classes by clustering objects according to ‘distance’ of variance in similarity of features. This would not produce good classes for neumatic symbols, however. For example, Figure 2 glyph L looks more like two puncta than it does a ‘prototypical’ clivis, and the left-hand ‘hook’ at the top of glyphs G, H, and I make them look like the typical virga (cf., Figure 8).

Of special difficulty are the Great Signs of Paleo-Byzantine notation (see, Figure 9). Although automated similarity-clustering of Great Signs may suggest classifications, the method cannot be validated, since none of the signs has a known meaning, nor is it even known yet which graphical features have significance.

![Figure 9. A Greek source (excerpt of two lines) with Great Signs above a chant text that has round-neume notation.](image)

An automated-classification scheme would have difficulty also with unpitched Western notations, that is, where there are no ‘staff’ lines or pitch letters in the neumation. Figure 10 shows an additional problem: the scribe ran out of space and was forced to continue the neumation on a diagonal into the margin.

![Figure 10. Neumation written off-axis would cause problems for classification of glyphs by an automaton.](image)

Paleographers have established (by analysis of handwriting and ink) that chant text and neumation usually were written by different scribes. (Neumatic calligraphy was an especially ‘high-tech’ skill, but copying of text could be delegated to any scribe.) Since one doesn’t easily lose one’s place in a text, but it is easy get lost in neumation, the text served as a guide: first, the chant text was written on the page; then the neumation was added to it. In all Western species, and in Byzantine neumations except ekphonetic and Great Signs, strings of neumatic symbols are tightly correlated to syllables of text: cantillation marks do not overlap syllable boundaries. In Figure 10, the text scribe did not leave enough space after the syllable “do”. The neumation scribe had to cram the neumation in-line, making it illegibly small, or else skew the neumation on a diagonal; he chose the latter. Automated recognition of individual glyphs is not the problem. The reconstruction of context that is what is hard, as in Figure 10.

Our solution to the taxonomy problem is to make a best first-guess of classes, and assign each glyph to a class with a certainty factor (CF). The default value is a CF of ‘definitely true’. Explicit CFs vary in a numerical range encompassing degrees of positive certainty, complete uncertainty, and negative certainty. (The latter is for defeating an assigned classification, and it isn’t used during
transcription but can arise as a side-effect of analysis.) The main benefit of this method is that, complex operations like search and comparison can be done without forcing unwarranted assumptions in classification of glyphs. When only a guess is warranted during transcription, class assignment is qualified by a CF. When CF-qualified data are involved with a pattern-matching or aggregation procedure, the result will be weighted according to our algorithm for combining CFs [1]. This uses an uncertainty preserving logic, which does not disguise speculations as established facts.

6. IMPLEMENTATION

6.1 NEUMES and NeumesXML

The name we gave to our (δ)-encoding data representation is NEUMES (an acronym for “Neumed and Ekphonetic Universal Manuscript Encoding Standard” [this replaces an earlier name due to a trademark restriction]) [2]. There is some controversy over the terms ‘markup’ and ‘meta-data’; for simplicity we shall refer to our entire (γ)-encoding as ‘meta-data’. We are using an XML format for the (γ)-encoding, which we call NeumesXML. It is the repository for non-prima-facie semantic content of a transcription, plus descriptive information about the source, a transcription log, the logical structure of the document, and so on. NeumesXML is defined by an XML Schema, which has powerful capabilities for data constraints that XML DTD lacks. Among other things, NeumesXML includes a regular-expression grammar that decides whether NEUMES transcriptions are ‘well-formed’. We remark that, although we recommend using NeumesXML as the ‘wrapper’ for NEUMES data, NEUMES can be ‘wrapped’ in other formats. (The role of NEUMES is analogous to the JPEG standard for images, in that it can be used by a broad variety of programs.) We are working also on a user-definable scheme for parallel mark-up.

6.2 The Prima Facie Razor

In the absence of an encoded whole population (statistically speaking), or a reliable way to select a random sample (that is, where each source in the whole population has an equal probability of being selected), we must find a discriminating method other than ‘frequency of change’. The method must select source features to be represented in the ‘low-level’ (δ)-encoding. Conversely, it must select features that can be represented in a tagged ASCII (γ)-encoding without scalability problems (viz., the computational feasibility of complex operations on large data sets). We wanted to minimise the re-coding of transcriptions, applications software, and documentation that would result from ‘trial-and-error’ representations. On the other hand, we didn’t want to settle on an ad hoc representation that would turn out to be inadequate. Rather than making a ‘judgement call’ for every class of features, we decided to settle the matter by rule, based on types of classes. Intuitively, this can be understood as a “prima-facie razor.” The end effect is that all, and only, what is visible on the face of a source without interpretative extrapolation, and which constitutes part of the ‘meaning’ of the chant (that is, the “prima facie semantic content” of the source) is represented in the (δ)-encoding. All other information in or about a transcription is represented in the (γ)-encoded meta-data.

This prima-facie razor makes a very clean logical cut. It is also consistent with the musicological concept of ‘the text’, which is a term musicologists use in a technical sense when speaking about a musical document — this sense is similar to our concept of “prima facie semantic content.” A benefit of this method is that the design and programming of software are much easier: a programmer knows ‘by rule’ which information belongs in the character data, and which belongs in the meta-data. For example, pitches that are not explicit on the face of a source (although they can be deduced) belong in the (γ)-encoding. Also, abbreviations that appear a chant text are expanded only in the (γ)-encoding.

6.3 Clarification by Analogy

To make this clearer, consider HTML (Hypertext Markup Language). A webpage resides on a server somewhere on the Internet, and users can access the page for display in a browser. Users aren’t seeing the actual HTML file; instead, the browser is interpreting the file to create a visualisation.

Let’s say the webpage includes this string: “<title>Hello World</title>”. “Hello World” is the character data, and the tags “<title>” and “</title>” are the markup (or, ‘meta-data’ in our terminology). The visualisation displays the character data, “Hello World” in some special manner, but the “title” markup tags are not displayed.

Character data are conveyed by the computer to the user without interpretation. Actually, the character data have to be interpreted, but the interpretation is left to the user (who is presumed to be literate and should know what the data mean — a kind of ‘contract’ between the author of the webpage and the user). Typically, the character data are encoded in ASCII, but they could also be in Unicode or another encoding. In the case of “Hello World”, the character data are in what we call a “first-order” data representation. By this, we mean that the characters data ‘stand for themselves’: no “mode switching” is necessary, and the character codespace is not overloaded. From the computer’s standpoint, the characters have just one meaning.

The reserved characters ‘<’ and ‘>’ trigger a mode-switch from a “first-order” to a “second-order” data representation, and then back again. When the second-order data representation is switched in, the characters in “title” no longer ‘stand for themselves’, but now must be interpreted. As we implied in § 4.1, the character string “title” is a named, semantic entity. In the HTML language, there exists a relation between this entity and a visualisation behaviour in Web browsers. We call this relation the ‘meaning’ or, more technically, the “semantics” of the entity.

Such relations also are ‘modal’: the meaning of an entity can depend on other relations that have already been instantiated. For example, consider the nested tags, “<div> ... <pre>Tenebre facete</pre></div>”. (The ellipsis ‘...’ indicates that more material can go in this position.) The character data are “Tenebre facete”, and the pair of “<pre>” “tags are nested inside a pair of “<div>” tags. To interpret the meaning of the “<pre>” tags, the interpreter must maintain state. Specifically, when it encounters the “<pre>” tag, it must ‘remember’ what its current state is: i.e., that it is already inside a “<div>” tag. The character data “Tenebre facete” are affected in the visualisation by both of these tags. We say that “<pre>” is ‘modal’ with respect to “<div>”.

Now, imagine that a program needed to spot-check the state of “Tenebre facete”; in other words, for some reason it needed to know what the semantics of the character data in terms of these relations. To do such a check, the HTML file would have to be read (serially or by DOM tree walk) from its beginning up to “Tenebre facete” so that the state can be determined properly. Typically, a Web browser interprets an HTML file just once, in sequential order, and so the semantics of character data do not need to be spot-checked by ‘random access’. Nevertheless, we anticipate that pattern-matching operations on NEUMES data (as
distinct from literal string matching) will be required during melodic search and comparison operations. Especially with unpitched sources, we expect that searching for a melody will be complex, not simply a matter of literal string matching.

We expect melodic pattern matching to involve what we call "complex traversal" of streamed data. In the computational model, a cursor is moved by 'random access' to various locations in the data in order to compute a degree-of-match using different start-end points and various freedoms of substitution, elaboration, and dropouts; this is an NP-complete problem. Furthermore, the overhead of disambiguating tagged data and maintaining state is such that even an 'engineering solution' is computationally infeasible for large data sets. (Performance with tagged data may be acceptable in 'toy' applications but it does not scale well to data sets involving potentially tens of thousands of files.) This is a core argument for why (6)-encoding is preferable for "prima facie" semantic content," instead of (γ)-encoding. The "prima-facie razor" assigns all neumation to the (6)-encoding where neither mode-switching nor state maintenance are needed. When a cursor is moved to an 'arbitrary' location in the streamed data, just a few adjacent characters (worst case) must be read in order to know the semantics at that location. Thus, the NEUMES data representation is optimised for pattern searches. In particular, this covers cases in which CEs (viz. uncertainty) may be present in the data.

In principle, a (γ)-encoding could be transformed on-demand into another data representation in which disambiguation and state maintenance are eliminated. Aside from the additional overhead of transformation, this just puts us back at our original problem: can a target representation be designed, such that melodic pattern-matching is computationally feasible? Scalability analysis looks at worst-case scenarios. In practice, however, the time required for melodic search can be lessened by narrowing the search space to a just a few transcriptions. This can be done by a filter-query on the meta-data (e.g., selecting only sources of one notational species and whose chant has a particular incipit). Querying the meta-data normally involve literal string matching, and so it is fast.

6.4 Transcription Example

Transcription Example 1 is an excerpt from a beta-version NEUMES/NeumesXML transcription, which we include here as an XML parser; they provide 'human-readable' for transcriptions. A neume may consist of one or more neumatic symbols; the boundaries of a neume can be a matter of judgement, and so these boundaries are recorded in the meta-data. The third line contains a CF of +6 ('CF_p06'), where the transcriber was uncertain about the identity of the source glyph (she guessed it is a clivis symbol).

6.5 Distributed vs Centralised

There is a latency from availability of software technologies to popular understanding of what the software does and 'means'. When the PC became a household commodity in the early 1980s, ‘computer literate’ meant knowing one’s way around WordStar or WordPerfect. Lotus 1-2-3 seemed to be mainly for accountants. Some clever users began adapting word-processing documents to act like spreadsheets (for which the string-search capability was especially handy). Gradually, users understood spreadsheets. As sophistication grew, people began using spreadsheets like flat-file databases: added to string-searches, they could produce automated calculations, customised reports, and colour charts. Eventually, dbase II was accepted into the mainstream, and today the notion of a relational database is widely understood by computer users.

None of the technologies mentioned above was new when it was introduced on the PC. Even the Internet was used for decades by military contractors and research laboratories before the Web came about. Although many people now use the Web, its potential ‘meaning’ for accomplishing work is not widely understood yet.

The musicological community seems just barely to be coming out of the database phase of awareness. The database idea of finding things by query is intuitively clear from the paradigms of card catalogues, book indexes, and so on. But, chant scholars seem to be stuck on the idea that they need a database of chant; it is difficult to persuade them about the distributed e-library idea.

Search engines like Google can be frustrating; some scholars have expressed alarm at the amount of misinformed ‘junk’ they find among the results of a Web search. Search engine developers are well aware of the inadequacy of literal string matching as a method for finding relevant content, and people are hard at work on creating better tools. Despite the exponential growth of Web content, we believe the relevance of content returned by search engines will improve as query options will become more flexible.

A distributed e-library is perhaps best explained as a huge, global database, where search engines or directory services act as the indexes to information (see, Figure 11). A server-side database on the Web (where one accesses information via a ‘portal’) has two main disadvantages. First, full texts are hidden from search engines — or, so-called ‘dark matter’ on the Web. Second, users must query several databases separately if the desired information might be in one of several places, unless these databases share a
common index. Some people feel that what is needed is that all medieval chant information should be stored in a central database at a particular university (possibly having mirror sites). This is, in our view, a non-starter: it fundamentally fails to understand the distributed nature of information on the Internet. One common rationale for the ‘portal’ model is to control access and thereby generate income. There are, however, many ways of generating income from the Web without restricting access to content.

Figure 11. A centralised index of distributed content on the Web, and offline usage of content retrieved from the Web.

In this architecture, transcriptions are ‘first-class objects’ on the Web, meaning that each transcription is accessible by a URL. No central authority will control what transcriptions may be put online. As with other kinds of Web content, users will give more credence to material from universities and other reliable sources. We expect, too, that specialised directories will provide links to known-reliable transcriptions. Also of interest are ‘Endorsement Systems’, where authorised submitters of content can ‘endorse’ other submitters, thus creating a self-moderating, online index.

7. CONCLUSIONS

The goal of this project is to produce new tools for making transcriptions, visualising them in different ways, adding value to them, and searching/comparing them automatically. Our larger purpose is a new electronic library of unique musical material for the public and scholars to browse, study and enjoy. A growing (and, we hope, eventually comprehensive) digital library of transcriptions of these artefacts will provide important, new research opportunities for graduate students and scholars, especially in the use of systematic analysis and mathematical methods. The eventual results of such study may include a better understanding of the origins, evolution, and interconnections of liturgical chant; more reliable interpretation of neumatic symbols and the open question of rhythm in chant; more accurate dating of manuscripts; and so forth. Long before such results are eventually realised, transcriptions and the associated tools will provide new resources for interactive teaching of medieval music.

8. ACKNOWLEDGEMENTS

Many people have contributed work and ideas to this project, especially the members of our technical team and Advisory Board, most notably Annalisa Doneda, Jacqueline Elemans, Constantine Terzopoulos, and Tillman Weyde. The project was supported in an earlier phase by a Scholarly Communications grant awarded to Harvard University by The Andrew W. Mellon Foundation [2]. Significant contributions to the early design of the NEUMES data representation were made by Ugo O. Gagliardi and Thomas Forrest Kelly. The project originated in work by Louis Barton in 1995 with James Grier and Peter J. Kindlimann at Yale University.

9. REFERENCES

[6] Ensemble Giles Binchois, Les Trois Maries, (EMI, CD ALL429). Hear the Kyrie Trope and Offertory tracks, to which Vellard gives especially mellifluous interpretation (excerpts online at barnesandnoble.com or amazon.com.) Vellard’s earlier recordings of chant are mostly out of print.
[7] Keyrouz, Sr Marie, Chant Byzantin; Passion et Resurrection, (Harmonia Mundi, CD 901315). Her renditions sound more like the Muslim Adhan, or call to prayer — characterized by quartertones, dissonance, decoration, and rapturous delivery — than to Gregorian chant as it is normally sung today.
[16] ———, With Voice and Pen: Coming to Know Medieval Song and How it was Made, (Oxford: University of Oxford, 2003); includes audio CD of musical examples.