

Music Information Retrieval and the future of Musicology

Tim Crawford, Goldsmiths College, University of London

*

Introduction

It hardly needs saying that musicology as a discipline will increasingly be affected by present and future advances in technology just as the practice of music – and, of course, the music industry – have already been so transformed. It may have been possible a few years ago to cling to the belief that unassisted ‘traditional’ methods were still adequate to maintain a comprehensive grasp of developments in a chosen sub-field of musicology together with the necessary general overview of the discipline, given a well-stocked music library provided with generous financial and human resources, and given the academic leisure to browse catalogues and read in depth where necessary.

But the steady increase in academic activity combined with what might be termed a general ‘dissemination imperative’ associated with today’s style of research funding has given rise to an explosion of published material (whether in music editions or scholarly literature). In some areas it is now almost impossible for any individual to keep up with the amount of reading. For example, it is said that a new book on J.S. Bach or his music appears each week; while this is probably an exaggeration it contains an element of the truth, and the same might well be said of several other major composers. This is to say nothing of the difficulties faced by librarians in making accessible to their demanding readers a comprehensive selection of current and out of print materials in a climate of increasing pressure on their financial and human resources.

Meanwhile, owing to a positive and welcome shift in attitudes towards musical performance over the last decade or so, a whole new class of primary materials has been added to the mix; audio and video recordings, of whatever vintage, are now, rightly, regarded as musical documents in their own right, and can no longer be ignored in the way they formerly were.

In some respects, the situation is the same as that in every discipline in the Humanities. The surge in published output, the simultaneous shrinking of library budgets and the increasing pressure on the amount of time academics have at their disposal to read around their subject, affect everyone equally. A vast array of new sources of textual data (editions, databases, online archives, genealogies, catalogues and the like) is growing with exponential vigour. Locating the small proportion of this data that is relevant – or potentially relevant – to one’s scholarly domain is a skill that academics are acquiring almost without noticing it. The reason for this is that, despite some real difficulties that still provide challenges in the field, text-based Information Retrieval (IR) is a highly-developed branch of computer science with a very successful commercial component.

One unique feature of music is its sheer universality. It cannot have escaped even the most blinkered academic musicologist that the Internet is the forum for a radical change in the way music is marketed and consumed (both aspects, of course, being ripe for academic investigation, it should be said). One hears that Internet downloads now exceed

CD sales; by the same token, the economic basis of the recording industry is undergoing seismic shift as new ways to make money are having to be devised in order to compensate for the lack of physical, concrete product. Some authorities predict that the CD in its jewel-case may soon be a fond memory; at least one leading record label specialising in classical music provides its entire catalogue as downloadable files on a web-site for a modest annual subscription.¹ More and more people habitually use online sources to locate their favourite music, generally by text-based IR methods such as Google, or the internal databases maintained by distribution companies such as Amazon. Almost without exception these use metadata such as title, artist's name and genre-categories to specify queries without direct reference to the recorded content of the music. The technical methodology behind these search-engines is in fact standard text IR, only marginally adapted to the special nature of music. (The extraction and/or encoding of the metadata used in the search-engines is another matter; this is usually provided by the makers of the recording, or by specialist teams of expert human cataloguers.)

For roughly the last decade, a growing community of research workers across a rather wide range of academic disciplines has been focussing on the problems of Music Information Retrieval (MIR), by which is meant the content-based application of IR techniques to process musical queries, instead of (or perhaps in conjunction with) using the proxy of conventional textual or numeric metadata. The motivation may vary depending on the home domain of the researcher from the purely commercial to the natural investigative response to an intriguing and challenging set of problems for computer scientists in Artificial Intelligence (AI) and/or for electronic engineers in Digital Signal Processing (DSP). Whatever the background of the researchers, it soon becomes clear to each of them that this field of study is above all multidisciplinary, to some extent involving everyone concerned in the pursuit of answers to a fundamental question, "What do we mean by musical similarity?"

Musicologists are among those who have a significant contribution to make to this effort. For they are very largely concerned with that same fundamental question whenever they analyse a piece of music, or talk about style, or discourse on the influence of one composer on another. But MIR brings into high profile the need to put the question in a rather different way than is usual in musicology, since one can assume no cultural background or experience when dealing with computers, and certainly one cannot expect them to make subjective judgements since they are – to put it bluntly – not human.

The good news this essay hopes to hint at a little is that not only will the task of achieving a working MIR system that is built on a sound, musicologically-informed basis help to establish some very useful methodological criteria for thinking about musical similarity, but also that such an emergent system will itself have the potential to become an extraordinarily powerful tool for musicological research.

Textual corpora and scholarship

¹ The Hong Kong-based company, Naxos; see: <http://www.naxos.com>

The accumulation of archives of text- and number-based material for scholarly purposes in machine-readable form has been in progress for several decades. This is a natural extension of the obvious need for digitisation of such data in the commercial, legal and political worlds, and requires little by way of special technology or methodology to be adapted for academic use. Online, or CD-ROM, publications of scholarly texts and materials are now a familiar part of the academic landscape and providing access to a large range of such resources is an expected part of the duties of university and public libraries. To take just two prominent examples of text archives, Project Gutenberg² and the Oxford Text Archive³ between them offer some 17,000 complete texts of public-domain works, mostly of literature, in a variety of languages and on a very wide range of topics.

Researchers wishing to explore the subjects covered by such archives, or the languages in which the contents are written, have a formidable armoury of resources to draw upon for assistance. Just to mention the study of English, a selective but wide-ranging web-site⁴ compiled to aid researchers gives several dozen links, including large and academically prestigious projects such as the *British National Corpus*⁵ of modern spoken and written English, through to some more esoteric sites, such as the one dedicated to codifying geographical variations in US usage of the terms ‘pop’ and ‘soda’ to describe carbonated soft drinks.⁶

Linguistic researchers and others can use a formidable array of low-level textual analysis tools⁷ to dig deep into the content of these collections. But first of all, they need to locate and retrieve what they need from the mass that is available. This is the role that is generally filled by tools developed within the computer-science field of Information Retrieval (IR). The distinction between text IR and textual analysis can often be somewhat blurred, especially as IR has been forced to become more sophisticated in its use of text-analysis methods (for instance, in word-stemming and phrase-parsing) as its users become more demanding. But for many purposes, from the domestic to the academic, commercial text-IR technologies provide a more-than-adequate means of navigating the truly vast amounts of information distributed world-wide in the explosively-expanding World Wide Web.

IR and scholarship

While to some extent the information explosion has been driven by the new technology, that technology has to a large extent provided at least an interim solution to the problem of managing large data resources and providing access tailored to the use of individual scholars. The use of Information Retrieval tools (such as Google, Yahoo or Ask Jeeves)

² <http://www.gutenberg.org/>

³ <http://ota.ahds.ac.uk/>

⁴ http://sitemaker.umich.edu/acurzan/electronic_resources_for_english_language_study

⁵ <http://thetis.bl.uk/lookup.html>

⁶ <http://www.popvssoda.com/>

⁷ See <http://www.textanalysis.info/> for a comprehensive coverage.

is now almost second nature to anyone who uses a computer attached to the Internet. (It is also for school and undergraduate students fast becoming a one-stop method of obtaining either unmediated data or – in the extreme case – fully written essays on exam topics, not always linked with warnings about the perils of plagiarism).

In the world of computer science, and in many other scientific disciplines, the Internet is already the de facto forum for sharing scholarly results. Papers published in conference proceedings are considered to carry the highest prestige, yet the print-runs of those volumes are usually small and they are hard to find, and it is considered a normal scholarly courtesy to make one's work, subject to copyright agreements with publishers of paper proceedings, available in the form of downloadable files. These are to an increasing extent automatically cross-referenced by bibliographical resources such as Citeseer.⁸

This is not yet the case in the humanities, where the scholarly article or monograph generally carries more weight than a conference paper. Conventional publishing therefore remains the principal mode of research dissemination. But the scene is changing; a few online-only journals are emerging,⁹ as are initiatives to provide instant access to out-of-print back issues of journals and other texts,¹⁰ which naturally tend to retain their currency in the humanities much longer than they do in science.

It is small wonder, therefore, that outside their immediate area of expertise scholars increasingly find themselves driven to use unconventional means – sometimes hardly suitable means, it should be said – to manage their access to the essential materials of their work. A Google search can often rapidly and effectively reveal the existence of an apparently authoritative and comprehensive source of information that might have taken hours to find on unfamiliar library shelves. This method can sometimes be surprisingly useful, but it has to be said that it is haphazard at best. Quality control as such does not exist on the Internet.

Musical Corpora

A few scholarly projects are dedicated to the maintenance of collections of music, the foremost of which is the MuseData collection of electronic scores at Stanford University,¹¹ which has some 5,000 movements in encoded form. This primarily concentrates on music of the 18th century, especially the works of J.S. Bach, in which it is particularly rich.

A different kind of collection, specialising in a single historical repertory, is represented by two AHRB funded projects. The Electronic Corpus of Lute Music (ECOLM)¹² aims to provide encodings of substantially complete sources of music in lute tablature, a form of

⁸ <http://citeseer.ist.psu.edu>

⁹ An example in the domain of musicology is the *Journal of Seventeenth-Century Music*; see: <http://sscm-jscm.press.uiuc.edu/jscm/>

¹⁰ <http://www.jstor.org/>

¹¹ <http://www.musedata.org/>

¹² <http://www.ecolm.org>

notation that is unintelligible to those who do not play the instrument; the tablature encodings can be processed to give MIDI playback or simple score-notation. The Digital Image Archive of Medieval Music (DIAMM)¹³ has a slightly different emphasis, being largely concerned with the virtual collection and preservation of sources that are widely scattered and sometimes at risk; this includes some badly-damaged music manuscript fragments recovered from book bindings whose images can be processed to improve legibility to a dramatic extent.

Many other digital collections can be found, with varying degrees of scholarly aspiration. A large collection of choral music in PDF (graphic) format can be found via the web-site of the Choral Public Domain Library¹⁴; the Mutopia project houses many MIDI files of classical music on its web-site.¹⁵ Projects such as these depend on contributions from friendly musicians – usually those with a technical, rather than musicological background – so it is not surprising that editorial standards are, to say the least, variable.

In principle, the huge non-scholarly collections of music, especially audio recordings, that are available on the Web should be subject to analysis in the same kind of way that texts are. Unfortunately, this is very far from being reality. Apart from the problems of restricted (legal) access due to the extreme sensitivity of music to rights issues, the tools for content-based retrieval and analysis of music are only just in the development stage, although it is likely this will change quite rapidly in the near future.

One reason for this lack is commercial, in that software companies have not yet seen an industrial motivation that would merit the necessary large-scale investment. But the main reasons are technical. ‘Music’ is not a single one-format entity like ASCII text and it appears in a variety of forms, none of which in general can be conveniently mapped to text so that conventional IR systems can handle it. To be really useful for the musicologist, the tools ideally need to be able to act with seamless efficiency on all manifestations of a piece of music: live performance, audio and video recordings and printed or written scores. Put starkly, this requirement is a set of extremely challenging technical problems, involving state-of-the-art research in music perception and cognition, electronic engineering, artificial intelligence and algorithm design, and will occupy researchers for many years to come. The annual conferences of the International Symposium on Music Information Retrieval (ISMIR)¹⁶ are a highly-active forum for research in this area.

Listening to music

Things would be complicated enough if the problem were restricted just to musical scores in encoded form, but the inclusion of recorded material as part of the domain of musical enquiry adds an extra dimension of complexity for musicologists without parallel in most sister disciplines. Dealing with the sheer amount of data involved in a digital recording of

¹³ www.diamm.ac.uk/

¹⁴ <http://www.cpd.org>

¹⁵ <http://www.mutopiaproject.org/>

¹⁶ <http://www.ismir.net>

a substantial piece of music is a daunting prospect, with each second (of mono sound) being represented by a stream of no fewer than 41,000 numbers stored in the normal CD standard (and likely to be superseded by even denser representations in the near future).

It is very important to understand that in some senses, ‘notes’ only exist as fixed, ‘atomic’, separate events within a piece of music in their manifestation as elements of a written or printed score. If a score does not exist – and, in general, performed music does not exist as a score (although it may *derive* from a score) – one can never be absolutely sure which ‘sound event’ corresponds to a particular note. Sometimes perfectly musical sound events cannot be described in terms of atomic notes at all: glissandi, portamenti, vibrato, ‘note-bends’ and other kinds of ornament in vocal or instrumental performance go well beyond what can be expressed in terms of conventional music notation.

Some instruments (often somewhat outside the normal territory of musicology), such as the Swanee whistle, the musical saw or the Theremin, produce what can only be thought of as a continuous contour of pitch, with no separate note event being distinguishable from any other. The same could be said for performances in the singing style known as ‘vocalise’, although in the famous examples – Rakhmaninoff, Glière – they are based on notated scores.

(This is quite apart from the problem of notating musical events of indeterminate pitch; while one might reasonably assume that all notes have pitch, it is hard to say what the pitch of a cymbal ‘note’ might be.)

This problem becomes especially acute in the field of ethnomusicology, where experts are likely to disagree how to transcribe certain ‘notes’ that are clearly an essential part of the music as performed, yet might be considered to exist on a different, ‘ornamental’ level than the ‘main’ notes of a melody. Even in scored-based historical musicology, this problem of ornaments cannot be ignored: different manuscript or printed sources of the same piece might record a melody with an ornament omitted altogether, indicated by sign or written out in standard notation.

The distinction that scores tend to impose between ‘notes’ and ‘ornaments’ is but one example of the way in which a notated score-tradition tends to impose, or at least privilege, a certain hierarchical way of hearing the music. Learning to hear music in a structure-based way is, of course, a fundamental part of the training of a musicologist. It makes use of the amazing facility of the human brain in selecting important ‘features’ from the vastly complex stream of sensory input to which it is subjected, and of making categorical judgements about those features almost instantaneously at an unconscious level. Coupled with that mysterious indexing system which we call ‘memory’ in which many musical processes seem to be lodged, the human perceptual/cognitive system equips us well to do things of which we are almost unaware. Among these are the ability to ‘hear’ missing notes in melodies, or to substitute correct notes for those which are out of tune, played late or early, or just plain wrong.

Without extremely well-designed processing methods, computer systems cannot ‘listen’ to music in this structure-based way; they hear everything at the same level of significance, as it were.

Navigating music

One thing musical scores can do very well is to provide the means to find our way in a piece of music; at the most basic level, the performance of a certain number of notes in a certain instrumental or vocal part takes one to a certain point in a score which corresponds to the elapsing of a certain amount of time in that performance, and which could be given some helpful label ('End of section A located just after note 74 in the first violin part, occurring 33 seconds after the beginning' for example). A more sophisticated labelling would involve higher-level features of the score; the most familiar of these is the bar- (measure-) number, sometimes combined with the beat-level within the bar ('... at bar 23, beat 3').

Sometimes special 'sign-post' labels such as rehearsal numbers are explicitly embedded within a score. These were originally devised for performers in an orchestra to coordinate their rehearsals more effectively without too much tedious bar-counting. Their position is often associated with some sectioning of the score based on convenience in performance rather than the music's form. Similar labels might, however, be used in formal analysis to indicate structural segmentation of a more principled kind ('A', 'A'', 'B', and so on.)

On the other hand, while such labels may indeed have structural significance in terms of the score and the abstract concept of the work that it represents, they can only have temporal meaning in terms of a single performance. Even given an explicit direction from the composer or editor about the tempo at which a piece *should* be performed we cannot predict with any certainty how long any performance might take to perform from one sign-post label to another. The reason is simple: performers have the right – some would say the duty – to decide for themselves how fast or slow to play, and how much the basic tempo is altered during the passage of the music. They are also under no obligation to play the same piece at the same tempo in different performances.

Search within score

All professional score-editing computer programs have the facility to insert rehearsal numbers, and most allow a user to jump directly within the score to a given rehearsal- or bar-number. None yet has the ability to respond to a request such as 'Go to the first appearance of the main theme in the dominant'; far less: 'Show me all allusions to the *Dies Irae* in this movement.'

Applying such search facilities to work on collections, or 'databases', of music changes the order of the problem, and introduces one difficulty in particular. As the term 'allusions' in the example just given is intended to show, it is almost useless to design a musical search system to recognise only identical matches to a query; musical similarity is something much more subtle than that. Often one can hear a similarity (in a certain context) where only a rhythmic identity is concerned, or when just the melodic contour of the query is matched. Thus we are forced to face the problem of what computer scientists call 'approximate matching', in which elements of the items to be matched are either missing or different from what is being sought from the query. But this raises a further difficulty: just how 'different' can a matched passage be from a query while still remaining 'like' it?

Musical similarity is a hot topic in MIR. It is notoriously hard to pin down in a quantitative manner that will allow fair comparative evaluation of MIR systems. The best that can be done is to assemble panels of human subjects, selected to represent a wide

variety of musical expertise, and ask for binary (Yes/No) choices as to whether a collection of responses to a set of standard queries, if used to search a standard database, is 'like' (according to their personal interpretation) each of those queries.

This, in fact, is the evaluation paradigm in most common use in the world of text IR, where the Text Retrieval Conference (TREC)¹⁷ has been using something like it for decades, and it is also being used in MIR evaluation projects such as IMIRSEL¹⁸. There are, however, not surprisingly, some differences between text and music concerning the nature of queries and what a 'correct' result from a search might be.

Normally in text IR a query is a list of *words* that express a *concept* on which one seeks information; most text IR systems return a ranked list of documents that contain those words, the ranking being some measure of the system's judgement of the 'relatedness' of the document to the query. (E.g it might relate to frequency of occurrence of the query words, or to the proximity of multiple query words, in the document, or any combination of such values.) At a more sophisticated level, an IR system might return a ranked list of documents that are judged to be 'about', or 'relevant to', the underlying concept expressed in the query, rather than merely to its word list.

The TREC evaluation methodology uses a set of standard relevance judgements, each of which is simply a list of documents from within a standard database of documents which is judged by human experts to be 'relevant' to that query, whether by strict word-occurrence criteria or by conceptual association. This is inevitably a subjective matter which gives rise to a good deal of controversy in the text IR field, but by using a sufficiently large enough team of experts statistical significance can be ensured.

The retrieval effectiveness of an IR system can be judged by scoring how well (in terms of the list-rankings) each query retrieves those documents that have been pre-judged as relevant to that query. Various statistical summary-values of these scores can be used to compare the relative performance of different systems – a valuable competitive spur to ongoing academic research.

Music queries

The same kind of approach can be used for music, though the question of 'relevance' is even more problematic. Just as with text files, relevance can only be judged by human topic-domain experts, so assessments of musical relevance will be subject to the biased opinions of the experts who make the judgements – and there is no reason to suppose that musical experts would be any less biased than text-experts. However, text IR has an advantage for IR system designers in that there is, at least at a basic ('dictionary') level, a general degree of agreement about the meaning of individual words, despite the obvious problems caused by ambiguities, mis-spellings by users, contextual alterations of meaning and all the other things that make human language such a rich means of communication. But music has no semantic unit which corresponds in any useful way to a word, and certainly no agreed vocabulary or ontology of 'meaning'. While it is possible to divorce the individual 'atomic' elements that make up the temporal stream that

¹⁷ <http://trec.nist.gov/>

¹⁸ <http://www.music-ir.org/evaluation/>

constitutes a piece of music from the stream itself, none of those atomic elements ‘makes sense’ on its own except in relation to the others around it. The elements which make up a piece of music convey its ‘message’ almost entirely by a succession of overlapping contexts, unfolding in time as the music is experienced, not by their individual absolute meaning. There is no sense in which an unordered collection of pitches or of duration-values, say, could make a useful search query in the way with which we are familiar in text IR.

Given a suitable user-interface, a very natural form of musicological query would be to ask a system to find passages in the music in its database (which may be all the music on the World Wide Web, a single collection of pieces or simply a single piece) which are similar to the query. (This is somewhat different from seeking complete pieces matched by a query as in the earlier discussion.) Queries could be built by using a familiar music-notation editor, by mouse-selecting in an on-screen score or other time-based visualisation, by humming, singing or whistling the query into a microphone, or by playing a polyphonic score at a MIDI keyboard or (one day) into a microphone.

A general problem that has to be tackled in MIR is the need to match across different musical domains, to recognise, for example, a query expressed in score notation within audio recordings as well as in other scores. (Cf. ‘Navigating Music’, above.) If this can be done, and current research is enjoying some success in this area, this opens up very useful possibilities for musicological research that engages with both domains, in particular in performance analysis. Probably the most important medium-term goal is to develop some form of music representation format that will enable a computer program to take advantage of the common features of different performances without being confused by their divergences (especially in performance timing).

With such a representation in place, all that is needed is to encode for each performance to be compared the deviations from some ‘standard’ or representative performance (one that may not actually ever have existed) rather than computing the differences between all possible pairs of performances.

Conclusions

“There is ... a genuine prospect that empirical musicology ... will be able in future to participate in the development of music analysis, though not – I think – to change its fundamental nature as an exercise of the musical imagination.”
(Anthony Pople, ‘Modelling Musical Structure’, in E. Clarke and N. Cook, eds, *Empirical Musicology* (Oxford and New York: O.U.P., 2004), p. 154)

I make no apology for ending this essay with a number of open questions. Questions are one of the pieces of apparatus in the mental gym that most agreeably help us to exercise our imaginations. I would like to think that the much-missed Anthony Pople would agree with me about this.

What we need is *tools* for MIR and for music analysis. And it’s hard for scientists to gain academic credit for the development of tools. Where is the ‘good science’ in assembling the best work of others into an applied system? This is where industry should step in, but is the commercial incentive there? Are the systems that are likely to appeal to domestic consumers also likely to be really anything like what musicologists need? Can MIR tools

as they exist – or will soon exist – actually be used for musicological investigation? What is the role of TREC-style evaluation in that context?

Whatever tools we develop should allow researchers to exercise their musical imagination. Their output should perhaps be treated as *suggestions* rather than as manifestations of the underlying *facts*, allowing a variety of interpretations of the available evidence rather than uncovering universal truths. There is a lot to be done, but at least a beginning is being made.