AHRC ICT Methods Network Expert Seminar on Music

MODERN METHODS FOR MUSICOLOGY: PROSPECTS, PROPOSALS AND REALITIES

Hosted by Tim Crawford and Andrew Wathey at Royal Holloway, University of London, 3 March 2006

Rapporteur's Report

David Meredith

Goldsmiths College, University of London

Introduction

The main purpose of this day-long seminar was to explore the ways in which information and communication technology (ICT) can be used to enhance research, teaching and learning in musicology. In his introductory address, Tim Crawford explained that, when conceiving the seminar, he intended the term 'musicology' to be understood to include fields such as music theory, analysis and performance analysis as well as traditional historical musicology. The original intention was to exclude consideration of composition on the grounds that ICT has already been much more extensively and fruitfully applied in this area than it has in other fields of musical study. Nevertheless, some consideration was given to the ways that ICT can be used in creative music practice (i.e., performance and composition). This seminar provided both a picture of the *realities* of how ICT is currently being used in musicology as well as *prospects* and *proposals* for how it could be fruitfully used in the future.

The day was split into two sessions, each consisting of four half-hour talks followed by a half-hour open discussion. The morning session began with a talk by Geraint Wiggins which focused on the problem of representing musical knowledge so that it can be effectively processed using computers. This was followed by two talks, the first by Frans Wiering and the second by John Rink, describing on-going projects that use ICT to enhance the presentation, manipulation and management of large and complex networks of digitized musical source materials. The final talk in the morning session was given by Celia Duffy, who presented a preliminary map of how ICT is currently used within creative music practice.

The afternoon session began with David Howard talking about the various ways in which computers have been used in singing for voice training, analysis and research. Michael Fingerhut then described software tools developed at IRCAM that can be used to facilitate and enhance musicological study. There followed a talk by Michael Casey which focused on tools for structural analysis and information retrieval in musical audio. Finally, Adam Lindsay described an on-going project to identify user needs and existing technology for processing time-based audio-visual media.

In the open discussions that ended the two sessions, the topics discussed included:

- the computational representation of musical information and knowledge and, in particular, the dichotomy between symbolic and audio music representations;
- visualizing musical information and the design of appropriate interfaces for music processing software;
- the need for greater trans-disciplinary awareness and collaboration among technologists and musicologists; and
- the ways in which the use of ICT is transforming musicological practice and whether this transformation should be sudden or gradual.

In the remainder of this report, I shall first summarize the eight talks and the two discussion sessions. I shall then attempt to identify the main themes and conclusions that emerged during the day and suggest some ways in which the results of the day could be developed further.

Summary of the talks

Geraint Wiggins: Computer representation of music in the research environment
Geraint Wiggins began by pointing out that, before using computers to do anything useful with music, the musical information needs to be represented in a way that properly supports the logical and mathematical properties of musical materials. He stressed that musical data only becomes information once it has been interpreted. According to Babbitt (1965), music exists in at least three different domains—the graphemic, acoustic and auditory—and a well-designed representational system for music should allow for information in all three of these domains to be manipulated easily and for corresponding elements in different domains to be mapped onto each-other. A music representation system should also allow a musically sophisticated (but possibly technologically naïve) user to manipulate the musical materials in an intuitive way without having to understand the implementational details of the system. Wiggins stressed the importance of defining, for any particular application, an appropriate *musical surface* which is the most detailed level of description that is interesting. Wiggins suggested that standard Western staff notation has evolved to become a highly effective representational system for Western tonal music, since it provides an appropriate musical surface for many applications and correctly characterizes many of the logical and mathematical properties of the elements of Western tonal music.

Frans Wiering: Digital critical edition of music – A multidimensional model
Frans Wiering discussed the ways in which ICT can be used to produce multidimensional, multi-media, electronic, critical and scholarly editions of music. He stressed that we are no longer limited to producing two-dimensional visual scores that simulate the printed page and proposed that our encoding systems should allow us not merely to emulate electronically what can already be done on paper, but also to exploit the possibility of having multidimensional digital representations of musical works. Following McGann (1995, 1997), Wiering proposed that 'HyperEditing' can produce a 'fully networked hypermedia archive' that naturally and fully represents the historical and influential relationships between not only the printed and manuscript sources but also the performances of a work. Such a network would allow us to generate various sonic and visual views on the musical information and thus represent the textual transmission of a musical work in a more 'user-friendly' way than a book. Such a network can also be easily updated as more information becomes available (e.g., if the dating of a source is corrected). Wiering proposed that musicologists should debate whether 'HyperEditing' should be made a central activity in musicological scholarship.

John Rink: The Online Chopin Variorum Edition – Music and musicology in new perspectives
John Rink described the ways in which new technologies for image comparison and music recognition are being used for managing and studying complicated source networks in the *Online Chopin Variorum Edition* (OCVE) (http://www.ocve.org.uk/) and *Chopin's First Editions Online* (CFEO) (http://www.cfeo.org.uk/). In particular, these new technologies facilitate and enhance the comparative analysis of both manuscript and printed editions, by allowing images of sources to be juxtaposed in flexible and dynamic ways. Rink proposed that this should encourage better understanding of musical works and their transmission histories. Rink suggested that ICT is not currently being exploited to its full potential in musicology because of a lack of awareness and a resistance to change among musicologists. Rink proposed that online archives such as the OCVE and CFEO should aim to represent the sum total of what defines musical works and pointed out the need to incorporate sound files into such online networked resources.

Celia Duffy: The music map – Towards a mapping of ICT in creative music practice
Celia Duffy talked about her attempt to map the use of ICT across music, with a particular emphasis on the ways that ICT is currently being used in creative music practice (i.e., performance and composition). She discussed the use of ICT in performance training, performance analysis, dissemination of musical resources and music making. Performance training software typically analyses a user's performance and gives feedback and guidance. For example, MakeMusic's *SmartMusic* software (http://www.smartmusic.com/) provides "intelligent" accompaniment as well as instant feedback and assessment. Performance analysis software facilitates the analysis of pre-existing recordings by providing tools for visualization, segmentation, mark-up, manipulation and measurement of digital audio files. For example, the HOTBED system (http://www.hotbed.ac.uk/index.php) is a web-based database of audio and video resources relating to otherwise difficult-to-access performances, which provides tools for looping and

mark-up of recordings. Duffy then discussed the way that ICT is being used for making recorded music widely available. She mentioned in particular an on-going project, funded by the JISC, to digitize 12000 items (nearly 4000 hours) from the British Library's Sound Archive (http://www.bl.uk/collections/sound-archive/archsoundrec.html). Finally, Duffy mentioned some novel ways in which ICT is being used to create music, highlighting the recent rise of "internet music" in which multiple users co-operate asynchronously via the internet on the creation of musical materials.

David Howard: The computer and the singing voice
David Howard described the various types of software that are increasingly being used by professional voice users for voice analysis, voice training and enhancing vocal performances. He also reviewed ways in which computers have been used for managing and analysing data in research on the singing voice. He described how, in the domain of voice analysis, techniques such as electrolaryngography and electroglottography have been used to track fundamental frequency in vocal groups and show that singers adjust their intonation so as to stay close to just tuning. The same techniques can be used to demonstrate interesting relationships between degree of training and the *closed quotient*, which is the proportion of time that the larynx is closed during singing. He also described his WinSingad software (http://www.winsingad.org/) which is designed to provide appropriate, immediate, visual feedback in singing lessons. In pilot studies, singing students and teachers at York and Guildford claimed that WinSingad enhanced their learning/teaching experience. Finally, Howard described some work on voice synthesis and, in particular, his own efforts to simulate the castrato voice by combining synthesized counter-tenor and mezzo-soprano voices.

Michael Fingerhut: Filling gaps between current musicological practice and computer technology at IRCAM
Michael Fingerhut began by stressing the degree to which the tools used within a research domain limit both the rate at which knowledge can be acquired and the nature of the discoveries that can be made. He then presented a map of the field of music information retrieval which attempted to characterize the relationships between the different levels and types of musical structure. Fingerhut described and demonstrated various software tools developed at IRCAM that can be used to enhance and facilitate musicological study. In particular, he talked about several components of the *Music Lab 2* (http://musiquelab.ircam.fr/) project including *ML-Annotation*, which can be used for score/audio synchronisation, comparison and summarization; and *ML-Maquette*, which is a tool for validating an analytical model of a musical work. He also described *OAI*, which is a protocol for accessing different types of multimedia metadata (e.g., audio, visual) in different organisations.

Michael Casey: Audio tools for music discovery and structural analysis
Michael Casey described his recent work on the development of tools for information retrieval and structural analysis in musical audio. In particular, he explained how similarity matrices can be used to visualize large-scale structural relationships both within and between audio recordings of musical works. He also described new techniques for rapid extraction and indexing of perceptually significant patterns in audio, thereby allowing very fast retrieval of information from internet-scale audio databases. He described how his *chroma matched filter* method can be used in conjunction with locality sensitive hashing (Datar *et al.*, 2004) (see http://www.mit.edu/~andoni/LSH/) to find, for example, occurrences of Leitmotivs or chord sequences in large audio collections. He also showed how these techniques can be used for automatic 'mashing'—that is, finding pairs of songs with the same harmonic structure that can be performed simultaneously. Casey pointed out that a downfall of many previous approaches to audio music information retrieval has been the false belief that effective retrieval is only possible after extracting complex features containing a great deal of structural information. His own recent systems use only relatively simple, low-level MPEG7 features such as chromagrams (see http://doc.gold.ac.uk/~mas01mc/mpeg7/).

Adam Lindsay: ICT tools for searching, annotation and analysis of audio-visual media
Adam Lindsay reported on progress made so far in a project entitled "ICT Tools for Searching, Annotation and Analysis of Audio-Visual Media" (ICT4AV). This is a two-phase project in which the purpose of Phase I is to identify the tools, technologies and research being explored by technologists; and the purpose of Phase II is to identify the needs and requirements of humanities researchers. Having completed Phase I,

one result of this project is that existing technologies can be divided into those that arise from a move to a digital workflow and those specialist tools for musical data analysis that are derived from contemporary research. Lindsay foresees that, in Phase II, challenges in communicating with non-technical users about ICT could include a general resistance to ICT tools in the humanities and an unrealistic expectation among such users as to the capabilities of ICT. Lindsay pointed out that digital rights issues represent a substantial obstacle to the free use of collections of digital media for research purposes. Lindsay then gave a brief overview of the field of music information retrieval (MIR) and a potted history of the highly successful ISMIR conferences (http://www.ismir.net).

Summary of the discussions

One of the main topics debated in the discussion sessions was the computational representation of musical information and knowledge and, in particular, the dichotomy between symbolic and audio music representations. Adam Lindsay began the debate by asking Geraint Wiggins if he thought that implementational details were unimportant. Lindsay suggested that it was surely important for the user to understand the underlying semantics of a system. Wiggins responded that a user should not have to understand the implementational details in order to use a system. It was not clear, however, whether Lindsay and Wiggins were referring to the *user* interface or the *application programmer's* interface. It is a well-established principle in object-oriented software engineering that the implementation of software objects should be hidden from application programmers who wish to use these objects to build new software. The programmer should only have access to the *application programmer's interface* (API) which essentially defines a contract between the programmer and the designer of an object, which stipulates what the object does and how the programmer can use the object, but does not place any constraints on *how* the object accomplishes its tasks. Wiggins stated that it was important to distinguish between the interface and the representation in any application.

Kia Ng then asked Wiggins what he thought about the possibility of using MusicXML (http://www.recordare.com/xml.html) as a universal format for exchanging musical information. Wiggins claimed that MusicXML was fundamentally flawed because it was impossible to use it to represent the parallel, intertwined and independent hierarchical structures that occur in music. He said that we need appropriate *abstract data types* (ADTs) that allow for interoperability, but we do not necessarily need a single standard file format for storage. He added that a very general representational system can sometimes be less useful for particular applications. Michael Casey also pointed out that abstractions need to be designed with specific test applications in mind—trying to design a "one-size-fits-all" representation is misguided unless one has a very clear idea as to the specific tasks to which the representation will be applied.

Alan Marsden suggested that computational representational systems for music should support a multiplicity of models. For example, there are various different methods for determining the key of a passage of music or for changing the tempo of an audio recording of a passage. He claimed that we need to employ cutting-edge probabilistic and machine-learning techniques in order to represent properly the multiplicity of models employed by musicians and musicologists. Wiggins agreed that different musical surfaces are required for different applications and stated that the sonic, score and perceptual manifestations of a musical work are simply different musical surfaces giving different perspectives on the Platonic, unattainable compositional idea of a musical work.

Denis Smalley disagreed with the claim made by Wiggins in his talk that notation is a good representation of music. Smalley echoed Celia Duffy's assertion that notes represent only a small part of what we perceive as important in music. Lisa Whistlecroft similarly suggested that representing a musical work at the note level resulted in a very bad approximation to the sonic intent of the composer. She stated that we need methods for analysing not just musical audio but also the composer's intent. Mark Plumbley pointed out that representing a musical work with notes results in only a very rough approximation to the sound. Wiggins responded by claiming that the notes, the sound and the intent of a musical work are just different musical surfaces. He suggested that, in evolving modern staff notation, musicians have done an excellent job in producing a representational system which provides a convenient musical surface that is structurally general enough and sufficiently expressively complete for a wide range of applications. Nicholas Cook agreed that the notated score of a musical work provides a convenient common point of reference, allowing

performers to negotiate the finer details of a work. He suggested that the same thing could be said about certain analytical techniques such as Schenkerian approaches (Schenker, 1979): such methods allow musicologists to distinguish between what is essential and what is not in a musical work. Frans Wiering pointed out that pitch is a perceptual entity, and suggested that note-oriented scholarship is too concerned with the production of scores and not concerned enough with the active consumption of music. He said that listeners perceive music *conceptually* as consisting of, not simply notes, but voices, chords, phrases, and so on. He suggested that our representational systems should account for this.

Geraint Wiggins remarked that, whereas the morning session had been concerned primarily with music at the symbolic level, the afternoon session had focused primarily on sub-symbolic, audio processing. I suggested that we need to be able to map seamlessly and transparently between representations at different levels of structure (i.e., between musical sounds, notes and perceptual entities such as groups, streams and chords). I also remarked that there no longer seems to be such a clear dichotomy between the symbolic and audio domains, as we are beginning to use similar techniques (e.g., chromagrams) on both types of data. However, Adam Lindsay claimed that audio engineers have basically given up trying to extract the notes from an audio recording. We can nevertheless, as Wiggins remarked, extract lots of other useful information from raw audio information, even if a perfect audio-to-score transcription system is an unattainable goal.

A second important topic discussed was the design of appropriate interfaces for music processing software and the visualization of musical information. Tim Crawford stressed the importance of interface design and the challenges involved in giving technologically naïve users access to the power of a system when there is complex underlying technology. As an example, Crawford mentioned the HUMDRUM tool kit (Huron, 1997) (see http://www.music-cog.ohio-state.edu/Humdrum/) which is a powerful set of software tools but which only has a "bare-bones" Unix command-line interface that a non-technical person used to graphical user interfaces would find very intimidating. Nicholas Cook said that the problem with HUMDRUM was that it was designed by technologists who did not take into account the huge learning curve that had to be negotiated before HUMDRUM could be operated effectively by a non-technical user. He re-iterated Crawford's point that we need software which is powerful but also easy to use. Suzanne Lodato claimed that the vast majority of people want a tool kit that they can just pick up and use without having to invest lots of time and effort in training. John Rink suggested that a user should be able to switch on and off various features of an interface and thus control aspects such as the degree to which processes are automated. Chris Banks pointed out that, in some disciplines, the user interface is considered to be a form of interpretation. Alan Marsden's claim that music software should support a multiplicity of models was also relevant to this discussion, since it implies the need for the user to be able to generate multiple views on the same data. Geraint Wiggins remarked that modern database technology is designed to be able to cope with this sort of multiplicity.

The third main topic considered was the need for greater trans-disciplinary awareness and collaboration among technologists and musicologists. Mark Plumbley claimed that there is now a new generation of people who have high-level skills in both music and technology and who are therefore not locked into traditional disciplines. It would therefore only be a matter of time before the problem of inter-disciplinarity in our field would disappear for good. John Bradley, however, claimed that people with interdisciplinary skills are often not good at building tools that can be used by users other than themselves.
Celia Duffy highlighted the danger of technologically naïve (but musically sophisticated) users making false conclusions from the output generated by analytical software tools. She said there was a danger of "de-skilling"—that is, of users putting too much faith in the results generated by software. David Howard agreed with this, pointing out that software can sometimes give expected results for the wrong reasons. Alan Marsden said that users could no longer afford to be naïve and that they must be aware of a multiplicity of methods for carrying out musicological tasks, such as, for example, key-tracking or tempo changing in audio data. Users must have a realistic appreciation of the limits and strengths of any software tool that they use and they should not be misled by exaggerated claims for the capabilities of (particularly commercial) software. Adam Lindsay concurred with this, emphasising the need for users to be trained properly in the use of software. Tim Crawford suggested that this seemed to imply that each of us had to be both a computer scientist and a musicologist. However, Marsden disagreed with this, claiming that it

was only necessary for musicologists to be able to frame questions in a way that is comprehensible to a computer scientist so that useful new tools can be developed.

Celia Duffy mentioned that one of her colleagues admitted to only using software as a last resort and David Howard mentioned that he had met singing teachers who used his software incorrectly but nonetheless obtained good results. Mark Plumbley suggested that we may be able to learn more about what is required by music professionals by studying the ways in which they misuse software than we can by doing a straightforward user needs survey. Suzanne Lodato suggested that it might be useful to present tools to non-technical musicologists and ask them what would be useful. She also pointed out, however, that such "naïve" users might not have a good sense of what is possible, and may only be able to make useful suggestions after sufficient exposure to the technology. I remarked that one, longer-term solution would be to give technology training a more central rôle in musicological education. Frans Wiering suggested that musicology needs to become a less isolated activity, that a new culture of collaboration needs to be nurtured within musicology that promotes both inter- and intra-disciplinary co-operation. Tim Crawford pointed out that the funding bodies are promoting inter-disciplinary collaboration by funding research projects which employ both computer scientists and musicologists. Frans Wiering remarked that some of the most interesting musicology is done outside of music departments. Lisa Whistlecroft proposed that the AHRC should establish a watch on the state-of-the-art in relevant technologies. She said we should aim to mitigate the risk of sophisticated tools being used in a simplistic manner by untrained users and that it was important that we should try to model thought processes rather than focus on the tools themselves. Geraint Wiggins added that there was an onus on technologists to explain their methods in a non-technical way.

Tim Crawford then highlighted the problem of musicologists and musicians in general being under-represented at the annual ISMIR conferences. Michael Fingerhut suggested that the MIR community is more concerned with popular music and commercially important problems which results in the tools developed often being inappropriate for musicological or academic use. This highlighted the need for improved communication between engineers and musicologists.

The final general topic debated during the discussion sessions was the ways in which the use of ICT is transforming musicological practice and whether this transformation should be sudden or gradual. John Bradley explained that the use of computing technology transforms the nature of musicology because it is so different from traditional methods. He warned that this transformation should be gradual, not sudden, if it is to be sustainable. He observed that, in text analysis, computer-based researchers had written themselves out of their communities by being too far outside traditional methods for their colleagues to be able to work with their results. He highlighted the difficulty for non-technical people to evaluate work based on technical methodology. Tim Crawford suggested that the possibilities of technology should be presented to musicologists as an opportunity for better working practices, not as a threat to traditional methods. Michael Casey, however, proposed that technology should be used, not merely to assist with tasks that can already be done by humans, but to do entirely new things that were impossible without it. Andrew Wathey remarked that one of the principal themes of the day had been the impact of technology on interdisciplinary transformation and the possibility for technology to be music-led. He stressed the need for users to be properly trained in the use of software.

Conclusions and future developments

A number of major themes and conclusions emerged throughout the course of the day. First, it appeared that the traditional dichotomy between symbolic and audio music representations in music informatics seems to be dissolving with similar techniques being used on both types of data. It became clear that representation systems for music must be able to cope, not just with notes, but also with the detailed structure of musical sounds, the composer's intent and other higher-level structures such as streams, groups and chords. Furthermore, it must become possible to map transparently between representations at these different structural levels. On the other hand, it would be a mistake to attempt to develop universal representational systems without having specific test applications in mind.

Another major theme of the day was the importance of appropriate interfaces and ways of visualizing music information. Technologically naïve but musically sophisticated users should be able to access the full power of a system by means of a graded interface which can be customized for use by both beginning and advanced users. Representations should support a multiplicity of views on the data and allow for multiple

methods to be applied. New web and database technologies should be exploited to produce multi-dimensional networked online archives containing both visual and audio digital musical materials.

Important issues arise from the inter-disciplinary nature of the field of computational musicology. It emerged that there is a great need for increased trans-disciplinary awareness: technologists need to be more in touch with what is required by music professionals and music professionals need to have a better understanding of what is technologically feasible. This suggests that training in relevant technology should be more central in music education, and professional users of ICT should be properly trained in its use. It also seems that, gradually, the "lone-scholar" culture in musicological research should be replaced with a more collaborative culture like the one that is typical in scientific disciplines.

Finally, there was some debate as to how computing technology can best transform musicological practice. There seemed to be general agreement that, for computing technology to have a positive and long-lasting effect on musicology, it has to transform practice within the discipline gradually, not suddenly. Nevertheless, there is no doubt that the new technology should be exploited to do things that were previously impossible without it.

A good way of building on the results of this seminar might be to hold a workshop for technologists and music professionals with the aim of pinning down more accurately the ways in which technology can help music professionals. This would enable the identification of worthwhile problems to be solved in future research projects in the field of computational musicology.

References

Babbitt, M. (1965). The use of computers in musicological research. *Perspectives of New Music*, 3(2), 74–83. Available online at <http://www.jstor.org>.

Datar, M., Indyk, P., Immorlica, N., and Mirrokni, V. (2004). Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the Symposium on Computational Geometry, 2004.*

Huron, D. (1997). Humdrum and Kern: Selective feature encoding. In E. Selfridge-Field, editor, *Beyond MIDI: The Handbook of Musical Codes*, pages 375–401. The MIT Press, Cambridge, Mass.

McGann, J. J. (1995). The rationale of hypertext. Available online at <http://jefferson.village.virginia.edu/public/jjm2f/rationale.html>.

McGann, J. J. (1997). The rationale of hypertext. In K. Sutherland, editor, *Electronic Text, Investigations in Method and Theory*, pages 19–46. Oxford University Press, Oxford.

Schenker, H. (1979). Free Composition (Der Freie Satz): Volume III of New Musical Theories and Fantasies. Schirmer, New York. Translated and edited by Ernst Oster.

Online resources

Online Chopin Variorum Edition (OCVE) <http://www.ocve.org.uk/>
Chopin's First Editions Online (CFEO) <http://www.cfeo.org.uk/>
MakeMusic's *SmartMusic* software <http://www.smartmusic.com/>
HOTBED system <http://www.hotbed.ac.uk/index.php>
British Library's Archival Sound Recordings Project <http://www.bl.uk/collections/sound-archive/archsoundrec.html>
WinSingad software <http://www.winsingad.org/>
Music Lab 2 <http://musiquelab.ircam.fr/>
Locality sensitive hashing <http://www.mit.edu/~andoni/LSH/>
MPEG7 <http://doc.gold.ac.uk/~mas01mc/mpeg7/>
ICT4AV <http://ict4av.lancs.ac.uk/>
ISMIR conferences <http://www.ismir.net>
MusicXML <http://www.recordare.com/xml.html>
Humdrum <http://www.music-cog.ohio-state.edu/Humdrum/>