

Purcell Plus: Exploring an eScience Methodology for Musicologists

Case for Support

(See also ‘Objectives’ and ‘Summary’)

Research Questions

Vast amounts of music are now available in digital form, most of it distributed across the Internet. This presents both opportunities and challenges for musicology, since the resources hold out the twin prospects of a vast expansion of the range of music that can be studied and ready-made corpora for such investigation. This defeats at a stroke the objection that musicology is perforce a data-poor discipline.¹ But the tools that have so far been developed by technologists are generally neither intended nor often really appropriate for use in musicology.

Little consideration has been given to the actual methodology of musicological research using these tools on the new corpora; still less on the impact such work might have on the traditional values and methods of the discipline. *Purcell Plus* addresses this issue by building a framework for the tools to help with eliciting a body of musical knowledge from the musical source material itself; this can then be assessed against the findings of traditional methods. Three basic domains of source material (encoded scores, audio performances and textual commentaries) can provide a rich context of associated labels, signifiers, statements and gestures by indexing their locations. This allows a scholar to draw conclusions using either traditional criteria or the inference techniques of AI.

The aim of *Purcell Plus* is to promote the use of the new technology in an enabling role, above all in supporting and strengthening musicological argument while opening up new horizons both in terms of the quantity of music that can be dealt with by computer-based methods and of the new quantitative modes of evaluation of evidence that such methods offer.² This project is designed to lay the foundations for future work in the field.

The *Purcell Plus* project is conceived within the context of two other parallel music research projects at Goldsmiths from which it will derive a good deal of benefit and to which it will in turn contribute: a) the MeTAMuSE musical score-encoding project³ funded by the Andrew W. Mellon Foundation (August 2006-July 2008) and b) OMRAS 2, a large eScience-oriented music-information-retrieval project funded by the EPSRC (January 2007-December 2009). The methodological issues on which it focuses, however, make it distinct from either of these, and we expect that our findings will be useful beyond the confines of musicology, applying in some degree to other humanistic scholarly disciplines needing to reconcile traditional modes of discussion with the content-based analysis of multimedia data.

Research Methods

We shall explore how certain aspects of the explicit knowledge and the tacit knowledge⁴ about a piece of music (or a group of pieces with some coherent integrity) – in our proof-of-concept study the Henry Purcell’s *Fantazies and In Nomines* (see below) – and its

¹ See Nicholas Cook, ‘Towards the Compleat Musicologist’, keynote address, ISMIR conference, London, September 2005, available online at:

<http://ismir2005.ismir.net/documents/Cook-CompleatMusicologist.pdf>

(accessed on 30th October 2006).

² Some pertinent issues are raised in the third of David Huron’s 1999 Bloch Lecture (UC, Berkeley), ‘The New Empiricism: Systematic Musicology in a Postmodern Age’, available online at:

<http://csm1.som.ohio-state.edu/Music220/Bloch.lectures/3.Methodology.html>

(accessed on 3rd October 2006).

³ Temporary MeTAMuSE web-site at: <http://doc.gold.ac.uk/~mas01tc/MMweb/>

⁴ http://en.wikipedia.org/wiki/Tacit_knowledge

performance might be elicited from three basic domains or bodies of evidence: a) notated musical source material, such as historical manuscripts and printed editions;⁵ b) modern recorded performances; and c) verbal commentaries by experts or others, as they might be expressed in scholarly articles, programme notes and concert or CD reviews.⁶

There is little precedent for such research, and we want significant musicological and methodological research outcomes. So the project has two closely-related strands, each of which could proceed independently but will achieve far more in a coordinated effort. A methodological eScience ‘framework’ will be built, founded on the distributed data and computing resources of OMRAS 2 and MeTAMuSE, whose tools will allow information to be retrieved, marshalled and presented from three linked domains of data (scores, audio performances, commentaries); this will be carried out by an RA working with and guided by the PI in the Intelligent Sound and Music Systems group (ISMS) in the Computing Department at Goldsmiths College. Making use of this technological framework as it develops, a PhD student, also at Goldsmiths, will investigate in depth possible methodologies for the exploitation of the empirical output of the low-level tools in the context of traditional musicological discourse; an important deliverable will thus be that student’s dissertation, though it will appear some two years after the rest of the project. In the doctoral study we would further expect specific musicological case-studies to be pursued to demonstrate the proof of concept by producing evidence of novel eScience-enabled research insight.

We shall adopt a pragmatic approach to the project, using technological tools where possible as they become available, but also traditional techniques where these are more successful. The progress of the project will be evaluated mainly informally by reference to an advisory team whose opinions and suggestions will be considered on a regular basis. This will be complemented by a survey of responses from the musicological community, linked with a workshop; the aim is to guide the direction of future work in the field, for which further funding will be sought. Throughout, we will be working with professional consultants in music publishing and software development, whose input is seen vital in designing products that musicologists can actually use.

The potential scope and complexity of such a project – even with a small collection of works – is very great. Therefore we shall not attempt to analyse *every* musical feature of the works, nor to elicit *every* aspect of musicological knowledge about them; the initial set of musical, verbal and performative features or gestures we hope to identify, extract, analyse and associate will be limited, though selected to provide a meaningful proof of concept.

The music that has been chosen is Henry Purcell’s *Fantazies and In Nomines* for instrumental ensemble, recognised as an early masterpiece by England’s greatest composer. There are recordings and several editions; a body of literature is devoted to this music, most recently a 2006 PhD dissertation,⁷ whose author will act as a consultant. Score-encodings will be done

⁵ Of course, a very large class of music has no manifestation at all in domain a; that is, music which has never been written down (much popular and traditional music, and all music that is improvised) or whose written form is at best a mere outline sketch (such as the lead-sheets used by jazz musicians). This is simply a special case where domain a is vanishingly data-poor; there are still likely to exist performances and commentaries on the music or those performances. For the purpose of the present discussion, however, we restrict ourselves to Western art music that *does* survive in notated form, i.e. the vast bulk of the classical repertory.

⁶ The verbal data may be modern or historical, and possibly accessed by remote linkage to a database project such as Simon McVeigh’s *Calendar of London Concerts 1750-1800 advertised in the London Daily Press* [McVeigh 2006] or to a resource such as *Grove Music Online* [www.grovemusic.com], given suitable online interfaces such as the Z39.50 protocol [Z39.50].

⁷ Alan Howard, ‘Purcell and the Poetics of Artifice: Compositional Strategies in the Fantasias and Sonatas’, unpublished PhD dissertation, King’s College, London, 2006

within the MeTAMuSE project before September 2007, providing an essential core element to *Purcell Plus* at no cost.⁸

Explicit and tacit musical knowledge

As a proof of concept of one type of explicit knowledge that is manifest across all three domains, the project uses recordings to see whether performance gestures do or do not reflect graphical signs in the score, or they can be identified with points raised in commentaries. Such gestures can be tacit: for example, performers may respond to structural features (such as cadences) as well as to markings in the score. Furthermore, as well as a change of tempo, we may find that the change is prepared by, perhaps, a slackening of the tempo *before* the marking; this is an example of tacit knowledge informing performance. Here, one is examining a *contextual* relationship between score signs and performance gestures. In the same way, one might look at implicit structural ‘signs’,⁹ the most obvious being harmonic cadences, to see if such gestures are detectable. Some gestures may emphasize analytical points that are mentioned in commentaries. Our method will capture the ways in which different performers interpret these signs.

The scores contain many musical features which could carry information for our body of knowledge, both explicit and tacit. We shall limit our enquiries to a) tempo-change labels, b) segmentation features, such as double-bars (explicit) and cadences (implicit), and c) other simple harmonic patterns. Where possible, these will be derived computationally directly from the score-encoding; otherwise we shall use more traditional methods.¹⁰

The methods and techniques developed in this project are not dependent on the nature of the musical work; a similar study could be done on any one of a number of musical works.¹¹ The research thus will benefit the entire musicological community in providing a methodological framework for a research approach using the new technology. The tools in development by MIR researchers¹² may prove more appropriate for rhythmically less variable styles of musical performance (e.g., most Western popular music); but the underlying methodology – relating written notations (if they exist), performances and commentaries – could apply in much the same way.

The research team and their roles

The **PI**, Mr Tim Crawford (CV attached), is experienced in managing complex research projects. He managed the UK effort (funded by JISC) for the OMRAS (Online Music Recognition and Searching) project (1999-2003),¹³ which has recently received further EPSRC funding as OMRAS 2 (Goldsmiths and Queen Mary), on which he will be employed

⁸ Funding for a parallel project, the digitisation and online provision of high-quality scanned colour images of Purcell’s manuscript, is being actively pursued by a consortium led by the Music Department of the British Library; but neither MeTAMuSE nor the present project significantly depends on this enhancement.

⁹ Using ‘sign’ in the semiotic sense of a signifying entity – in this case not a single graphical sign.

¹⁰ A similar approach is adopted in the tempo-tracking method employed in the AHRC-funded Mazurkas project, where a ‘reverse-conducting’ procedure, in which a researcher’s multiple attempts to tap a computer keyboard in time with a performance in which the tempo varies, is later enhanced by a specialised audio-analysis program which aligns the ‘human’ beats to precise note-onsets found in the audio. See <http://mazurka.org.uk/> (accessed 3rd October 2006).

¹¹ For example, we hope (subject to progress and further funding if necessary) to initiate a similar study based on some of Beethoven’s string quartets, which will also be encoded within MeTAMuSE and have been the subject of a recent JISC-funded audio-digitisation effort at the British Library, National Sound Archive.

¹² For example, as represented by the ISMIR conferences (<http://www.ismir.net>)

¹³ <http://www.omras.org>

from Jan 2007 working with professional users such as musicologists. He also conceived, led and managed the AHRB project ECOLM (Electronic Corpus of Lute Music),¹⁴ which was funded initially as a Research Grant (PI: Prof. Laurence Dreyfus, KCL, 1999-2002) and subsequently under the Resource Enhancement Scheme (PI: Prof. Geraint Wiggins, City Univ. and Goldsmiths, 2003-2006), and was praised by reviewers in glowing terms:

The project has more than matched its original ambitions, building a corpus of material of inestimable value to future research in the field. A very remarkable matching of musical with computing expertise to the benefit of participants at all levels of the project. This has been very good value for money, issuing in a considerable range of publications in conventional form and has developed not only the work of the individual researchers but aspects of musical bibliography, notation, and an awareness of the repertoire and its history.

The **Co-I**, Prof. Geraint Wiggins (CV attached), has worked in research since 1984. He has published extensively in computational linguistics, computational logic, computational music cognition and computational creativity. He is educated to doctoral level in both science (Computational Linguistics) and humanities (Musical Composition). Within music cognition, he has published in music representation, automated composition and interactive performance, musical meaning and melodic perception. He has edited five major conference proceedings, and three special issues of major journals. He is Associate Editor (English) of the international music cognition journal *Musicae Scientiae* and is Editor-in-Chief of the *AISB Journal*. He has been or is PI on ten research grants to date.

The PI and Co-I are currently working together on the MeTAMuSE project, funded by the Andrew W. Mellon Foundation, which is building a framework for building corpora of encoded musical scores. This will incorporate a musical knowledge-inference system based on musical features of the encoded scores developed by Prof. Wiggins, which can be linked directly with related parts of Purcell Plus.

The **RA** we intend to employ, Mr David Lewis (CV attached), a KCL-trained musicologist and expert programmer, worked for more than three years on the ECOLM project, for which he designed a database incorporating several features that are being adopted within MeTAMuSE and will be useful in Purcell Plus. He currently works with Prof. Wiggins on the EPSRC-funded project 'Modelling Melodic Memory and the Perception of Musical Similarity'. In his work he will be supervised and advised on a daily basis by the PI and Co-I.

The **Consultants** to the project are considered vital to its success, contributing both to its actual substance and to its likely impact on the musicological community.

Alan Howard will contribute a hypertextual expert musicological commentary to the Purcell *Fantazies*, based on his recent PhD thesis (see above); this will enable the project to demonstrate proof of concept for the methodological framework. He is currently employed as RA on the AHRC-funded project 'Musical Creativity in Restoration England' (Manchester; PI: Rebecca Herissone).

David Smith is Head of Research at Newport School of Art, Media & Design (University of Wales, Newport) and will advise on aspects of tacit knowledge elicitation; his recent experience of studying the tacit knowledge shared within communities of craft workers in rural India and the impact of newly-introduced technology will be of particular relevance.

David Streatfield runs Information Management Associates, which operates as a series of consultancy, training and research teams focusing on information management and professional development support in education, with particular attention to information strategies, information services and libraries. His contribution will be in assessing the likely

¹⁴ <http://www.ecolm.org>

impact of such technologies as used in Purcell Plus on the academic and libraries sector; to this purpose, a dedicated workshop is planned within the project.

Simon Rae-Scott runs **ConnectWorks** Limited, which specialises in publishing arts sites and database-driven information sites for corporate and public sector clients. He will negotiate with software developers to produce pilot/demo versions of an interactive web-site that would allow musicologists to engage with Purcell Plus in a productive manner; this could lead to future developments of products that will benefit academic and other musicians by building musician-friendly systems over the latest innovations in music analysis and retrieval technologies.

ConnectWorks has worked closely in the past with **Bill Hunt**, a founder member of the viol consort Fretwork, whose music-publishing business, Fretwork Editions,¹⁵ has produced many editions of music closely related to the Purcell Fantazies, including a forthcoming edition of the fantasies of John Jenkins, which will be made available to Purcell Plus in electronic, searchable form.

An **Advisory Group** will be set up, consisting primarily of musicologists and musical technologists, to advise on various aspects of the project as they arise.

¹⁵ See <http://publishing.fretwork.co.uk/>

Scheme of work (milestones/outputs in bold)

Schedule (months)	Task	Participants (time in months) NB PI supervises throughout						
		RA	PhD	CoI	CW	BH	DS	AH
1-3	Align audio performances and tag audio with score data; store aligned audio and score data in a simple relational database structure	3						
	Basic web-interface to database							
4-5	Build vocabulary of musical objects in the three domains of information; build vocabulary of structural/compositional processes and qualities, expressed as relationships between music objects	2	2	✓				
6-7	Build vocabulary of performance gestures that can be detected by audio analysis	2	2	✓				
8-10	Formalise vocabulary-items (where suitable) as ontology using OWL; combine with OMRAS music ontology	3		✓				
	Extended OMRAS music ontology							
11	Journal article	1	1	✓				
8-11	Establish vocabulary of verbal discourse referencing specific locations within the music		4					
12-13	Analyse scores and performances with special focus on the regions within them to which reference is made in the verbal discourse	2	3					
14-15	Identify implicit structural features of the scores and investigate how these have elicited verbal response in the commentaries or gestures in the performances	2	4					
16	? Encode Jenkins Fantasias (selection)	1				✓		
	Hypertextual commentary on Purcell <i>Fantazies & In nomines</i>					✓		✓
17	Journal article	1	1	✓		✓		✓
18-19	Develop data-model in a relational database to store the associations between items of information identified above	2						
	Pilot musical knowledge-base for Purcell <i>Fantazies & In nomines</i>							
20	DRH 2008 conference paper	1	1	✓				
	Develop interactive web-interface				✓			
	User survey report				✓		✓	
	Workshop	✓	✓	✓	✓	✓	✓	✓
21	ISMIR 2008 conference paper	1	1	✓	✓?	✓?	✓?	
	Regular meetings of whole team with advisors	✓	✓	✓	✓	✓	✓	✓
22-24	Final report	3	1	✓	✓	✓	✓	✓
	PhD dissertation		48					

Dissemination & Knowledge Transfer

The planned workshop, scheduled toward the end of the two-year project, will allow us to communicate with musicologists, and thus to incorporate their suggestions in our proposals for future funding.

We envisage submitting papers to at least two international conferences, (ISMIR 2008 and DRH 2008), and intend to produce various other scholarly papers and a journal article, possibly for *Computing in Musicology* or *Journal for New Music Research*. (See scheme of work, above.)

Extensive **Knowledge Transfer** results from our participation in the activities and contributions of our external consultants (see above). We feel that this engagement with non-academic professionals is of great mutual benefit and ensures that the work we do will have a presence outside the narrow confines of academic music and computing departments.