Overview of Personal Research Goals

David Meredith

January 12, 2006

1 Introduction

My personal research interests lie primarily in the field of music informatics. In particular, my goals are to develop robust systems for

1. automatic transcription of music from MIDI to notated scores,
2. content-based music information retrieval,
3. automatic music analysis and modelling music cognition.

These goals will be discussed in more detail below.

In order to evaluate and compare such systems, one needs access to large, 'gold-standard' collections of encodings of musical scores and expert music analyses that can provide a solid ground truth against which the output of such systems may be compared. I am therefore also interested in the development of better techniques for building publicly available, large, error-free collections of encodings of musical scores and expert music analyses.

In principle, it should be possible to integrate the best music-processing algorithms into a single system in which the various algorithms use the same core data structures for representing musical information. These core data structures should allow for the transparent mapping between corresponding elements in audio, MIDI and score representations of the same musical work. Ideally, it should be possible to use these same data structures for effectively manipulating and creating new musical material on both the audio and symbolic levels of structure. I am therefore also interested in the design of flexible, extensible and multi-level data structures and representations for processing music information.

2 Automatic transcription of music from MIDI to notated scores

A challenging problem in music informatics is that of computing a correctly notated score of a passage of music from a MIDI file representation of the passage, in which the onset time, duration and MIDI note number of each note in the passage are given. Typically, if the MIDI file is generated from a human performance on a MIDI-enabled instrument, then the onset times and durations of the notes in the passage are not strictly proportional to the correct notated values. The performer may also play wrong notes or miss notes out. A MIDI-to-notation

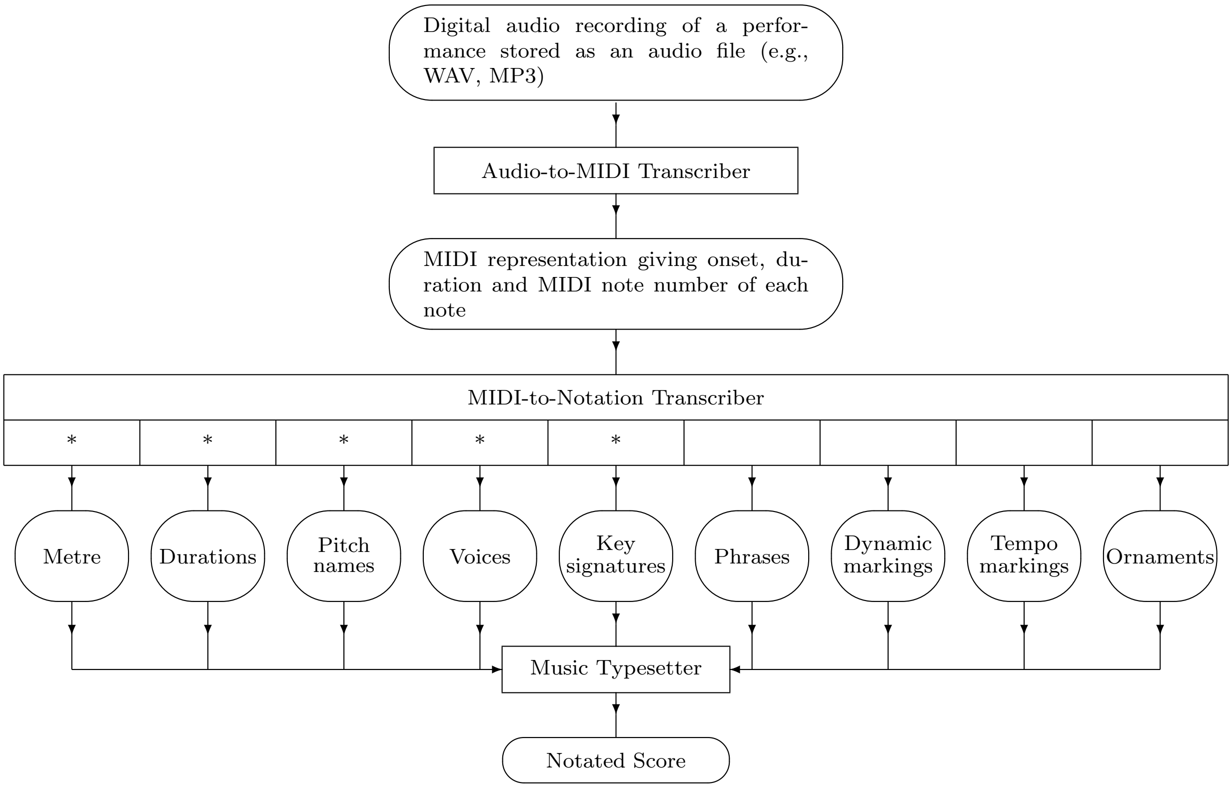

Figure 1: Flow chart illustrating the process of transcribing music from audio to a correctly notated score.

transcription system must therefore be robust to various types of 'noise' in the input data, if it is to cope with MIDI files generated from human performances.

Existing audio music transcription systems generate not notated scores but MIDI-like, 'piano roll' representations as output (Abdallah and Plumbley, 2004; Davy and Godsill, 2003; Plumbley, Abdallah, Bello, Davies, Monti, and Sandler, 2002; Walmsley, 2000). So, if one needs to produce a notated score from a digital audio recording, one needs not only an audio transcription system but also a MIDI-to-notation transcription system which is robust to the types of errors typically made by audio music transcription systems.

The flow-chart in Figure 1 illustrates the process of transcribing music from audio to a correctly notated score. A MIDI-like piano-roll representation is typically generated from an audio file using an audio-to-MIDI transcriber. The MIDI file is then given as input to a MIDI-to-notation transcriber which computes the information required to generate a correctly notated score. For the output to be a well-formed score, the MIDI-to-notation transcriber must at least compute the metrical structure, notated durations, pitch names, voices and key signatures. The components of the MIDI-to-notation transcriber that compute this essential information are indicated with asterisks.

A number of authors have proposed systems for automatic 'beat-tracking' or metrical structure analysis of MIDI data (for a recent review, see Temperley, 2004). The metrical structure of a passage has to be analysed in order to compute the time-signature (e.g., 3/4, 6/8, 4/4, etc.) that is in effect at each point in the passage. Computing the notated duration of each note is related to the 'quantization problem' (Desain and Honing, 1992) and involves associating the onset and offset of each note with a particular beat at some level in the metrical structure.

A number of algorithms have been proposed for computing the pitch name of each note in a passage of tonal music (Cambouropoulos, 1996, 1998, 2001, 2003; Chew and Chen, 2003a,b, 2005; Longuet-Higgins, 1987; Meredith and Wiggins, 2005; Meredith, 2003, 2005; Temperley, 2001). Most experts have assumed that the pitch name of each note depends on the local key and voice-leading. However, my research in this area (Meredith, 2006) suggests that the effect of local key is much more important than that of voice-leading and I have proposed a new algorithm, PS13s1, which predicts pitch names by modelling the listener's perception of local key. PS13s1 is the most accurate pitch spelling algorithm proposed to date.

PS13s1 can also be used to track the local key throughout a passage of tonal music and thus determine the appropriate key-signature at each point. A number of other key-finding algorithms have also been proposed in the literature (Chuan and Chew, 2005; Holtzmann, 1977; Krumhansl, 1990; Raphael and Stoddard, 2003; Shmulevich, Yli-Harja, Coyle, Povel, and Lemström, 1999; Temperley, 1999; Vos and van Geenen, 1996). However, evaluating and comparing key-finding algorithms is difficult because experts do not always agree on the key that is in effect at each point in a passage of tonal music.

In order to generate a well-formed score from a MIDI file, it is also necessary to identify the voice to which each note belongs. Algorithms for carrying out this task have been proposed by Temperley (2001), Chew and Wu (2004), Cambouropoulos (2000) and Kilian and Hoos (2002).

My goal is to produce an optimal system for carrying out MIDI-to-notation transcription by evaluating, comparing, optimising, improving and combining the algorithms that are currently available for carrying out the sub-tasks involved in the complete transcription process.

3 Content-based music information retrieval

The need for efficient and effective systems for managing large quantities of digital music information has increased steadily over the past decade, fuelled by, for example, the emergence of online music stores (e.g., iTunes) and the demand for intelligent music recommender systems (e.g., Amazon, MediaUnbound). Development in this area will undoubtedly accelerate rapidly with the emergence of new large-scale digital music services such as Microsoft and MTV's "URGE". The success of the ISMIR conferences (⟨`http://www.ismir.net`⟩) also bears witness to the growing demand for effective and efficient music information retrieval systems in education and academia as well as in the worlds of entertainment and e-commerce.

Fast and effective systems exist for content-based information retrieval in text databases (see, for example, the proceedings of the Text Retrieval Conferences (TREC), ⟨http://trec.nist.gov⟩). However, there are still fundamental problems that need to be solved before comparably efficient and effective systems are available for content-based retrieval of information from music data. Both audio and symbolic music data are fundamentally different from text data in that they are multi-dimensional and numerical. This means that the efficient string algorithms that are typically used in text retrieval cannot be straightforwardly adapted for use on polyphonic music data. A more promising approach is to employ techniques derived from fields such as computational geometry, computer vision and point-set pattern matching.

Between 1999 and 2003, I worked in collaboration with Kjell Lemström and Geraint Wiggins on developing multi-dimensional point-set pattern matching and pattern discovery algorithms for polyphonic music information retrieval (Lemström, Wiggins, and Meredith, 2001; Meredith, Wiggins, and Lemström, 2001; Meredith, Lemström, and Wiggins, 2002, 2003; Wiggins, Lemström, and Meredith, 2002). These algorithms were then further developed by Ukkonen, Lemström, and Mäkinen (2003) and Lubiw and Tanur (2004). The main problem with these algorithms is that even the fastest versions of the pattern discovery and matching algorithms are quadratic in the worst case. Within the past few months, I have worked with Raphaël Clifford and Manolis Christodoulakis at King's College, London, on improving the time complexities of these algorithms by using techniques including FFT, universal hashing and randomized projection.

Unfortunately, to be practical for use on large collections of music data, retrieval algorithms must work in sub-linear time, which involves the development of better indexing techniques. This provides a practical motivation for developing systems for analysing the motivic/thematic structure of musical works (see next section).

4 Automatic music analysis and modelling music cognition

A musicologist analyses a musical work in order to gain a deeper understanding of it. In particular, he or she attempts to "determine the structural elements and discover the functions of those elements" (Bent and Drabkin, 1987, p. 5). Although many different methods of analysis have been proposed, it is usual to begin the analysis of a tonal work by determining its harmonic and tonal structure: that is, by determining the sequence of keys and chords implied by the music. A number of algorithms have been proposed for carrying out harmonic analysis (e.g., Maxwell, 1992; Raphael and Stoddard, 2003; Temperley, 2001; Winograd, 1968, 1993). A harmonic analysis of a passage of tonal music is essentially a string of symbols. A multi-dimensional representation of a passage of polyphonic music can therefore potentially be reduced to a one-dimensional string by analysing its harmonic structure. Such strings would then be amenable to being processed using fast string algorithms. One possible application of harmonic analysis is therefore in music information retrieval.

Another well-established method of music analysis is motivic/thematic analysis (Nattiez, 1975; Ruwet, 1972; Schoenberg, 1967) in which the analyst identifies the perceptually significant repeated patterns in a work. Only a relatively small number of systems have been proposed for automatically determining the significant repeated patterns in a passage of polyphonic music (Meredith *et al.*, 2002). Unfortunately, the best of these systems works in quadratic time which is impractically slow for analysing large works such as symphonic movements. The results of a motivic/thematic analysis could be used to build indexes for music collections that could allow for retrieval in sub-linear time. However, no algorithm yet exists that can effectively analyse the motivic/thematic structure of a large polyphonic musical work in a practical running time. Another of my research goals is therefore to develop such an algorithm, possibly by specializing the algorithms described by Meredith *et al.* (2002).

References

Abdallah, S. A. and Plumbley, M. D. (2004). Polyphonic transcription by non-negative sparse coding of power spectra. In *Proceedings of the Fifth International Conference on Music Information Retrieval (ISMIR 2004)*, Universitat Pompeu Fabra, Barcelona, Spain.

Bent, I. and Drabkin, W. (1987). *Analysis*. New Grove Handbooks in Music. Macmillan.

Cambouropoulos, E. (1996). A general pitch interval representation: Theory and applications. *Journal of New Music Research*, **25**(3), 231–251.

Cambouropoulos, E. (1998). *Towards a General Computational Theory of Musical Structure*. Ph.D. thesis, University of Edinburgh.

Cambouropoulos, E. (2000). From MIDI to traditional musical notation. In *Proceedings of the AAAI 2000 Workshop on Artificial Intelligence and Music: Towards Formal Models for Composition, Performance and Analysis 17th National Conference on Artificial Intelligence (AAAI'2000), 30 July–3 August 2000*, Austin, TX. Available online at `<http://users.auth.gr/~emilios/papers/aaai2000.pdf>`.

Cambouropoulos, E. (2001). Automatic pitch spelling: From numbers to sharps and flats. In *VIII Brazilian Symposium on Computer Music (SBC&M 2001)*, Fortaleza, Brazil.

Cambouropoulos, E. (2003). Pitch spelling: A computational model. *Music Perception*, **20**(4), 411–429.

Chew, E. and Chen, Y.-C. (2003a). Determining context-defining windows: Pitch spelling using the spiral array. In *Fourth International Conference on Music Information Retrieval (ISMIR 2003) (October 26–30)*, Baltimore, MD.

Chew, E. and Chen, Y.-C. (2003b). Mapping MIDI to the Spiral Array: Disambiguating Pitch Spellings. In H. K. Bhargava and N. Ye, editors, *Computational Modeling and Problem Solving in the Networked World - Proceedings of the 8th INFORMS Computer Society Conference*, pages 259–275, Chandler, AZ. Kluwer.

Chew, E. and Chen, Y.-C. (2005). Real-time pitch spelling using the Spiral Array. *Computer Music Journal*, **29**(2), 61–76.

Chew, E. and Wu, X. (2004). Separating voices in polyphonic music: A contig mapping approach. In *Second International Symposium on Computer Music Modeling and Retrieval (CMMR 2004) (May 26–29, 2004)*, Aalborg University Esbjerg, Esbjerg, Denmark.

Chuan, C.-H. and Chew, E. (2005). Fuzzy analysis in pitch-class determination for polyphonic audio key finding. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005)*, pages 296–303, Queen Mary, University of London.

Davy, M. and Godsill, S. J. (2003). Bayesian harmonic models for musical signal analysis (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume VII. Oxford University Press.

Desain, P. and Honing, H. (1992). The quantization problem: traditional and connectionist approaches. In M. Balaban, K. Ebcioğlu, and O. Laske, editors, *Understanding Music with AI: Perspectives on Music Cognition*, pages 448–463. M.I.T. Press.

Holtzmann, S. R. (1977). A program for key determination. *Interface*, **6**, 26–56.

Kilian, J. and Hoos, H. H. (2002). Voice separation—a local optimisation approach. In *Proceedings of the Third International Conference on Music Information Retrieval, ISMIR 2002, IRCAM, Centre Pompidou, Paris (October 13–17, 2002)*, pages 39–46. Available online at `<http://ismir2002.ismir.net/proceedings/02-FP01-6.pdf>`.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*, volume 17 of *Oxford Psychology Series*. Oxford University Press, New York and Oxford.

Lemström, K., Wiggins, G. A., and Meredith, D. (2001). A three-layer approach for music retrieval in large databases. In *Second Annual International Symposium on Music Information Retrieval (ISMIR 2001) (October 15–17)*, pages 13–14, Indiana University, Bloomington, Indiana. Poster and paper available online at `<http://www.titanmusic.com/papers.html>`.

Longuet-Higgins, H. C. (1987). The perception of melodies. In H. C. Longuet-Higgins, editor, *Mental Processes: Studies in Cognitive Science*, pages 105–129. British Psychological Society/MIT Press, London, England and Cambridge, Mass.

Lubiw, A. and Tanur, L. (2004). Pattern marching in polyphonic music as a weighted geometric translation problem. In *Fifth International Conference on Music Information Retrieval, Barcelona (ISMIR 2004, 10–15 October 2004)*. Available online at `<http://ismir2004.ismir.net/proceedings/p054-page-289-paper154.pdf>`.

Maxwell, H. J. (1992). An expert system for harmonic analysis of tonal music. In M. Balaban, K. Ebcioğlu, and O. Laske, editors, *Understanding Music with AI: Perspectives on Music Cognition*, pages 335–353. AAAI Press/MIT Press, Cambridge, MA.

Meredith, D. and Wiggins, G. A. (2005). Comparing pitch spelling algorithms. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005)*, pages 280–287, Queen Mary, University of London. Available online at `<http://ismir2005.ismir.net/proceedings/1004.pdf>`.

Meredith, D., Wiggins, G. A., and Lemström, K. (2001). Pattern induction and matching in polyphonic music and other multidimensional datasets. In N. Callaos, X. Zong, C. Vergez, and J. R. Pelaez, editors, *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI2001), July 22-25*, volume X, pages 61–66, Orlando, FL. Paper and slides available online at `<http://www.titanmusic.com/papers.html>`.

Meredith, D., Lemström, K., and Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, **31**(4), 321–345. Available online at `<http://taylorandfrancis.metapress.com/link.asp?id=yql23xw0177lt4jd>`.

Meredith, D., Lemström, K., and Wiggins, G. A. (2003). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. In *Cambridge Music Processing Colloquium (28 March 2003)*, Cambridge University Engineering Department. Available online at `<http://www.titanmusic.com/papers/public/cmpc2003.pdf>`.

Meredith, D. (2003). Pitch spelling algorithms. In R. Kopiez, A. C. Lehmann, I. Wolther, and C. Wolf, editors, *Proceedings of the Fifth Triennial ESCOM Conference (ESCOM5) (8-13 September 2003)*, pages pp. 204–207, Hanover University of Music and Drama, Hanover, Germany. Available online at `<http://www.epos.uos.de/music/books/k/klww003/pdfs/080_Meredith_Proc.pdf>`.

Meredith, D. (2005). Comparing pitch spelling algorithms on a large corpus of tonal music. In U. K. Wiil, editor, *Computer Music Modeling and Retrieval, Second International Symposium, CMMR 2004, Esbjerg, Denmark, May 26–29, 2004, Revised Papers*, volume Vol. 3310 of *Lecture Notes in Computer Science (LNCS)*, pages 173–192, Berlin. Springer. Available online at `<http://www.springerlink.com/link.asp?id=b3vd6fpumrmpkqww>`. (Draft and slides available at `<http://www.titanmusic.com/papers.html>`).

Meredith, D. (2006). Computing Pitch Names in Tonal Music: An Analysis and Comparison of Pitch Spelling Algorithms. D.Phil. dissertation, Faculty of Music, University of Oxford. Available online at `<http://www.titanmusic.com/papers/private/meredith-dphil.pdf>`. To obtain username and password, send request to `<dave@titanmusic.com>`.

Nattiez, J.-J. (1975). *Fondements d'une sémiologie de la musique*. Union Générale d'Éditions, Paris.

Plumbley, M., Abdallah, S., Bello, J., Davies, M. E., Monti, G., and Sandler, M. (2002). Automatic music transcription and audio source separation. *Cybernetics and Systems*, **33**(6), 603–627.

Raphael, C. and Stoddard, J. (2003). Harmonic analysis with probabilistic graphical models. In *Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, Maryland, USA. Available online at `<http://ismir2003.ismir.net/papers/Raphael.PDF>`.

Ruwet, N. (1972). *Langage, Musique, Poésie*. Éditions du seuil, 27, rue Jacob, Paris VI.

Schoenberg, A. (1967). *Fundamentals of Musical Composition*. Faber and Faber, London. Edited by Gerald Stand with the collaboration of Leonard Stein.

Shmulevich, I., Yli-Harja, O., Coyle, E., Povel, D.-J., and Lemström, K. (1999). Perceptual issues in music pattern recognition—complexity of rhythm and key finding. In *Proceedings of the AISB'99 Symposium on Musical Creativity*, pages 64–69.

Temperley, D. (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, **17**(1), 65–100.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA.

Temperley, D. (2004). An evaluation system for metrical models. *Computer Music Journal*, **28**(3), 28–44.

Ukkonen, E., Lemström, K., and Mäkinen, V. (2003). Sweepline the music! In R. Klein, H.-W. Six, and L. Wegner, editors, *Computer Science in Perspective (LNCS 2598)*, pages 330–342. Springer.

Vos, P. G. and van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception*, **14**, 185–224.

Walmsley, P. J. (2000). *Signal Separation of Musical Instruments*. Ph.D. thesis, Signal Processing Group, Department of Engineering, University of Cambridge.

Wiggins, G. A., Lemström, K., and Meredith, D. (2002). SIA(M)ESE: An algorithm for transposition invariant, polyphonic, content-based music retrieval. In *3rd International Symposium on Music Information Retrieval (ISMIR 2002), 13–17 September 2002*, pages 283–284, IRCAM, Centre Pompidou, Paris, France. Available online at `<http://ismir2002.ismir.net/proceedings/03-SP03-4.pdf>`.

Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory*, **12**, 2–49. Republished as Winograd (1993).

Winograd, T. (1993). Linguistics and the computer analysis of tonal harmony. In S. M. Schwanauer and D. A. Levitt, editors, *Machine Models of Music*, pages 113–153. M.I.T. Press, Cambridge, Mass. Published earlier as Winograd (1968).