In T. Crawford & L. Gibson (Eds.), Modern Methods for Musicology: Prospects, Proposals, and Realities (pp. 93–113). London: Ashgate.

Chapter 7
On the Use of Computational Methods for Expressive Music Performance

Werner Goebl and Gerhard Widmer

1. Introduction

The general availability of more and more powerful computers over the past two decades and the advent of a standardized symbolic communication protocol between music instruments (MIDI) have led to a downright boom in quantitative research on expressive music performance.[1] The number of papers exploring the various aspects of (predominantly piano) performance even increased towards the millennium.[2] This trend towards the constitution of an entire new field is also reflected in comprehensive compilations of pertinent studies.[3] In order to bridge the gap between theoretical and practical approaches and to connect knowledge from science and music practice, Parncutt and McPherson[4] collected a wide range of contributions on various aspects of music teaching, learning and performance, each written jointly by a researcher and an active musician. Similarly, the contributions collected by Williamon,[5] which deal with the various aspects of achieving excellence in music performance through refined techniques of practice, strongly refer to the current scientific literature.

Parallel to these advances in more 'conventional' music research, extensive work on computational modelling of expressive music performance has been

1 R. Kopiez, 'Interpretationsforschung mit Hilfe des Computerflügels. Eine Studie zur Wahrnehmung von Interpretationsmerkmalen [Interpretation Research with the Help of Computer Pianos: An Analysis of the Perception of Features of Music Interpretation]', in K.E. Behne, G. Kleinen and H. d. la MotteHaber (eds), *Musikpsychologie. Empirische Forschungen, ästhetische Experimente*, vol. 10, *Jahrbuch der Deutschen Gesellschaft für Musikpsychologie* (Wilhelmshaven: Noetzel, 1994), pp. 7–23; C. Palmer, 'Music Performance', *Annual Review of Psychology*, 48 (1997): 115–38; A. Gabrielsson, 'Music Performance Research at the Millennium', *Psychology of Music*, 31/3 (2003): 221–72.

2 Ibid.

3 J. Rink (ed.), *The Practice of Performance: Studies in Musical Interpretation* (Cambridge: Cambridge University Press, 1995); J. Rink (ed.), *Musical Performance. A Guide to Understanding* (Cambridge: Cambridge University Press, 2002).

4 R. Parncutt and G. McPherson (eds), *The Science and Psychology of Music Performance. Creating Strategies for Teaching and Learning* (New York: Oxford University Press, 2002).

5 A. Williamon (ed.), *Musical Excellence: Strategies and Techniques to Enhance Performance* (Oxford: Oxford University Press, 2004).

carried out, with partly astonishing results – for example, the automatic discovery of fundamental performance principles and rules, computer recognition of individual performers, and even artificial performances rendered by computer.[6] Another recent development in the scientific landscape that may help the study of music performance in the future, is the advent of the research field of 'music information retrieval', which has a strong focus on new methods for intelligent music and audio analysis.

As measured data on music performance quickly reaches enormous dimensions (just a single Mozart sonata contains around 8,000 notes, each associated with performance properties such as onset, offset and loudness, to name only the most obvious), the use of computers for data processing is virtually inevitable. Measuring, managing and making sense of these data requires several processing steps (e.g. onset detection, score-performance alignment, error correction). For each of these steps various computational solutions have been proposed in the literature; for some of them even freely available software has been provided on the internet. In the following, we report on the state of the art of such computational methods in order to provide an overview for both the musicologist with a strong technical interest, and the technical developer with an interest in helping the former. Naturally, a certain emphasis will be given to work that has been performed in recent years by our music groups in Vienna and Linz.[7]

For the purpose of this chapter, we define the term 'computational method' to include computational approaches to retrieving data from recorded performances as well as tools for the abstract display, visualization and automatic analysis of such data. However, we exclude studies from our survey that, for example, simply use a waveform display to manually read off onset times of certain tones, not because such studies might not be highly valuable for research, but because our focus here is on more advanced and autonomous computational methods.

Despite some excellent attempts to provide introductory texts for empirical research in classical musicology[8] and to bridge the gap between computational research and music education,[9] it is still not common to apply advanced tools from information technology in everyday musicological research or music education. We want this chapter to be helpful in this respect by describing current technologies and proposing ways of using them in real musical applications. We will set out to describe computational means for establishing access to music performance

6 For exhaustive overviews, see G. Widmer and W. Goebl, 'Computational Models of Expressive Music Performance: The State of the Art', *Journal of New Music Research*, 33/3 (2004): 203–16; G. De Poli, 'Methodologies for Expressiveness Modelling of and for Music Performance', *Journal of New Music Research*, 33/3 (2004): 189–202.

7 The first author was formerly affiliated with the Austrian Research Institute for Artificial Intelligence in Vienna.

8 E.F. Clarke and N. Cook, *Empirical Musicology: Aims, Methods, and Prospects* (Oxford: Oxford University Press, 2004).

9 Parncutt and McPherson (eds), *The Science and Psychology of Music Performance*.

data mainly through audio recordings (section 2). In order to understand (and eventually to interact with) performance data, it is indispensable to think about ways to display and visualize data that are intuitive and informative at the same time (section 3). Finally, section 4 will report on recent developments in the computational modelling of music performance.

2. Access to music performance

Expressive music performance is the process of singing or playing a musical instrument in order to create musical output. This process can be captured in multiple ways: the most common is to record its acoustic outcome with microphones; the recordings can be stored on different media such as tapes or computer hard disk. The digitized audio data are a good representation of the acoustical signal, but in order to understand the properties of the performance such as tempo variations, dynamics or articulation, these data have to be processed. A common first step is to distinguish entities such as notes or chords from the audio stream, determine the times of their onsets and offsets, and estimate their loudness values. Several computer programs have been developed to support such operations. We will discuss some of them in section 2.1 below.

Another common way of capturing the process of music performance is to record the movement of the parts of the musical instrument that are involved in tone production. A good example here is the action of a piano as a mechanical interface between the musician and tone production. Computer-monitored pianos that measure the speed and the timing of the piano hammers and output this information in a symbolic format (usually MIDI) have been developed exclusively for research purposes, such as the 'Iowa Piano Camera'[10] or Henry Shaffer's Photocell Bechstein.[11] Today, instruments such as the Disklavier by Yamaha or the computer-controlled pianos by Bösedorfer[12] are commercially available, not to mention the many digital pianos and synthesizer keyboards that are frequently used as well.

We called expressive music performance a 'process of singing or playing a musical instrument', a definition that logically entails the involvement of a human individual who produces all sorts of movements during performance that can be the

10 C.E. Seashore (ed.), *Objective Analysis of Musical Performance*, *University of Iowa Studies in the Psychology of Music*, vol. 4 (Iowa City: Iowa University Press, 1936).

11 See L.H. Shaffer, 'Analysing Piano Performance', in G.E. Stelmach and J. Requin (eds), *Tutorials in Motor Behavior* (Amsterdam: North Holland, 1980); L.H. Shaffer, 'Performances of Chopin, Bach and Bartòk: Studies in Motor Programming', *Cognitive Psychology*, 13/3 (1981): 326–76.

12 W. Goebl and R. Bresin, 'Measurement and Reproduction Accuracy of Computer Controlled Grand Pianos', *Journal of the Acoustical Society of America*, 114/4 (2003): 2273–83.

subject of research.[13] These movements may contain large amounts of expressivity that can even override the acoustic information.[14] Thus, they should be regarded as an integral part of a music performance. In order to capture performers' movements, other monitoring techniques are required. They range from conventional video cameras or webcams[15] to complex three-dimensional motion capture systems.[16] While movement analysis will more and more become an integral part of music performance research, the present chapter will focus on performance analysis at a symbolic and acoustic level.

In the following, we describe ways of automatically or at least semi-automatically retrieving performance data from audio recordings (annotation), and ways of relating one performance to a score or multiple performances to each other (alignment). At this point we want to refer the reader to an outstanding introductory chapter on empirical methods for studying music performance[17] that discusses basic principles of data measurement and display as well as problems and shortcomings of quantitative as compared to qualitative methods.

2.1 Annotation

By music annotation, we understand the process of retrieving performance data from audio recordings or labelling audio files with content-based metadata. In the domain of music information retrieval, annotation refers not only to expressive performance data such as tone onsets, offsets or loudness, but also to more general aspects of transcription such as harmony or instrumentation.

Any automated system for music transcription or beat-tracking will produce errors, even if only a small number. However, to use performance data for analysis, one requires data that is completely correct. A common solution is to provide a (graphical) user interface that allows the user to go through the audio recordings

---

13 J.W. Davidson and J.S. Correia, 'Body Movement', in R. Parncutt and G. McPherson (eds), *The Science and Psychology of Music Performance: Creating Strategies for Teaching and Learning* (Oxford: Oxford University Press, 2002), pp. 237–50.

14 K.E. Behne, '"Blicken Sie auf die Pianisten?!" Zur bildbeeinflußten Beurteilung von Klaviermusik im Fernsehen', *Medienpsychologie*, 2/2 (1990): 115–31; J.W. Davidson, 'What Type of Information is Conveyed in the Body Movements of Solo Musician Performers?', *Journal of Human Movement Studies*, 26/6 (1994): 279–301.

15 A. Camurri, G. Volpe G. De Poli and M. Leman, 'Communicating Expressiveness and Affect in Multimodal Interactive Systems', *IEEE Multimedia*, 12/1 (2005): 43–53.

16 See, for example, M.M. Wanderley, B. Vines, N. Middleton, C. McKay and W. Hatch, 'The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field', *Journal of New Music Research*, 34/1 (2005): 97–11; W. Goebl and C. Palmer, 'Anticipatory Motion in Piano Performance', *Journal of the Acoustical Society of America*, 120/5 (2006): 3002.

17 E.F. Clarke, 'Empirical Methods in the Study of Performance', in E.F. Clarke and N. Cook (eds), *Empirical Musicology: Aims, Methods, and Prospects* (Oxford: Oxford University Press, 2004), pp. 77–102.

and check for possible errors. Recently, a number of systems have been proposed that support the semi-automatic gathering of performance data from musical audio signals.

A first system that provided a graphical front-end to check and manipulate the output of a sophisticated beat-tracking algorithm was BeatRoot.[18] It allows the user to derive the onset times of 'beats' relatively quickly from any type of musical audio file.[19] BeatRoot gives the user graphical and aural feedback concerning the placement of each beat (markers in the waveform on the screen, and click sounds in parallel with audio playback). When detecting errors, the user can correct them and re-run the beat-tracking algorithm from that point. The system then updates its beat hypothesis, taking into account the corrected information. By repeating these steps iteratively, the user can go through an audio file quite quickly. The determined beat times can be exported for further analysis in a text-based MIDI file format that can easily be imported into other software packages. The beat level at which a given piece is tracked can be chosen by the user. With this procedure, it is possible to obtain timing data from very regularly timed music (e.g. pop or jazz) virtually automatically, and from classical music in a fairly short time. It has already proved useful in several studies of music performance.[20] Gouyon et al.[21] implemented a subset of BeatRoot as a plugin into the free audio editor WaveSurfer,[22] which was originally designed for speech signals. BeatRoot is available in the JAVA

18 S. Dixon, 'An Interactive Beat Tracking and Visualisation System', in A. Schloss, R. Dannenberg and P. Driessen (eds), *Proceedings of the 2001 International Computer Music Conference* (San Francisco: International Computer Music Association, 2001a), 215–18; S. Dixon, 'Automatic Extraction of Tempo and Beat from Expressive Performances', *Journal of New Music Research* 30/1 (2001b): 39–58.

19 A comprehensive description of this system can be found in G. Widmer, S. Dixon, W. Goebl, E. Pampalk, and A. Tobudic, 'In Search of the Horowitz Factor', *AI Magazine*, 24/3 (2003): 111–30.

20 S. Dixon, W. Goebl and G. Widmer, 'Real Time Tracking and Visualisation of Musical Expression', in C. Anagnostopoulou, M. Ferrand, and A. Smaill (eds), *Proceedings of the Second International Conference on Music and Artificial Intelligence* (Berlin: Springer, 2002), Lecture Notes in Artificial Intelligence 2445, 58–68; W. Goebl, E. Pampalk and G. Widmer, 'Exploring Expressive Performance Trajectories: Six Famous Pianists Play Six Chopin Pieces', in S.D. Lipscomp, R. Ashley, R.O. Gjerdingen and P. Webster (eds), *Proceedings of the 8th International Conference on Music Perception and Cognition*, CD-ROM (Adelaide: Causal Productions, 2004): 505–509; E. Stamatatos and G. Widmer, 'Automatic Identification of Music Performers with Learning Ensembles', *Artificial Intelligence*, 165/1 (2005): 37–56.

21 F. Gouyon, N. Wack and S. Dixon, 'An Open Source Tool for Semiautomatic Rhythmic Annotation', in *Proceedings of the 7th International Conference on Digital Audio Effects* (Naples, 2004), pp. 193–6.

22 K. Sjölander and J. Beskow, 'WaveSurfer – An Open Source Speech Tool', in B. Yuan, T. Huang, and X. Tang (eds), *Proceedings of the International Conference on Spoken Language Processing* (Bejing, 2000), pp. 464–7.

1.5 programming language, which runs on multiple platforms;[23] it comes with an improved onset detection algorithm[24] and can be freely downloaded.[25] BeatRoot won the annual MIREX 2006 contest on audio beat-tracking.[26]

A software framework for annotation of musical audio signals is CLAM (C++ Library for Audio and Music), which contains an annotator tool that allows the user to modify any pre-processed low-level frame descriptor.[27] Low-level descriptors provided in the sample files are signal energy, centroid, flatness and the like, as well as higher-level descriptors referring to harmony (chord, note) and structure (chorus, verse, etc.). CLAM reads and writes its descriptors in standard XML text files so that a simple import of an audio file beat-tracked by BeatRoot would be easy to accomplish. At the time of writing this chapter (January 2007, CLAM 0.97), there was no tool for beat-tracking or transcription provided in the CLAM framework. However, due to its modular concept and its input and output in standard XML, any custom-made descriptor could be loaded and manipulated in this annotator tool. CLAM is available for multiple platforms at <http://clam.iua.upf.edu/> (accessed 5 September 2007).

Another advanced and flexible tool for annotating musical signals is the Sonic Visualiser[28] developed at the Centre for Digital Music at Queen Mary, University of London. It employs a multi-layer architecture that is designed to stack multiple analyses of the audio signal on top of each other (just as Adobe's Photoshop does for image processing) so that multiple views that are synchronized in time produce a comprehensive image of the audio signal for analysis and understanding. The graphical user interface provides rich facilities for visual data display (waveform and spectral representations), annotation (e.g. a time instants layer), as well as aural display of selected features. All displayed annotation layers can be manipulated, saved to and loaded from standard XML.

In order to expand the features of the Sonic Visualiser, custom-made plugins can be loaded and programmed (Vamp Plugins). At the time of writing, the authors

---

23 Dixon, S., 'MIREX 2006 Audio Beat Tracking Evaluation: BeatRoot' (2006). Available online at <http://www. musicir.org/mirex2006>, accessed 5 September 2007.

24 S. Dixon, 'Onset Detection Revisited', in *9th Int. Conference on Digital Audio Effects* (Montreal, 2006), pp. 133–7.

25 <http://www.elec.qmul.ac.uk/people/simond/beatroot/>, accessed 5 September 2007.

26 <http://www.music-ir.org/mirex2006/index.php/Main_Page>, accessed 5 September 2007.

27 X. Amatriain, J. Massaguer, D. Garcia and I. Mosquera, 'The CLAM Annotator: A Crossplatform Audio Descriptors Editing Tool', in 6*th International Conference on Music Information Retrieval* (London, UK, 2005), pp. 426–9.

28 C. Cannam, C. Landone, M. Sandler and J.P. Bello, 'The Sonic Visualiser: A Visualisation Platform for Semantic Descriptors of Musical Signals', in *Proceedings of 7th International Conference on Music Information Retrieval* (Victoria, 2006).

found plugins for beat-tracking,[29] onset detection, and various spectral processors (from the Mazurka Project web page[30]). The beat-tracker did not perform well on the tested slow-paced classical pieces (like most other beat-trackers); however, since there is no re-track function built in, as in BeatRoot, getting a piece beat-tracked takes basically as long as a purely manual annotation. The onset detection plugin from the Mazurka Project produced very satisfying results although it missed some of the onsets, but not necessarily soft ones in general. Like the other packages described, Sonic Visualiser runs on multiple computer platforms and is freely available online.[31]

2.2 Alignment

Unlike a human music listener, none of the computational tools described above has any deeper knowledge of the music that it processes. A way to enhance the performance of beat-trackers and other feature extractors would be to give them access to, for example, symbolic score information. Such an approach would involve some sort of score-to-performance alignment procedure that – if working in real time – could be used for automatic accompaniment.

An alternative approach would be to use existing knowledge (e.g. previously annotated audio files) and to transfer these metadata to other performances of the same piece. A system that matches two audio files on to each other is MATCH (Music Alignment Tool CHest[32]). It finds the optimal alignment between pairs of recordings of the same piece of music and provides a time-warping function that has a pointer usually every 20 milliseconds. MATCH compares the recordings frame by frame, based on a spectral similarity measure, and computes the warping function with a dynamic programming algorithm that operates in linear time (as opposed to other algorithms whose processing time grows quadratically with the length of the files to be processed[33]).

An example of four automatically aligned performances can be seen in Figure 7.1. The multiple lines between the individual performances indicate the time-warping relations as output by MATCH (for the sake of clarity, a line is plotted only every 100 milliseconds). We invite the reader to compare this automatically produced output with the manually annotated 'ground truth' as marked by the

29 M. Davies and M. Plumbley, 'Beat Tracking with a Two State Model', in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Philadelphia, 2005), vol. 3, 241–4.

30 <http://www.mazurka.org.uk/>, accessed 5 September 2007.

31 <http://www.sonicvisualiser.org/> (accessed 5 September 2007).

32 S. Dixon and G. Widmer, 'MATCH: A Music Alignment Tool Chest', in *6th International Conference on Music Information Retrieval* (London, 2005), pp. 492–7.

33 See S. Dixon, 'An Online Time Warping Algorithm for Tracking Musical Performances', in *International Joint Conference on Artificial Intelligence* (Edinburgh, 2005a), pp. 1727–8.

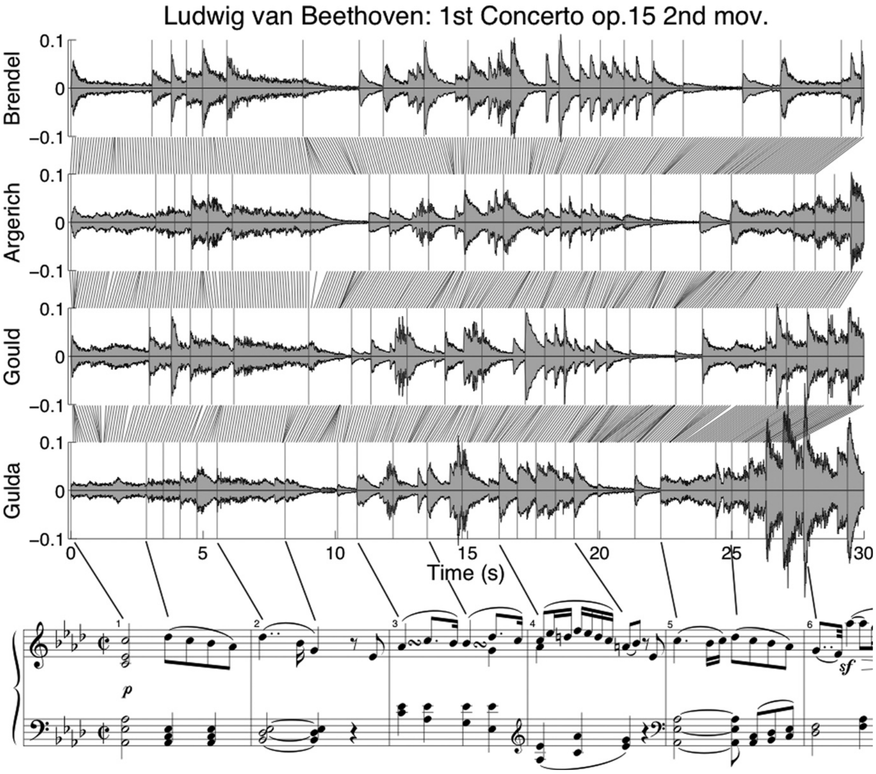

Figure 7.1 The waveforms show the first 30 seconds of four performances of four performances of Beethoven's first piano concerto Op. 15.

Note: The lines superimposed on the waveforms indicate manually annotated onsets at the eighth note level quasi as a 'ground truth'. The solid lines between the panels show the time-warping functions between two neighbouring performances as automatically tracked by MATCH. In this figure, only warping lines every 100 ms are plotted, although the default window size of MATCH is 20 cms. Please compare the manually annotated onsets with the output of the algorithm that does not know anything about music. (As we plotted the same amount of time for each performance, a different relative to the scorer is shown due to different performance tempi. We indicated the link between Gulda's performance and the music score shown by a line every half bar.

solid lines superimposed on the waveforms. Although it gives slightly incorrect estimates at some points, it performs astonishingly well given the fact that neither onset detection nor any other higher-level processing or knowledge is involved.

Given this, a procedure that might be more efficient than direct beat-tracking or annotation, might be to analyse one recording with the tools described above and then use MATCH to project these annotations automatically on to an arbitrary number of other recordings of the same piece. The output can then be reviewed and – if necessary – corrected with the software tools mentioned above. Another application that can be realized with MATCH is score-performance alignment in real time. The user would have to synthesize a score into an audio file which would then be aligned with the (live) input of a microphone listening to a performance of that score.[34] That could be used for real-time accompaniment and similar applications. MATCH runs on multiple platforms and can be freely downloaded on the internet.[35]

In some ways technically similar to the online version of MATCH, several automatic accompaniment systems have been proposed[36] that work on symbolic (MIDI) data as well as directly on audio.[37] They have been either commercialized or are not available from the researchers' homepages, so their application for performance research could not be directly assessed by the authors.

Other common demands from music researchers involve matching recorded expressive data to a score at a symbolic (usually MIDI) level. This problem has been addressed several times;[38] however, to date no conveniently working software has been provided that would offer symbolic score-performance matching interactively combined with a user interface for error correction.[39]

34 S. Dixon, 'Live Tracking of Musical Performances Using Online Time Warping', in *8th International Conference on Digital Audio Effects* (Madrid, 2005b), pp. 92–7.

35 <http://www.elec.qmul.ac.uk/people/simond/match/index.html> (accessed 5 September 2007).

36 R. Dannenberg, 'An Online Algorithm for Real-time Accompaniment', in *Proceedings of the 1984 International Computer Music Conference* (San Francisco: International Computer Music Association, 1984), pp. 193–8; N. Orio and F. Déchelle, 'Score Following Using Spectral Analysis and Hidden Markov Models', in *2001 International Computer Music Conference* (Havana: International Computer Music Association, 2001), pp. 151–4; C. Raphael, 'A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores', in *Proceedings of the 5th International Conference on Music Information Retrieval* (Barcelona, 2004), pp. 387–94.

37 C. Raphael, 'Aligning Music Audio with Symbolic Scores Using a Hybrid Graphical Model', *Machine Learning*, 65/2–3 (2006): 389–409.

38 H. Heijink, L. Windsor, and P. Desain, 'Data Processing in Music Performance Research: Using Structural Information to Improve Score-performance Matching', *Behavior Research Methods, Instruments and Computers*, 32/4 (2000): 546–54; E.W. Large, 'Dynamic Programming for the Analysis of Serial Behaviors', *Behavior Research Methods, Instruments and Computers*, 25/2 (1993): 238–41; for an overview see H. Heijink, P. Desain, H. Honing and L. Windsor, 'Make Me a Match: An Evaluation of Different Approaches to Score-performance Matching', *Computer Music Journal* 24/1 (2000): 43–56.

39 Ed Large and colleagues have developed an interactive user interface to the system described in (Large, 'Dynamic Programming for the Analysis of Serial Behaviors'), but it is presently being beta tested and not yet available online. Also, the Department of Computational Perception at the University of Linz is currently developing such a tool, which should be made available online some time in 2007.

2. Performance visualization and interaction

After having retrieved and successfully post-processed (corrected) the performance data, one would want to see, explore and finally make sense of the collected information. In the following, we will present some ideas and techniques for visualizing music performance data with the help of computers. A crucial property of music performance is that it evolves over time. Therefore, one requirement of a visualization is that it should reflect or recreate this temporal process.

One such technique, developed and presented by Langner and Goebl,[40] combines the two most important performance parameters into one animated display. Instant tempo and loudness are displayed in a two-dimensional space in which a little disc indicates the current state of a performance at a particular moment in time. With time, the disc moves inside this space according to tempo and loudness measurements of a performance, leaving behind a trace that fades with time. The trace can be viewed as a performance trajectory that is specific for a particular performance. For obvious reasons, this display has elsewhere also been called the 'Performance Worm'.[41] This display technique has proven to be an excellent tool for performance comparisons to lay-audiences and musicians alike. The two-dimensional trajectory representation has also been used for computational analysis of larger corpora of performance data,[42] as well as to characterize the individual style of performers.[43]

To illustrate, we plot the performance trajectories of the first four bars of Alfred Brendel's and Glenn Gould's performances of the second movement of Beethoven's first piano concerto, Op. 15 (Figure 7.2; for the score, see Figure 7.1). Both trajectories are stopped at the beginning of bar 5 (cf. the number inside the circle) for comparison purposes. The two black discs within the red tail indicate the phrase boundaries at bars 5 and 3. While Brendel shapes the first phrase (the first two bars) in a typical way by speeding up and slowing down in combination with a crescendo–decrescendo pattern, Gould does just the opposite: he slows down in the initial eights notes and shortens the break before the second phrase. In the second part of this excerpt, our two pianists agree more in their interpretations: they both employ an up-down pattern in their dynamics, while monotonically slowing down – probably in order to shape the turns more carefully.

Of course, the advantage that we claimed for the described display – its account of the temporal nature of performance via animation – is not visible in

40 J. Langner and W. Goebl, 'Visualizing Expressive Performance in Tempo–Loudness Space', *Computer Music Journal*, 27/4 (2003): 69–83.

41 S. Dixon, W. Goebl and G. Widmer, 'The Performance Worm: Real Time Visualisation Based on Langner's Representation', in M. Nordahl (ed.), *Proceedings of the 2002 International Computer Music Conference* (San Francisco: International Computer Music Association, 2002), pp. 361–4; Widmer et al., 'In Search of the Horowitz Factor'.

42 Goebl et al.'Exploring Expressive Performance Trajectories'.

43 Widmer et al., 'In Search of the Horowitz Factor'.

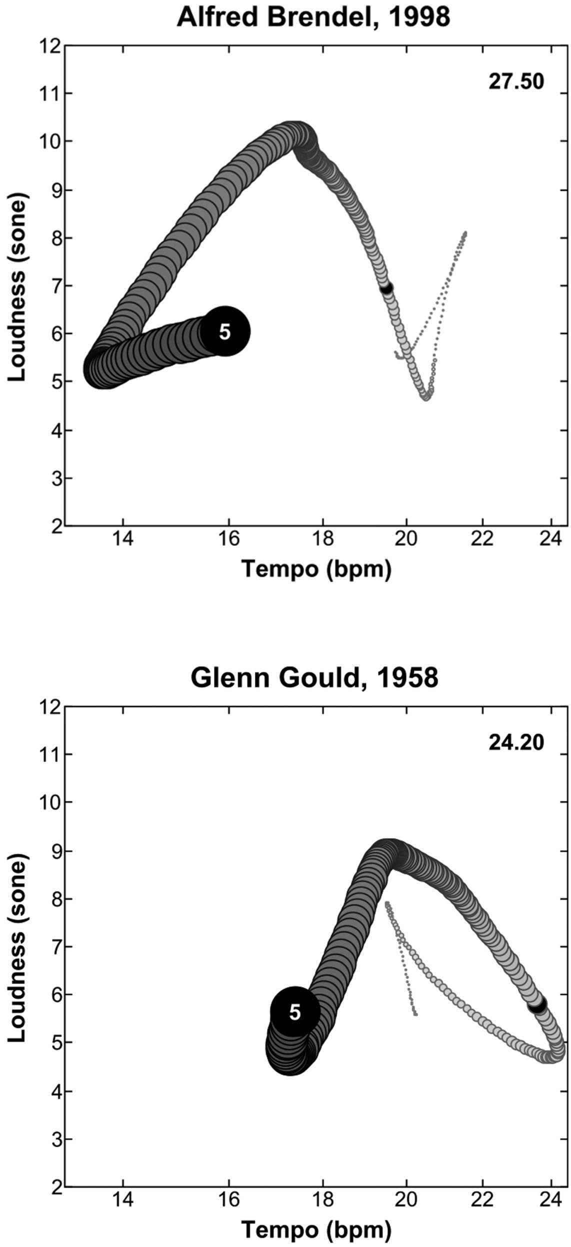

Figure 7.2 Tempo–loudness trajectories of the first four bars of the second movement of Beethoven's Op. 15 as played by (a) Alfred Brendel and (b) Glenn Gould. The black discs indicate the beginning of bar 3 (smaller) and bar 5

this static screen shot. Animated movies of this display can be downloaded from the internet.[44] Limitations of this way of showing performance are the loss of detail due to data smoothing and the absence of performance measures other than expressive tempo and (overall) loudness.

This tempo–loudness space has also been used to control expressive performance – a ‘reversed worm’, as it were. The user manipulates the current position of the worm either with a computer mouse, or – more intuitively – by moving a hand within the two antennae of a computer-monitored theremin, and the computer software shapes the performance accordingly in real time.[45] Since this interface is only able to control overall tempo and loudness, the template performances to be used for playing with the system are ‘flattened’ real performances with all the small-scale, local expression (chord asynchronies, dynamics of individual voices, etc.) still in them. Thus, the (lay-) user becomes a conductor in their own right with the help of computer software.

Appropriate and possibly interactive data visualization can be used in music education or for practising an instrument. There have been several attempts to provide feedback on expression in music performance. We refer the reader to a very elaborate educational system for the singing voice that is audio-based and involves fundamental pitch trackers,[46] and also to the chapter by David Howard in this volume. For piano performances, the use of MIDI-based pianos makes data collection easy. Such instruments are not as widely available as conventional computers equipped with microphones are. Nevertheless, there are reports by piano teachers that computer-monitored acoustic pianos combined with simple piano-roll displays were used successfully in piano instruction.[47] A current research initiative is carried out at the Piano Pedagogy Lab in Ottawa, Canada, where researchers are developing a software tool (the MIDIator)[48] that analyses and displays performance data relatively quickly after the performance has been recorded.[49] This research initiative is especially promising as it is situated in an explicitly pedagogical environment where the impact and usefulness of such technologies

44 <http://www.ofai.at/~werner.goebl/animations/>, accessed 5 September 2007.

45 S. Dixon, W. Goebl and G. Widmer, ‘The “Air Worm”: An Interface for Real-time Manipulation of Expressive Music Performance’, in *Proceedings of the 2005 International Computer Music Conference* (San Francisco: International Computer Music Association, 2005), pp. 614–17.

46 VOXed, G.F. Welch, E. Himonides, D.M. Howard and J. Brereton, ‘VOXed: Technology as a Meaningful Teaching Aid in the Singing Studio’, in R. Parncutt, A. Kessler and F. Zimmer (eds), *Conference on Interdisciplinary Musicology* (Graz: University of Graz, 2004).

47 E.g. K. Riley-Butler, ‘Teaching Expressivity: An Aural–Visual Feedback-replication Model’, in *ESCOM 10th Anniversary Conference on Musical Creativity* (Liège: Université de Liège, 2002), CD-ROM.

48 <http://www.piano.uottawa.ca/>, accessed 5 September 2007.

49 S. Shirmohammadi, A. Khanafar and G. Comeau, ‘MIDIator: A Tool for Analysing Students’ Piano Performance’, *Recherche en Éducation Musicale*, 24 (2006): 35–48.

can immediately be tested and evaluated in practice.[50] A similar path is taken by a European research consortium, which developed an interactive multimedia music tuition system[51] that is aimed specifically at beginners. However, just like other computational teaching or visualization approaches,[52] the MIDIator and IMUTUS work offline – that is, the data can only be viewed after the performance is completed.

A real-time approach was taken by Goebl and Widmer.[53] The proposed interfaces, called Practice Tools, are designed to display performance as it comes from a MIDI piano, and to keep running and adapting automatically to the performer's needs without explicit interaction. They feature an extended ('acoustic') piano-roll display that incorporates all aspects of the piano sound, including the pedals, tone decay and their interaction. In addition to this, separate displays are provided that show special sub-features of the performance. For instance, a chord display shows the relative timing and dynamic differences between the tones of a chord whenever a chord occurs. With this tool, pianists can practise to balance the sound of their chords or train to bring out individual voice more deliberately.

Also operating in real time is PracticeSpace, a tool for drummers interactively to improve the timing of complex rhythms.[54] First assessments demonstrate improved learning with this visualization system.[55]

With further advances in research on intelligent online music (audio) processing and the parallel growth in technological solutions, we may expect several of the approaches mentioned above to be combined into compact and lightweight software packages that will be available to everybody. In particular, improved audio processing algorithms will eliminate the need for costly computer-

50 G. Comeau, 'Recherche scientifique et pédagogie du piano', *Recherche en Éducation Musicale*, 24 (2006): 1–11.

51 IMUTUS, see S. Raptis, A. Askenfelt, D. Fober et al., 'IMUTUS – An Effective Practicing Environment for Music Tuition', in *2005 International Computer Music Conference* (Barcelona, 2005), pp. 383–6.

52 As, for example, S.W. Smoliar, J.A. Waterworth and P.R. Kellock, 'pianoFORTE: A System for Piano Education beyond Notation Literacy', in *Proceedings of the ACM International Conference on Multimedia, San Francisco* (New York: ACM Press, 1995), pp. 457–65; R. Hiraga, and N. Matsuda, 'Visualization of Music Performance as an Aid to Listener's Comprehension', in *Proceedings of the Working Conference on Advanced Visual Interfaces* (Gallipoli: ACM Press, 2004), pp. 103–106.

53 W. Goebl and G. Widmer, 'Unobtrusive Practice Tools for Pianists', in *Proceedings of the 9th International Conference on Music Perception and Cognition* (Bologna, 2006), pp. 209–14.

54 A. Brandmeyer, D. Hoppe, M. Sadakata, R. Timmers and P. Desain, 'PracticeSpace: A Platform for Real-time Visual Feedback in Music Instruction', in *Proceedings of the 9th International Conference on Music Perception and Cognition* (Bologna, 2006).

55 D. Hoppe, A. Brandmeyer, M. Sadakata, R. Timmers and P. Desain, 'The Effect of Realtime Visual Feedback on the Training of Expressive Performance Skills', in *Proceedings of the 9th International Conference on Music Perception and Cognition* (Bologna, 2006).

controlled musical instruments like the Disklavier. Readily and instantaneously available visualization tools will be crucial for the further spread and acceptance of advanced computational technology in music practice and research.

3. Modelling expressive performance

Data visualization can provide researchers with valuable insights, and practitioners (musicians) with useful feedback. But if the goal of music performance research is to build a thorough understanding of the phenomenon, quantitative analyses and studies must be performed, with verifiable results. Naturally, computers also play an important role in this.

An obvious first step is statistical analysis of the measurement data, which permits the researcher to verify or falsify various hypotheses about the data, and to ascertain the significance of the findings. However, in this chapter we are interested in more sophisticated uses of computers, viz. as the carriers or embodiments of computational models of (aspects of) expressive performance. As discussed by Widmer and Goebl,[56] the purpose of computational performance models is 'to specify precisely the physical parameters defining a performance (e.g. onset timing, inter-onset intervals, loudness levels, note durations, etc.), and to postulate (quasi)systematic relationships between certain properties of the musical score, the performance context, and an actual performance of a given piece' – in short, to represent a hypothesis about music performance in the form of an operational computer program. Computational models are predictive in the sense that they can be made to produce 'expressive' performances, which opens up new possibilities for directly testing hypotheses via listening and quantitative comparison between model predictions and real measured performances.

Widmer and Goebl[57] reference a substantial number of computational modelling approaches and describe four of them in more detail. In the context of the present chapter, we focus on two of these, in order to highlight two ways in which computers can play a central role in performance analysis, beyond pure statistical data analysis.

Example 1 – the system of performance rules developed over decades at KTH, Stockholm, by Johan Sundberg and co-workers – illustrates how computational models can be used actively to explore hypotheses via simulation and analysis. Example 2 – a massively data-driven, bottom-up approach based on artificial intelligence and machine learning developed at our own institute in Vienna – shows how the computer can play a much more active role in the research process, as an active, autonomous discoverer. Both of these projects have been published extensively in the literature. In this section, we merely want to highlight those

56 Widmer and Goebl, 'Computational Models of Expressive Music Performance'.

57 Ibid.

aspects that demonstrate the specific advantages of computer-based approaches to performance research.

The KTH rule system was developed at the Royal Institute of Technology (KTH) in Stockholm over more than twenty years of research, starting with Sundberg et al.[58] Over the years, it has been extended in many ways. A fairly comprehensive description is given in Friberg[59] and more recently in Friberg et al.[60] The KTH model consists of a set of performance rules, each of which predicts, or prescribes, a specific aspect of timing, dynamics or articulation based on the local musical context. For instance, one particular rule ('Duration Contrast') is concerned with modifying the duration ratio of two successive notes; another ('Harmonic Charge') with changing the dynamics of a note depending on the harmonic context, etc. All rules are parameterized with a varying number of parameters.

Here, the computer plays the role of an interpreter of performance rules. Rules – that is, individual, partial hypotheses about some aspect of performance – can be formulated and added to the system one by one, and their impact on performances generated by the model can be analysed by listening tests, or by comparing the computer performances to real ones. Information from these tests then feeds back into the modelling process, prompting the researchers to modify rules, change parameter settings, etc. This incremental, iterative modelling process has been termed analysis by synthesis by Sundberg and colleagues[61] and is made possible by the modular nature of the model: rules produce their effects independently of other rules in the system. While modularity offers practical advantages in the modelling process, the deeper question of whether it adequately reflects the nature of the phenomenon under study – the factors governing expressive performance decisions – is more difficult to answer.

Modularity does, however, have undeniable advantages when it comes to practical applications of such models. The KTH rules have been implemented in the program Director Musices[62] that comes with different predefined rule sets and parameter settings ('rule palettes') that are intended to model different basic emotions, for example fear, anger, happiness, sadness, tenderness, and

58 J. Sundberg, A. Askenfelt and L. Frydén, 'Musical Performance. A Synthesis-by-Rule Approach', *Computer Music Journal*, 7 (1983): 37–43.

59 A. Friberg, 'A Quantitative Rule System for Musical Performance', Doctoral dissertation, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 1995.

60 A. Friberg, R. Bresin and J. Sundberg, 'Overview of the KTH Rule System for Musical Performance', *Advances in Cognitive Psychology*, 2/2–3 (2006): 145–61.

61 J. Sundberg, L. Frydèn and A. Askenfelt, 'What Tells You the Player is Musical? An Analysis-by-Synthesis Study of Music Performance', in J. Sundberg (ed.), *Studies of Music Performance* (Stockholm: Royal Swedish Academy of Music, 1983b), vol. 39, 61–75.

62 <http://www.speech.kth.se/music/performance/download/>, accessed 5 September 2007.

solemnity.[63] Director Musices has recently been combined with an interactive real-time conducting system based on the Radio Baton,[64] which has been used in numerous public performances. More recently, Friberg[65] has provided a system that allows manipulation of the rule parameters in real time. There have also been efforts to combine Director Musices with computational research on expressive intentions performed at the University of Padova;[66] the resulting computer system, Expressive Director,[67] permits real-time control of music performance synthesis, in particular regarding expressive and emotional aspects.

The increasing interest of the KTH group in modelling emotional aspects of performance via the rule system has even led to practical, commercial applications: 'emotional' ring tones ('moodies') for mobile phones produced with the help of some expression rules[68] are now being sold by a Swedish company.

The second approach to performance modelling that we want to discuss briefly here takes the role of the computer one step further, from a modelling machine to an autonomous discovery machine. This is a new approach to performance research that was developed at the Austrian Research Institute for Artificial Intelligence in Vienna.[69] The basic idea is to start from large amounts of measurement data and to develop computer programs that autonomously discover significant regularities and patterns in the data, via machine-learning and data-mining techniques. In other words, the predictive performance model is built by the machine in such a way that it 'explains' the given data as well as possible. The potential advantages of this approach are that it is firmly rooted in large amounts of empirical data and that, in principle, the machine, free from human biases, may discover truly novel and unexpected things. On the other hand, it is not that straightforward to incorporate

63 R. Bresin and A. Friberg, 'Emotional Coloring of Computer-controlled Music Performances', *Computer Music Journal*, 24/4 (2000): 44–63.

64 M.V. Mathews, A. Friberg, G. Bennett et al., 'A Marriage of the Director Musices Program and the Conductor Program', in R. Bresin (ed.), *Proceedings of the Stockholm Music Acoustics Conference* (Stockholm: Department of Speech, Music, and Hearing, Royal Institute of Technology, 2003), vol. 1, pp. 13–15.

65 A. Friberg, 'pDM: An Expressive Sequencer with Realtime Control of the KTH Music Performance Rules', *Computer Music Journal*, 30/1 (2005): 37–48.

66 E.g. S. Canazza, G. De Poli, C. Drioli, A. Rod'a and A. Vidolin, 'An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance', *Journal of New Music Research*, 32/3 (2003a): 281–94.

67 S. Canazza, A. Rodá, P. Zanon and A. Friberg, 'Expressive Director: A System for the Realtime Control of Music Performance Synthesis', in R. Bresin (ed.), *Proceedings of the Stockholm Music Acoustics Conference*, 2nd edn (Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, 2003), vol. 2, pp. 521–4.

68 R. Bresin and A. Friberg, 'Expressive Musical Icons', in *Proceedings of the 2001 International Conference on Auditory Display* (Espoo, 2001).

69 G. Widmer, 'Studying a Creative Act with Computers: Music Performance Studies with Automated Discovery Methods', *Musicae Scientiae*, 9/1 (2005): 11–30.

musicological knowledge and hypotheses into the discovery and model building process.

In numerous studies, we have demonstrated that machine-learning algorithms do have the potential to make interesting discoveries. For instance, in a set of experiments based on a very large set of performance data (recordings of 13 complete Mozart piano sonatas by a Viennese concert pianist), the computer discovered a small set of 17 quite simple and succinct rules that predict certain aspects of local timing and articulation surprisingly well. The complete set of rules is analysed in great detail by Widmer,[70] where also the generality of the rules is quantified, based on independent reference performances. Indeed, it was shown that some of the rules seem to describe very general (if simple) performance principles, as they carried over, with little loss in predictive accuracy, to performances by other pianists and even music of a different style. One of the interesting aspects is that some of the rules discovered by the machine turned out to bear a strong resemblance to some rules in the KTH model. In this way, the machine-learning approach provides further circumstantial evidence for the relevance and validity of the KTH model.

This rule model was later extended to a multilevel model of expressive timing and dynamics, again via machine learning. Widmer and Tobudic[71] describe how the computer learns to apply extended expressive tempo and dynamics gestures to entire musical phrases, at several levels of the structural hierarchy, via a kind of case-based reasoning. In other words, the computer learns to transfer, in a suitable way, appropriate timing and dynamics patterns from reference performances to new pieces. An 'expressive' Mozart performance generated by this model won second prize at the International Performance Rendering Contest (RENCON 2002) in Kyoto.[72] Again, this shows that computers can, in principle, extract musically relevant patterns from performance data. Unfortunately, the structure-level model is not readily amenable to interpretation, being based as it is on the direct transfer of expressive patterns between performances based on a measure of phrase similarity. Developing algorithms for learning human-interpretable high-level models remains a challenge.

Recent research has also produced first indications that machine learning may help in getting a first grasp on the elusive notion of individual artistic performance

70 G. Widmer, 'Machine Discoveries: A Few Simple, Robust Local Expression Principles', *Journal of New Music Research*, 31/1 (2002): 37–50.

71 G. Widmer and A. Tobudic, 'Playing Mozart by Analogy: Learning Multilevel Timing and Dynamics Strategies', *Journal of New Music Research*, 32/3 (2003): 259–68.

72 G. Widmer and A. Tobudic, 'Playing Mozart by Analogy: Learning Multilevel Timing and Dynamics Strategies', in R. Hiraga (ed.), *Proceedings of the ICAD 2002 Rencon Workshop on Performance Rendering Systems* (Kyoto: Rencon Steering Group, 2002), pp. 28–35.

style. Stamatatos and Widmer[73] and Saunders et al.[74] managed to show that computers can learn to identify different concert pianists, based only on timing and dynamics patterns extracted from their performances. A machine-learning study presented in (Tobudic and Widmer, 2005) demonstrated, in precise quantitative terms, that to a certain extent, learning algorithms seem to be able to learn artist-specific performance strategies from recordings of famous pianists. And Madsen and Widmer (2006) describe a computational study where, using string-matching algorithms and evolutionary computing methods, computers could compute a quantitative measure of something like the relative stylistic consistency of famous pianists.

Other researchers have also experimented with machine learning as a tool for expressive performance research. For instance, in the context of jazz standards, Arcos and López de Mántaras (2001) describe a system that transforms saxophone phrases in expressive ways, also by analogy to a corpus of known performed phrases, and Grachten et al. (2006) present a system named TempoExpress that learns to transform expressively played jazz phrases to different tempi in musically meaningful ways.

In the field of machine learning, there is currently a strong trend towards probabilistic models. This is also reflected by some recent work on music performance analysis. For instance, Grindlay and Helmbold (2006) use Hierarchical Hidden Markov Models (HH-MMs) empirically to model statistical relationships between musical scores and expressive timing. Probabilistic models (such as Bayesian Networks, Hidden Markov Models, Conditional Random Fields, etc.), along with new powerful learning algorithms, promise to become very useful for performance research, provided that they can be extended not just to reveal probabilistic patterns, but also to yield intelligible descriptions that can be readily understood by musicologists. That is one of the challenges we will be working on in a new research project.

What all this is intended to illustrate is that computers are becoming much more than mere computation assistants – in music performance research as in many other fields of scientific investigation. Combined with refined and more efficient methods for extracting performance information from real recordings (see section 2 above), computational modelling and learning will contribute to a growing trend towards large-scale quantitative, data-based investigations in performance research (as exemplified, for example, by the Mazurka project[75] in the context of the CHARM Resarch Centre in the UK[76]).

73 'Automatic Identification of Music Performers with Learning Ensembles',

74 C. Saunders, D.R. Hardoon, J. ShaweTaylor and G. Widmer, 'Using String Kernels to Identify Famous Performers from their Playing Style', in *Proceedings of the 15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases* (Pisa, 2004).

75 <http://www.mazurka.org.uk/>, accessed 5 September 2007.

76 <http://www.charm.rhul.ac.uk/>, accessed 5 September 2007.

4. Final remarks

We have reviewed recent computational methods for data retrieval, display and modelling of expressive music performance, mainly for two purposes: first, to point the technically interested scholar and musician to potentially useful technology, and second, to encourage the computer researcher and developer to contribute algorithms or software solutions that might help enhance our access to, and understanding of, phenomena connected with music performance.

As the application of computational technology in the wide field ranging from music education to quantitative music research is relatively young, one might speculate about its future. In what domains will the use of such technology be fruitful, where will it make less sense?

One promising domain where the use of computational methods and technologies will and should play an increasingly important role is in music education. As reported in section 3, a number of steps have already been taken to explore this field of application. However, there are limitations as to where and when technology can be applied. Probably the most important aspect is that the widespread application of computers in music education will only be possible and successful when computer programs are well designed and easy to use, do not require complex hardware and are freely available. If information on a performance can be sensibly assessed while playing or immediately afterwards with a relatively short amount of interaction, the performer can still relate the information to an experience that is still fresh in their mind. So, computer programs should reach a level where they can handle arbitrary performance data autonomously in order to provide appropriate, usually visual feedback. While visualization can enhance cognition and help in reasoning regarding the subject under discussion (and also in teacher–student interaction), the fundamental mode of learning in music will always be based on an auditory feedback loop, not a visual one. Any technological tools should therefore be used with care and knowledge, and certainly cannot replace conventional methods of teaching and practice. But they could assist occasionally when needed in order to enhance awareness – just as a metronome should not be switched on the whole time during a practice day, but can help a great deal in certain selected situations.

Another domain of promising applications are real-time interaction systems that allow interactive manipulation and control of music performances and recordings, as we have sketched briefly in section 3. A new generation of researchers are working on novel interfaces for controlling musical expression, for example by using optical gesture trackers based on simple webcam setups. With such technology, one may even be able to manipulate recordings of great performers with intuitive gestural interaction according to personal taste. This would contribute to redefining the notion of music ‘consumption’ from one of passive exposure to a more active, and perhaps creative, process.

With respect to the role of computers in performance research, here there is also still ample room for further work. Current work in computational performance modelling – interesting results though it may have produced – is severely limited by some of the basic assumptions it makes (or must make, given the data it has gathered and the methods it uses). One serious problem is the tacit assumption of a deterministic (or at least probabilistic) mapping from musical structure to performance, which is certainly not the case in real life. A performer may well choose different dynamics, phrasing or other variations to shape a repeat of the same part of the music in a different way. Such variability often manifests itself in rather subtle nuances, which are, nevertheless, clearly discernible and understandable for an informed audience. As David Huron (2006) puts it, it is the micro-emotions that make music interesting and not so much the bold, big emotions (fear, anger, sadness) that are the topic of much of the current research on musical emotion. Here, one would have to gather very precise data in rather subtle experimental setups in order to get at the fine-grained level of these phenomena.

Another source of new opportunities might be the emerging alliance between musicology and the field of music information retrieval. Musicologists are beginning to make substantial efforts to digitize and put online large amounts of historical material (e.g. the entire works of Mozart[77]). Such enormous databases could be fertile sources for musical data mining, which might lead to even more well-founded and significant empirical findings than the results reported above. One initiative to gather together the diverse research approaches on computational modelling of music performance is the RENCON initiative,[78] which organizes annual workshops on performance rendering research, including a competition to determine the 'best' artificially performed piece of music. However, the goal (as stated on the RENCON web page) to win the Chopin competition in 2050 with a computer program is, if not intended to be tongue-in-cheek, simply a misleading approach. The aim of modelling expressive artistic behaviour should not be to replace or compete with a human in this domain, but to create models with which the complex nature of expressive performance can be better investigated and understood.

Finally, it is challenging to probe the limits of quantifiability in expressive music performance. What aspects of it cannot be modelled at all? Even if research into expression of movement, motor control and biomechanics advances, how would it ever be possible to measure, quantify and finally model the performer's 'mind and body in the "heroic struggle" to express and communicate'?[79]

---

77 <http://dme.mozarteum.at/>, accessed 5 September 2007.

78 R. Hiraga, R. Bresin, K. Hirata and H. Katayose, 'Rencon 2004: Turing Test for Musical Expression', in *Proceedings of the 2004 Conference on New Interfaces for Musical Expression* (Hamamatsu, 2004), pp. 120–23; <http://shouchan.ei.tuat.ac.jp/~rencon/>, accessed 5 September 2007.

79 As so aptly described by E.F. Clarke, 'Listening to Performance', in J. Rink (ed.), *Musical Performance: A Guide to Understanding* (Cambridge: Cambridge University Press, 2002), pp. 185–96.

Acknowledgements

This work is supported by the Austrian Science Fund (the Erwin Schrödinger Fellowship J2526 and the research project P19349N15) and the Vienna Science and Technology Fund ('Interfaces to Music', I2M, CI010). The Austrian Research Institute for Artificial Intelligence acknowledges basic support by the Austrian Federal Ministries BMBWK and BMFIT.