Goebl, W., Dixon, S., & Schubert, E. (2014). Quantitative methods: Motion analysis, audio analysis, and continuous response techniques. In D. Fabian, R. Timmers & E. Schubert (Eds.), Expressiveness in Music Performance – Empirical Approaches Across Styles and Cultures (pp. 221-239). Oxford, U.K.: Oxford University Press.

Chapter 13

Quantitative Methods: Motion Analysis, Audio Analysis, and Continuous Response Techniques

Werner Goebl, Simon Dixon, and Emery Schubert

> Measurement of performance has dominated performance research.
> *(Gabrielsson 2003)*

This chapter summarizes recent quantitative measurement and analysis techniques of three domains of musical expressiveness: body motion, musical sound, and listeners' continuous response to musical sound. We outline computational methods to quantitatively assess expressive aspects of the body movements of the performing musicians, to extract expressive information from the musical sound itself, and finally to examine the perception of expressiveness through self-report continuous response methods. We also expand and add to recent overviews of performance analysis techniques (e.g. Timmers and Honing 2002; Goebl *et al.* 2008; Goebl and Widmer 2009; Windsor 2009).

When considering methods of investigating expressiveness in performance, we find it helpful to think of music expression as occurring in three "worlds" in the Popperian sense (see Popper and Eccles 1977; Parncutt 2011): first, the physical world, secondly, the experiential world, and thirdly, the world of thought and knowledge. This chapter considers Worlds 1 and 2—the physical and experiential—as being viewed through the lens of World 3—the measurement tools and ideas that aid understanding of Worlds 1 and 2. Therefore, by physical world (World 1) we refer to the measurable parameters of movement and music, be they the movement of the body parts or properties of the recorded sound from which World 2 psychoacoustic fundamentals such as the tempo, pitch, loudness, articulation, and so on may be extracted.[1] The experiential world (World

[1] Although perception of pitch, loudness, tempo, and so on may appear to be "World 2" experienced properties, the scientific reporting of these variables is usually taken directly from abstract (mathematical) manipulations for physical properties (fundamental frequency, intensity, timing, and so on, respectively), based on models which aim to mimic general, rather than specific, human response. It serves our purpose, therefore, to treat these psychophysical and musical variables as being directly linked to, or drawn from, the physical world (World 1). The experiential world is more to do with thoughts and feelings that result from, say (in this case), music perception, performance, and processing. In fact, the assumption of a direct translation from the physical to the experiential world is itself a World 3 phenomenon, since mathematical relationships are called forth when converting a frequency to a pitch or intensity to loudness (products of ideas). For further discussion, see Archer and Elder-Vass (2012).

2) is at the crux of the phenomenon under investigation—since it is the sensation and construction of the expressiveness of a performance that ultimately drives what is and is not expressive, and that drives many investigators to examine what physical world aspects cause this experience. Interactions with World 3 are a necessary part of the discussion, since in the present overview this will refer to the language, measurement, and analytic tools used to understand musical expressiveness in Worlds 1 and 2, such as musical notation, spectral representation of sound, motion capture hardware, timing plots, statistics, and so on.

Importantly, developments in psychological and engineering techniques and tools have enabled researchers to understand the moment-by-moment changes in expressive parameters, in some cases in real time. This chapter reports some of these "continuous measurement" techniques. Although continuous measures have been commonplace in physical world measures of expression, the area is quite new in studies regarding the experiential world. Traditional approaches have relied on "post-listening" ratings of the musical expression perceived by the listener (see Chapter 16), and the present chapter investigates how such measurement can be done in real time. Before that, we give an overview of measurement methods for body motion and musical sound.

Physical world expression measurement: motion analysis

Most musical sounds are the result of the musician's body movements during performance. To give some examples, this can range from the finger and hand movements of a pianist, all the way to the arm, torso, or even whole-body movements of a violinist or a clarinet player. Some of the movements are required to produce the intended sound (e.g. pianists' finger motion), other movements may convey particular meanings to the listener (e.g. head nodding), some may be executed to help the performer to perform the music (e.g. foot tapping), while others may not necessarily be required, but are usually executed by the performer (e.g. torso sway; Davidson 1993). Likewise, recent scholarly accounts identify four different, and partly overlapping, kinds of movement (Dahl *et al.* 2010; Jensenius *et al.* 2010), namely sound-producing, communicative, sound-facilitating, and sound-accompanying movements (or gestures). A particular movement may well belong to more than one category (e.g. a bowing movement might simultaneously cue fellow musicians).

Sound-producing movements (or effective gestures; Delalande 1988) are those required to create or control the sound (e.g. the bowing movements create the sound on a violin, and the left-hand fingering controls pitch and intonation). They are the most fundamental to music performance, and are also quite constrained as the goals of the actions (the sound) determine their execution. Their execution changes with the performance requirements of the score, such as the tempo (Goebl and Palmer 2009a) or dynamics (Dalla Bella and Palmer 2011), and also with biomechanical factors (Loehr and Palmer 2008) and the particular instrument played (Dahl 2004; Schoonderwaldt 2009), but not with communicative intentions (Goebl and Palmer 2009b).

All other movements of the performing musicians may be called **ancillary** or **accompanist** (Wanderley *et al.* 2005), but depending on their intention they may be differentiated further into **communicative movements** and **sound-facilitating movements** (Dahl *et al.* 2010; Jensenius *et al.* 2010). The communicative content of musicians' movements has been shown to strongly influence and even overrule the auditory information (Davidson 1993; Broughton and Stevens 2009; Behne and Wöllner 2011). The visually transmitted movement information even alters the usually stable perception of durations of single tones, when the production movement is changed (Schutz and Lipscomp 2007). This ties to Alfred Brendel's famous advice (Brendel 1976) to employ suggestive gestures when attempting to perform a crescendo on a single held chord on the piano (which is acoustically impossible, as a tone inevitably decays once it is played). Even in the absence of sound, listeners could identify basic emotions and levels of expressiveness well from the movements only (Dahl and Friberg 2007). Communication via movements also extends to communication between ensemble members (Williamon and Davidson 2002), or more clearly to conductor–orchestra communication (Luck and Sloboda 2009). The other category of ancillary movements—sound-facilitating movements—involves movements that strengthen performance success (such as a constant body sway or particular breathing patterns) or that support the performer in their expressive statement (e.g. by head shakes or gazes; see Dahl *et al.* 2010, p. 54).

Sound-accompanying movements are usually generated by a person who is not involved in the process of music production, but is listening to and watching the music. Typical examples are dance, but also sound tracing (Leman *et al.* 2009) and continuous response movements (discussed in the next section).

There is not always a clear distinction between these four categories of movements, as they strongly overlap, so that a movement might belong to multiple categories at the same time. The categorization of movements may particularly depend on the intended purpose of the movement by the performer—a variable that is usually unknown or hard to assess empirically (referred to as "intention" by Godøy and Leman 2010). Next, we explain methods to measure the physical movement of the musicians' bodies during performance ("extension"), and we describe experimental designs aimed at disentangling sound-producing movements from the other categories.

Methods and analysis techniques

To assess the physical movements of performing musicians quantitatively, there are numerous technical options at one's disposal that will be briefly summarized here. Recent overviews of the technological possibilities for domains other than music are given by Zhou and Hu (2007) and Burdea and Coiffet (2003).

Video-based approaches

Video footage is easy to acquire, as current laptop computers have a webcam built in with adequate frame rate, resolution, and image quality; also, video cameras are inexpensive and video data care be easily transferred to a personal computer for playback and analysis. The most straightforward method of assessing musicians' movements is to videotape them and analyze the footage through ocular inspection. This process may be supported and made more reliable through video annotation software such as Anvil,[2] ELAN,[3] VARS,[4] or more versatile commercial tools as Atlas.ti,[5] Observer[6] by Noldus, or nVivo[7] by QSR International. Particular gestures or movement patterns are identified by the researcher and labelled in the software for later analysis.

[2] www.anvil-software.de/

[3] http://tla.mpi.nl/tools/tla-tools/elan/

[4] http://vars.sourceforge.net/

[5] www.atlasti.com/

[6] www.noldus.com/human-behavior-research

[7] www.qsrinternational.com/

Video data may also be used for quantitative analysis of the recorded motion. Several approaches aim to extract complex kinematic data from video recordings (Camurri and Moeslund 2010). Advanced algorithms perform motion tracking from video and output, among others, various kinematic measures such as the range of motion or even the continuous two-dimensional position of particular points. Such video-based methods have been used to recognize the gestures of a conductor (Kolesnik 2004).

Three-dimensional motion capture systems

Motion capture systems measure the position of particular body markers in two or three dimensions. Extensively used in the gaming and film industry, and for gait analysis in rehabilitation and the military, this technology has developed tremendously over the past years. Common systems work optically with infrared light using either passive reflective markers that are lit and filmed by multiple cameras (passive motion capture, as used, for example, by Vicon, Qualisys, or Optitrack by Natural Point) or active markers that are connected by cables and emit light themselves (such as Optotrack Certus by Northern Digital or VisualEyez by Phoenix Technologies). Motion capture systems deliver discrete three-dimensional data for the markers that can be used for analysis. However, they are quite expensive and require refined technical knowledge.

Passive motion capture systems have been used for music research (Wanderley *et al.* 2005; Goebl and Palmer 2009b, 2013; Schoonderwaldt 2009; Dalla Bella and Palmer 2011). They are versatile and relatively unobtrusive due to their small markers (down to about 4 mm in diameter), not restricted by the number and size of markers, and provide accurate three-dimensional position data. To label each marker trajectory, the vendors provide various software solutions. Problems occur when markers become occluded due to an interrupted line of sight, making the marker trajectories discontinuous. Regardless of how well the automated marker labelling works, passive systems might require several post-processing steps to obtain reliable three-dimensional data.

Active systems, on the other hand, involve markers that emit infrared light that is seen by multiple cameras. Such systems advantageously reduce the laborious labelling step in data post-processing, because the system is able to identify each marker through its unique time point of light emission. The output data are correctly labelled even across missing data, and are immediately ready for subsequent analysis (e.g. Wanderley 2002; Goebl and Palmer 2009b), but the number of markers is limited by the overall sampling rate of the system (the higher the number of markers, the lower the sampling rate).

As all optical systems require line of sight of each marker to a minimum of three cameras to triangulate their position, many music applications will run into the problem of marker occlusion. In the example of finger and hand motion capture of piano performance, the fingertip markers in particular are "lost" when they curl in during performance (which occurs regularly). There is no work-around available except for using other equipment such as magnetic motion capture (e.g. Liberty by Polhemus or MotionStar by Ascension Technologies), which is usually restricted by small capture volumes (the space in which the measurements are taken), or assisting optical motion capture with accelerometers or gyroscopes.

The use of motion-capture data may range from analyzing large-scale body movements (metrical embodiment; see Toiviainen *et al.* 2010), where sampling rates of 60 frames per second (fps) are sufficient, to detailed investigations of sound-producing gestures (e.g. touch in piano performance; Goebl and Palmer 2008). When recording limb movements that contain impacts with rigid bodies (such as the drum stick on the membrane or the pianist's finger on the key surface), much higher sampling rates are required to monitor the sudden changes in movements resulting in large acceleration peaks (e.g. 400 fps; Dahl 2004). As there is a trade-off between capture volume and sampling rate, researchers have to find compromises (Bouënard *et al.* 2011). Toiviainen and Burger (2011) have provided a MoCap Toolbox for Matlab featuring several standardized analysis steps which also imports data from Wii devices.

Sensor-based approaches

As an alternative to the techniques just mentioned, researchers have used various kinds of sensors mounted on musicians' limbs. The earliest experiment was by Otto Ortmann (1929), who constructed a complicated lever system to draw a pianist's leap movements on paper, and used a vibrating tuning fork to capture continuous key movements over time. Other detectors include accelerometers (MacDougall and Moore 2005), gyroscopes that measure orientation and position in space (such as InertiaCube by InterSense or systems by XSens), or devices that combine different sensors (such as the CyberGlove used for finger analysis in piano performance; Furuya *et al.* 2011).

Physical world expression measurement: audio analysis

Various mechanical and electrical devices have been employed for measurement of expressive performance parameters such as timing, dynamics, and articulation, as reviewed by Goebl *et al.* (2008). Special-purpose hardware such as a computer-monitored piano or an electro-laryngograph can provide highly accurate measurements, but in many cases the direct monitoring of musical performances is not a viable option, whether due to the intrusive nature of the method, the limitations of the hardware (e.g. requiring a specific instrument which is not the instrument of choice of the performer), or merely because the performance of interest took place in the past. Given the availability and modest cost of audio recordings, now numbering millions and covering more than a century of musical performances, there is considerable interest in analyzing the expression represented in these recorded performances (see Chapter 4). This section therefore focuses on the measurement of expressive performance parameters from audio recordings. Audio analysis software enables research on recordings from many sources, including commercial CDs and archives of historical and ethnomusicological research. However, it is not without its limitations, and several methodological and technical challenges must be addressed in order to obtain useful data for performance analysis.

The main issue is that of obtaining reliable measurements, for each performed tone, of parameters such as timing, amplitude, and pitch, which are the main attributes investigated in performance research (Gabrielsson 2003). From these measurements, other properties such as tempo, dynamics, intonation, articulation, and chord asynchrony can be derived. Typically the measurements are obtained by some combination of human judgement and the use of automated audio analysis tools in the form of computer programs that vary in sophistication from waveform visualization to score–audio synchronization software.

The simplest method is to inspect the audio waveform with computer software such as Sonic Visualiser (Cannam *et al.* 2006, 2010), and manually annotate the desired note onsets. This approach is labour-intensive and only suitable for simple (e.g. monophonic) musical textures in which onsets do not mask each other, and where the instruments of interest are percussive and thus have well-defined onset times. Also it is not clear how accurate this method is. Povel (1977) claimed a temporal precision of 1–2 ms when determining note onsets "by eye" from oscillograms of recordings of J. S. Bach's C Major Prelude from the *Well-Tempered Clavier Book 1*. Examining performances of the same piece, Cook (1987) estimated the timing resolution more conservatively at 10 ms, using a computational system with manual correction of its output. Studies of solo piano music were performed by Repp (1990, 1992), who read off note onset times from waveform displays, using audio playback of short excerpts to resolve unclear cases. Excerpts leading up to the onset were chosen, and the end point was varied to find the latest point for which the succeeding note was not audible. For repeated measurements, mean absolute errors of 6.5 ms (Repp 1990) and 4.3 ms (Repp 1992) were reported. For other instruments, precision values of 3 ms for cymbals (Friberg and Sundström 2002), 3–5 ms for jazz melody instruments and double bass (Ashley 2002), and 2 ms for trumpet (Collier and Collier 2002) have been reported.

For larger-scale studies, a range of algorithms is available for automatic analysis of audio files. Ideally, to extract performance-related data directly from audio recordings, we would need a fully automatic transcription system, but state-of-the-art systems are not yet sufficiently robust to provide the precision required for expression research (Klapuri 2004). In cases where a partial transcription (e.g. of beat times only) gives sufficient information, existing algorithms and systems can be used. In other cases, extra information (e.g. from scores, other recordings, or user interaction) is employed to bridge the gap between algorithm performance and required accuracy.

Many automatic onset detection algorithms exist (e.g. Bello *et al.* 2005; Dixon 2006), several of which are also available in Sonic Visualiser. The accuracy of onset detection methods varies greatly depending on the instruments playing in the recording and the choice of parameter settings, with reported detection rates ranging from around 50% for the solo singing voice to well over 90% for pitched and non-pitched percussive instruments. For example, on a large set of solo piano recordings (over 100 000 onsets), 96% of the onsets were detected with an average error of 8.8 ms (Dixon 2006).

In some cases, not all onsets are required. For example, to analyze the evolution of the tempo, it is sufficient to estimate the timing of notes corresponding to downbeat locations, and ignore the remaining notes.[8] One way to achieve this is by recording a listener tapping along with the recording, using for example a MIDI drum pad or a keyboard (Cook 1995; Sapp 2007). This method is a relatively fast way to obtain rough timing data for a single metrical level. However, some biases have been observed in tapping studies, where participants underestimate abrupt tempo changes or systematic variations, even after repeated attempts on the same short excerpt (Dixon *et al.* 2006).

An alternative approach is to use an automatic beat tracking system such as BeatRoot (Dixon 2001a, b) to estimate the beat times, followed by manual correction, as discussed later in this chapter. BeatRoot produces a list of beat times, from which tempo curves and other representations can be computed. Although it has its drawbacks, such as failure to model higher-level characteristics of the music (e.g. metric hierarchy), this system has been used extensively in studies of musical expression (Dixon *et al.* 2002; Widmer *et al.* 2003; Goebl *et al.* 2004; Flossmann *et al.* 2009; Grachten *et al.* 2009).

Once onset or beat times have been established, other parameters can be estimated from the segments of the recording between identified events. For example, dynamics can be approximated by computing the root-mean-square (RMS) energy of the signal over the given segment (Repp 1999), but this is only suitable for monophonic excerpts, as it does not distinguish the individual contributions of simultaneous tones. For the polyphonic case, a technique such as score-informed analysis (Scheirer 1995) could prove helpful, but the natural variation in musical tones precludes accurate

[8] This is due to the perceptual smoothing of tempo which occurs because tempo is integrated across a certain time, so that sub-beat events have little influence on the perceived tempo.

estimation of individual dynamics from polyphonic mixtures (Repp 1993). Pitch estimation is a widely studied problem, and many algorithms have been proposed. For the monophonic case, YIN (de Cheveigné and Kawahara 2002) is a robust and precise algorithm. Algorithms for polyphonic analysis are reviewed by de Cheveigné (2006). Most approaches are developed for applications where a resolution of a semitone is considered sufficient (e.g. transcription to standard western notation), but studies of temperament measure tones with accuracies of 1 to 2 cents (hundredths of a semitone) (Dixon *et al.* 2012). For estimating time-varying aspects of pitch, such as vibrato, Wen and Sandler (2007) propose an interactive approach that allows the user to select a tone on a spectrogram, after which the system calculates the frequency trajectories of all partials and displays the parameters of the selected tone, also allowing the user to modify the audio via changes to the parameters. The extraction of further parameters from audio, such as pedal information, tone duration, and articulation, are considered unsolved signal-processing problems (McAdams *et al.* 2004).

Much music analysis software exists in the form of a research prototype demonstrating an algorithm, but lacking an interface for interactive editing of partially correct outputs, without which most software is not significantly more efficient to use than manual annotation. Exceptions include BeatRoot (Dixon 2001a), an automatic beat-tracking system with a graphical user interface for visualizing (and sonifying) the beat times and underlying audio, allowing the user to edit the output and retrack the audio data based on the corrections. A similar methodology was applied in the development of JTranscriber (Dixon 2004), written as a front end for an existing transcription system (Dixon 2000). The graphical interface shows a spectrogram scaled to a semitone frequency scale, with the transcribed notes superimposed over the spectrogram in piano-roll notation. The automatically generated output can be edited with simple mouse-based operations, and monitored with audio playback of the original and the transcription, together or separately. A more recent development is Melodyne,[9] a commercial audio editor that includes transcription software, producing western music notation output of audio recordings. Sonic Visualiser (Cannam *et al.* 2006, 2010) provides a platform for analysis algorithms implemented as plug-ins, allowing a range of visualization, sonification, and editing options. Songle[10] is a web service that allows users to upload music and obtain an automatic analysis of segmentation, metrical structure, melody line, and chords, also providing visualization and editing options. Other programs that group together suites of audio analysis software include PRAAT[11] and PsySound (Cabrera *et al.* 2007). One danger with visual feedback is that the type of feedback can bias the data (Dixon *et al.* 2006), as has also been noted by Leech-Wilkinson (2009): "it's all too easy to be led into hearing things one can see on a computer screen but can't perceive without one" (chapter 8.1, paragraph 25).

The degree of manual interaction required to create accurate annotations sets a limit on the scale of studies that can be performed with semi-automatic methods. To improve automatic analysis, and thus reduce human effort in the data preparation stage, systems can take advantage of any existing knowledge that might be available, such as scores or other performances of the same work. Since a score is often available for the performances being analyzed, Scheirer (1995) recognized that he could obtain better results by incorporating score information into the audio analysis algorithm. An alternative approach, suitable for analyzing multiple performances of the same work, is to annotate one performance semi-automatically, and then align the audio files and transfer the annotations automatically from the first performance to the corresponding time points in the other recordings,

[9] www.celemony.com/cms/

[10] http://songle.jp

[11] www.fon.hum.uva.nl/praat/

using software such as MATCH (Dixon and Widmer 2005). This approach is more efficient than direct annotation of all files, as the audio alignment step is generally more accurate than techniques for direct extraction of expressive information, so the amount of subsequent correction for each matched file is greatly reduced (as applied, for example, by Flossmann *et al.* 2009).

Taking this idea one step further, the initial annotation phase can be avoided entirely if a digital score is available, in which case a mechanical performance can be synthesized from the score and then matched to the audio recordings. Then it is relatively straightforward to compute performance parameters such as tempo from the relationship between actual (performed) and nominal (score) durations. Several score-performance alignment systems have been developed for various types of music (Cano *et al.* 1999; Soulez *et al.* 2003; Turetsky and Ellis 2003; Shalev-Shwartz *et al.* 2004), as well as systems that can track live performances (Orio *et al.* 2003; Dixon and Widmer 2005), enabling real-time visualization of performance expression or driving page-turning devices (Arzt and Widmer 2010). Niedermayer and Widmer (2010) consider means for improving time resolution in score-to-audio alignment by identifying points of high confidence, called anchor notes, which can direct the alignment. Another development of these tracking algorithms is the "Complete Classical Music Companion" (Arzt *et al.* 2012), a computer program that, while listening to a piano through, for example, a laptop microphone, is able to identify within milliseconds what piece is being played by accessing a database of symbolic scores (currently the complete Mozart and Chopin works).

Apart from analyzing audio for music expression, there is a large body of work studying music expression from symbolic data (i.e. MIDI or similar formats). Recently, Widmer and co-workers analyzed a large corpus of performances that were recorded by a computer-controlled grand piano live on concert stage by the renowned pianist Nikita Magaloff (Flossmann *et al.* 2010b). They established new ways of processing the raw performance data of over 330 000 notes or some 10 hours of music (Flossmann *et al.* 2010a), to be able to perform detailed large-scale analysis of particular performance aspects, such as tempo rubato (Goebl *et al.* 2010). Such large performance corpora are used to train computational models on stylistic particularities of individual performers (Widmer and Goebl 2004) that are able to generate their own renditions of previously unknown scores, which can be listened to, as a direct way of validating the computational models (Widmer and Tobudic 2003; Flossmann *et al.* 2011; Flossmann 2012). One of those algorithms, YQX, has dominated the international Performance Rendering Contest Rencon,[12] being awarded first prize in 2008 and 2011 (Widmer *et al.* 2009; Flossmann *et al.* 2011).

Despite their imperfections, audio analysis tools are becoming part of the standard equipment of empirical musicologists (Cook 2004; Leech-Wilkinson 2009), enabling performance research to extend to larger data sets than it was previously feasible to examine. Current research topics include the joint estimation of multiple musical parameters and the incorporation of higher-level musical knowledge (including knowledge of scores, musical styles, and music theory) into analysis systems. Advances in these areas would increase the robustness and accuracy of systems, and require less human effort in extracting performance data from audio.

Experiential world expression measurement: self-reported continuous response

Imagine that you are listening to a piece of music, and as the music unfolds you are asked to report at each moment in time, when possible, how expressive the performance appears to be (e.g. by

[12] http://renconmusic.org/

moving a slider accordingly). Collecting these responses while the piece unfolds is referred to as **self-reported continuous response**, and is an alternative to more traditional ways of collecting explicit self-reported perceptions of musical expression, where a rating is given along some scale at the end of the piece, retrospectively (e.g. Lucas *et al.* 1996; Crist 2000; Kinney 2004; Price and Chang 2005; Fabian and Schubert 2009; Morrison *et al.* 2009; Fabian *et al.* 2010; Napoles 2013).

Explicit attempts to continuously rate expressiveness in music are just beginning to be explored (see Chapter 16), although they have their origins in the work of Clifford Madsen and colleagues in the late 1980s, who used a continuous response digital interface that could be configured to record ratings of a wide range of scale and category responses (Madsen 1990; Geringer and Madsen 1996; Madsen *et al.* 2007). Devices like these require the listener to rate a parameter or general aspect of the music over time by moving a slider while listening. Manfred Clynes' sentograph (Clynes 1989, 1995) was a device that also measured expression, although not necessarily through self-report. Finger movements were thought to be automatic, implicit, and biologically pre-programmed (Clynes 1973). Antonio Camurri developed an analogous tool using a hand-held laser light whose fluctuations were measured while a participant moved the light in response to the music or dance stimulus (Camurri *et al.* 2006). The majority of continuous response approaches record ratings of emotional responses (happiness, sadness, valence, arousal, etc.) (Nettheim 1999; Schubert 1999; Cowie *et al.* 2000; Demany and Semal 2002; Nagel *et al.* 2007; Stevens *et al.* 2009), and not explicitly musical expressiveness (for a review, see Schubert 2010).

In this section, we shall focus our attention on the explicit rating of expressiveness of musical interpretation. With this focus, we would expect that at certain points in a performance of some piece of music, expressivity is rated as quantifiably different to that at other points in the music because of some interpretative decision made by the performer/performers. It might be because a belting singer is providing unusual ornaments on a held note, or a violinist exercises a greater than expected rallentando, followed by a sudden pick up in tempo. Therefore the variations in rated expressiveness made by the listener could be related back to the physical parameters of the performance. Because of the lack of explicit self-reported continuous response research on musical expressiveness, this section of the chapter is necessarily speculative, but points to some initial results, and particularly to the seminal study by Seiichiro Namba and colleagues (Namba *et al.* 1991).

The utility of continuous expressiveness rating

The small amount of literature that can be classified as direct self-reported continuous rating of expression suggests that this approach to data gathering can provide insights into the locations in an unfolding piece of music where expressive content is varying, high or low, and thus can provide explicit data on where a performance is more or less expressive. The self-report approach provides relief from reliance on a music score against which the expressive norm is usually compared. Here the expressive norm would be the underlying central tendency of participant responses across a wide range of performances of the same piece at each point in time. The typical (as estimated from the continuous responses of several participants) "expressiveness rating" time series for one performance of a piece will then provide a direct experiential-world measure of expressiveness when compared with the overall mean time series used to generate the "norm reference" time series based on ratings of "all" available performances of that same piece.

As with retrospective ratings of expressiveness, the underlying central tendency time series can also be used to determine whether one performance is more expressive (i.e. the expressiveness times series is overall higher) than the time series of another performance. The relationship between the overall-continuous and post-performance ratings is also an area of interest. Current thinking suggests that post-performance ratings will be most influenced by and therefore correspond to high points and concluding (recency) ratings of the time-series response. In other words, many of the impressions made during listening seem to be forgotten when making the post-performance response—a phenomenon referred to as duration neglect (Fredrickson and Kahneman 1993; Rozin *et al.* 2004; for further discussion, see Schubert 2010).

An early continuous self-report measure of **emotion** study is now described because it does provide valuable information on **musical** expressive playing. It is a pioneering study by Namba and co-workers (Namba *et al.* 1991), in which keys on a computer keyboard representing various emotions were pressed by the listening participant as the music was playing, and the location and identity of the key press in time provided information about the emotion expressed in the music. This study could be classified as rating **musical** *expression* continuously because Namba and colleagues collected continuous ratings of several interpretations of the Promenade from Mussorgsky's *Pictures at an Exhibition*. Different performances produced different emotional responses, suggesting that the physical performance parameters had been altered because the composition remained the same. In the study by Namba and Kuwano (1990), time-series plots of the adjective "votes" made by participants demonstrate how points of high expression and the nature of the expression can be identified. Figure 13.1 shows two such time-series plots comparing ratings of two performances of Promenade 1. A high point of "expression" (large number of votes in the time series) can be located in both performances in bars 11 and 12. However, for the recording conducted by Ashkenazy, the largest proportion of responses suggest that the expression is described by the term "brilliant," whereas for the performance conducted by Karajan the term "smooth" is used most frequently to describe the same passage. Hence, in addition to the

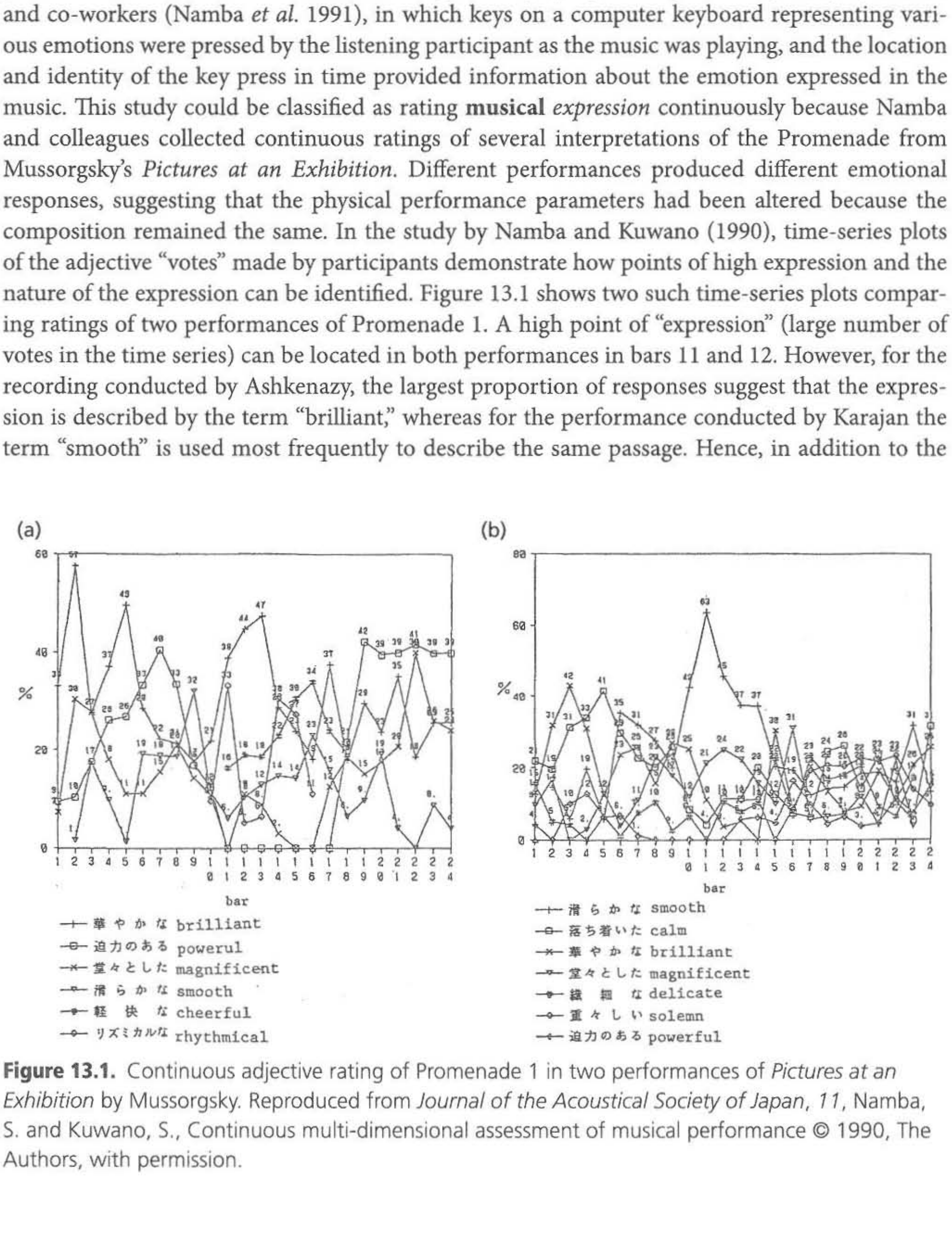

Figure 13.1. Continuous adjective rating of Promenade 1 in two performances of *Pictures at an Exhibition* by Mussorgsky. Reproduced from *Journal of the Acoustical Society of Japan*, *11*, Namba, S. and Kuwano, S., Continuous multi-dimensional assessment of musical performance © 1990, The Authors, with permission.

other benefits of continuous ratings, the techniques developed by Namba and colleagues can be used to assess where a large amount of expressive activity takes place in a piece, even though it is based on selection of an emotion adjective.

Of course, Namba's technique was not devised explicitly to measure self-reported musical expression. Studies that require the rating of "expressiveness" continuously while the music is playing are rare (e.g. Peddell 2008). An example is reported in Chapter 16. Inspection of that time-series data presented interesting differences across recordings in expressiveness ratings at each point in time, but interpretation of such data needs to be done with caution (Schubert 2001; Upham 2011). That is, musical expressiveness is a dynamic process, but parts of the process (e.g. "serial correlation") can be hidden in the time series, meaning that statistically each response at a point in time in the piece is not necessarily independent of a response that was made earlier. Although time-series techniques are more involved than the analysis reported in Chapter 16 (see, for example, Schubert 2010), the results nevertheless indicated that "direct" musical expressiveness rating of performances revealed differences that would be hard to discern with single-sample retrospective ratings. Retrospective ratings must be influenced by some kind of averaging over time, selective recall, or duration neglect. Expressiveness is not, therefore, solely a post-listening effect. Judgements of locations where musical expressiveness is higher or lower in one performance than in another can be ascertained by recording continuous expressiveness ratings.

Some technical issues

The technical details of self-report continuous response methods as applied to music are beyond the scope of this chapter, and are discussed elsewhere (Dean and Bailes 2010; Schubert 2010; Pearce 2011). These techniques can easily be applied to the rating of expressiveness. However, it is worth expanding here on a methodological problem that is peculiar to continuous response measurement of expression. This is the matter of synchronizing two or more performances of the same piece or section of music. As we have seen in the measures of physical properties, one of the most frequently researched aspects of music expression is fluctuation in timing. Performers use microstructural variations in tempo to produce expressive effects (Repp 1992; Palmer 1997; Gabrielsson 2003; Widmer and Goebl 2004). As a result, different interpretations of the same piece will show net fluctuation in overall and small-scale durations.

Comparison of two time-series plots of the same piece will therefore not align correctly when laid on top of each other using the same time scale. This can lead to misleading results (the same time point in two different performances will correspond to different positions in the score). Two common ways of dealing with this problem are temporal registration with time scaling (dilation or compression), and short-time signal-to-signal alignment (STSTSA) applying dynamic time warping algorithms as used to compare multiple physical-world sound files (e.g. Dixon and Widmer 2005; see previous section).

Registration involves locating points in the music, usually with the aid of a music score, where the two performances might be expected to coincide, such as a section boundary, the start of a phrase, or at a marked change in tempo. McAdams and colleagues refer to this approach as "landmark-registration" (McAdams *et al.* 2004) and is similar to anchoring, previously described. One of the performances can be treated as a reference (nominal performance), against which others are adjusted section by section (i.e. landmark by landmark), compressing or dilating (stretching) their duration to match the duration of the reference section. A common landmark-registration approach is to map expressiveness ratings to the respective bar (measure) in the music in which the rating occurs, as in the Namba example mentioned earlier (Figure 13.1). An even simpler and cruder form of this approach is to treat the entire performance as a single unit, thus registering the start and end time points of each performance, and providing a proportional stretch or compression of all notes for each performance. The dilation or compression applied within segments is usually linear, which has the disadvantage of losing alignment in fine temporal differences (Vines *et al.* 2003). However, since the researcher can determine what instances of the performances should be aligned (such as the beginning of a phrase or the start of a new section), the landmark registration technique has a musicological validity associated with it. STSTSA is the other extreme of registration—where sound recordings are broken up into many small consecutive segments of several tens of milliseconds that are matched to each other using an algorithm, rather than manually registering landmarks, in order to compute the alignment path (Dixon *et al.* 2005; Macrae and Dixon 2010). Such techniques have been applied to identifying variations in different performances of contemporary dance (using video, in a manner analogous to audio only), involving time- and frequency-based techniques (Ferguson *et al.* 2009; Stevens *et al.* 2009).

Expression ratings provide an additional complication because they are likely to be reported immediately after the musically expressive event took place. This lag in response can be diagnosed by procedures such as cross-correlation (Nettheim 1999; Schubert 1999; Snyder and Krumhansl 2001; Toiviainen *et al.* 2010) to allow effects of lag in expressiveness response to be adjusted accordingly. Another approach is to perform STSTSA on the expressiveness ratings time series themselves, but here the problem is that the expressiveness rating is (falsely) assumed not to change significantly across performances, which, as we saw in the example of Namba and colleagues (Figure 13.1), is not likely to happen at all points in time, and defeats the purpose of identifying important, distinct, and unique expressive events.

Even though STSTSA algorithms do not require further high-level knowledge about the musical piece (form, structure, phrasing, and so on), and computational power is usually no longer a limiting factor, landmark-registration approaches are still more commonly used than STSTSA in self-reported expression research. However, the conceptually simpler and more pragmatic approach (at least from a musical perspective) of landmark registration is likely to find a place in further research.

Conclusions

Researchers of music expressiveness are faced with a vast array of options for measuring and analyzing musical expression, from sound files to live performance, from ocular inspection of videos through to sophisticated motion-capture techniques, and self-report methods collected after a piece has been heard, or while it is unfolding. This chapter has outlined some of the techniques and tools available. The literature demonstrates that audio analysis provides a vast range of options, from basic manual analysis using freely available software, through to highly sophisticated, automated extraction of various psychoacoustic estimates of parameters such as tempo, pitch, and dynamics from audio. Motion capture is a relatively new technology that has supported a growing interest in looking at how the physical intra-player activity of the performer affects expressive communication and music perception via this non-auditory channel (for a discussion of between-player effects, see Chapter 15). Most motion capture systems require considerable post-processing prior to meaningful analysis. Self-reported ratings use the human participants' statistically processed responses to provide a simple way of identifying which performances of a piece and sections of music are more or less expressive.

Our main aim in this chapter was to introduce the reader to developing and state-of-the-art tools for measuring musical expressivity. By applying Popper's worlds, we also aimed to draw to researchers' attention the critical philosophical implications of making measurements of expressiveness, specifically in making the distinction between measuring World 1 (physical) aspects, such as motion and musical characteristics, and World 2 (experiential) aspects—the actual sensation of expressiveness experienced by the perceiver. The key philosophical point that unifies our review is that we understand each of these worlds through Popper's World 3—that of ideas, definitions, and, most importantly, measuring instruments of physical and psychometric signals. In many ways we (and researchers in particular) are trapped in World 3, but our instruments in the future may help us to get closer to the physical and experiential worlds. Furthermore, we have limited our review to essentially western thinking about what expressiveness is, without consideration of the way that other cultures and conventions may view expressiveness in music.

References

Archer, M. S. and **Elder-Vass, D.** (2012). Cultural system or norm circles? An exchange. *European Journal of Social Theory*, **15**(*1*), 93–115.

Arzt, A. and **Widmer, G.** (2010). Simple tempo models for real-time music tracking. In: *Proceedings of the 7th Sound and Music Computing Conference (SMC 2010)*, 21–24 July 2010, Barcelona, Spain.

Arzt, A., Widmer, G., Böck, S., Sonnleitner, R., and **Frostel, H.** (2012). Towards a Complete Classical Music Companion. In: *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, 27–31 August 2012, Montpellier, France.

Ashley, R. (2002). Do[n't] change a hair for me: the art of jazz rubato. *Music Perception*, **19**(*3*), 311–32.

Behne, K-E. and **Wöllner, C.** (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, 15(*3*), 324–42.

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and **Sandler, M.** (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, **13**(*5*), 1035–47.

Bouënard, A., Wanderley, M. M., Gibet, S., and **Marandola, F.** (2011). Virtual gesture control and synthesis of music performances: qualitative evaluation of synthesized timpani exercises. *Computer Music Journal*, **35**(*3*), 57–72.

Brendel, A. (1976). *Musical Thoughts and Afterthoughts*. London: Robson Books.

Broughton, M. and **Stevens, C.** (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, **37**(*2*), 137–53.

Burdea, G. C. and **Coiffet, P.** (2003). *Virtual Reality Technology*, 2nd edn. Hoboken, NJ: John Wiley & Sons.

Cabrera, D., Ferguson, S., and **Schubert, E.** (2007). Psysound3: Software for acoustical and psychoacoustical analysis of sound recordings. In: *Proceedings of the 13th International Conference on Auditory Display (ICAD)*, Montreal, Canada.

Camurri, A. and **Moeslund, T. B.** (2010). Visual gesture recognition. In: R. I. Gody and M. Leman (Eds), *Musical Gestures: Sound, movement, and meaning* (pp. 238–63). New York: Routledge.

Camurri, A., Castellano, G., Ricchetti, M., and **Volpe, G.** (2006). Subject interfaces: measuring bodily activation during an emotional experience of music. *Gesture in Human–Computer Interaction and Simulation, 3881*, 268–79.

Cannam, C., Landone, C., Sandler, M., and **Bello, J. P.** (2006). The Sonic Visualiser: a visualisation platform for semantic descriptors of musical signals. In: *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, 8–12 October 2006, Victoria, Canada.

Cannam, C., Landone, C., and **Sandler, M.** (2010). Sonic Visualiser: an open source application for viewing, analysing, and annotating music audio files. In: *Proceedings of the ACM Multimedia 2010 International Conference, October 2010, Firenze, Italy* (pp. 1467–8).

Clynes, M. (1973). Sentics: biocybernetics of emotion communication. *Annals of the New York Academy of Sciences*, **220**(*3*), 57–88.

Clynes, M. (1989). Methodology in sentographic measurement of motor expression of emotion—two-dimensional freedom of gesture essential. *Perceptual and Motor Skills*, **68**(*3*), 779–83.

Clynes, M. (1995). Microstructural musical linguistics: composers' pulses are liked most by the best musicians. *Cognition*, **55**(*3*), 269–310.

Collier, G. L. and **Collier, J. L.** (2002). A study of timing in two Louis Armstrong solos. *Music Perception*, **19**(*3*), 463–83.

Cook, N. (1987). Structure and performance timing in Bach's C major prelude (WTC I): an empirical study. *Music Analysis*, **6**(*3*), 100–14.

Cook, N. (1995). The conductor and the theorist: Furtwängler, Schenker and the first movement of Beethoven's Ninth Symphony. In: J. Rink (Ed.), *The Practice of Performance* (pp. 105–25). Cambridge: Cambridge University Press.

Cook, N. (2004). Computational and comparative musicology. In: E. F. Clarke and N. Cook (Eds), *Empirical Musicology. Aims, methods, and prospects* (pp. 103–26). Oxford: Oxford University Press.

Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., and **Schröder, M.** (2000). FEELTRACE: An instrument for recording perceived emotion in real time. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion* (Newcastle, UK: Co. Down), 19–24.

Crist, M. R. (2000). The effect of tempo and dynamic changes on listeners' ability to identify an expressive performance. *Contributions to Music Education*, **27**, 63–77.

Dahl, S. (2004). Playing the accent: comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica*, **90**(*4*), 762–76.

Dahl, S. and **Friberg, A.** (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, **24**(*5*), 433–54.

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., and **Poggi, I.** (2010). Gestures in performance. In: R. I. Godøy and M. Leman (Eds), *Musical Gestures: Sound, movement, and meaning* (pp. 36–68). New York: Routledge.

Dalla Bella, S. and **Palmer, C.** (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS ONE*, **6**(*6*), e20518.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, **21**(*2*), 103–13.

de Cheveigné, A. (2006). Multiple F0 estimation. In: D. L. Wang and G. J. Brown (Eds), *Computational Auditory Scene Analysis: Principles, algorithms and applications* (pp. 45–79). Piscataway, NJ: IEEE Press/Wiley.

de Cheveigné, A. and **Kawahara, H.** (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, **111**(*4*), 1917–30.

Dean, R. T. and **Bailes, F.** (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review*, 5(*4*), 152–75.

Delalande, F. (1988). La gestique de Gould. In: G. Guertin (Ed.), *Glenn Gould: Pluriel* (pp. 85–111). Verdun, Quebec, Canada: Louise Courteau.

Demany, L. and **Semal, C.** (2002). Limits of rhythm perception. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, **55a**(*2*), 643–57.

Dixon, S. (2000). On the computer recognition of solo piano music. *Mikropolyphonie*, **6**, 31–7.

Dixon, S. (2001a). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, **30**(*1*), 39–58.

Dixon, S. (2001b). An interactive beat tracking and visualisation system. In: A. Schloss, R. Dannenberg, and P. Driessen (Eds), *Proceedings of the 2001 International Computer Music Conference, Havana, Cuba* (pp. 215–8). San Francisco, CA: International Computer Music Association.

Dixon, S. (2004). Analysis of musical content in digital audio. In: J. DiMarco (Ed.), *Computer Graphics and Multimedia: Applications, problems, and solutions* (pp. 214–35). Hershey, PA: Idea Group.

Dixon, S. (2006). Onset detection revisited. In: *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx'06), 18–20 September 2006, Montreal, Canada* (pp. 133–7).

Dixon, S., Goebl, W., and **Widmer, G.** (2002), 'The Performance Worm: Real time visualisation based on Langner's representation. In: M. Nordahl (ed.), *Proceedings of the 2002 International Computer Music Conference, Göteborg, Sweden* (pp. 361–64). San Francisco, CA: International Computer Music Association.

Dixon, S., Goebl, W., and **Widmer, G.** (2005). The "Air Worm:" An interface for real-time manipulation of expressive music performance. In: *Proceedings of the 2005 International Computer Music Conference, Barcelona, Spain* (pp. 614–7). San Francisco, CA: International Computer Music Association.

Dixon, S. and **Widmer, G.** (2005). MATCH: A music alignment tool chest. In: *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* (pp. 492–7).

Dixon, S., Goebl, W., and **Cambouropoulos, E.** (2006). Perceptual smoothness of tempo in expressively performed music. *Music Perception*, **23**(*3*), 195–214.

Dixon, S., Mauch, M., and **Tidhar, D.** (2012). Estimation of harpsichord inharmonicity and temperament from musical recordings. *Journal of the Acoustical Society of America*, **131**(*1*), 878–87.

Fabian, D. and **Schubert, E.** (2009). Baroque expressiveness and stylishness in three recordings of the D minor Sarabanda for solo violin (BWV 1004) by JS Bach. *Music Performance Research*, **3**, 36–55.

Fabian, D., Schubert, E., and **Pulley, R.** (2010). A Baroque Träumerei: The Performance and Perception of two Violin Renditions', *Musicology Australia*, **32**(*1*), 27–44.

Ferguson, S., Stevens, C. J., and **Schubert, E.** (2009). *Using dynamic time warping and manual video annotation to compare the time progression of dance performance.* Paper presented at the HCSNet Workshop on Motion Capture, Macquarie University.

Flossmann, S., Goebl, W., and **Widmer, G.** (2009). Maintaining skill across the life span: Magaloff's entire Chopin at age 77. In: A. Williamon, S. Pretty, and R. Buck (Eds), *Proceedings of the International Symposium on Performance Science 2009, 15–18 December 2009, Auckland, New Zealand* (pp. 119–24). Utrecht, The Netherlands: European Association of Conservatoires (AEC).

Flossmann, S., Goebl, W., and **Widmer, G.** (2010a). The Magaloff corpus: an empirical error study. In: S. M. Demorest, S. J. Morrison, and P. S. Campbell (Eds), *International Conference on Music Perception and Cognition (ICMPC11)* (pp. 469–73). Adelaide, Australia: Causal Productions.

Flossmann, S., Goebl, W., Grachten, M., Niedermayer, B., and **Widmer, G.** (2010b). The Magaloff Project: an interim report. *Journal of New Music Research*, **39**(*4*), 363–77.

Flossmann, S., Grachten, M., and **Widmer, G.** (2011). *Expressive performance with Bayesian networks and linear basis models.* Paper presented at the Rencon Workshop 2011: Musical Performance Rendering Competition for Computer Systems, Padova, Italy.

Fredrickson, B. L. and **Kahneman, D.** (1993). Duration Neglect in Retrospective Evaluations of Affective Episodes. *Journal of Personality and Social Psychology*, **65**(*1*), 45–55.

Friberg, A. and **Sundström, A.** (2002). Swing ratios and ensemble timing in jazz performance: evidence for a common rhythmic pattern. *Music Perception*, **19**(*3*), 333–49.

Furuya, S., Flanders, M., and **Soechting, J. F.** (2011). Hand kinematics of piano playing. *Journal of Neurophysiology*, **106**, 2849–64.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, **31**(*3*), 221–72.

Godøy, R. I. and **Leman, M.** (Eds). (2010). *Musical Gestures: Sound, movement, and meaning.* New York: Routledge.

Goebl, W. and **Palmer, C.** (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, **186**(*3*), 471–9.

Goebl, W. and **Palmer, C.** (2009a). Finger motion in piano performance: touch and tempo. In: A. Williamon, S. Pretty, and R. Buck (Eds), *Proceedings of the International Symposium on Performance Science, 15–18 December 2009, Auckland, New Zealand* (pp. 65–70). Utrecht, The Netherlands: European Association of Conservatoires (AEC).

Goebl, W. and **Palmer, C.** (2009b). Synchronization of timing and motion among performing musicians. *Music Perception*, **26***(5)*, 427–38.

Goebl, W., and **Palmer, C.** (2013). Termporal control and hand movement efficiency in skilled music performance, PLOS ONE, 8(1), e50901.

Goebl, W. and **Widmer, G.** (2009). On the use of computational methods for expressive music performance. In: T. Crawford and L. Gibson (Eds), *Modern Methods for Musicology: Prospects, Proposals, and Realities* (pp. 93–113). London: Ashgate.

Goebl, W., Dixon, S., De Poli, G., Friberg, A., Bresin, R., and **Widmer, G.** (2008). "Sense" in expressive music performance: data acquisition, computational studies, and models. In: P. Polotti and D. Rocchesso (Eds), *Sound to Sense—Sense to Sound: A state of the art in sound and music computing* (pp. 195–242). Berlin: Logos.

Goebl, W., Flossmann, S., and **Widmer, G.** (2010). Investigations into between-hand synchronisation in Magaloff's Chopin. *Computer Music Journal*, **34***(3)*, 35–44.

Goebl, W., Pampalk, E., and **Widmer, G.** (2004). Exploring expressive performance trajectories: Six famous pianists play six Chopin pieces. In: S.D. Lipscomp, *et al.* (Eds), *Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL, 2004 (ICMPC8)*(pp.505–09) Adelaide, Australia: Causal Productions.

Grachten, M., Goebl, W., Flossmann, S., and **Weidmer, G.** (2009). Phase-plane representation and visualization of gestural structure in expressive timing. *Journal of New Music Research*, **38***(2)*, 183–95.

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and **Leman, M.** (2010). Musical gestures. Concepts and methods in research. In: L. R. I. Godøy and M. Leman (Eds), *Musical Gestures: Sound, movement, and meaning* (pp. 12–35). New York: Routledge.

Kinney, D. W. (2004). The effect of performing ensemble participation on the ability to perform and perceive expression in music. *International Journal of Music Education*, **56**, 322–37.

Klapuri, A. (2004). *Signal processing methods for the automatic transcription of music*. PhD thesis. Tampere, Finland: Tampere University of Technology. www.cs.tut.fi/sgn/arg/klap/phd/klap_phd.pdf

Kolesnik, P. (2004). *Conducting gesture recognition, analysis and performance system*. Masters thesis. Montreal, Canada: McGill University.

Leech-Wilkinson, D. (2009). *The Changing Sound of Music: Approaches to studying recorded musical performances*. London: Centre for the History and Analysis of Recorded Music (CHARM).

Leman, M., Desmet, F., Styns, F., van Noorden, L., and **Moelants, D.** (2009). Sharing musical expression through embodied listening: a case study based on Chinese Guqin music. *Music Perception*, **26***(3)*, 263–78.

Loehr, J. D. and **Palmer, C.** (2008). Sequential and biomechanical factors constrain timing and motion in tapping. *Journal of Motor Behavior*, **41***(2)*, 128–36.

Lucas, K. V., Hamann, D. L., and **Teachout, D. J.** (1996). Effect of perceptual mode on the identification of expressiveness in conducting. *Southeastern Journal of Music Education*, **8**, 166–75.

Luck, G. and **Sloboda, J. A.** (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception*, **26***(5)*, 465–73.

McAdams, S., Depalle, P., and **Clarke, E. F.** (2004). Analyzing musical sound. In: E. F. Clarke and N. Cook (Eds), *Empirical Musicology. Aims, methods, and prospects* (pp. 157–96). Oxford: Oxford University Press.

MacDougall, H. G. and **Moore, S. T.** (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology*, **99**, 1164–73.

Macrae, R. and **Dixon, S.** (2010). Accurate real-time windowed time warping. In: J. S. Downie and R. C. Veltkamp (Eds), *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 423–8). International Society for Music Information Retrieval.

Madsen, C. K. (1990). Measuring musical response. *Music Educators Journal*, 77(*3*), 26–8.

Morrison, S. J., Price, H. E., Geiger, C. G, and **Cornacchio, R. A.** (2009). The effect of conductor expressivity on ensemble performance evaluation. *Journal of Research in Music Education*, **57**(*1*), 37–49.

Nagel, F., Kopiez, R., Grewe, O., and **Altenmüller, E.** (2007), EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, **39**(*2*), 283–90.

Namba, S. and **Kuwano, S.** (1990). Continuous multi-dimensional assessment of musical performance. *Journal of the Acoustical Society of Japan*, **11**, 43–51.

Namba, S., Kuwano, S., Hatoh, T., and **Kato, M.** (1991). Assessment of musical performance by using the method of continuous judgment by selected description. *Music Perception*, **8**(*3*), 251–75.

Napoles, J. (2013). The influences of presentation modes and conducting gestures on the perceptions of expressive choral performance of high school musicians attending a summer choral camp. *International Journal of Music Education*, **31**(*3*), 321–30.

Nettheim, N. (1999). The statistics of Schubert's keys. *The Schubertian*, **26**, 2–3.

Niedermayer, B. and **Widmer, G.** (2010). Strategies towards the automatic annotation of classical piano music. In: *Proceedings of the 7th Sound and Music Computing Conference* (pp. 118–25). Barcelona: Music Technology Group of the Universitat Pompeu Fabra.

Orio, N., Lemouton, S., and **Schwarz, D.** (2003). Score following: state of the art and new developments. In: *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada* (pp. 36–41).

Ortmann, O. (1929). *The Physiological Mechanics of Piano Technique*. London: Kegan Paul, Trench, Trubner, E. P. Dutton.

Palmer, C. (1997). Music performance, *Annual Review of Psychology*, **48**, 115–38.

Parncutt, R. (2011). The tonic as triad: key profiles as pitch salience profiles of tonic triads. *Music Perception*, **28**(*4*), 333–66.

Pearce, M. T. (2011). Time-series analysis of music: Perceptual and information dynamics. *Empirical Musicology Review*, **6**(*2*), 125–30.

Peddell, L. T. (2008). Factors influencing listeners' perception of expressiveness for a conducted performance. *Bulletin of the Council for Research in Music Education*, **178**, 47–61.

Popper, K. R. and **Eccles, J. C.** (1977). *The Self and its Brain*. Berlin: Springer-Verlag.

Povel, D.-J. (1977). Temporal structure of performed music: some preliminary observations. *Acta Psychologica*, **41**(*4*), 309–20.

Price, H.E. and **Chang, E.C.** (2005). Conductor and ensemble performance expressivity and state festival ratings. *Journal of Research in Music Education*, **53**(*1*), 66–77.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, **88**(*2*), 622–41.

Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America*, **92**(*5*), 2546–68.

Repp, B. H. (1993). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America*, **93**(*2*), 1136–44.

Repp, B. H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, **105**(*3*), 1972–88.

Rozin, A., **Rozin, P.**, and **Goldberg, E.** (2004). The feeling of music past: How listeners remember musical affect. *Music Perception*, **22**(*1*), 15–39.

Sapp, C. (2007). Comparative analysis of multiple musical performances. In: *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), 23–27 September 2007, Vienna, Austria* (pp. 497–500). Vienna: Austrian Computer Society.

Scheirer, E. D. (1995). *Extracting expressive performance information from recorded music*. Masters thesis. Cambridge, MA: Massachusetts Institute of Technology.

Schoonderwaldt, E. (2009). The player and the bowed string: coordination of bowing parameters in violin and viola performance. *Journal of the Acoustical Society of America*, **126**(*5*), 2709–20.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, **51**(*3*), 154–65.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In: P. N. Juslin and J. A. Sloboda (Eds), *Music and emotion: Theory and research* (pp. 393–414). Oxford: Oxford University Press.

Schubert, E. (2010). Continuous self-report methods. In: P. N. Juslin and J. A. Sloboda (Eds), *Handbook of Music and Emotion: Theory, research, applications*. (pp. 223–53). Oxford: Oxford University Press.

Schutz, M. and **Lipscomp, S. D.** (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, **36**(*8*), 888–97.

Snyder, J, and Krumhansl, C.L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception*, **18**(*4*), 455–89.

Stevens, C. J., **Schubert, E.**, **Wang, S.**, **Kroos, C.**, and **Halovic, S.** (2009). Moving with and without music: scaling and lapsing in time in the performance of contemporary dance. *Music Perception*, **26**(*5*), 451–64.

Timmers, R. and **Honing, H.** (2002). On music performance, theories, measurement and diversity. *Cognitive Processing (International Quarterly of Cognitive Sciences)*, **1**(*2*), 1–19.

Toiviainen, P. and **Burger, B.** (2011). *MoCap Toolbox Manual*. Jyväskylä, Finland: University of Jyväskylä.

Toiviainen, P., **Luck, G.**, and **Thompson, M. R.** (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. *Music Perception*, **28**(*1*), 59–70.

Upham, F. (2011). Quantifying the temporal dynamics of music listening: A critical investigation of analysis techniques for collections of continuous responses to music, Master of Arts (McGill University).

Vines, B. W., **Wanderley, M. M.**, **Krumhansl, C. L.**, **Nuzzo, R. L.**, and **Levitin, D. J.** (2003). Performance gestures of musicians: what structural and emotional information do they convey? *Gesture-Based Communication in Human-Computer Interaction*, **2915**, 468–78.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In: I. Wachsmuth and T. Sowa (Eds), *Gesture and Sign Language in Human–Computer Interaction* (pp. 241–53). Berlin: Springer.

Wanderley, M. M., **Vines, B.**, **Middleton, N.**, **McKay, C.**, and **Hatch, W.** (2005). The musical significance of clarinetists' ancillary gestures: an exploration of the field. *Journal of New Music Research*, **34**(*1*), 97–113.

Wen, X. and **Sandler, M.** (2007). New audio editor functionality using harmonic sinusoids. In: *Proceedings of the AES 122nd Convention, 5–8 May 2007, Vienna*.

Widmer, G., **Dixon, S.**, **Goebl, W.**, **Pampalk, E.**, and **Tobudic, A.** (2003). In search of the Horowitz factor. *AI Magazine*, **24**(*3*), 111–30.

Widmer, G. and **Goebl, W.** (2004). Computational models of expressive music performance: the state of the art. *Journal of New Music Research*, **33**(*3*), 203–16.

Widmer, G., Flossmann, S., and **Grachten, M.** (2009). YQX plays Chopin. *AI Magazine*, **30**(*3*), 35–48.

Williamon, A. and **Davidson, J. W.** (2002). Exploring co-performer communication. *Musicae Scientiae*, **6**(*1*), 53–72.

Windsor, W.L. (2009), Measurement and models of performance. In: S. Hallam, I. Cross, and M. Thaut (Eds), *The Oxford Handbook of Music Psychology* (pp. 323–32). Oxford: Oxford University Press.

Zhou, H. and **Hu, H.** (2007). Human motion tracking for rehabilitation—a survey. *Biomedical Signal Processing and Control*, **3**, 1–18.