Chapter 1

'Sense' in Expressive Music Performance: Data Acquisition, Computational Studies, and Models

Werner Goebl[1], Simon Dixon[1], Giovanni De Poli[2],
Anders Friberg[3], Roberto Bresin[3], and Gerhard Widmer[1,4]

[1]Austrian Research Institute for Artificial Intelligence (OFAI), Vienna
[2]Department of Information Engineering, University of Padova;
[3]Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm;
[4]Department of Computational Perception, Johannes Kepler University, Linz

This chapter gives an introduction into basic strands of current research in expressive music performance. A special focus is given on the various methods to acquire performance data either during a performance (e.g., through computer-monitored instruments) or from audio recordings. We then overview the different computational approaches to formalise and model the various aspects in expressive music performance. Future challenges and open problems are tackled briefly at the end of this chapter.

1.1 Introduction

Millions of people are regularly attending live music events or listening to recordings of music performances. What drives them to do so is hard to pin down with certainty, and the reasons for it might be manifold. But while enjoying the music, they are all listening to (mostly) human-made music that contains a specific human expression, whatever kind it might be — what they hear makes sense to them. Without this expressivity the music would not attract people; it is an integral part of the music.

Given the central importance of expressivity (not only in music, but in all communication modes and interaction contexts), it is not surprising that human expression and expressive behaviour have become a domain of intense scientific study. In the domain of music, much research has focused on the act of *expressive music performance*, as it is commonly and most typically found in classical music: the deliberate shaping of the music by the performer, the imposing of expressive qualities onto an otherwise 'dead' musical score via controlled variation of parameters such as intensity, tempo, timing, articulation, etc. Early attempts at quantifying this phenomenon date back to the beginning of the 20th century, and even earlier than that.

If we wish to precisely measure and analyse every detail of an expressive music performance (onset timing, timbre and intensity, duration, etc), we end up with huge amounts of data that quickly become unmanageable. Since the first large-scale, systematic research into expression in music performance (usually of classical music) in the 1930s, this has always been a main problem in this field that was controlled either by reducing the amount of music investigated to some seconds of music, or by limiting the number of performances studied to one or two. Recent approaches try to overcome this drawback by using modern computational methods in order to study, model, and understand this complex interaction of performed events and other information of the performance (e.g., the score and the music structure in the case of "classical music").

In the past ten years, quite some very comprehensive overview papers have been published on the various aspects of music performance research. The probably most cited is Alf Gabrielsson's chapter in Diana Deutsch's book "Psychology of Music" (Gabrielsson, 1999) in which he reviewed over 600 papers in this field published until approximately 1995. In a follow-up paper, he added and discussed another 200 peer-reviewed contributions that appeared until 2002 (Gabrielsson, 2003). A cognitive-psychological review has been contributed by Palmer (1997) summarising empirical research that concentrate on cognitive aspects of music performance such as memory retrieval, anticipatory planning, or motor control. The musicologist's perspective is represented by two major edited books devoted exclusively to music performance research (Rink, 1995, 2002). Lately, more introductory chapters highlight the various methodological issues of systematic musicological performance research (Rink, 2003; Clarke, 2004; Cook, 2004; Windsor, 2004). Two recent contributions surveyed the diversity of computational approaches in modeling expressive music performance (De Poli, 2004; Widmer and Goebl, 2004). Parncutt and McPherson (2002) attempted to bridge the gap between the research on music performance and the music practice by bringing together two authors from each of the two sides for each chapter of this book.

Having this variety of overview papers in mind, we aim in this chapter to give a systematic overview on the more technological side of accessing, measuring, analysing, studying, and modeling expressive music performances. As an outset, we screened the current literature of the past century on the various ways of obtaining expressive data of music performances. Then, we review current computational models for expressive music performance. In a final section we briefly sketch possible future strands and open problems that might be tackled by future research in this field.

1.2 Data Acquisition and Preparation

This section is devoted to very practical issues of obtaining data of various kinds on expressive performance and the basic processing therof. We can distinguish basically two different kinds of obtaining information on music performance. The first is to monitor performances during the production process with various measurement devices (MIDI pianos, accelerometers, movement sensors, video systems, etc.). Specific performance parameters can be accessed directly (hammer velocity of each played tone, bow speed, fingering, etc.). The other way is to extract all these relevant data from the recorded audio signal. This method has the disadvantage that some information easily to extract during performance is almost impossible to gain from the audio domain (think for instance of the right pedal at the piano). The advantage, however, is that we have now over a century of recorded music at our disposal that could serve as valuable sources for various kinds of scientific investigation. In the following sub-sections, we discuss the various approaches for monitoring and measuring music performance.

1.2.1 Using Specially Equipped Instruments

Before computers and digital measurement devices were invented and easily available for everyone, researchers employed a vast variety of mechanical and electric measurement apparati to capture all sorts of human or mechanical movements at musical instruments.

Historical Measurement Devices

Mechanical and Electro-Mechanical Setups Of the first to record the movement of piano keys were Binet and Courtier (1895) who used a 6-mm caoutchouc rubber tube placed under the keys that was connected to a cylindric graphical recorder that captured continuous air pressure resulting from striking different keys on the piano. They investigated some basic pianistic tasks such as playing trills, connecting tones, or passing-under of the thumb in scales with exemplary material. In the first of the two contributions of this study, Ebhardt (1898) mounted metal springs on a bar above the strings that closed a electrical shutter when the hammer was about to touch the strings. The electric signal was recorded with a kymograph and timed with a 100-Hz oscillator. He studied the timing precision of simple finger tapping and playing scales. Further tasks with binary and ternary metrum revealed lengthening the IOI of an accentuated onset. Onset and offset timing of church hymn performances were investigated by Sears (1902). He equipped a reed organ with mercury contacts that registered key depression of 10 selected keys. This information was recorded on four tracks on the surface of a smoked kymograph drum. He studied several temporal aspects of performances by four organ players, such as duration of the excerpts, bars, and individual note values, accent behavior, or note overlap (articulation).

A multitude of mechanical measurement devices introduced Ortmann (1925, 1929) in studies on physiological determinants of piano playing. To investigate the different behaviors of the key, he mounted a tuning fork aside one piano key that scribe wave traces into smoked paper that vary with the speed of the key. With this setup, he was one of the first to study the response of the key on different pianistic playing techniques. For assessing finger movements, Ortmann (1929, p. 230) used a purpose-built mechanical apparatus that comprises non-flexible aluminum strips that are on one side connected to either the finger (proximal phalanx) or the key surface and on the other side they write on a revolving drum. With this apparatus continuous displacement of finger and key could be recorded and analysed. Another mechanical system was the "Pantograph" (Ortmann, 1929, p. 164), a parallelogram lever construction to record lateral arm movement. For other types of movement, he used active optical systems. The motion of a tiny light bulb attached to the wrist or the finger leaves a clear trace on a photo plate (the room in very subdued light), when the shutter of the photo camera remains open for entire duration of the movement.

Similar active markers mounted on head, shoulder, elbow, and wrist were used by Bernstein and Popova in their important 1930 study (reported by Kay et al., 2003) to study the complex interaction and coupling of the limbs in piano playing. They used their "kymocyclographic camera" to record the movements of the active markers. A rotating shutter allows the light of the markers to impinge on the constantly moving photographic film. With this device they could record up to 600 instances of the movement per second.

Piano Rolls as Data Source A source of expression data are piano rolls for reproducing pianos that exist from different manufacturers (e.g., Welte-Mignon, Hupfeld, Aeolian Duo-Art, Ampico) and of performances of a manifold of renowned pianists (Bowers, 1972; Hagmann, 1984). They were the first means to record and store artistic music performances before the gramophone has been invented. Starting in the late 1920s, scientists took advantage of this source of data and investigated various aspects of performance. Heinlein (1929a,b, 1930) used Duo-Art rolls of the Aeolian company to study pedal use of four pianists playing Schumann's *Träumerei*. Rolls of the same company were the basis of Vernon's 1936 study. He investigated vertical synchronisation of the tones in a chord (see Goebl, 2001). Hartmann (1932) used Hupfeld "Animatic Rolls" and provided a very detailed study on tone and bar durations as well as note onset asynchronies in two recordings of the first movement of Beethoven's Op. 27 No. 2 (Josef Pembaur, Harold Bauer). Since the precise recording procedures by these companies are still unknown because they deliberately were hold back for commercial reasons, the authenticity of these rolls is sometimes questionable (Hagmann, 1984; Gottschewski, 1996). For example, the Welte-Mignon system were able to simultaneously control dynamics only for keyboard halves. Hence, to emphasise the melody note and to play the rest of the chord tones softer was only possible on such a system when the melody tone was played at a different point in time than the others (Gottschewski, 1996, pp. 26–42). Although we know today that pianists anticipate melody notes (Palmer, 1996b; Repp, 1996a; Goebl, 2001), the Welte-Mignon rolls cannot be taken literally as a source for studying note asynchronies (as done by Vernon, 1936). The interpretation of piano rolls need to carefully performed having in mind the conditions of their production. There are currently some private attempts to systematically scan in piano rolls and transform them into standard symbolic format (e.g., MIDI). However, we are not aware of any scientific project concerned with this.

The Iowa Piano Camera During the 1930's, Carl E. Seashore guided a research group that focused on different aspects of music performance, namely the singing voice, violin playing, and piano performance (Seashore, 1932, 1936a,b). They developed various measurement setups for scientific investigation, among those most prominently the "Iowa Piano Camera" (Henderson et al., 1936) that captured optically onset and offset times and hammer velocity of each key and additionally the movement of the two pedals. It was therefore a complete and comparably very precise device that was not topped until the present days computer-controlled pianos (such as the Disklavier or the SE, see Goebl and Bresin, 2003). Each hammer is equipped with a shutter that controls light exposure onto a moving film. The hammer shutter interrupts (as in later computer-control reproducing pianos) twice the light exposure on the film: a first time from 24 to 12 mm before the hammer touches the strings and a second time at hammer–string contact. The average hammer speed of the last 12 mm of the hammer's travel can be inferred from the distance on the film between these two interrupts (today's computer-controlled pianos take the average speed of the final 5 mm). According to Skinner and Seashore (1936), the temporal resolution goes down to 10 ms. The hammer velocity gets quantised into 17 dynamics categories (Henderson, 1936). With this system, the IOWA group performed several studies with professional pianists. Henderson (1936) had two professionals playing the middle section of Chopin's Nocturne Op. 15 No. 3. In this very comprehensive study, they examine temporal behavior, phrasing, accentuation, pedalling, and chord asynchronies. Skinner and Seashore (1936) analysed repeated performances of pieces by Beethoven and Chopin and found high timing consistency within the pianists.

Contemporary Measurement Devices

Henry Shaffer's Photocell Bechstein After the efforts of Seashore's research group at Iowa, it took over 40 years before a new initiative included modern technology to capture piano performance. It was L. Henry Shaffer at Exeter who equipped each of the 88 tones of a Bechstein grand piano with pairs of photocells and the two pedals to capture the essential expressive parameters of piano performance (Shaffer, 1980, 1981, 1984; Shaffer et al., 1985; Shaffer and Todd, 1987; Shaffer, 1992). The optical registration of the action's movements had the advantage not to affect the playability of the piano. The photocells were mounted into the piano action in pairs, each capturing the moment of the hammer's transit. One was placed to register the instant of hammer–string contact, the other one the resting position of the hammer. The position of the two pedals were monitored by micro switches and stored as 12-bit words on the computer. Each such event was assigned a time stamp rounded to the nearest microsecond and stored on a computer. The sensor at the strings yielded the note onset time, the one at the hammer's resting position (when the hammer returns) the note offset time. The time difference between the two sensors is an inverse estimate of the force at which the key was depressed. Already then, the introduced technology was in principle identical to the commercially available computer-monitored pianos until now (e.g., the Yamaha Disklavier series or the Bösendorfer SE). This device was used also by other members of that laboratory (e.g., Clarke, 1982, 1985; Todd, 1985, 1989, 1992)

Studies with Synthesiser Keyboards or Digital Pianos Before computer-monitored acoustic pianos became widely distributed and easily available, simple synthesiser keyboards or digital pianos were used to capture expressive data from music performances. These devices provide timing and loudness data for each performed event through the standardised digital communications protocol MIDI (Musical Instrument Digital Interface) that can be stored in files on computer hard-disks (Huber, 1999) and used as an ideal data source for expression. However, such keyboards do not provide a realistic performance setting for advanced pianists, because the response of the keys is very different from an acoustic piano and the synthesised sound (especially with extensive use of the right pedal) does not satisfy trained ears of highly-skilled (classical) pianists.

Such electronic devices were used for various general expression studies (e.g., Palmer, 1989, 1992; Repp, 1994a,b, 1995a; Desain and Honing, 1994). Bruno Repp repeated two of his studies that were first performed with data from a digital piano (one concerned with legato articulation, Repp, 1995a, the other with the use of the right pedal, Repp, 1996b) later on a computer-controlled grand piano (Repp, 1997a,d, respectively). Interestingly, the results of both pairs of studies were similar to each other, even though the acoustic properties of the digital piano were considerably different from the grand piano.

The Yamaha Disklavier System Present performance studies dealing with piano performances make generally use of commercially available computer-controlled acoustic pianos. Apart from systems that can be built into a piano (e.g., Autoklav, Pianocorder, see Coenen and Schäfer, 1992), the most common is the Disklavier system by Yamaha. The first computer-controlled grand pianos was available from 1989 onwards (e.g., MX100A/B, DGP); a revised version was issued in 1992 (e.g., MX100II, DGPII, all informations derived from personal communication with Yamaha Rellingen, Germany). The Mark II series was retailed since 1997, the Mark III series followed approximately in 2001. Currently, the Mark IV series can be purchased that includes also a computer with screen and several high-level functions such as an automatic accompaniment system. From 1998, Yamaha introduced their high-end PRO series of Disklaviers that involves an extended MIDI format to store more than 7-bit velocity information (values from 0 to 127) and information on key release.

There were few attempts to assess the Disklavier's accuracy of recording and reproducing performances. Coenen and Schäfer (1992) compared various reproducing systems (among them a Disklavier DG2RE and a SE225) on their applicability for compositorical purposes (reproducing compositions for mechanical instruments). Maria (1999) had a Disklavier DS6 Pro at his disposal and tested its precision in various ways. More systematic tests on recording and reproduction accuracy were performed by Goebl and Bresin

(2001, 2003) using accelerometer registration to inspect key and hammer movements during recording and reproduction.

Yamaha delivers both upright and grand piano versions of its Disklavier system. One of the first to investigate an early upright Disklavier (MX100A) was Bolzinger (1995) who found a logarithmic relationship between MIDI velocity values and sound pressure level (dB). This upright model was used for several performance studies (Palmer and van de Sande, 1993; Palmer and Holleran, 1994; Repp, 1995b,c, 1996a,c,d, 1997b,c).

The Yamaha Disklavier grand piano was even more widely used in performance research. Moore (1992) combined data from a Disklavier grand piano with electromyographic recordings of the muscular activity of four performers playing trills. Behne and Wetekam (1994) recorded student performances of the theme of Mozart's K.331 on a Disklavier grand piano and studied systematic timing variations of the Siciliano rhythm. As mentioned above, Repp repeated his work on legato and pedalling on a Disklavier grand piano Repp (1997a,d, respectively). Juslin and Madison (1999) used a Disklavier grand piano to record and play back different (manipulated) performances of two melodies to assess listeners' ability to recognise simple emotional categories. Bresin and Battel (2000) analysed multiple performances recorded on a Disklavier grand piano of Mozart's K.545 in terms of articulation strategies. Clarke and Windsor (2000) used recordings made on a Disklavier grand piano for perceptual evaluation of real and artificially created performances. A short piece by Beethoven was recorded on a Disklavier grand piano played by either one professional pianist (Windsor et al., 2001) or by 16 professional pianists (Timmers et al., 2002; Timmers, 2002) in different tempi. Timing characteristics of the different types of grace notes were investigated. Riley-Butler (2002) used a Disklavier grand piano in educational settings. She showed piano roll representations of student's performances to them and observed considerable increase of learning effectivity with this method.

Bösendorfer's SE System The SE ("Stahnke Electronics") System dates back to the early 1980s when the engineer Wayne Stahnke developed a reproducing system in cooperation with the MIT Artificial Intelligence Laboratory built into a Bösendorfer Imperial grand piano (Roads, 1986; Moog and Rhea, 1990). A first prototype was ready in 1985; the system had been officially sold by Kimball (at that time owner of Bösendorfer) starting from summer 1986. This system was very expensive and only few academic institution could afford it. Until the end of its production, only about three dozen of these systems have been built and sold. The SE works in principle like the Disklavier system (optical sensors register hammershank speed and key release and linear motors reproduce final hammer velocity, see for details Goebl and Bresin, 2003). However, its recording and reproducing capabilities are superior even compared with other much younger systems (Goebl and Bresin, 2003). Despite its rare occurrence in academic institutions, it was used for performance research in some cases.

Palmer and Brown (1991) performed basic tests on the relation of hammer velocity and peak amplitude of the outcoming sound. Repp (1993) tried to estimate peak sound level of piano tones from the two lowest partials as measured in the spectrogram and compared a digital piano, a Disklavier MX100A upright piano with the Bösendorfer SE. Studies in music performance were accomplished at Ohio State University (Palmer and van de Sande, 1995; Palmer, 1996b,a), at Musichochschule Karlsruhe (e.g., Mazzola and Beran, 1998; Mazzola, 2002, p. 833), or at the grand piano located at the Bösendorfer company in Vienna (Goebl, 2001; Widmer, 2001, 2002b, 2003; Goebl and Bresin, 2003; Widmer, 2005).

Currently (June 2005), the Bösendorfer company in Vienna is developing a new computer-controlled reproducing piano called "CEUS" (personal communication with Bösendorfer Vienna) that introduces among other features sensors that register the continuous motion of each key. This data might be extremely valuable for performance studies into the pianists' touch and tone control.

1.2.2 Measuring Audio By Hand

In contrast to measuring music expression during performance through any kind of sensors placed in or around the performer or the instrument (see previous section), the other approach is to analyse the recorded sound of music performances. It has the essential advantage that any type of recording may serve as a basis for investigation, e.g., commercially available CDs, historic recordings, or recordings from ethnomusicological research. One has simply to go into a record store and buy all the famous performances by the great pianists of the past century.[1]

However, to extract discrete performance information from audio is difficult and sometimes impossible. The straight-forward method is to inspect the wave form of the audio signal with computer software and mark manually with a cursor the onset times of selected musical events. Though this method is time consuming, it delivers timing information with a reasonable precision. To extract data on dynamics is a bit more complicated (e.g., by reading peak energy values from the root-mean-square of the signal averaged over a certain time window), but only possible for overall dynamics. We are not aware of a successful procedure to extract individual dynamics of simultaneous tones (for an attempt, see Repp, 1993). Many other signal processing problems have not been solved as well (e.g., extracting pedal information, tone length, etc., see also McAdams et al., 2004).

First studies that extracted timing information directly from sound used oscillogram filming (e.g., Bengtsson and Gabrielsson, 1977, for more references see Gabrielsson, 1999, p. 533). Povel (1977) analysed gramophone records of three performances of Johann Sebastian Bach's first prelude of WTC I. He determined the note onsets "by eye" from two differently obtained oscillograms of the recordings (that were transferred on analog tape). He reported a temporal precision of 1–2 ms (!). Recordings of the same piece were investigated by Cook (1987) who obtained timing (and intensity) data already through a computational routine. The onset detection was automated by a threshold procedure applied to the digitised sound signal (8 bit, 4 kHz) and post corrected by hand. He reported

[1] In analysing recordings the researcher has to be aware that almost all records are glued together from several takes so the analysed performance might never have taken place in this particular rendition (see also Clarke, 2004, p. 88).

a timing resolution of 10 ms. He also stored intensity values, but did not specify in more detail what exactly was measured here.

Gabrielsson et al. (1983) analysed timing patterns of performances from 28 different monophonic melodies played by 5 performers. The timing data were measured from the audio recordings with a precision of ±5 ms (p. 196). In a later study, Gabrielsson (1987) extracted both timing and (overall) intensity data from the theme of Mozart's K.331. In this study, a digital sampling system was used that allowed a temporal precision of 1–10 ms (p. 87). The dynamics were estimated by reading peak amplitudes of each score event (in voltages). Nakamura (1987) used a Brüel & Kjær level recorder to register dynamics of solo performances played on a violin, oboe, and recorder. He analysed the produced dynamics in relation to the perceived intensity of the music.

The first larger corpus of recordings was measured by Repp (1990) who fed 19 recordings of the third movement of Beethoven's Op. 31 No. 3 into a VAX 11/780 computer and read off the note onsets from waveform displays. In cases of doubt, he played the sound until the onset and moved the cursor stepwise back in time, until the following note was no longer audible (Repp, 1990, p. 625). He measured the performances on quarter-note level[2] and reported an absolute mean error of 6.5 ms for repeated measurements (equivalent to 1% of the inter-onset intervals, p. 626). In a further study, Repp (1992) had 28 recordings of Schumann's "Träumerei" by 24 renowned pianists at his disposal. This time, he used a standard waveform editing program to hand-measure the 10-kHz sampled audio files. The rest of the procedure was identical (aural control of ambiguous onsets). He reported an average absolute measurement error of 4.3 ms (or less than 1%). In his later troika on the "microcosm of musical expression" (Repp, 1998, 1999a,b), he applied the same measurement procedure on 115 performances of the first five bars of Chopin's Op. 10 No. 3 Etude collected from libraries and record stores. He used "SoundEdit16" software to measure the onset on sixteenth note level. In addition to previous work, he extracted overall intensity information as well (Repp, 1999a) by taking the peak sound levels (pSPL in dB) extracted from the root-mean-square (RMS) integrated sound signal (over a rectangular window of 30 ms).

Nettheim (2001) measured parts of recordings of four historical performances of Chopin e-minor Nocturne Op. 72 No. 1 (Pachmann, Godowsky, Rubinstein, Horowitz). He used a time-stretching software ("Musician's CD Player," par. 8) to reduce the playback speed by factor 7 (without changing the pitch of the music). He then simply took the onset times from a time display during playback. Tone onsets of all individual tones were measured with this method.[3] In repeated measurements, he reported accuracy of the order of 14 ms. In addition to note onset timing, he assigned arbitrary intensity values to each tone ranging from 1 to 100 by ear (par. 11). He reports about the difficulties arising from that approach.

In recent contributions on timing and synchronisation in Jazz performances, the timing of the various instruments of Jazz ensembles were investigated. Friberg and Sundström

[2]In the second part of this paper, he measured and analysed eight-note and sixteenth-note values as well.

[3]Obviously, the chosen excerpts were slow pieces with a comparatively low note density.

(2002) measured cymbal onsets from spectrogram displays with a reported precision of ±3 ms. Ashley (2002) studied the synchronisation of the melody instruments with the double bass line. He repeatedly measured onsets of both lines from wave form plots of the digitised signal with usual differences between the measurements of 3-5 ms. About the same consistency (typically 2 ms) was achieved by Collier and Collier (2002) through a likewise measurement procedure ("CoolEdit 96," manual annotation of physical onsets in trumpet solos). They exemplified an equivocal situation where the trumpet tone "emerges from the band" (p. 468). In those cases, they aurally determined the onset. Lisboa et al. (2005) used "Pro Tools" wave editor to extract onset timing oc Cello solo performances; Moelants (2004) made use of a speech transcription software ("Praat") to assess trill and ornament timing in solo string performances.

In a recent commercial enterprise, John Q. Walker and colleagues have been trying to extract the complete performance information out of historical (audio) recordings in order to play them back on a modern Disklavier.[4] Their commercial aim is to re-sell old recordings with modern sound quality or live performance feel. They computationally extract as much performance information as possible and add the missing information (e.g., tone length, pedalling) to an artificially created MIDI file. They use it to control a modern Disklavier grand piano and compare this performance to the original recording. Then they modify the added information in the MIDI files and play it back again and repeat this process iteratively until the Disklavier's reproduction sounds identical to the original recording (see also Midgette, 2005).

Another way of assessing temporal content of recordings is by repeatedly tapping along with the music recording e.g., on a MIDI drum pad or the like and recording this information (Cook, 1995; Bowen, 1996; Bachmann, 1999). This is a comparably fast method to gain rough timing data on a tappable beat level. However, perceptual studies on tapping along with expressive music showed that tappers — even after repeatedly tapping along with the same short piece of music — still underestimate abrupt tempo changes or systematic variations (Dixon et al., 2005).

1.2.3 Computational Extraction of Expression from Audio

Several approaches exist for the extraction of expression from audio data, or equivalently, annotating audio data with content-based metadata. The most general approach is to attempt to extract as much musical information as possible, using an automatic transcription system, but such systems are not robust enough to provide the level of precision and accuracy required for analysis of expression (Klapuri, 2004). Nevertheless, some systems were developed with the specific goal of expression extraction, in an attempt to relieve some of the painstaking effort of manual annotation (e.g., Dixon, 2000). Since the score is often available for the musical performances being analysed, Scheirer (1997) recognised that much better performance could be obtained by incorporating score information into the audio analysis algorithms, but the system was never developed to be sufficiently general

[4] http://www.zenph.com

or robust to be used in practice. One thing that was lacking from music analysis software was an interface for interactive editing of partially correct automatic annotations, without which the use of the software was not significantly more efficient than manual annotation.

The first system with such an interface was BeatRoot (Dixon, 2001a,b), an automatic beat tracking system with a graphical user interface which visualised (and auralised) the audio and derived beat times, allowing the user to edit the output and retrack the audio data based on the corrections. BeatRoot produces a list of beat times, from which tempo curves and other representations can be computed. Although it has its drawbacks, this system has been used extensively in studies of musical expression (Goebl and Dixon, 2001; Dixon et al., 2002; Widmer, 2002a; Widmer et al., 2003; Goebl et al., 2004). Recently, Gouyon et al. (2004) implemented a subset of BeatRoot as a plugin for the audio editor WaveSurfer (Sjölander and Beskow, 2000).

A similar methodology was applied in the development of JTranscriber (Dixon, 2004), which was 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, with audio playback of the original and the transcription, together or separately, possible at any time.

These tools provide a better approach than manual annotation, but since they have no access to score information, they still require a significant amount of interactive correction, so that they are not suitable for very large scale studies. An alternative approach is to use existing knowledge, such as from previous annotations of other performances of the same piece of music and transfer the metadata after aligning the audio files. The audio alignment system MATCH (Dixon and Widmer, 2005) finds optimal alignments between pairs of recordings, and is then able to transfer annotations from one recording to the corresponding times in the second. This proves to be a much more efficient method of annotating multiple performances of the same piece, since manual annotation needs to be performed only once. Further, audio alignment algorithms are much more accurate than techniques for direct extraction of expressive information from audio data, so the amount of subsequent correction for each matched file is much less.

Taking this idea one step further, the initial step of annotation can be avoided entirely if the musical score is available in a symbolic format, by synthesising a mechanical performance from the score and matching the audio recordings to the synthetic performance. For analysis of expression in audio, e.g. absolute measurements of tempo, the performance data must be matched to the score, so that the relationship between actual and nominal durations can be computed. Several score-performance alignment systems have been developed for various classes of music (Cano et al., 1999; Soulez et al., 2003; Turetsky and Ellis, 2003; Shalev-Shwartz et al., 2004).

Other relevant work is the on-line version of the MATCH algorithm, which can be used for tracking live performances with high accuracy (Dixon, 2005b,a). This system is being developed for real time visualisation of performance expression. The technical issues are similar to those faced by score following systems, such as those used for automatic accompaniment (Dannenberg, 1984; Orio and Déchelle, 2001; Raphael, 2004), although the goals are somewhat different. Matching involving purely symbolic data has also been explored. Cambouropoulos developed a system for extracting score files from expressive performances in MIDI format (Cambouropoulos, 2000). After manual correction, the matched MIDI and score files were used in detailed studies of musical expression. Various other approaches to symbolic score-performance matching are reviewed by Heijink et al. (2000b,a).

1.2.4 Extracting Expression from Performers Movements

While the previous sections dealt with the extraction of expression contained in music performances, this section is devoted to expression as represented in all kinds of movements that occur when performers interact with their instruments during performance (for an overview, see Davidson and Correia, 2002; Clarke, 2004). Performers' movements are a powerful communication channel of expression to the audience, sometimes even overriding the acoustic information (Behne, 1990; Davidson, 1994).

There are several ways to monitor performers' movements. One possibility is to connect mechanical devices to the playing apparatus of the performer (e.g., Ortmann, 1929) that has the disadvantage to inhibit the free execution of the movements. More common are optical tracking systems that either simply video-tape performers movements or record special passive or active markers placed on particular joints of the performers' body. We already mentioned an early study by Berstein and Poppova (1930) who introduced an active photographical tracking system (Kay et al., 2003). These systems use light-emitting makers placed on the various limbs and body parts of the performer. They are recorded by video cameras that are connected to software that extracts the position of the markers (e.g., the Selspot System, as used by Dahl, 2004, 2005). The disadvantage of those systems is that the participants need to be cabled which is a time-consuming process and the cables might inhibit the participants to move as they would move normally. Passive systems use reflective markers that are illuminated by external lamps. In order to create a three-dimensional picture of movement, the data from several cameras are coupled by software (e.g., Palmer and Dalla Bella, 2004).

Even less intrusive are video systems that simply record performance movements without any particular marking of the performer's limbs. Elaborated software systems are able to track defined body joints directly from the plain video signal (e.g., EyesWeb[5], see Camurri et al., 2004, 2005 or Camurri and Volpe, 2004 for an overview in gesture-related research). Perception studies on communication of expression through performers gestures use simpler point-light video recordings (reflective markers on body joints recorded in a darkened room) to present them to participants for ratings (Davidson, 1993).

1.2.5 Extraction of Emotional Content from MIDI and Audio

For listeners and musicians, an important aspect of music is its ability to express emotions (Juslin and Laukka, 2004). An important research question has been to investigate

[5]`http://www.megaproject.org`

the coupling between emotional expression and the underlying musical parameters. Two important distinctions have to be made. The first distinction is between perceived emotional expression ("what is communicated") and induced emotion ("what you feel"). Here, we will concentrate on the perceived emotion which has been the focus of most of the research in the past. The second distinction is between compositional parameters (pitch, melody, harmony, rhythm) and performance parameters (tempo, phrasing, articulation, accents). The influence of compositional parameters has been investigated during a long time starting with the important work of Hevner (1937). A comprehensive summary is given in Gabrielsson and Lindström (2001). The influence of performance parameters has recently been investigated in a number of studies (for overviews see Juslin and Sloboda, 2001; Juslin, 2003). These studies indicate that for basic emotions such as happy, sad or angry, there is a simple and consistent relationship between the emotional description and the parameter values. For example, a sad expression is characterised by slow tempo, low sound level, legato articulation and a happy expression is characterised by fast tempo, moderate sound level and staccato articulation.

Predicting the emotional expression is usually done using a two-step process (see also Lindström et al., 2005): (1) Parameter extraction The first step extracts the basic parameters from the incoming signal. The selection of parameters is a trade-off between what is needed in terms of emotion mapping and what is possible. MIDI performances are the simplest case in which the basic information in terms of notes, dynamics and articulation is already available. From this data it is possible to deduce for example the tempo using beat-tracking methods as described above. Audio from monophonic music performances can also be analyzed on the note-level giving similar parameters as for the MIDI case (with some errors). In addition, using audio a few extra parameters are available such as the spectral content and the attack velocity. The CUEX algorithm by Friberg et al. (2005), including a real-time version (Friberg et al., 2002), was specifically designed for prediction of emotional expression yielding eight different parameters for each recognised note. Polyphonic audio is the most difficult case which has only recently been considered. Due to the analysis difficulty several approaches can be envisioned. One possibility is to first make a note extraction using the recent advances in polyphonic transcription mentioned above (e.g., Klapuri, 2004) and then extract the parameters. Due to the lack of precision of polyphonic transcription there will be many errors. However, this may not be too important for the prediction of the emotion in the second step below since preferably the mapping is redundant and insensitive to small errors in the parameters. A more straight-forward approach is to extract overall parameters directly from audio, such as using auditory-based measures for pitch, rhythm and timbre (Leman et al., 2004; Liu et al., 2003). (2) Mapping The second step is the mapping from the extracted parameters to the emotion character. The selection of method is dependent on the use (research or real time control) and the desired behaviour of the output data. A typical data-driven method is to use listener ratings (the "right" answer) for a set of performances to train a model. Common statistical/mathematical models are used such as regression (Leman et al., 2004; Juslin, 2000), bayesian networks (Canazza et al., 2003), or hidden markov models (Dillon, 2003). An alternative approach more suitable for real time control is to directly implement qualitative data from previous studies using a fuzzy logic model (Seif El-Nasr et al., 2000; Friberg, 2005), see also Section **??**.

1.3 Computational Models of Music Performance

Models describe relations among different kinds of observable (and often measurable) information about a phenomenon, discarding details that are felt to be irrelevant. They serve to generalise the findings and have both a descriptive and predictive value. Often the information is quantitative and we can distinguish input data, supposedly known, and output data, which are inferred by the model. In this case, inputs can be considered as the causes and output the effect of the phenomenon. When a model can be implemented on a computer, it is called computational model and it allows deducing the values of output data corresponding to the provided values of inputs. This process is called simulation and it is widely used to predict the behaviour of the phenomenon in different circumstances and can be used to validate the model, by comparing the predicted results with actual observations.

In music performance modelling, the information that can be considered is not only quantitative, as *physical information*, e.g. timing or performer's movements. We have also *symbolic information* that refers more to a cognitive organization of the music than to an exact physical value and *expressive information* more related to the affective and emotional content of the music. Recently computer science and engineering started paying attention to expressive information and developing suitable theories and processing tools giving rise to the field of affective computing and Kansei information processing. Music and music performance in particular, attracted the interest of researchers for developing and testing such tools. Music indeed is the more abstract of the arts and has a long tradition of formalization. Moreover it combines in an interesting way all these aspects.

1.3.1 Modeling Strategies

We may distinguish some strategies in developing the structure of the model and in finding its parameters. The most prevalent ones are analysis-by-measurement and analysis-by-synthesis. Recently some methods from artificial intelligence started being developed: machine learning and case based reasoning. We may distinguish local models, that acts at note level and try to explain the observed facts in a local context, and global models that take into account the higher level of the musical structure or more abstract expression pattern. The two approaches often require different modelling strategies and structures. In certain cases, it is possible to devise a combination of both approaches with the purpose being to obtain better results. The composed models are built by several components, each one aiming to represent the different sources of expression. However, a good combination of the different parts is still quite challenging.

Analysis By Measurements

The first strategy, analysis-by-measurements, is based on the analysis of deviations from the musical notation measured in recorded human performances. The analysis aims to recognise regularities in the deviation patterns and to describe them by means of a mathematical model, relating score to expressive values (see Gabrielsson 1999 and Gabrielsson 2003 for an overview of the main results). The method starts by selecting the performances to be analyzed. Often rather small set of carefully selected performances are used. Then the physical properties of every note are measured using the methods seen in section 1.2 and the data so obtained are checked for reliability and consistency. The most relevant variables are selected and analyzed by statistical methods. The analysis assumes an interpretation model that can be confirmed or modified by the results of the measurements. Often the hypothesis that deviations deriving from different patterns or hierarchical levels can be separated and then added is implicitly assumed. This hypothesis helps the modelling phase, but may be oversimplified. Several methodologies of approximation of human performances were proposed using neural network techniques or fuzzy logic approach or using a multiple regression analysis algorithm or linear vector space theory. In these cases, the researcher devises a parametric model and then estimates its parameters that best approximate a set of given performances.

Many models address very specific aspects of expressive performance, for example, the final ritard and its relation to human motion (Kronman and Sundberg, 1987; Todd, 1995; Friberg and Sundberg, 1999; Sundberg, 2000; Friberg et al., 2000b; Hong, 2003); the timing of grace notes (Timmers et al., 2002); vibrato (Desain and Honing, 1996; Schoonderwaldt and Friberg, 2001); melody lead (Goebl, 2001, 2003); legato (Bresin and Battel, 2000); or staccato and its relation to local musical context (Bresin and Widmer, 2000; Bresin, 2001).

A global approach was pursued by Todd in his phrasing model (Todd, 1992, 1995). This model assumes that the structure of a musical piece can be decomposed in a hierarchical sequence of segments, where each segment is on its turn decomposed in a sequence of segments. The performer emphasises the hierarchical structure by an accelerando-ritardando pattern and by a crescendo-decrescendo pattern for each segment. These patterns are superimposed (summed) onto each other and describe from the global variation over the whole to local fluctuations at the note level.

Analysis By Synthesis

While analysis by measurement develop models that best fit quantitative data, the analysis-by-synthesis paradigm takes into account the human perception and subjective factors. First, the analysis of real performances and the intuition of expert musicians suggest hypotheses that are formalised as rules. The rules are tested by producing synthetic performances of many pieces and then evaluated by listeners. As a result the hypotheses are refined, accepted or rejected. This method avoids the difficult problem of objective comparison of performances, including subjective and perceptual elements in the development loop. On the other hand, this method depends too much on the personal competences and taste of few experts.

The most important one is the KTH rule system (Friberg, 1991, 1995; Friberg et al., 1998, 2000a; Sundberg et al., 1983, 1989, 1991). In the KTH system, the rules describe quantitatively the deviations to be applied to a musical score, in order to produce a more attractive and human-like performance than the mechanical one that results from a literal playing of the score. Every rule tries to predict (and to explain with musical or psychoacoustic principles) some deviations that a human performer is likely to insert. Many rules are based on low-level structural analysis of the text. The KTH rules can be grouped according to the purposes that they apparently have in music communication. Differentiation rules appear to facilitate categorization of pitch and duration, whereas grouping rules appear to facilitate grouping of notes, both at micro and macro level.

Machine Learning

In the traditional way of developing models, the researcher normally makes some hypothesis on the performance aspects s/he want to model and then s/he tries to establish the empirical validity of the model by testing it on real data or on synthetic performances. A different approach, pursued by Widmer and coworkers (Widmer, 1995a,b, 1996, 2000, 2002b; Widmer and Tobudic, 2003; Widmer, 2003; Widmer et al., 2003; Widmer, 2005; Tobudic and Widmer, 2005), instead tries to extract new and potentially interesting regularities and performance principles from many performance examples, by using machine learning and data mining algorithms. The aim of these methods is to search for and discover complex dependencies on very large data sets, without any preliminary hypothesis. The advantage is the possibility of discover new (and possibly interesting) knowledge, avoiding any musical expectation or assumption. Moreover, these algorithms normally allow describing discoveries in intelligible terms. The main criteria for acceptance of the results are generality, accuracy, and simplicity.

Models were developed to predict local, note-level expressive deviations and higher-level phrasing patterns. Moreover, these two types of models can be combined to yield an integrated, multi-level model of expressive timing and dynamics.

Case-Based Reasoning

An alternative approach, much closer to the observation-imitation-experimentation process observed in humans, is that of directly using the knowledge implicit in human performances samples. Case-based reasoning (CBR) is based on the idea of solving new problems by using (often with some kind of adaptation) similar previously solved problems. An example in this direction is the SaxEx system for expressive performance of Jazz ballads (Arcos et al., 1998; López de Mántaras and Arcos, 2002) which predicts expressive transformations to saxophone phrases recordings by looking at how other, similar phrases were played by a human musician. The success of this approach greatly depends on the availability of a large amount of well-distributed previously solved problems, that are not easy to collect.

Mathematical Theory Approach

A rather different model based mainly on mathematical considerations is the Mazzola model (Mazzola, 1990; Mazzola and Zahorka, 1994; Mazzola et al., 1995; Mazzola, 2002; Mazzola and Göller, 2002). This model basically consists of an analysis part and a performance part. The analysis part involves computer-aided analysis tools, for various aspects of the music structure, that assign particular weights to each note in a symbolic score. The performance part, that transforms structural features into an artificial performance, is theoretically anchored in the so-called Stemma Theory and Operator Theory (a sort of additive rule-based structure-to-performance mapping). It iteratively modifies the performance vector fields, each of which controls a single expressive parameter of a synthesised performance.

1.3.2 Perspectives

Comparing Performances

A problem that normally arises in performance research is how performances can be compared. In subjective comparison often a supposed ideal performance is taken as reference by the evaluator. In other cases, an actual reference performance can be assumed. Of course subjects with different background can have dissimilar preferences that are not easily made explicit.

However when we consider computational models, objective numerical comparisons would be very appealing. In this case, performances are represented by a set of values. Sometimes the adopted strategies compare absolute or relative values. As measure of distance the mean of the absolute differences can be considered, or the Euclidean distance (square root of difference squares) or maximum distance (i.e., take the maximal difference component). It is not clear how to weight the components, nor which distance formulation is more effective. Different researchers employ different measures.

More basically it is not clear how to combine time and loudness distances for a comprehensive performance comparison. For instance as already discussed, the emphasis of a note can be obtained by lengthening, dynamic accent, time shift, timbre variation. Moreover, it is not clear how perception can be taken into account, nor how to model subjective preferences. How are subjective and objective comparisons related? The availability of good and agreed methods for performance comparison would be very welcome in performance research. A subjective assessment of objective comparison is needed. More research effort on this direction is advisable.

Modeling Different Expressive Intentions

The models discussed in the previous sections aim at explaining and simulating performances which is played accordingly to appropriate rules imposed by a specific musical praxis. The focus is on aspects that most performances have in common. Recently research started paying attention to aspects that differentiate performances and performers styles (Repp, 1992; Widmer, 2003). The same piece of music can be performed trying to convey different expressive intentions (Gabrielsson and Lindström, 2001), changing the style of the performance. The CARO model (Canazza et al., 2004) is able to modify a neutral performance (i.e. played without any specific expressive intention) in order to convey different expressive intentions. Bresin and Friberg (2000) developed some macro rules for selecting appropriate values for the parameters of the KTH rule system in order to convey different emotions.

Expression Recognition Models

The methods seen in the previous sections aim at explaining how expression is conveyed by the performer and how it is related to the musical structure. Recently these accumulated research results started giving rise to models that aim to extract and recognise expression from a performance (Dannenberg et al., 1997; Friberg et al., 2002; Mion and De Poli, 2004).

1.4 Open Problems and Future Paths

Although computational modelling of expressive human performance has been developing quickly during the past decade, there is ample room for further research, and the field of computational performance modelling continues to be active. However, the idea of a creative activity being predictable and, more specifically, the notion of a direct "quasi-causal" relation between the musical score and the performance is quite problematic. The person and personality of the artist as a mediator between music and listener is totally neglected in basically all models discussed above. There are some severe general limits to what any predictive model can describe. For instance, very often performers intentionally play the repetition of the same phrase or section totally differently the second time around. Being able to predict this would presuppose models of aspects that are outside the music itself, such as performance context, artistic intentions, personal experiences, listeners' expectations, etc.

Although it might sound quaint, there are concrete attempts to elaborate computational models of expressive performance to a complexity so that they are able to compete with human performers. Since 2002, a scientific initiative brings together scientists from all over the world for a competition of artificially created performances (RENCON, contest for performance rendering systems, the next one to be held at the ICMC'05 in Barcelona[6]). Their aim is to construct computational systems that are able to pass ax expressive performance Turing Test (that is an artificial performance sounds indistinguishable to a human performance, Hiraga et al., 2004). One ambitious goal is a computer system to win the Chopin competition in 2050 (Hiraga et al., 2004).

It is very hard to imagine that this will ever be possible, not only because the organisers of such a competition wont accept a computer to participate, but also because a computational model would have to take into account the complex social and cognitive contexts in

[6]http://www.icmc2005.org

which, like any human intellectual and artistic activity, a music performance is situated. But even if complete predictive models of such phenomena are strictly impossible, they advance our understanding and appreciation of the complexity of artistic behaviour, and it remains an intellectual and scientific challenge to probe the limits of formal modelling and rational characterisation.

Acknowledgements

This research is supported by the European Union (project FP6 IST-2004-03773 S2S2 “Sound to Sense, Sense to Sound”); the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF; START project Y99-INF “Computer-Based Music Research: Artificial Intelligence Models of Musical Expression”), and the Viennese Science and Technology Fund (WWTF; project CI010 “Interfaces to Music”). The Austrian Research Institute for Artificial Intelligence (OFAI) acknowledges basic financial support by the Austrian Federal Ministries for Education, Science, and Culture, and for Transport, Innovation and Technology.

Bibliography

Arcos, J. L., López de Mántaras, R., and Serra, X. (1998). SaxEx: A case-based reasoning system for generating expressive performances. *Journal of New Music Research*, 27(3):194–210.

Ashley, R. (2002). Do[n't] change a hair for me: The art of Jazz Rubato. *Music Perception*, 19(3):311–332.

Bachmann, K. (1999). *Das Verhältnis der Tempi in mehrsätzigen Musikwerken: ein Beitrag zur musikalischen Aufführungsanalyse am Beispiel der Symphonien Ludwig van Beethovens*. Unpublished doctoral thesis, Institut für Musikwissenschaft, Universität Salzburg, Salzburg.

Behne, K.-E. (1990). "Blicken Sie auf die Pianisten?!" Zur bildbeeinflußten Beurteilung von Klaviermusik im Fernsehen. *Medienpsychologie*, 2(2):115–131.

Behne, K.-E. and Wetekam, B. (1994). Musikpsychologische Interpretationsforschung: Individualität und Intention. In Behne, K.-E., Kleinen, G., and la Motte-Haber, H. d., editors, *Musikpsychologie. Empirische Forschungen, ästhetische Experimente*, volume 10, pages 24–32. Noetzel, Wilhelmshaven.

Bengtsson, I. and Gabrielsson, A. (1977). Rhythm research in Uppsala. In *Music, Room, Acoustics*, volume 17, pages 19–56. Publications issued by the Royal Swedish Academy of Music, Stockholm.

Binet, A. and Courtier, J. (1895). Recherches graphiques sur la musique. *L'Année Psychologique*, 2:201–222. Available also in a German translation by Schmitz, H.-W. (1994), Das Mechanische Musikinstrument 61, 16–24.

Bolzinger, S. (1995). *Contribution a l'étude de la rétroaction dans la pratique musicale par l'analyse de l'influence des variations d'acoustique de la salle sur le jeu du pianiste*. Unpublished doctoral thesis, Institut de Mécanique de Marseille, Université Aix-Marseille II, Marseille.

Bowen, J. A. (1996). Tempo, duration, and flexibility: Techniques in the analysis of performance. *Journal of Musicological Research*, 16(2):111–156.

Bowers, Q. D. (1972). *Encyclopedia of Automatic Musical Instruments*. Vestal Press Ltd., New York, 13th edition.

Bresin, R. (2001). Articulation rules for automatic music performance. In Schloss, A. and Dannenberg, R., editors, *Proceedings of the 2001 International Computer Music Conference, Havana, Cuba*, pages 294–297. International Computer Music Association, San Francisco.

Bresin, R. and Battel, G. U. (2000). Articulation strategies in expressive piano performance. *Journal of New Music Research*, 29(3):211–224.

Bresin, R. and Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal*, 24(4):44–63.

Bresin, R. and Widmer, G. (2000). Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results. *Speech, Music, and Hearing. Quarterly Progress and Status Report*, 2000(4):1–6.

Cambouropoulos, E. (2000). Score Extraction from MIDI Files. In *In Proceedings of the 13th Colloquium on Musical Informatics (CIM'2000)*. L'Aquila, Italy.

Camurri, A., Krumhansl, C. L., Mazzarino, B., and Volpe, G. (2004). An exploratory study of aniticipation human movement in dance. In *Proceedings of the 2nd International Symposium on Measurement, Analysis, and Modeling of Human Functions*. Genova, Italy.

Camurri, A. and Volpe, G., editors (2004). *Gesture-Based Communication in Human-Computer Interaction*. Springer, Berlin. LNAI 2915.

Camurri, A., Volpe, G., De Poli, G., and Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, 12(1):43–53.

Canazza, S., De Poli, G., Drioli, C., Rodà, A., and Vidolin, A. (2004). Modeling and control of expressiveness in music performance. *Proceedings of the IEEE*, 92(4):686–701.

Canazza, S., De Poli, G., Mion, G., Rodà, A., Vidolin, A., and Zanon, P. (2003). Expressive classifiers at CSC: An overview of the main research streams. In *Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003) May 8–10*. Firenze.

Cano, P., Loscos, A., and Bonada, J. (1999). Score-performance matching using HMMs. In *Proceedings of the International Computer Music Conference*, pages 441–444. International Computer Music Association.

Clarke, E. F. (1982). Timing in the performance of Erik Satie's 'Vexations'. *Acta Psychologica*, 50(1):1–19.

Clarke, E. F. (1985). Some aspects of rhythm and expression in performances of Erik Satie's "Gnossienne No. 5". *Music Perception*, 2:299–328.

Clarke, E. F. (2004). Empirical methods in the study of performance. In Clarke, E. F. and Cook, N., editors, *Empirical Musicology. Aims, Methods, and Prospects*, pages 77–102. University Press, Oxford.

Clarke, E. F. and Windsor, W. L. (2000). Real and simulated expression: A listening study. *Music Perception*, 17(3):277–313.

Coenen, A. and Schäfer, S. (1992). Computer-controlled player pianos. *Computer Music Journal*, 16(4):104–111.

Collier, G. L. and Collier, J. L. (2002). A study of timing in two Louis Armstrong solos. *Music Perception*, 19(3):463–483.

Cook, N. (1987). Structure and performance timing in Bach's C major prelude (WTC I): An empirical study. *Music Analysis*, 6(3):100–114.

Cook, N. (1995). The conductor and the theorist: Furtwängler, Schenker and the first movement of Beethoven's Ninth Symphony. In Rink, J., editor, *The Practice of Performance*, pages 105–125. Cambridge University Press, Cambridge, UK.

Cook, N. (2004). Computational and comparative Musicology. In Clarke, E. F. and Cook, N., editors, *Empirical Musicology. Aims, Methods, and Prospects*, pages 103–126. University Press, Oxford.

Dahl, S. (2004). Playing the accent – comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica*, 90(4):762–776.

Dahl, S. (2005). Movements and analysis of drumming. In Altenmüller, E., Kesselring, J., and Wiesendanger, M., editors, *Music, Motor Control and the Brain*, page in press. University Press, Oxford.

Dannenberg, R. (1984). An on-line algorithm for real-time accompaniment. In *Proceedings of the International Computer Music Conference*, pages 193–198.

Dannenberg, R., Thom, B., and Watson, D. (1997). A machine learning approach to musical style recognition. In *Proceedings of the 1997 International Computer Music Conference*, pages 334–347, San Francisco, CA. International Computer Music Association.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21(2):103–113.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, 26(6):279–301.

Davidson, J. W. and Correia, J. S. (2002). Body movement. In Parncutt, R. and McPherson, G., editors, *The Science and Psychology of Music Performance. Creating Strategies for Teaching and Learning*, pages 237–250. University Press, Oxford.

De Poli, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, 33(3):189–202.

Desain, P. and Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research*, 56:285–292.

Desain, P. and Honing, H. (1996). Modeling continuous aspects of music performance: Vibrato and Portamento. In Pennycook, B. and Costa-Giomi, E., editors, *Proceedings of the 4th International Conference on Music Perception and Cognition (ICMPC'96)*. Faculty of Music, McGill University, Montreal, Canada.

Dillon, R. (2003). A statistical approach to expressive intention recognition in violin performances. In Bresin, R., editor, *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03), August 6–9, 2003*, pages 529–532. Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden.

Dixon, S. (2000). Extraction of musical performance parameters from audio data. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia*, pages 42–45, Sydney. University of Sydney.

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

Dixon, S. (2001b). Learning to detect onsets of acoustic piano tones. In Buyoli, C. L. and Loureiro, R., editors, *MOSART Workshop on current research directions in computer music, November 15–17, 2001*, pages 147–151. Audiovisual Institute, Pompeu Fabra University, Barcelona, Spain.

Dixon, S. (2004). Analysis of musical content in digital audio. In DiMarco, J., editor, *Computer Graphics and Multimedia: Applications, Problems, and Solutions*, pages 214–235. Idea Group, Hershey PA.

Dixon, S. (2005a). Live tracking of musical performances using on-line time warping. Submitted.

Dixon, S. (2005b). An on-line time warping algorithm for tracking musical performances. In *Proceedings of the International Joint Conference on Artificial Intelligence*. to appear.

Dixon, S., Goebl, W., and Cambouropoulos, E. (2005). Smoothed tempo perception of expressively performed music. *Music Perception*, 23:in press.

Dixon, S., Goebl, W., and Widmer, G. (2002). The Performance Worm: Real time visualisation based on Langner's representation. In Nordahl, M., editor, *Proceedings of the 2002 International Computer Music Conference, Göteborg, Sweden*, pages 361–364. International Computer Music Association, San Francisco.

Dixon, S. and Widmer, G. (2005). MATCH: A music alignment tool chest. In *6th International Conference on Music Information Retrieval*, page Submitted.

Ebhardt, K. (1898). Zwei Beiträge zur Psychologie des Rhythmus und des Tempo. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 18:99–154.

Friberg, A. (1991). Generative rules for music performance. *Computer Music Journal*, 15(2):56–71.

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

Friberg, A. (2005). A fuzzy analyzer of emotional expression in music performance and body motion. In Sundberg, J. and Brunson, B., editors, *Proceedings of Music and Music Science, October 28–30, 2004*. Royal College of Music in Stockholm, Stockholm. CD-ROM.

Friberg, A., Bresin, R., Frydén, L., and Sundberg, J. (1998). Musical punctuation on the mircolevel: Automatic identification and performance of small melodic units. *Journal of New Music Research*, 27(3):271–292.

Friberg, A., Colombo, V., Frydén, L., and Sundberg, J. (2000a). Generating musical performances with Director Musices. *Computer Music Journal*, 24(3):23–29.

Friberg, A., Schoonderwaldt, E., and Juslin, P. N. (2005). CUEX: An algorithm for extracting expressive tone variables from audio recordings. *Acoustica united with Acta Acoustica*, in press.

Friberg, A., Schoonderwaldt, E., Juslin, P. N., and Bresin, R. (2002). Automatic Real-Time Extraction of Musical Expression. In *Proceedings of the 2002 International Computer Music Conference, Göteborg, Sweden*, pages 365–367. International Computer Music Association, San Francisco.

Friberg, A. and Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, 105(3):1469–1484.

Friberg, A., Sundberg, J., and Frydén, L. (2000b). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, 29(3):199–210.

Friberg, A. and Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19(3):333–349.

Gabrielsson, A. (1987). Once again: The Theme from Mozart's Piano Sonata in A Major (K.331). In Gabrielsson, A., editor, *Action and Perception in Rhythm and Music*, volume 55, pages 81–103. Publications issued by the Royal Swedish Academy of Music, Stockholm, Sweden.

Gabrielsson, A. (1999). Music Performance. In Deutsch, D., editor, *Psychology of Music*, pages 501–602. Academic Press, San Diego, 2nd edition.

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychology of Music*, 31(3):221–272.

Gabrielsson, A., Bengtsson, I., and Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology*, 24:193–213.

Gabrielsson, A. and Lindström, E. (2001). The influence of musical structure on emotional expression. In Juslin, P. N. and Sloboda, J. A., editors, *Music and Emotion: Theory and Research*, pages 223–248. Oxford University Press, New York.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110(1):563–572.

Goebl, W. (2003). *The Role of Timing and Intensity in the Production and Perception of Melody in Expressive Piano Performance*. Doctoral thesis, Institut für Musikwissenschaft, Karl-Franzens-Universität Graz, Graz, Austria. available online at `http://www.ofai.at/music`.

Goebl, W. and Bresin, R. (2001). Are computer-controlled pianos a reliable tool in music performance research? Recording and reproduction precision of a Yamaha Disklavier grand piano. In Buyoli, C. L. and Loureiro, R., editors, *MOSART Workshop on Current Research Directions in Computer Music, November 15–17, 2001*, pages 45–50. Audiovisual Institute, Pompeu Fabra University, Barcelona, Spain.

Goebl, W. and Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America*, 114(4):2273–2283.

Goebl, W. and Dixon, S. (2001). Analyses of tempo classes in performances of Mozart piano sonatas. In Lappalainen, H., editor, *Proceedings of the Seventh International Symposium on Systematic and Comparative Musicology, Third International Conference on Cognitive Musicology, August 16–19, 2001*, pages 65–76. University of Jyväskylä, Jyväskylä, Finland.

Goebl, W., Pampalk, E., and Widmer, G. (2004). Exploring expressive performance trajectories: Six famous pianists play six Chopin pieces. In Lipscomp, S. D., Ashley, R., Gjerdingen, R. O., and Webster, P., editors, *Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL, 2004 (ICMPC8)*, pages 505–509. Causal Productions, Adelaide, Australia. CD-ROM.

Gottschewski, H. (1996). *Die Interpretation als Kunstwerk. Musikalische Zeitgestaltung und ihre Analyse am Beispiel von Welte-Mignon-Klavieraufnahmen aus dem Jahre 1905.* Freiburger Beiträge zur Musikwissenschaft, Bd. 5. Laaber-Verlag, Laaber.

Gouyon, F., Wack, N., and Dixon, S. (2004). An open source tool for semi-automatic rhythmic annotation. In *Proceedings of the 7th International Conference on Digital Audio Effects*, pages 193–196.

Hagmann, P. (1984). *Das Welte-Mignon-Klavier, die Welte-Philharmonie-Orgel und die Anfänge der Reproduktion von Musik*. Europäische Hochschulschriften: Reihe 35, Musikwissenschaft, Bd. 10. Peter Lang, Bern, Frankfurt am Main, New York. available at `http://www.freidok.uni-freiburg.de/volltexte/608/`.

Hartmann, A. (1932). Untersuchungen über das metrische Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie*, 84:103–192.

Heijink, H., Desain, P., Honing, H., and Windsor, L. (2000a). Make me a match: An evaluation of different approaches to score-performance matching. *Computer Music Journal*, 24(1):43–56.

Heijink, H., Windsor, L., and Desain, P. (2000b). Data processing in music performance research: Using structural information to improve score-performance matching. *Behavior Research Methods, Instruments and Computers*, 32(4):546–554.

Heinlein, C. P. (1929a). A discussion of the nature of pianoforte damper-pedalling together with an experimental study of some individual differences in pedal performance. *Journal of General Psychology*, 2:489–508.

Heinlein, C. P. (1929b). The functional role of finger touch and damper-pedalling in the appreciation of pianoforte music. *Journal of General Psychology*, 2:462–469.

Heinlein, C. P. (1930). Pianoforte damper-pedalling under ten different experimental conditions. *Journal of General Psychology*, 3:511–528.

Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In Seashore, C. E., editor, *Objective Analysis of Musical Performance*, volume IV of *University of Iowa Studies in the Psychology of Music*, pages 281–305. University Press, Iowa City.

Henderson, M. T., Tiffin, J., and Seashore, C. E. (1936). The Iowa piano camera and its use. In Seashore, C. E., editor, *Objective Analysis of Musical Performance*, volume IV, pages 252–262. University Press, Iowa City.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, 49:621–630.

Hiraga, R., Bresin, R., Hirata, K., and Katayose, H. (2004). Rencon 2004: Turing Test for musical expression. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME04)*, pages 120–123. Hamamatsu, Japan.

Hong, J.-L. (2003). Investigating expressive timing and dynamics in recorded cello performances. *Psychology of Music*, 31(3):340–352.

Huber, D. M. (1999). *The MIDI Manual*. Butterworth-Heinemann, Boston, MA.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1797–1813).

Juslin, P. N. (2003). Studies of music performance: A theoretical analysis of empirical findings. In Bresin, R., editor, *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03), August 6–9, 2003*, volume II, pages 513–516. Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden.

Juslin, P. N. and Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3):217–238.

Juslin, P. N. and Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, 17(2):197–221.

Juslin, P. N. and Sloboda, J. A. (2001). *Music and Emotion: Theory and Research*. Oxford University Press, New York.

Kay, B. A., Turvey, M. T., and Meijer, O. G. (2003). An early oscillator model: Studies on the biodynamics of the piano strike (Bernstein & Popova, 1930). *Motor Control*, 7(1):1–45.

Klapuri, A. (2004). Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–282.

Kronman, U. and Sundberg, J. (1987). Is the musical retard an allusion to physical motion? In Gabrielsson, A., editor, *Action and Perception in Rhythm and Music*, volume 55, pages 57–68. Publications issued by the Royal Swedish Academy of Music, Stockholm, Sweden.

Leman, M., Vermeulen, V., L., D. V., Taelman, J., Moelants, D., and Lesaffre, M. (2004). Correlation of gestural musical audio cues and perceived expressive qualities. In Camurri, A. and Volpe, G., editors, *Gesture-based Communication in Human-Computer Interaction*, pages xx–yy. Springer, Berlin. LNAI 2915.

Lindström, E., Camurri, A., Friberg, A., Volpe, G., and Rinman, M.-L. (2005). Affect, attitude and evaluation of multi-sensory performances. *Journal of New Music Research*, in press.

Lisboa, T., Williamon, A., Zicari, M., and Eiholzer, H. (2005). Mastery through imitation: A preliminary study. *Musicae Scientiae*, 9(1).

Liu, D., Lie, L., and Zhang, H.-J. (2003). Automatic mood detection from acoustic music data. In *Proceedings of the International Symposium on Music Information Retrieval*.

López de Mántaras, R. and Arcos, J. L. (2002). AI and music: From composition to expressive performances. *AI Magazine*, 23(3):43–57.

Maria, M. (1999). Unschärfetests mit hybriden Tasteninstrumenten. In Enders, B. and Stange-Elbe, J., editors, *Global Village – Global Brain – Global Music. KlangArt Kongreß 1999*. Osnabrück, Germany.

Mazzola, G. (1990). *Geometrie der Töne. Elemente der Mathematischen Musiktheorie*. Birkhäuser Verlag, Basel.

Mazzola, G., editor (2002). *The Topos of Music — Geometric Logic of Concepts, Theory, and Performance*. Birkhäuser Verlag, Basel.

Mazzola, G. and Beran, J. (1998). Rational composition of performance. In Kopiez, R. and Auhagen, W., editors, *Controlling Creative Processes in Music*, volume 12 of *Schriften zur Musikpsychologie and Musikästhetik, Bd. 12*, pages 37–68. Lang, Frankfurt/M.

Mazzola, G. and Göller, S. (2002). Performance and interpretation. *Journal of New Music Research*, 31(3):221–232.

Mazzola, G. and Zahorka, O. (1994). Tempo curves revisited: Hierarchies of performance fields. *Computer Music Journal*, 18(1):40–52.

Mazzola, G., Zahorka, O., and Stange-Elbe, J. (1995). Analysis and performance of a dream. In Friberg, A. and Sundberg, J., editors, *Proceedings of the KTH Symposion on Grammars for Music Performance*, pages 59–68. Department of Speech Communication and Music Acoustics, Stockholm, Sweden.

McAdams, S., Depalle, P., and Clarke, E. F. (2004). Analyzing musical sound. In Clarke, E. F. and Cook, N., editors, *Empirical Musicology. Aims, Methods, and Prospects*, pages 157–196. University Press, Oxford.

Midgette, A. (2005). Play it again, Vladimir (via computer). *The New York Times*.

Mion, L. and De Poli, G. (2004). Expressiveness detection of music performances in the kinematics energy space. In *Proc. Sound and Music Computing Conf. (JIM/CIM 04)*, pages 257–261, Paris.

Moelants, D. (2004). Temporal aspects of instrumentalists' performance of tremolo, trills, and vibrato. In *Proceedings of the International Symposium on Musical Acoustics (ISMA'04)*, pages 281–284. The Acoustical Society of Japan, Nara, Japan.

Moog, R. A. and Rhea, T. L. (1990). Evolution of the keyboard interface: The Bösendorfer 290 SE recording piano and the Moog multiply-touch-sensitive keyboards. *Computer Music Journal*, 14(2):52–60.

Moore, G. P. (1992). Piano trills. *Music Perception*, 9(3):351–359.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception and Psychophysics*, 41(6):525–533.

Nettheim, N. (2001). A musical microscope applied to the piano playing of Vladimir de Pachmann. `http://users.bigpond.net.au/nettheim/pachmic/microsc.htm`.

Orio, N. and Déchelle, F. (2001). Score following using spectral analysis and hidden Markov models. In *Proceedings of the International Computer Music Conference*, pages 151–154.

Ortmann, O. (1925). *The Physical Basis of Piano Touch and Tone*. Kegan Paul, Trench, Trubner; J. Curwen; E. P. Dutton, London, New York.

Ortmann, O. (1929). *The Physiological Mechanics of Piano Technique*. Kegan Paul, Trench, Trubner, E. P. Dutton, London, New York. Paperback reprint: New York: E. P. Dutton 1962.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15(12):331–346.

Palmer, C. (1992). The role of interpretive preferences in music performance. In Jones, M. R. and Holleran, S., editors, *Cognitive Bases of Musical Communication*, pages 249–262. American Psychological Association, Washington DC.

Palmer, C. (1996a). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13(3):433–453.

Palmer, C. (1996b). On the assignment of structure in music performance. *Music Perception*, 14(1):23–56.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.

Palmer, C. and Brown, J. C. (1991). Investigations in the amplitude of sounded piano tones. *Journal of the Acoustical Society of America*, 90(1):60–66.

Palmer, C. and Dalla Bella, S. (2004). Movement amplitude and tempo change in piano performance. *Journal of the Acoustical Society of America*, 115(5):2590.

Palmer, C. and Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception and Psychophysics*, 56(3):301–312.

Palmer, C. and van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2):457–470.

Palmer, C. and van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5):947–962.

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

Povel, D.-J. (1977). Temporal structure of performed music: Some preliminary observations. *Acta Psychologica*, 41(4):309–320.

Raphael, C. (2004). A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proceedings of the 5th International Conference on Musical Information Retrieval*, pages 387–394.

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):622–641.

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):2546–2568.

Repp, B. H. (1993). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America*, 93(2):1136–1144.

Repp, B. H. (1994a). On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22:157–167.

Repp, B. H. (1994b). Relational invariance of expressive microstructure across global tempo changes in music performance: an exploratory study. *Psychological Research*, 56(4):269–284.

Repp, B. H. (1995a). Acoustics, perception, and production of legato articulation on a digital piano. *Journal of the Acoustical Society of America*, 97(6):3862–3874.

Repp, B. H. (1995b). Detectability of duration and intensity increments in melody tones: a partial connection between music perception and performance. *Perception and Psychophysics*, 57(8):1217–1232.

Repp, B. H. (1995c). Expressive timing in Schumann's "Träumerei:" An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98(5):2413–2427.

Repp, B. H. (1996a). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100(6):3917–3932.

Repp, B. H. (1996b). Pedal timing and tempo in expressive piano performance: A preliminary investigation. *Psychology of Music*, 24(2):199–221.

Repp, B. H. (1996c). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14(2):161–184.

Repp, B. H. (1996d). The dynamics of expressive piano performance: Schumann's "Träumerei" revisited. *Journal of the Acoustical Society of America*, 100(1):641–650.

Repp, B. H. (1997a). Acoustics, perception, and production of legato articulation on a computer-controlled grand piano. *Journal of the Acoustical Society of America*, 102(3):1878–1890.

Repp, B. H. (1997b). Expressive timing in a Debussy Prelude: A comparison of student and expert pianists. *Musicae Scientiae*, 1(2):257–268.

Repp, B. H. (1997c). The Aesthetic Quality of a Quantitatively Average Music Performance: Two Preliminary Experiments. *Music Perception*, 14(4):419–444.

Repp, B. H. (1997d). The effect of tempo on pedal timing in piano performance. *Psychological Research*, 60(3):164–172.

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 104(2):1085–1100.

Repp, B. H. (1999a). 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):1972–1988.

Repp, B. H. (1999b). A microcosm of musical expression: III. Contributions of timing and dynamics to the aesthetic impression of pianists' performances of the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 106(1):469–478.

Riley-Butler, K. (2002). Teaching expressivity: An aural–visual feedback–replication model. In *ESCOM 10th Anniversary Conference on Musical Creativity, April 5–8, 2002*. Université de Liège, Liège, Belgium. CD-ROM.

Rink, J., editor (1995). *The Practice of Performance: Studies in Musical Interpretation*. University Press, Cambridge UK.

Rink, J., editor (2002). *Musical Performance. A Guide to Understanding*. Cambridge University Press, Cambridge, UK.

Rink, J. (2003). In respect of performance: The view from Musicology. *Psychology of Music*, 31(3):303–323.

Roads, C. (1986). Bösendorfer 290 SE computer-based piano. *Computer Music Journal*, 10(3):102–103.

Scheirer, E. (1997). Using musical knowledge to extract expressive performance information from audio recordings. In Okuno, H. and Rosenthal, D., editors, *Readings in Computational Auditory Scene Analysis*. Lawrence Erlbaum.

Schoonderwaldt, E. and Friberg, A. (2001). Towards a rule-based model for violin vibrato. In Buyoli, C. L. and Loureiro, R., editors, *MOSART Workshop on Current Research Directions in Computer Music, November 15–17, 2001*, pages 61–64. Audiovisual Institute, Pompeu Fabra University, Barcelona, Spain.

Sears, C. H. (1902). A contribution to the psychology of rhythm. *American Journal of Psychology*, 13(1):28–61.

Seashore, C. E., editor (1932). *The Vibrato*, volume I of *University of Iowa Studies in the Psychology of Music*. University Press, Iowa City.

Seashore, C. E., editor (1936a). *Objective Analysis of Musical Performance*, volume IV of *University of Iowa Studies in the Psychology of Music*. University Press, Iowa City.

Seashore, C. E., editor (1936b). *Psychology of the Vibrato in Voice and Instrument*, volume III of *University of Iowa Studies. Studies in the Psychology of Music Volume III*. University Press, Iowa City.

Seif El-Nasr, M., Yen, J., and Iorger, T. R. (2000). FLAME – Fuzzy logic adaptive mode of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3):219–257.

Shaffer, L. H. (1980). Analysing piano performance. In Stelmach, G. E. and Requin, J., editors, *Tutorials in Motor Behavior*. North-Holland, Amsterdam.

Shaffer, L. H. (1981). Performances of Chopin, Bach and Bartòk: Studies in motor programming. *Cognitive Psychology*, 13(3):326–376.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 36A(4):577–595.

Shaffer, L. H. (1992). How to interpret music. In Jones, M. R. and Holleran, S., editors, *Cognitive Bases of Musical Communication*, pages 263–278. American Psychological Association, Washington DC.

Shaffer, L. H., Clarke, E. F., and Todd, N. P. M. (1985). Metre and Rhythm in Pianoplaying. *Cognition*, 20(1):61–77.

Shaffer, L. H. and Todd, N. P. M. (1987). The interpretative component in musical performance. In Gabrielsson, A., editor, *Action and Perception in Rhythm and Music*, volume 55, pages 139–152. Publications issued by the Royal Swedish Academy of Music, Stockholm, Sweden.

Shalev-Shwartz, S., Keshet, J., and Singer, Y. (2004). Learning to align polyphonic music. In *5th International Conference on Music Information Retrieval*, pages 381–386.

Sjölander, K. and Beskow, J. (2000). WaveSurfer – an open source speech tool. In *Proceedings of the International Conference on Spoken Language Processing*.

Skinner, L. and Seashore, C. E. (1936). A musical pattern score of the first movement of the Beethoven Sonata, Opus 27, No. 2. In Seashore, C. E., editor, *Objective Analysis of Musical Performance*, volume IV of *Studies in the Psychology of Music*, pages 263–279. University Press, Iowa City.

Soulez, F., Rodet, X., and Schwarz, D. (2003). Improving polyphonic and poly-instrumental music to score alignment. In *4th International Conference on Music Information Retrieval*, pages 143–148.

Sundberg, J. (2000). Four years of research on music and motion. *Journal of New Music Research*, 29(3):183–185.

Sundberg, J., Askenfelt, A., and Frydén, L. (1983). Musical performance. A synthesis-by-rule approach. *Computer Music Journal*, 7:37–43.

Sundberg, J., Friberg, A., and Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review*, 3:89–109.

Sundberg, J., Friberg, A., and Frydén, L. (1991). Threshold and preference quantities of rules for music performance. *Music Perception*, 9(1):71–92.

Timmers, R. (2002). *Freedom and Constraints in Timing and Ornamentation*. Shaker Publishing, Maastricht.

Timmers, R., Ashley, R., Desain, P., Honing, H., and Windsor, L. W. (2002). Timing of ornaments in the theme of Beethoven's Paisello Variations: Empirical data and a model. *Music Perception*, 20(1):3–33.

Tobudic, A. and Widmer, G. (2005). Relational IBL in classical music. *Machine Learning*, to appear.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception*, 3(1):33–58.

Todd, N. P. M. (1989). A computational model of Rubato. *Contemporary Music Review*, 3:69–88.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91(6):3540–3550.

Todd, N. P. M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97(3):1940–1949.

Turetsky, R. and Ellis, D. (2003). Ground-truth transcriptions of real music from force-aligned MIDI syntheses. In *4th International Conference on Music Information Retrieval*, pages 135–141.

Vernon, L. N. (1936). Synchronization of chords in artistic piano music. In Seashore, C. E., editor, *Objective Analysis of Musical Performance*, volume IV of *Studies in the Psychology of Music*, pages 306–345. University Press, Iowa City.

Widmer, G. (1995a). A machine learning analysis of expressive timing in pianists' performances of Schumann's "Träumerei". In Friberg, A. and Sundberg, J., editors, *Proceedings of the KTH Symposion on Grammars for Music Performance*, pages 69–81. Department of Speech Communication and Music Acoustics, Stockholm, Sweden.

Widmer, G. (1995b). Modeling rational basis for musical expression. *Computer Music Journal*, 19(2):76–96.

Widmer, G. (1996). Learning expressive performance: The structure-level approach. *Journal of New Music Research*, 25(2):179–205.

Widmer, G. (2000). Large-scale induction of expressive performance rules: first quantitative results. In Zannos, I., editor, *Proceedings of the 2000 International Computer Music Conference, Berlin, Germany*, pages 344–347. International Computer Music Association, San Francisco.

Widmer, G. (2001). Using AI and machine learning to study expressive music performance: Project survey and first report. *AI Communications*, 14(3):149–162.

Widmer, G. (2002a). In search of the Horowitz factor: Interim report on a musical discovery project. In *Proceedings of the 5th International Conference on Discovery Science (DS'02), Lübeck, Germany*. Springer, Berlin.

Widmer, G. (2002b). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research*, 31(1):37–50.

Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148.

Widmer, G. (2005). Studying a creative act with computers: Music performance studies with automated discovery methods. *Musicae Scientiae*, 9(1):11–30.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.

Widmer, G. and Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3):203–216.

Widmer, G. and Tobudic, A. (2003). Playing Mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research*, 32(3):259–268.

Windsor, L. (2004). Data collection, experimental design, and statistics in musical research. In Clarke, E. F. and Cook, N., editors, *Empirical Musicology. Aims, Methods, and Prospects*, pages 197–222. University Press, Oxford.

Windsor, L., Desain, P., Aarts, R., Heijink, H., and Timmers, R. (2001). The timing of grace notes in skilled musical performance at different tempi. *Psychology of Music*, 29(2):149–169.