Computational Models of Expressive Music Performance: The State of the Art

Article *in* Journal of New Music Research · September 2004

DOI: 10.1080/09298210420003178O4

CITATIONS
198

READS
630

2 authors:

Gerhard Widmer
Johannes Kepler University Linz
425 PUBLICATIONS **10,769** CITATIONS
SEE PROFILE

Werner Goebl
Universität für Musik und darstellende Kunst Wien
116 PUBLICATIONS **2,008** CITATIONS
SEE PROFILE

Journal of New Music Research
2004, Vol. 33, No. 3, pp. 203–216

Routledge
Taylor & Francis Group

Computational Models of Expressive Music Performance: The State of the Art

Gerhard Widmer[1,2] and Werner Goebl[2]

[1]Department of Computational Perception, Johannes Kepler University Linz, Austria; [2]Austrian Research Institute for Artificial Intelligence, Vienna, Austria

Abstract

This contribution gives an overview of the state of the art in the field of computational modeling of expressive music performance. The notion of predictive computational model is briefly discussed, and a number of quantitative models of various aspects of expressive performance are briefly reviewed. Four selected computational models are reviewed in some detail. Their basic principles and assumptions are explained and, wherever possible, empirical evaluations of the models on real performance data are reported. In addition to these models, which focus on general, common principles of performance, currently ongoing research on the formal characterisation of differences in individual performance style are briefly presented.

1. Introduction

Music performance as the act of interpreting, structuring, and physically realising a piece of music is a complex human activity with many facets – physical, acoustic, physiological, psychological, social, artistic. Since the early investigations by Seashore and colleagues (Seashore, 1938), the field of performance research has made great strides towards understanding this complex behaviour and the artifacts it produces. Rather than attempting to do justice to all the great research that has been done in this area in the past 100 years, we refer the reader to papers by Gabrielsson (1999, 2003) for an excellent overview (with some 800 literature references!).

The present article will focus on one specific aspect, namely, *expressive* music performance, i.e., the deliberate shaping by performers of parameters like tempo, timing, dynamics, articulation, etc. More precisely, the goal is to give an overview of the current state of the art of *quantitative* or *computational modelling* of expressive performance, and to give the reader an idea of some of the current research directions in this area. Computational modelling is an attempt at formulating hypotheses concerning expressive performance in such a precise way that they can be empirically verified (or disproved) on real measured performance data.

In the following, the notion of computational model is briefly discussed in the context of expressive performance. A number of models of various specialised aspects of expressive performance are briefly reviewed or at least mentioned, and four specific models are then discussed in more detail in the rest of this article: the rule-based performance model developed at KTH (Sundberg et al., 1983, 1989; Friberg & Sundberg, 1987; Friberg, 1991, 1995a) and implemented in the *Director Musices* system (Friberg et al., 2000; henceforth called the *KTH Model*); the structure-level models of timing and dynamics advocated by Neil Todd (1985, 1989a, 1992; collectively called the *Todd Model*); the mathematical model of musical structure and expression by Guerino Mazzola (Mazzola, 1990; Mazzola & Zahorka, 1994a; Mazzola et al., 1995; Mazzola, 2002; Mazzola & Göller, 2002; the *Mazzola Model*); and our own recent model that combines note-level rules with structure-level expressive patterns and was induced automatically by machine learning systems (Widmer, 2002; Widmer & Tobudic, 2003; the *Machine Learning Model*).

These four models, as indeed most of the modelling attempts in performance research, try to capture common performance principles, that is, they focus on *commonalities* between performances and performers. A final section of this article briefly reports on a recent research project that also

Correspondence: Gerhard Widmer, Department of Computational Perception, Johannes Kepler University Linz, Austria. E-mail: gerhard.widmer@jku.at

DOI: 10.1080/0929821042000317804

tries to quantify and characterize, if not yet model in a predictive way, the *differences* between artists, that is, aspects of personal artistic performance style.

2. Computational modelling of expressive music performance

The *Encyclopedia Britannica* defines a scientific model fairly generally as a "familiar structure or mechanism used as an analogy to interpret a natural phenomenon". In domains of study that are open to observation and measurement, *mathematical models* postulate quantitative (possibly causal) relationships, deterministic or probabilistic, between various variables that are used to define the state of some system or phenomenon of interest, possibly dependent on the setting of some model parameters. Mathematical models might thus more appropriately be called abstract descriptions, rather than analogies, of natural phenomena. *Computational modelling* involves embodying mathematical models in computer programs that can immediately be applied to given data for testing and prediction purposes. A mathematical or computational model is *predictive* in the sense that, assuming a specific fixed setting of all parameters, the model predicts the values of a specific set of variables from the values of the other variables.

The purpose of computational models of expressive music performance is thus to specify precisely the physical parameters defining a performance (e.g., onset timing, inter-onset intervals, loudness levels, note durations, etc.), and to quantify (quasi-)systematic relationships between certain properties of the musical score, the performance context, and an actual performance of a given piece. Of course, models in human or artistic domains cannot be expected to be "correct" in the sense that their predictions will always correspond to the behaviour observed in humans. Finding criteria for testing and evaluating such models is thus a research question in its own right. Still, computational models are useful even in "soft" domains; by being able to compare a model's predictions with real performances and quantify the discrepancies, the shortcomings of the underlying assumptions and performance theories can be measured and pinpointed (Windsor & Clarke, 1997).

A lot of quantitative research on detailed aspects of expressive performance, based on measurements of timing, dynamics, etc., has been done in recent decades (e.g., Shaffer, 1981, 1984; Shaffer et al., 1985; Clarke, 1985; Gabrielsson, 1987; Palmer, 1989, 1996a,b; Repp, 1992, 1995, 1998, 1999; Goebl, 2001, to name but a few). Some of this work is descriptive rather than predictive, focusing on measuring performance details and describing classes of common patterns, often with the help of statistical analysis. Other work has led to quantitative and computational models of some very specific aspects of expressive performance, for example, the *final ritard* and its conspicuous relation to human motion (Kronman & Sundberg, 1987; Todd, 1995; Friberg & Sundberg, 1999; Sundberg, 2000; Friberg et al., 2000; Honing, 2003); the timing of *grace notes* (Timmers et al., 2002); *vibrato* (Desain & Honing, 1996; Schoonderwaldt & Friberg, 2001); *legato* (Bresin & Battel, 2000); or *staccato* and its relation to local musical context (Bresin & Widmer, 2000; Bresin, 2001). Dannenberg and Derenyi (1998) described a model of articulation in trumpet playing (derived from actual acoustic performances) that was used to drive an instrument model which synthesises trumpet performances.

A step towards a general model of expressive timing and dynamics can be found in Clarke (1988), where Clarke proposed a small set of nine generative rules that express the grouping structure of the music through tempo and dynamics changes. The underlying assumption of a strong systematic link between musical structure and structure of performances is also the basis for the computational models that will be discussed in more detail in the following sections.

Somewhat controversial is Manfred Clynes' *Composer's Pulse* theory, which postulates particular timing and accent patterns for specific composers (Haydn, Mozart, Beethoven, etc.) that repeat within a short period (e.g., one second; see Clynes, 1983, 1986, 1987). These composer-specific pulse patterns were derived from pressure curves produced by professional musicians (e.g., Pablo Casals, Rudolf Serkin, and Clynes himself) on a sentograph while listening to or imagining the music of a particular composer. Different patterns were extracted for double and triple meters. In the model, they repeat hierarchically at different metrical levels and are said to form the "composer-specific pulse microstructure". Another aspect of this model is a set of rules to link adjacent (sinusodial) tones ("predictive amplitude shaping", cf. Clynes, 1983, 1987), again in a composer-specific way.

Attempts at verifying or refuting the theory have led to quite diverging results and an extended scientific controversy (Repp, 1989; Thompson, 1989; Clynes, 1990; Repp, 1990a,b,c). The "best" perceptual evaluation of Clynes' pulse model was provided by Clynes himself (Clynes, 1995); he found famous pianists, music students, and non-musicians to prefer performances with the "correct" pulse over such with any other pulse. Other authors found it difficult to reproduce such results. Generally, the theoretical background of the composer-specific pulse model (Becking, 1928; – see Clynes, 1995) is rather vague and most likely impossible to verify or disprove through human expressive performances.

There are also some more "implicit" computational models that base their predictions on analogy or case-based reasoning. An example of that is the Saxex system by Arcos and López de Mántaras (Arcos et al., 1998; Mántaras & Arcos, 2002), which predicts expressive transformations to jazz (saxophone) phrases by looking at how other, similar phrases were played by a human musician.

In the following, we will take a closer look at four of the more comprehensive models that have been published in recent years. They are rather comprehensive in the sense that each of them could potentially produce an expressive interpretation of a piece, at least in terms of expressive timing

and dynamics, and they have been evaluated more or less extensively.

3. The KTH model

This well-known rule system has been developed at the Royal Institute of Technology in Stockholm in more than 20 years of research, starting with (Sundberg et al., 1983). In the meantime it has been extended and analysed in many ways (e.g., Sundberg et al., 1989; Friberg, 1991; Sundberg, 1993; Friberg et al., 1998) and applied also to contemporary music (Friberg et al., 1991). A comprehensive description is given in Friberg (1995a).

3.1 The basic model

The KTH model consists of a set of performance rules that predict or prescribe aspects of timing, dynamics, and articulation, based on local musical context. The rules refer to a limited class of musical situations and theoretical concepts (simple duration ratios between successive notes, ascending lines, melodic leaps, melodic and harmonic charge, phrase structure, etc.). Most of the rules operate at a rather low level, looking only at very local contexts and affecting individual notes, but there are also rules that refer to entire phrases. The rules are parameterised with a varying number of parameters.

The KTH model was developed using the so-called "analysis-by-synthesis approach" (Sundberg et al., 1983; Friberg & Sundberg, 1987; Sundberg et al., 1991a) that involves a professional musician directly evaluating any tentative rule brought forward by the researcher. Musician and researcher are in a constant feed-back loop trying to find the best formulation and parameter settings for each rule. This method has the advantage of modelling a kind of performer–listener interaction (as the authors argue, see Friberg, 1995a), but falls potentially short by placing high demands on the evaluating expert musician(s) and by not permitting reliable evaluations due to the small number of judgements. The analysis-by-synthesis method was inspired by and adapted from speech synthesis methods (Friberg, 1995a).

To illustrate, we mention here one particular rule: DURATION CONTRAST (Friberg, 1995a). This rule modifies the ratio between successive note durations in order to enhance, and thus make more easily perceivable, differences in duration. Most commonly it changes the durations of a long and a short one so that the longer becomes a bit longer, and the shorter becomes shorter. As for every rule, there is a quantity control parameter (k) to be set by the researcher. This parameter is designed so as to give a fairly good result when its value is around 1. With a k value of 0, a particular rule can be switched off; when k is set to a negative value, the rules are inverted (in our example of the DURATION CONTRAST, it would blur the rhythmic contrasts), which can produce bizarrely unmusical results. An important aspect of the model is that it is additive. For instance, there are several rules that influence the duration of a note, and the effects of the individual rules are added cumulatively to give the final duration. This additivity is a particular problem when trying to fit the parameters to collections of real recordings (see below).

More recently, some additional model components have been published (e.g., musical punctuation rules, Friberg et al., 1998; articulation rules, Bresin, 2001, 2002; or swing ratio in Jazz, Friberg & Sundström, 2002). We lack the space to recapitulate all of these here.

3.2 Empirical evaluation

The KTH Model features a large number of free parameters that govern the relative strengths of the individual notes and a number of other aspects. To turn the rule set into a truly predictive model, these parameters must be set to fixed values. Several attempts have been made to find appropriate parameter settings. Although the "analysis-by-synthesis" approach involves perceptual judgements from highly skilled individual musicians as a basic method, it still needs independent empirical evaluation on real performances (Gabrielsson, 1985).

Already in the early stages of the model's development, several perceptual evaluations were reported by the Sundberg group and guests (Thompson et al., 1989; Sundberg et al., 1991b) that brought forward evidence for the model's validity and reliability. In two listening tests, Sundberg and colleagues evaluated the perceptual responsiveness of musicians and non-musicians to expressive variations (Sundberg et al., 1991b). In the first experiment, they determined the perceptual threshold for the individual rule quantities (k values) at which listeners are able to detect differences between two examples. Musicians were more sensitive than non-musicians (who sometimes guessed when they did not hear the intended variation); the quantity thresholds depended strongly on the type of rule. In a second experiment, skilled musicians adjusted the k values for six rules. In one rule (DURATION CONTRAST), they adjusted the value to zero, in others close to the threshold of perceptibility.

In experiments with a single piece, Friberg (1995b) used a simple greedy search method to fit the parameters of a single rule (the PHRASE ARCH rule) to the timing data of 28 performances of the first nine bars of Schumann's *Träumerei* (as measured by Bruno Repp, 1992). Even with this restriction to only one rule, a total of 18 parameters had to be fitted. Parameter settings were found that produced a reasonable fit, but it was not tested whether these parameter settings generalised to other pieces. Thus the experiment does not say much about the general validity or usefulness of the particular parameter values found. It does, however, indicate that the model as such is a useable description language for expressive performance. Other approaches on the same test piece used artificial neural networks (Bresin & Vecchio, 1995; Bresin, 1998) and fuzzy logic (Bresin et al., 1995) to approximate parameter settings that partly matched human expressive performances.

In a recent study by Sundberg and colleagues, it was examined how well the KTH model can be fitted to a

particular performance of a slow Mozart sonata movement (K.332, first 20 bars of the *Adagio*) manually, by trial and error (Sundberg et al., 2003). They focussed exclusively on expressive timing of the melody (in terms of deviations from the score) and determined similarity between the performance and the model with the correlation coefficient. The PHRASE ARCH rule (shaping phrases on different structural levels) seemed to dominate the performance, although they could not find one single *k* parameter for these initial bars (cf. Kroiss, 2000, see below). Therefore, they tested *k* values for each phrase individually. Here again, it was mostly the PHRASE ARCH rule that got the highest correlations, as well as HARMONIC CHARGE and negative DURATION CONTRAST.

Also outside KTH, there have recently been several attempts to evaluate the KTH model using various computational approaches. In a more extended study, Kroiss (2000) used genetic algorithms to fit a set of *k* parameter values onto a large number of pieces. The test data were performances of several complete Mozart piano sonatas performed by a professional Viennese pianist (cf. Widmer, 2001, 2002). He could not find a single set of parameter values that would produce a fit better than the baseline.

More recently, Zanon and De Poli attempted to fit the model to real-world data, first with fixed parameter values (Zanon & De Poli, 2003a), then with time-varying ones (Zanon & De Poli, 2003b). In the first approach (Zanon & De Poli, 2003a), several human performances of the first 24 bars of Beethoven's *Sonatine* in G major were used that involved some specific emotional intentions (as for example anger, fear, happiness, etc.). It was found that most of the KTH rules were quasi-orthogonal to each other with only few exceptions (e.g., MELODIC and HARMONIC CHARGE as well as DURATION CONTRAST). The different emotional intentions as represented in the performances were differentiated by particular rules, i.e., the PUNCTUATION rule evidently separated fearful or sad performances from happy or neutral ones (Zanon & De Poli, 2003a, p. 39).

In a second study, Zanon & De Poli (2003b) tried to overcome one basic restriction of the KTH model, namely, the fixed *k* parameter values over the duration of a piece. They tried to infer optimal *k* parameters for windows of approximately 1–2 bars, shifting the windows notewise along with the score. The degrees of freedom became too many, so that they restricted the number of included rules in a first estimation phase. In a second phase, both the window size and the number of considered rules were increased. The two test pieces were the Beethoven *Sonatine* as before and the slow movement of Mozart's K.332 piano sonata, as in Sundberg et al. (2003). Particular rules received typical *k* parameters for certain structural parts (phrases); however, the overall results became somewhat confusing, except for the Mozart excerpt, where the parameter estimation was more successful. The time-varying *k* parameter estimation was very sensitive to artefacts, such as the timing of grace notes (Zanon & De Poli, 2003b, p. 308). The parameter setting for the Mozart excerpt yielded an ambiguous emotional expression according to Bresin and Friberg (2000).[1]

The KTH model has also been used to model certain emotional colourings that might not be immediately inherent in the music structure (Gabrielsson & Juslin, 1996; Bresin & Friberg, 2000). For example, the DURATION CONTRAST rule was found to especially distinguish between a sad and a happy performance. Certain specifically selected subsets of rules and *k* parameter settings ("emotional rule palettes", see Bresin & Friberg, 2000) were derived from measured performances in order to model emotional colouring of arbitrary performances.

This extension to emotionality has led to a more comprehensive computational model of expressive performances (the *GERM model*, Juslin et al., 2002). This model includes, besides the generative KTH model ("G") and the emotional models reported above ("E"), also random variability ("R") within certain bounds, and analogies to physical motion ("M"). Recently, it was proposed to extend it with a factor of "stylistic unexpectedness" ("S," Juslin, 2003). Such a combination of several somewhat orthogonal model approaches expand the degrees of freedom to a rather large size, so that it will most likely be very difficult to perform stringent empirical evaluations of this complex model.

In summary, there is evidence that the KTH rule model is a viable representation language for describing expressive performance. To what extent it can account for the observed variations in large collections of performances of truly complex music is still an open issue.

4. The Todd model

In contrast to the KTH model that used the unique "analysis-by-synthesis" approach, the following models may be summarised under the notion of "analysis-by-measurement", because they obtain their empirical evidence directly from measurements of human expressive performances.

In the late 1980s and early 1990s, Neil Todd proposed a number of structure-level models of expressive timing (Todd, 1985, 1989a,b) and dynamics (Todd, 1992), based on earlier work by researchers like Shaffer and co-workers (Shaffer et al., 1985; Shaffer & Todd, 1987), Clarke (1985), Gabrielsson (1987), and Repp (1990c, 1992).

4.1 The basic model

The essence of these models is the assumption that there is a direct link between certain aspects of the musical structure

[1] The *k* parameter settings of Zanon and De Poli (2003b) were considerably different from the settings found by Sundberg et al. (2003) for the same performance of the same Mozart piece. Although the computational study claimed to estimate time-varying parameters, they had only two sub-sections on which parameters were independently fitted, whereas the analysis-by-synthesis study involved a total of 16 subsections.

(i.e., grouping structure) and the performance and, secondly, that this relation can be modeled by one single, simple rule. The grouping structure, as defined in the theoretical framework of Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music*, identifies points of various degrees of stability in the music, at different structural levels simultaneously (cf. the "time-span reduction" component of Lerdahl & Jackendoff's theory). It is assumed by Todd (1985) that a performer slows down at those points of stability in order to enable the listener to perceive the hierarchical structure of the music. This is consistent with measurements of expressive performances that observed tempo to be minimal at phrase boundaries and maximal between them.[2]

The first approaches modelled production (Todd, 1985, 1989a) and perception (Todd, 1989b) of expressive timing exclusively. Todd's last contribution included the interaction of timing and dynamics as well (Todd, 1992). In Todd (1989a), a recursive look-ahead procedure was introduced that reduced the (theoretically) high demands on working memory introduced by the first model (Todd, 1985). The interaction between timing and dynamics was modelled by the simple relation "the faster, the louder", where intensity is proportional to the squared tempo (Todd, 1992, p. 3544). Thus, the hierarchical grouping structure of the music directly controls the instantaneous tempo (the more closure and thus, the more important the phrase boundary, the slower the performed tempo) as well as the expressive dynamics (in terms of hammer velocity at a computer-controlled piano). At each level of the hierarchy, the model increases tempo and dynamics towards the middle of phrases and vice versa. A basic assumption in these models is the analogy to physical motion (see also Kronman & Sundberg, 1987; Todd, 1992, 1995) that may even have a (neuro-)physiological basis in the vestibular system (Todd, 1992, p. 3549).

4.2 Empirical evaluation

The simplistic nature of Todd's model has the potential advantage that its theoretical assumptions can be tested relatively easily. In his own contributions, Todd (1985, 1989a, 1992) compared the model's output with the tempo and dynamics curves from one or two performances of selected pieces by Haydn (the beginning of the Adagio of the Sonata No. 59), Mozart (the theme of the K.331 Sonata), and Chopin (the A♭ major Nouvelle Étude No. 3 and the F♯ minor Prelude op. 28 No. 8, as reported in Shaffer & Todd, 1987). The real and algorithmic curves look partly similar in the figures provided; however, no quantitative evaluations were performed.

In a more recent empirical study (Clarke & Windsor, 2000), a panel of human listeners evaluated both human performances and algorithmic performances created with the Todd model. In a pilot experiment, two different phrase interpretations of the first four bars of Mozart's K.331 (see also Gabrielsson, 1987) were realised by the Todd model and by two professional pianists. In the human performances, listeners identified the two phrasing versions best when they listened to dynamics changes only and to both timing and dynamics (articulation was eliminated from this experiment). The artificial performances could not convey the two phrase interpretations. The Todd model and the pianists sometimes did the opposite: while the humans emphasised the unexpected eighth note as an upbeat to a new phrase, the Todd model stubbornly played it softest, as the beginning of a new phrase. Also, the model did not shape the *Siciliano* pattern in the typical systematic way as pianists usually realise it (see also Gabrielsson, 1983, 1987). The model did perform better in communicating two phrase interpretations of a simple eight-bar melody to musically trained participants. A general finding in these experiments was that expressive timing and dynamics did not relate to one another in the simple manner suggested by the model (Todd, 1992). This finding (faster–louder is too simple) is also supported by other empirical data (Palmer, 1996a; Windsor & Clarke, 1997; Repp, 1999).

In another empirical study, Windsor and Clarke (1997) tuned the model's parameters to human performances of the initial measures of Franz Schubert's G♭ major Impromptu, D. 899, No. 3. They created several different artificial versions using different level weightings of the Todd (1992) model and evaluated them with regression analysis against two repeated human performances by one professional pianist. The best fit was always between the two repetitions of the pianist; the best algorithmic performances was one with different level weights for timing and dynamics (the so-called "hybrid performance" – Windsor & Clarke, 1997, p. 141). Timing required more emphasis on lower structural levels, whereas dynamics on higher levels (these results are similar to findings by Widmer & Tobudic, 2003, where it was found that intensity is better modeled by quadratic polynomials than tempo). Further examining the differences between model and human performances, Windsor and Clarke introduced the notion of "residuals", which give more detailed insight into the details of a performance. As they argue, the most interesting is what is *not* explained by the model (Clarke & Windsor, 2000, p. 312). In this way the Todd model was used as an analysis tool to assess the idiosyncrasies of human performances. Similar results were recently obtained in a comparison, with animated two-dimensional tempo–dynamics visualisations, between the hybrid performance by Windsor and Clarke (1997) and Alfred Brendel's professional performance (Langner & Goebl, 2003).

5. The Mazzola model

A rather different model based mainly on mathematical considerations is the "Mazzola model". The Swiss mathematician and Jazz pianist Guerino Mazzola developed his

[2] This basic principle of the Todd model is also reflected in the PHRASE ARCH rules of the KTH model.

mathematical music theory and performance model from the 1980s up to the present (Mazzola, 1990; Mazzola & Zahorka, 1994a; Mazzola et al., 1995; Mazzola & Göller, 2002). His most recent book (Mazzola, 2002) extends over more than 1300 pages and gives a supposedly comprehensive survey of the vast theoretical background as well as its computational implementation. While previous implementations of this model required special computer hardware (NeXTStep platform, see Mazzola & Zahorka, 1994b), the most recent version of the software package runs on Mac OS X (Müller, 2002) and is freely available under the GNU public license at http://www.rubato.org.

5.1 The basic model

The Mazzola model builds on an enormous theoretical background, namely, the "mathematical music theory" (Mazzola, 1990) that not only covers various aspects of music theory and analysis through a highly complex mathematical approach, but involves all sorts of philosophical, semiotic, and aesthetic considerations as well. Every step of the theory is explained in specifically mathematical terms and with a special terminology that is greatly different from what is commonly used in performance research. The whole theory is rather "hermetic", if one may say so. Therefore, we restrict ourselves here to the more concrete computational facets as they have been reported in the literature.

The Mazzola 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 as, e.g., meter, melody, or harmony. Each of these is implemented in particular plugins, the so-called RUBETTEs, 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.

To illustrate, we report on an application of the MetroRUBETTE (Fleischer et al., 2000). The (inner) metrical analysis was simply performed by computing all possible combinations of equally-spaced sequences of note onsets in a musical score with varying inter-onset intervals (e.g., from the smallest note value occurring, i.e., a sixteenth note, up to a full bar) and adding up the amount of participation in such regular patterns for each note. These sums are referred to as the so-called (inner) metrical weights. In the example brought forward by Fleischer et al. (2000, Fig. 2), this (inner) metrical analysis produced a result that is totally different from a conventional (outer) metrical structure (e.g., Lerdahl & Jackendoff, 1983), partly because it ignored rests, which may coincide with metrically strong positions. Other examples from this study were more convincing, although the basic concept of such an analytical approach did not become explicit. This study used a linear mapping between metrical weight and tone intensity to generate artificial performances. Unfortunately, the artificial performances were not compared with real performance data (Fleischer et al., 2000).

The software package also includes a plugin for analysing expressive performance data (the EspressoRUBETTE). It analyses MIDI-like data input, performs score-to-performance matching, and extracts vector fields for a given human performance (Mazzola, 2002, pp. 903–929). These operations are summarised here by the term "inverse performance theory". The plugin visualises the extracted performance data in several ways; alongside classical pianoroll notation (pitch against time or score time) it displays the extracted performance vector fields as two-dimensional colour contour plots. However, since these visualisations lack labels, legends, or explanation (e.g., Mazzola, 2002, p. 924), their meaning remains rather unclear to the reader.

5.2 Empirical evaluation

Unfortunately, we could find no empirical investigations outside the "Zürich School"[3] that tried to systematically evaluate the model. One contribution that used RUBATO to generate various artificial performances of parts of J. S. Bach's "Kunst der Fuge" is (Stange-Elbe, 1999). However, no comparisons with real performances were made or even intended. Similar experiments were reported in Mazzola (2002, pp. 833–852), again without any quantitative evaluation with empirical data.

In an experiment carried out by Jan Beran (reported in Mazzola, 2002, pp. 871–901), a multiple regression analysis was conducted on the tempo curves of 28 performances of the "Träumerei" as measured by Bruno Repp (1992). The metrical, harmonic, and melodic weights as provided by the RUBATO software served as independent variables. The overall model could explain 84% of the average tempo curve of the 28 performances, each of the three analytical components contributing about equally to the model. Although the mapping between analytical weights and actual performance parameters is claimed to be extremely complex (Mazzola, 2002), Beran could explain a large portion of variance with a simple linear mapping.

6. The machine learning model

An alternative way of building computational models of expressive performance is to start from large amounts of empirical data – precisely measured performances by skilled musicians – and to have the computer autonomously discover significant regularities in the data, via *inductive machine*

[3] The term "Zurich School" was introduced by Thomas Noll (see Mazzola, 2002, p. 744); it involves, besides Mazzola's group at Zürich University, the *Research Group for Mathematical Music Theory (MaMuTh)* at the Technical University in Berlin.

learning and *data mining* techniques. Some of these learning algorithms produce general performance rules that can be interpreted and used directly as predictive computational models. This is the approach that has been developed and pursued by our research group in Vienna over the past years (e.g., Widmer, 1995a,b, 1996, 2000, 2002, 2003; Widmer et al., 2003). The following two subsections report on machines learning to predict both local, note-level expressive deviations and higher-level phrasing patterns, and show how these two types of models can be combined to yield an integrated, multi-level model of expressive timing and dynamics.

6.1 The note-level model

In a first attempt to find general rules of performance, we developed a new inductive rule learning algorithm (Widmer, 2003) and applied it to the task of learning note-level rules for timing, dynamics, and articulation, where by "note-level" we mean rules that predict how a pianist is going to play a particular note in a piece – slower or faster than notated, louder or softer than its predecessor, staccato or legato. This should be contrasted with higher-level expressive strategies like the shaping of an entire musical phrase (e.g., with a gradual ritardando towards the end), which will be addressed in the next section.

The training data used for the experiments consisted of recordings of 13 complete piano sonatas by W. A. Mozart (K.279–284, 330–333, 457, 475, and 533), performed by a Viennese concert pianist on a Bösendorfer 290SE computer-controlled piano, from which every detail of timing, dynamics, and articulation could be computed. The resulting dataset comprises more than 106000 performed notes and represents some four hours of music.

The experiments were performed on the melodies (usually the soprano parts) only, which gives an effective training set of 41116 notes. Each note was described by 29 attributes that represent both intrinsic properties (such as scale degree, duration, metrical position) and some aspects of the local context (e.g., local melodic contour around the note). From these 41116 examples of played notes, the computer learned a small set of 17 quite simple classification rules that predict a surprisingly large number of the note-level choices of the pianist. For instance, four rules were discovered that together correctly predict almost 23% of all the situations where the pianist lengthened a note relative to how it was notated (which corresponds to a local slowing down of the tempo). To illustrate, the following is an example of a particularly simple and general rule that was found by the computer:

RULE TL2:
abstract_duration_context = equal-longer
& metr_strength ≤ 1
⇒ lengthen

"Given two notes of equal duration followed by a longer note, lengthen the note (i.e., play it more slowly) that precedes the final, longer one, if this note is in a metrically weak position ('metrical strength' ≤ 1)".

This is an extremely simple principle that turns out to be surprisingly general and precise: rule TL2 correctly predicts 1894 cases of local note lengthening, which is 14.12% of all the instances of significant lengthening observed in the training date. The number of incorrect predictions is 588 (2.86% of all the counterexamples).

The complete set of rules is described in detail in (Widmer, 2002), where also the generality and robustness of the rules is quantified, based on extensive experiments with real data. Interestingly, some of the rules discovered by the machine bear a strong resemblance to performance rules postulated in the KTH model. In this way, the machine learning approach provides further circumstantial evidence for the relevance and validity of the KTH model.

6.2 The multi-level model

Referring as it does to single notes and their local context, the above-mentioned rule-based model can only be expected to account for a rather small part of the expressive patterns observed in real performances. Musicians understand the music in terms of a multitude of more abstract structures (e.g., motifs, groups, phrases), and they use tempo, dynamics, and articulation to "shape" these structures. Music performance is a multi-level phenomenon, with musical structures and performance patterns at various levels embedded within each other.

Accordingly, recent work at our laboratory has focussed on inductively learning multi-level models of expressive timing an dynamics from recordings. The goal is for the computer to learn to predict what kind of elementary tempo and dynamics "shapes" (like a gradual crescendo–decrescendo) a performer will apply to a given musical phrase in a given context, at a given level of the phrase hierarchy. Underlying the model are a number of rather simplistic (but necessary) *assumptions*: one, that the expressive timing or dynamics gestures applied to a phrase by a performer can be reasonably approximated by a family of (quadratic) curves (this assumption is similar to the assumption underlying the Todd model – see above); two, that a complete multi-level performance can be reasonably represented as a linear combination of such expressive shapes at different hierarchical levels; and three, that, all other things being equal, similar phrases will tend to be played similarly by pianists.

Clearly, none of these assumptions can be expected to be entirely true, but they provide the foundation for an operational multi-level model of expressive phrasing that is embodied in a machine learning system (Widmer & Tobudic, 2003). The system takes as input a set of example performances by musicians, represented by the musical score, a hierarchical phrase analysis of the music, plus tempo and dynamics curves that describe the timing and dynamics aspects of the expressive performances. It decomposes the

given performance curves by fitting quadratic approximation functions to the sections of the curves associated with the individual phrases, in a levelwise fashion, thus associating an elementary expressive "shape" with each phrase at each level. It then predicts elementary expressive shapes for phrases in new pieces, based on perceived similarity to known phrases, again at multiple levels, and combines the individual shapes in a linear fashion into complex composite expression (tempo and dynamics) curves. Learning is thus done in a case-based fashion, based on analogies between pieces. There is no explicit prediction model; the essence of the model is embodied in the three assumptions mentioned above.

There is a natural way of combining the phrase-level prediction model with the rule-based learning approach described above: after fitting quadratic approximation polynomials to a given tempo or dynamics curve and "subtracting" the resulting approximation from the original curve, what is left is what Windsor and Clarke (1997) called the "residuals", i.e., those low-level, local timing and dynamics deviations that cannot be explained by reference to extended structural entities like phrases. The rule learning algorithm described in the previous section can be used to learn a rule-based model of these local effects. Combining the two models then yields a predictive computational model of expressive timing and dynamics that takes into account both the hierarchical structure of the music an local musical context. Details of the entire procedure and extended experimental results can be found in (Widmer & Tobudic, 2003).

6.3 Empirical evaluation

Both the note-level rule model and the multi-level model have been extensively tested on real performances. In (Widmer, 2002), coverage and precision on the large training set of 41116 played notes are listed in detail for each discovered rule, thus giving a very detailed picture of the relative generality and reliability of the rules. Also, quantitative experiments with large numbers of new test performances are described that show that the rules carry over to other performers and even music of a different style with virtually no loss of precision. For example, the rules were tested on performances of quite different music (Chopin), and quite surprisingly, some of them exhibited significantly higher prediction accuracy than on the original (Mozart) data they had been learned from. The machine seems to have discovered some fundamental, though mostly simple, local performance principles.

The predictive performance of the multi-level model is quantified in (Widmer & Tobudic, 2003), again by measuring how well it manages to predict the details of a pianist's performances. More precisely, it was measured how closely the tempo and dynamics curves predicted by the model match the curves extracted from actual human performances (of new, previously unseen pieces). Experiments with a substantial number of extended performances by one particular concert pianist show that on average, the curves predicted by the model fit the actual performances better than chance and, in particular, better than straight lines that would correspond to strictly mechanical performances. The results are better for dynamics than for timing and tempo. More detailed investigations revealed that the poor performance of the model in the tempo domain is largely due not to problems of the learning algorithm, but to fundamental problems of approximating real tempo curves with hierarchies of quadratic functions. In other words, quadratic functions may not be a good modeling language for expressive timing.

Quantitative improvements over these first results are reported in Tobudic and Widmer (2003a), where the case-based learning algorithm was optimised. The model was then extended towards an explicit modelling of the hierarchical context of phrases, using a first-order logic representation language and a newly developed measure of structural similarity (Tobudic & Widmer, 2003b); again, this led to some quantitative improvements. Currently ongoing research suggests that the model's predictive accuracy can be improved still further by refining the definition and representation of musical context.

Also, an "expressive" Mozart performance generated by the multi-level model after learning from real performances of other Mozart pieces won Second Prize in a "computer performance rendering contest" in Tokyo in 2002, where computer interpretations of classical music were rated by listeners. That also indicates that what the machine extracts from the performance data seem to be at least "reasonable" performance patterns.

7. Current research: quantification of individual style

Predictive models like those presented above generally focus on fundamental, common performance principles, that is, on those aspects that most performances have in common. In this section, we will briefly address some ongoing research that also tries to quantify and characterize, if not yet model in a predictive way, the *differences* between artists, that is, aspects of personal artistic performance style.

Of course, there have also been attempts at measuring, quantifying, and describing stylistic performance differences. To name just one, Bruno Repp (1992) has presented a striking demonstration of systematic differences in the styles of different famous pianists. In a study involving the timing curves extracted from 28 performances of Robert Schumann's *Träumerei* by 24 famous pianists, he showed that while there was strong agreement between the performances at a global level – all pianists more or less observed the major ritardandi in the piece and clearly expressed the large-scale phrase structure of the piece through their timing – the differences between the pianists increased at lower levels of the structural hierarchy. A statistical analysis revealed a number of characteristic and distinctive phrasing behaviours, some of which could be associated (in a statistical sense) with certain pianists.

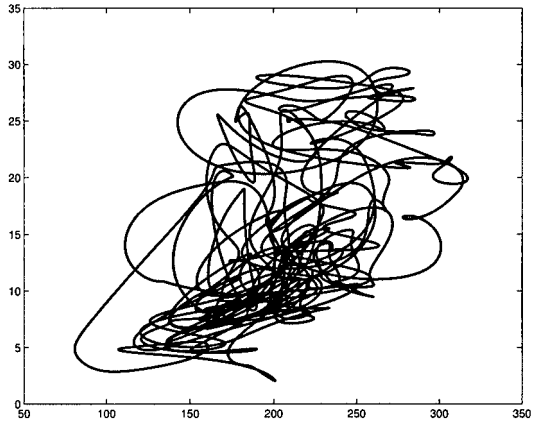

Fig. 1. Smoothed tempo–loudness trajectory representing a performance of Frédéric Chopin's Ballade op.47 in Ab major by Artur Rubinstein. Horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in *sone*.

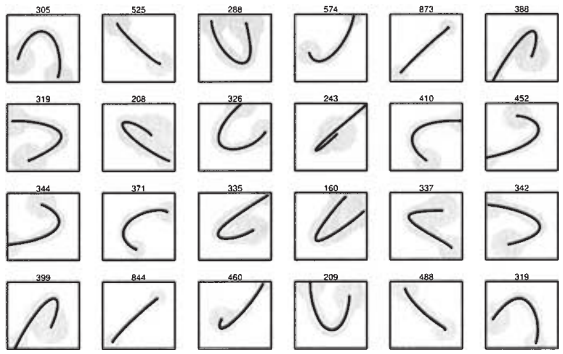

Fig. 2. A "Mozart performance alphabet" (cluster prototypes) computed by segmentation, mean and variance normalization, and clustering, from performances of Mozart piano sonatas by six pianists (Daniel Barenboim, Roland Batik, Vladimir Horowitz, Maria João Pires, András Schiff, Mitsuko Uchida). To indicate directionality, dots mark the end points of segments. Shaded regions indicate the variance within a cluster.

Due to the huge efforts involved in manually measuring details of expressive timing from audio recordings, Repp's analyses were limited to one particular piece. In our institute in Vienna, a large-scale project is currently being undertaken which aims at analysing truly large amounts of empirical performance data derived from recordings (Widmer et al., 2003). With the help of new computational methods that support the semi-automatic measurement of timing and dynamics from audio recordings, hundreds of recordings are being measured and characterised in terms of beat-level timing and global loudness changes.

7.1 Visualisation: performance trajectories

The resulting performance data – beat-level tempo and dynamics curves – can be represented in an integrated way as trajectories in a tempo – loudness space that show the joint development of tempo and dynamics over time (Langner & Goebl, 2003). Figure 1 shows a complete trajectory representing a performance of a Chopin Ballade by Artur Rubinstein. The line is produced by interpolating between the measured tempo and dynamics points, and smoothing the result with a Gaussian window to make the general trends visible. The degree of smoothing controls the amount of local variation that becomes visible.

A first intuitive analysis of high-level strategies characterising individual performances is facilitated by an interactive visualisation system called the Performance Worm (Dixon et al., 2002) that computes and visualises such performance trajectories via computer animation. But the trajectory representation also provides the basis for more detailed quantitative analysis, with data analysis (data mining) methods from the field of Artificial Intelligence. Various avenues towards the characterisation of individual performance style are being followed, and we will briefly introduce some of these in the following subsections.

7.2 Characterisation: performance alphabets

The performance trajectories must first be converted into a form that is accessible to the automated data analysis machinery provided by data mining. To that end, the trajectories are cut into short segments of fixed length, e.g., two beats, which are then optionally subjected to various normalisation operations. The resulting segments can be grouped into classes of similar patterns via *clustering*. The centers of these clusters – the cluster prototypes – represent a set of typical elementary tempo – loudness patterns that can be used to approximately reconstruct a "full" trajectory (i.e., a complete performance). In that sense, they can be seen as a simple *alphabet* of performance, restricted to tempo and dynamics. Figure 2 displays such an alphabet computed from a set of Mozart sonata recordings by different artists.

Such performance alphabets support a variety of quantitative analyses. A first useful step consists in the visualisation of the distribution of performance patterns over pianists, pieces, musical styles, etc. (Pampalk et al., 2003). That provides a very global view of aspects of personal style, such as "pianist A tends to use abrupt tempo turns combined with rather constant dynamics" or "pianist B combines a crescendo with a ritardando much more often than other pianists". An example of such a visualisation can be found in (Widmer et al., 2003). An extensive study along these lines using Chopin performances by several famous pianists has recently revealed a number of characteristic performance strategies (Goebl et al., 2004).

A more direct way of studying individual performance style is to search for specific extended patterns in performance trajectories that are somehow typical of a particular pianist. We are currently developing data mining algorithms that do this by searching for frequent and discriminative substrings in performance trajectories that are coded as sequences of performance alphabet "letters". To illustrate,

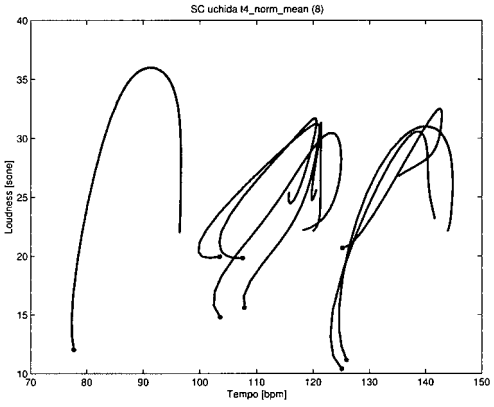

Fig. 3. A set of similar performance shapes possibly characteristic of Mitsuko Uchida, discovered in tempo–loudness trajectories of Mozart sonata recordings. To indicate directionality, a dot marks the *end point* of a segment.

Figure 3 shows several instances of a particular performance pattern found in various Mozart performances by Mitsuko Uchida, and rarely found in any of the other pianists mentioned in Figure 2. The pattern represent a particular way of combining a crescendo–decrescendo (the vertical movement of the trajectory) with a slowing down during the loudness maximum and afterwards. Assessing the statistical and, above all, musical significance of such discovered patterns is a rather difficult problem that we are currently working on.

7.3 Classification: automatic identification of performers

Another way of trying to quantify individual performance style is to develop computer programs that attempt to identify artists on the basis of their performance characteristics. In initial experiments with state-of-the-art machine learning algorithms we showed that a computer could learn to partly differentiate several pianists from the Vienna University of Music, given only one test recording by each pianist (Stamatatos & Widmer, 2002). The recordings were made on a Bösendorfer 290Se computer-controlled piano, so that very precise and detailed performance information was available.

Recently, we have managed to show that a computer can also learn, to some extent, to distinguish famous pianists based only on aspects of their high-level performance trajectories (Zanon & Widmer, 2003b). For instance, in pairwise identification experiments (Zanon & Widmer, 2003a), computer programs trained to distinguish between two particular famous pianists achieved correct recognition rates, on new recordings not used during training, of 80% and above for certain pianist pairs (e.g., András Schiff vs. Maria João Pires). The results are still very preliminary, and we have not yet managed to pinpoint precisely which features are the most distinguishing ones, but the current results do indicate that there is relevant information contained in this high-level representation, and that the machine may help us in getting a firmer grip on the elusive notion of personal performance style.

8. Conclusions

This article has tried to give a comprehensive overview of the state of the art of computational modelling of expressive music performance. Four somewhat complementary models and approaches were presented in some detail and, wherever possible, empirical evaluations of the models on real performance data were reported. In addition, currently ongoing research on the formal characterisation of individual performance style was briefly presented.

The results clearly show that there is still ample room for further research, and the field of computational performance modelling continues to be active. One recent trend is a research focus on basic emotions (or "expressive intentions") and the way they are expressed and controlled in performed music (e.g., De Poli et al., 1998). Knowledge gleaned from such studies inspires research on new control spaces and devices for the rendering and control of emotional aspects of music performance (e.g., Bresin & Friberg, 2000; Canazza et al., 2003).

Generally, the idea of a creative activity being predictable and, more specifically, the notion of a direct "quasi-causal" relation between musical score and performance is rather 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. Like any human intellectual activity, music performance is a complex social and cognitive phenomenon with a rich context. 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 a very generous START Research Prize by the Austrian Federal Government, administered by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung (FWF)* (project no. Y99-INF), and the project "Interfaces to Music" (WWTF project no. CI010). Additional support for our research on AI, machine learning, scientific discovery, and music is provided by the European project HPRN-CT-2000-00115 (MOSART) and the EU COST Action 282 (Knowledge Exploration in Science and Technology). The Austrian Research Institute for

Artificial Intelligence acknowledges basic financial support by the Austrian Federal Ministry for Education, Science, and Culture, and the Austrian Federal Ministry for Transport, Innovation and Technology. The authors would like to thank Simon Dixon, Elias Pampalk, and Asmir Tobudic for their cooperation and for helpful comments on this article.

References

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

Becking, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle*. Augsburg, Germany: Filser.

Bresin, R. (1998). Artificial neural networks based models for automatic performance of musical scores. *Journal of New Music Research*, *27*, 239–270.

Bresin, R. (2001). Articulation rules for automatic music performance. In: A. Schloss, & R. Dannenberg (Eds), *Proceedings of the 2001 International Computer Music Conferenc, Havana, Cuba* (pp. 294–297). San Fransisco, CA: International Computer Music Association.

Bresin R. (2002). Importance of note-level control in automatic music performance. In: R. Hiraga (Ed.), *Proceedings of the ICAD 2002 Rencon Workshop on Performance Rendering Systems* (pp. 1–6). Kyoto, Japan: The Rencon Steering Group.

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

Bresin, R., De Poli, G., & Ghetta, R. (1995). Fuzzy performance rules. In: A. Friberg, & J. Sundberg (Eds), *Proceedings of the KTH Symposion on Grammars for Music Performance* (pp. 15–36). Stockholm, Sweden: Department of Speech Communication and Music Acoustics.

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

Bresin, R., & Vecchio, C. (1995). Neural networks play Schumann. In: A. Friberg, & J. Sundberg (Eds), *Proceedings of the KTH Symposion on Grammars for Music Performance* (pp. 5–14). Stockholm, Sweden: Department of Speech Communication and Music Acoustics.

Bresin, R., & 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, 1–6.

Canazza, S., De Poli, G., Drioli, C., Rodà, A., & Vidolin, A. (2003). An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research*, 32, 281–294.

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. (1988). Generative principles in music performance. In: J.A. Sloboda (Ed.), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* (pp. 1–26). Oxford: Clarendon Press.

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

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In: J. Sundberg (Ed.), *Studies of Music Performance* (Vol. 39, pp. 76–181). Stockholm: Publications issued by the Royal Swedish Academy of Music.

Clynes, M. (1986). Generative principles of musical thought: Integration of microstructure with structure. *Communication and Cognition AI*, *3*, 185–223.

Clynes, M. (1987). What can a musician learn about music performance from newly discovered microstruture principles (PM or PAS)? In: A. Gabrielsson (Ed.), *Action and Perception in Rhythm and Music* (Vol. 55, pp. 201–233). Stockholm, Sweden: Publications issued by the Royal Swedish Academy of Music.

Clynes, M. (1990). Some guidelines for the sunthesis and testing of pulse microstructure in relation to musical meaning. *Music Perception*, *7*, 403–422.

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

Dannenberg, R.D., & Derenyi, I. (1998). Combining instrument and performance models for high-quality music synthesis. *Journal of New Music Research*, *27*, 211–238.

De Poli, G., Rodà, A., & Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *Journal of New Music Research*, *27*, 293–321.

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

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

Fleischer, A., Mazzola, G., & Noll, T. (2000). Computergestützte Musiktheorie. Zur Konzeption der Software RUBATO für musikalische Analyse und Performance. *Musiktheorie*, *15*, 314–325.

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

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

Friberg, A. (1995b). Matching the rule parameters of Phrase Arch to performances of "Träumerei:" A preliminary study. In: A. Friberg, & J. Sundberg (Eds.), *Proceedings of the KTH Symposion on Grammars for Music Performance*

(pp. 37–44). Stockholm, Sweden: Department of Speech Communication and Music Acoustics.

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

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

Friberg, A., Frydén, L., Bodin, L., & Sundberg, J. (1991). Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, *15*, 49–55.

Friberg, A., & Sundberg, J. (1987). How to terminate a phrase. An analysis-by-synthesis experiment on a perceptual aspect of music performance. In: A. Gabrielsson (Ed.), *Action and Perception in Rhythm and Music* (Vol. 55, pp. 49–55). Stockholm, Sweden: Publications issued by the Royal Swedish Academy of Music.

Friberg, A., & 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*, 1469–1484.

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

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

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

Gabrielsson, A. (1985). Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception*, *3*, 59–86.

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

Gabrielsson, A. (1999). Music Performance. In: D. Deutsch (Ed.), *Psychology of Music* (2nd edn., pp. 501–602). San Diego, CA: Academic Press.

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

Gabrielsson, A., & Juslin, P.N. (1996). Emotional expression in music performance: Between the performer's intention and the listeners experience. *Psychology of Music*, *24*, 68–91.

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

Goebl, W., Pampalk, E., & Widmer, G. (2004). Exploring expressive performance trajectories: Six famous pianists play six Chopin pieces. In: *Proceedings of the 8th International Conference on Music Perception and Cognition (ICMPC'04)*. Evanston, Illinois.

Honing, H. (2003). The final ritard: On music, emotion, and kinematic models. *Computer Music Journal*, *27*, 66–72.

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

Juslin, P.N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae*, *Special issue 2001–2002*, 63–122.

Kroiss, W. (2000). *Parameteroptimierung für ein Modell des musikalischen Ausdrucks mittels Genetischer Algorithmen*. Master's thesis, Department of Medical Cybernetics and Artificial Intelligence, University of Vienna, Vienna, Austria.

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

Langner, J., & Goebl, W. (2003). Visualizing expressive performance in tempo–loudness space. *Computer Music Journal*, *27*, 69–83.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge (MA), London: MIT Press.

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

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

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

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

Mazzola, G., & Zahorka, O. (1994b). The RUBATO performance workstation on NeXTStep. In: *Proceedings of the 1994 International Computer Music Conference*, *Århus*, *Denmark* (pp. 102–108). San Fransisco, CA: International Computer Music Association.

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

Müller, S. (2002). Computer-aided musical performance with the Distributed Rubato enironment. *Journal of New Music Research*, *31*, 233–237.

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

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

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

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

Pampalk, E., Goebl, W., & Widmer, G. (2003). Visualizing changes in the inherent structure of data for exploratory feature selection. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 157–166). Washington, DC: ACM.

Repp, B.H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composer's pulses. *Music Perception*, *6*, 243–274.

Repp, B.H. (1990a). Composer's pulses: Science or art. *Music Perception*, *7*, 423–434.

Repp, B.H. (1990b). Further perceptual evaluations of pulse microstructure in computer performances of classical piano music. *Music Perception*, *8*, 1–33.

Repp, B.H. (1990c). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, *88*, 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*, 2546–2568.

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

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*, 1085–1100.

Repp, B.H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, *105*, 1972–1988.

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

Seashore, C.E. (1938). *Psychology of Music*. New York: McGraw-Hill. (reprinted 1967 by Dover Publications, New York.)

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

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

Shaffer, L.H., Clarke, E.F., & Todd, N.P.M. (1985). Metre and rhythm in pianoplaying. *Cognition*, *20*, 61–77.

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

Stamatatos, E., & Widmer, G. (2002). Music performer recognition using an ensemble of simple classifiers. In: F.V. Harmelen (Ed.), *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI'2002)*, *July 21–26*, *2003*, *Lyon* (pp. 335–339). Amsterdam: IOS Press.

Stange-Elbe, J. (1999). Computer-assisted analysis and performance: A case study using RUBATO. In: I. Zannos (Ed.), *Music and Signs*. *Semiotic and Cognitive Studies in Music* (pp. 231–241). Bratislava: ASCO: Art and Science.

Sundberg, J. (1993). How can music be expressive. *Speech Communication*, *13*, 239–253.

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

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

Sundberg, J., Friberg, A., & Bresin, R. (2003). Attempts to reproduce a pianist's expressive timing with Director Musices performance rules. *Journal of New Music Research*, *32*, 317–325.

Sundberg, J., Friberg, A., & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review*, *8*, 89–109.

Sundberg, J., Friberg, A., & Frydén, L. (1991a). Common secrets of musicians and listerners – An analysis-by-synthesis study of musical performance. In: P. Howell, R. West, & I. Cross (Eds.), *Representing Musical Structure* (pp. 161–197). London: Academic Press.

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

Sundberg, J., Frydén, L., & Askenfelt, A. (1983). What tells you the player is musical? An analysis-by-synthesis study of music performance. In: J. Sundberg (Ed.), *Studies of Music Performance* (Vol. 39, pp. 61–75). Stockholm, Sweden: Publication issued by the Royal Swedish Academy of Music.

Thompson, W.F. (1989). Composer-specific aspects of musical performance: An evaluation of Clynes's theory of pulse for performances of Mozart and Beethoven. *Music Perception*, *7*, 15–42.

Thompson, W.F., Sundberg, J., Friberg, A., & Frydén, L. (1989). The use of rules for expression in the performance of melodies. *Psychology of Music*, *17*, 63–82.

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

Tobudic, A., & Widmer, G. (2003a). Playing mozart phrase by phrase. In: K.D. Ashley, & D.G. Bridge (Eds.), *Proceedings of the 5th International Conference on Case-based Reasoning (ICCBR'03)*, *Trondheim*, *Norway* (pp. 552–566). Berlin: Springer.

Tobudic, A., & Widmer, G. (2003b). Relational ibl in music with a new structural similarity measure. In: T. Horváth, & A. Yamamoto (Eds.), *Proceedings of the 13th International Conference on Inductive Logic Programming (ILP'03)*, *Szeged*, *Hungary* (pp. 365–382). Berlin: Springer.

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

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

Todd, N.P.M. (1989b). Towards a cognitive theory of expression: The performance and perception of Rubato. *Contemporary Music Review*, *4*, 405–416.

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

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

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

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

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

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

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

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

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

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

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

Windsor, W.L., & Clarke, E.F. (1997). Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Perception*, *15*, 127–152.

Zanon, P., & De Poli, G. (2003a). Estimation of parameters in rule systems for expressive rendering in musical performance. *Computer Music Journal*, *27*, 29–46.

Zanon, P., & De Poli, G. (2003b). Time-varying estimation of parameters in rule systems for music performance. *Journal of New Music Research*, *32*, 295–315.

Zanon, P., & Widmer, G. (2003a). Learning to recognize famous pianists with machine learning techniques. In: R. Bresin (Ed.), *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, *August 6–9*, *2003* (Vol. 2, pp. 581–584). Stockholm, Sweden: Department of Speech, Music, and Hearing, Royal Institute of Technology.

Zanon, P., & Widmer, G. (2003b). Recognition of famous pianists using machine learning algorithms: First experimental results. In: *Proceedings of the 14th Colloquium on Musical Informatics (CIM'2003)* (pp. 84–89). Florence, Italy.