In Search of the Horowitz Factor

Gerhard Widmer, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic

■ The article introduces the reader to a large interdisciplinary research project whose goal is to use AI to gain new insight into a complex artistic phenomenon. We study fundamental principles of expressive music performance by measuring performance aspects in large numbers of recordings by highly skilled musicians (concert pianists) and analyzing the data with state-of-the-art methods from areas such as machine learning, data mining, and data visualization. The article first introduces the general research questions that guide the project and then summarizes some of the most important results achieved to date, with an emphasis on the most recent and still rather speculative work. A broad view of the discovery process is given, from data acquisition through data visualization to inductive model building and pattern discovery, and it turns out that AI plays an important role in all stages of such an ambitious enterprise. Our current results show that it is possible for machines to make novel and interesting discoveries even in a domain such as music and that even if we might never find the "Horowitz Factor," AI can give us completely new insights into complex artistic behavior.

The potential of AI, particularly machine learning and automated discovery, for making substantial discoveries in various branches of science has been convincingly demonstrated in recent years, mainly in the natural sciences ([bio]chemistry, genetics, physics, and so on) (Hunter 1993; King et al. 1992; Muggleton, King, and Sternberg 1992; Shavlik, Towell, and Noordewier 1992; Valdés-Pérez 1999, 1996, 1995). However, can AI also facilitate substantial discoveries in less easily quantifiable domains such as the arts? In this article, we want to demonstrate that

it can. We report on the latest results of a long-term interdisciplinary research project that uses AI technology to investigate one of the most fascinating—and at the same time highly elusive—phenomena in music: expressive music performance.[1] We study how skilled musicians (concert pianists, in particular) make music come alive, how they express and communicate their understanding of the musical and emotional content of the pieces by shaping various parameters such as tempo, timing, dynamics, and articulation.

Our starting point is recordings of musical pieces by actual pianists. These recordings are analyzed with intelligent data analysis methods from the areas of machine learning, data mining, and pattern recognition with the aim of building interpretable quantitative models of certain aspects of performance. In this way, we hope to obtain new insight into how expressive performance works and what musicians do to make music sound like music to us. The motivation is twofold: (1) by discovering and formalizing significant patterns and regularities in the artists' musical behavior, we hope to make new contributions to the field of musicology, and (2) by developing new data visualization and analysis methods, we hope to extend the frontiers of the field of AI-based scientific discovery.

In this article, we take the reader on a grand tour of this complex discovery enterprise, from the intricacies of data gathering—which already require new AI methods—through novel approaches to data visualization all the way to automated data analysis and inductive learning. We show that even a seemingly intangible phenomenon such as musical expression can

be transformed into something that can be studied formally and that the computer can indeed discover some fundamental (and sometimes surprising) principles underlying the art of music performance. It turns out that AI plays an important role in each step of this complex, multistage discovery project.

The title of the article refers to the late Vladimir Horowitz (1903–1989), Russian pianist, legendary virtuoso, and one of the most famous and popular pianists of the twentieth century, who symbolizes, like few others, the fascination that great performers hold for the general audience.

Formally explaining the secret behind the art and magic of such a great master would be an extremely exciting feat. Needless to say, it is not very likely, and no "Horowitz Factor" will be revealed in this article. Still, we do hope the following description of the project and its results will capture the reader's imagination.

Expressive Music Performance

Expressive music performance is the art of shaping a musical piece by continuously varying important parameters such as tempo and dynamics. Human musicians do not play a piece of music mechanically, with constant tempo or loudness, exactly as written in the printed music score. Rather, they speed up at some places, slow down at others, stress certain notes or passages by various means, and so on. The most important parameter dimensions available to a performer (a pianist, in particular) are *timing* and continuous tempo changes, *dynamics* (loudness variations), and *articulatio*n (the way successive notes are connected). The precise parameter changes are not specified in the written score, but at the same time, they are absolutely essential for the music to be effective and engaging. The expressive nuances added by an artist are what makes a piece of music come alive (and what makes some performers famous).

Expressive variation is more than just a deviation from, or a distortion of, the original (notated) piece of music. In fact, the opposite is the case: The notated music score is but a small part of the actual music. Not every intended nuance can be captured in a limited formalism such as common music notation, and the composers were and are well aware of this. The performing artist is an indispensable part of the system, and expressive music performance plays a central role in our musical culture. That is what makes it a central object of study in the field of musicology (see Gabrielsson [1999] for an excellent overview of pertinent research in the field).

Our approach to studying this phenomenon

is data driven: We collect recordings of performances of pieces by skilled musicians;[2] measure aspects of expressive variation (for example, the detailed tempo and loudness changes applied by the musicians); and search for patterns in these tempo, dynamics, and articulation data. The goal is to find interpretable models that characterize and explain consistent regularities and patterns, if such should indeed exist. This requires methods and algorithms from machine learning, data mining, and pattern recognition as well as novel methods of intelligent music processing.

Our research is meant to complement recent work in contemporary musicology that has largely been hypothesis driven (for example, Friberg [1995]; Sundberg [1993]; Todd [1992, 1989]; Windsor and Clarke [1997]), although some researchers have also taken real data as the starting point of their investigations (for example, Palmer [1988]; Repp [1999, 1998, 1992]). In the latter kind of research, statistical methods were generally used to verify hypotheses in the data. We give the computer a more autonomous role in the discovery process by using machine learning and related techniques.

Using machine learning in the context of expressive music performance is not new. For example, there have been experiments with case-based learning for generating expressive phrasing in jazz ballads (Arcos and López de Mántaras, 2001; Lopez de Mántaras and Arcos 2002). The goal of that work was somewhat different from ours; the target was to produce phrases of good musical quality, so the system makes use of musical background knowledge wherever possible. In our context, musical background knowledge should be introduced with care because it can introduce biases in the data analysis process. In our own previous research (Widmer 1998, 1995), we (re-)discovered a number of basic piano performance rules with inductive learning algorithms. However, these attempts were extremely limited in terms of empirical data and, thus, made it practically impossible to establish the significance of the findings in a statistically well-founded way. Our current investigations, which are described here, are the most data-intensive empirical studies ever performed in the area of musical performance research (computer-based or otherwise) and, as such, probably add a new kind of quality to research in this area.

Two Basic Questions: Commonalities and Differences

The starting points for the following presentation are two generic types of questions regard-

Inductive Learning of Classification Rules and the PLCG Algorithm
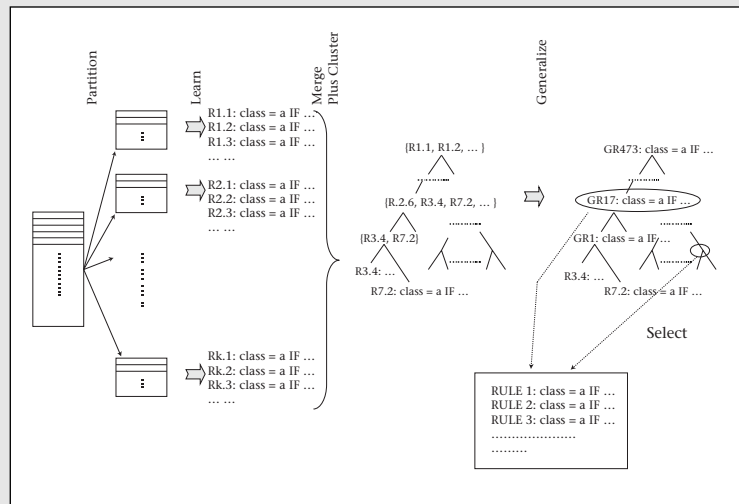

Figure A. The PLCG Learning Algorithm: Main Stages.

The induction of classification rules is one of the major classes of learning scenarios investigated in machine learning. Given a set of examples, each described by a well-defined set of descriptors and labeled as belonging to one of n disjoint classes $c_1 \ldots c_n$, the task is to induce a general model that is (more or less) consistent with the training examples and can predict the class of new, previously unseen examples. One class of models are classification rules of the form class = c_i IF <condition1> AND <condition2> AND.... If there are only two classes a and b, one of which (a, say) is the class we want to find a definition for, one usually speaks of concept learning and refers to instances of a as *positive examples* and instances of b as *negative examples*. Sets of classification rules are commonly referred to as *theories*.

The most common strategy for learning classification rules in machine learning is known as *sequential covering,* or separate-and-conquer (Fürnkranz 1999). The strategy involves inducing rules one by one and, after having learned a rule, removing all the examples covered by the rule so that the following learning steps will focus on the still-uncovered examples. In the simplest case, this process is repeated until no positive examples are left that are not covered by any rule. A single rule is learned by usually starting with the most general rule (a rule with no conditions) that would cover all given examples, positive and negative, and then refining the rule step by step by adding one condition at a time so that many negative examples are excluded, and many positive ones remain covered. The process of selecting conditions is usually guided by heuristics such as weighted information gain (Quinlan 1990) that assess the discriminatory potential of competing conditions. In the simplest case, rule refinement stops when the current rule is *pure*, that is, covers no more negative examples. However, in real life, data can be *noisy* (that is, contain errors), and the given data and rule representation languages might not even permit the formu-

lation of perfectly consistent theories; so, good rule learning algorithms perform some kind of pruning to avoid overfitting. They usually learn rules that are not entirely pure, and they stop before all positive examples have been covered. The short presentation given here is necessarily simplified in various ways, and the reader who wishes to learn the whole story is referred to Fürnkranz (1999).

At the heart of our new learning algorithm PLCG is such a sequential covering algorithm. However, "wrapped around" this simple rule learning algorithm is a metaalgorithm that essentially uses the underlying rule learner to induce several partly redundant theories and then combines these theories into one final rule set. In this sense, PLCG is an example of what is known in machine learning as ensemble methods (Dietterich 2000). The PLCG algorithm proceeds in several stages (figure A); it is here we can explain the acronym PLCG. PLCG stands for partition + learn + cluster + generalize. In a first step, the training examples are partitioned into several subsets (partition). From each of these subsets, a set of classification rules is induced (learn). These rule sets are then merged into one large set, and a hierarchical clustering of the rules into a tree of rule sets is performed (cluster), where each set contains rules that are somehow similar. Each of these rule sets is then replaced with the least general generalization of all the rules in the set (generalize). The result is a tree of rules of varying degrees of generality. Finally, a heuristic algorithm selects the most promising rules from this generalization tree and joins them into the final rule set that is then returned.

The motivation for this rather complex procedure is that in this way the advantage of *ensemble learning*—improved prediction accuracy by selectively combining the expertise of several classifiers—is combined with the benefit of inducing comprehensible theories. In contrast to most common ensemble learning methods such as bagging (Breiman 1996), stacking (Wolpert 1992), or boosting (Freund and Shapire 1996), which only combine the predictions of classifiers to improve prediction accuracy, PLCG combines the theories directly and produces one final rule set that can be interpreted. This is important in the context of our project, where the focus is on the discovery of interpretable patterns.

Systematic large-scale experiments have shown that PLCG consistently learns more precise theories than state-of-the-art rule-learning algorithms such as RIPPER (Cohen 1995), but these theories also tend to be much simpler and also more specific; that is, they cover, or explain, fewer of the positive examples. To put it in simple terms, PLCG only learns rules for those parts of the target concept where it is quite sure it can competently predict; this feature is quite desirable in our discovery context. More detail on algorithm and experimental results can be found in Widmer (2003).

ing expressive music performance. First, are there general, fundamental principles of music performance that can be discovered and characterized? Are there general (possibly unconscious and definitely unwritten) rules that all or most performers adhere to? In other words, to what extent can a performer's expressive actions be predicted? Second, is it possible to formally characterize and quantify aspects of individual artistic style? Can we formally describe what makes the special art of a Vladimir Horowitz, for example?

The first set of questions thus relates to similarities or commonalities between different performances and different performers, and the second focuses on the differences. The following project presentation is structured according to these two types of questions. The section entitled "Studying Commonalities" focuses on the commonalities and briefly recapitulates some of our recent work on learning general performance rules from data. The major part of this article is presented in Studying Differences, which describes currently ongoing (and very preliminary) work on the discovery of stylistic characteristics of great artists.

Both of these lines of research are complex enterprises and comprise a number of important steps—from the acquisition and measuring of pertinent data to computer-based discovery proper. As we see, AI plays an important role in all these steps.

Studying Commonalities: Searching for Fundamental Principles of Music Performance

The question we turn to first is the search for commonalities between different performances and performers. Are there consistent patterns that occur in many performances and point to fundamental underlying principles? We are looking for general rules of music performance, and the methods used will come from the area of inductive machine learning. This section is kept rather short and only points to the most important results because most of this work has already been published elsewhere (Widmer 2003, 2002b, 2001; Widmer and Tobudic 2003).

Data Acquisition: Measuring Expressivity in Performances

The first problem is data acquisition. What we require are precise measurements of the tempo, timing, dynamics, and articulation in a performance of a piece by a musician. In principle, we need to measure exactly when and how long

and how loud each individual note was played and how these measurements deviated from the nominal values prescribed in the written musical score. Extracting this information with high precision from sound recordings is not possible for basic signal processing reasons. Instead, our main source of information are special pianos that precisely record each action by a performer. In particular, the Bösendorfer SE290 is a full-concert grand piano with a special mechanism that measures every key and pedal movement with high precision and stores this information in a format similar to MIDI. (The piano also features a mechanical reproduction facility that can reproduce a recorded performance with very high accuracy.) From these measurements, and by comparing them to the notes as specified in the written score, every expressive nuance applied by a pianist can be computed.

These nuances can be represented as expression curves. For example, figure 1 shows dynamics curves—the dynamics patterns produced by three different pianists in performing the same piece. More precisely, each point represents the relative loudness with which a particular melody note was played (relative to an averaged standard loudness); a purely mechanical, unexpressive rendition of the piece would correspond to a perfectly flat horizontal line at $y = 1.0$. Variations in tempo and articulation can be represented in an analogous way.

Figure 1 exhibits some clear common patterns and tendencies in the three performances. Despite individual differences between the recordings, there seem to be common strategies, or "rules," that are followed by the pianists, consciously or unconsciously. Obviously, there is hope for automated discovery algorithms to find some general principles.

Induction of Note-Level Performance Rules

Some such general principles have indeed been discovered with the help of a new inductive rule-learning algorithm named PLCG (Widmer 2003) (see page 113). PLCG was applied to the task of learning note-level performance rules; 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. Such rules should be contrasted with higher-level expressive strategies such as the shaping of an entire musical phrase (for example, with a gradual slowing toward the end), which is addressed later. The training data used for the experiments consisted of recordings of 13 complete piano sonatas by W. A.
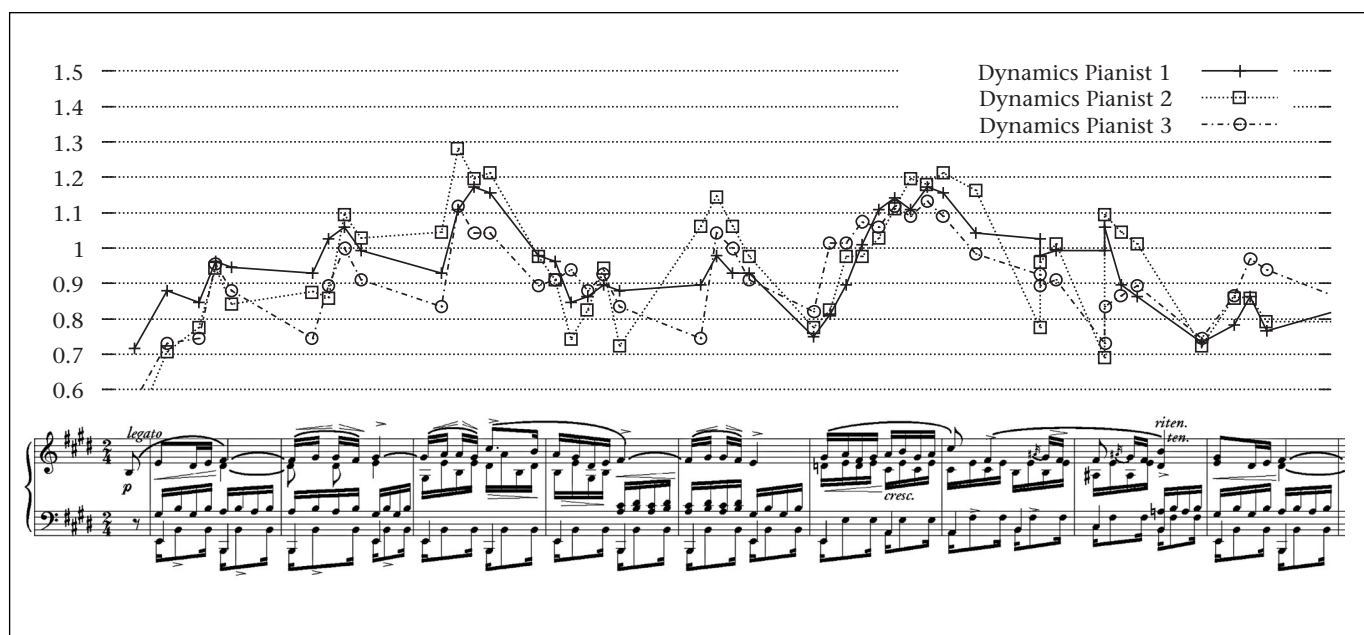
Figure 1. Dynamics Curves (relating to melody notes) of Performances of the Same Piece (Frédéric Chopin, Etude op.10 no.3, E major) by Three Different Viennese Pianists (computed from recordings on a Bösendorfer 290SE computer-monitored grand piano).

Mozart (K.279-284, 330-333, 457, 475, and 533), performed by the Viennese concert pianist Roland Batik. The resulting data set comprises more than 106,000 performed notes and represents some 4 hours of music. The experiments were performed on the melodies (usually the soprano parts) only, which gives an effective training set of 41,116 notes. Each note was described by 29 attributes (10 numeric, 19 discrete) that represent both intrinsic properties (such as scale degree, duration, metrical position) and some aspects of the local context (for example, melodic properties such as the size and direction of the intervals between the note and its predecessor and successor notes, rhythmic properties such as the durations of surrounding notes and so on, and some abstractions thereof).

From these 41,116 examples of played notes, PLCG learned a small set of 17 quite simple classification rules that predict a surprisingly large number of the note-level choices of the pianist. The rules have been published in the musicological literature (Widmer 2002b) and have created some interest. The surprising aspect is the high number of note-level actions that can be predicted by very few (and mostly very simple) rules. For example, 4 rules were discovered that together correctly predict almost 23 percent of all the situations where the pianist lengthened a note relative to how it was notated (which corresponds to a local slowing of the tempo).[3] To give the reader an impression of the simplic-

ity and generality of the discovered rules, here is an extreme example:

> 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 (that is, 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: Rule TL2 correctly predicts 1,894 cases of local note lengthening, which is 14.12 percent of all the instances of significant lengthening observed in the training data. The number of incorrect predictions is 588 (2.86 percent of all the counterexamples), which gives a precision (percentage of correct predictions) of .763. It is remarkable that one simple principle like this is sufficient to predict such a large proportion of observed note lengthenings in complex music such as Mozart sonatas.

To give the reader an impression of just how effective a few simple rules can be in predicting a pianist's behavior in certain cases, figure 2 compares the tempo variations predicted by our rules to the pianist's actual timing in a performance of the well-known Mozart Sonata K.331 in A major (first movement, first sec-
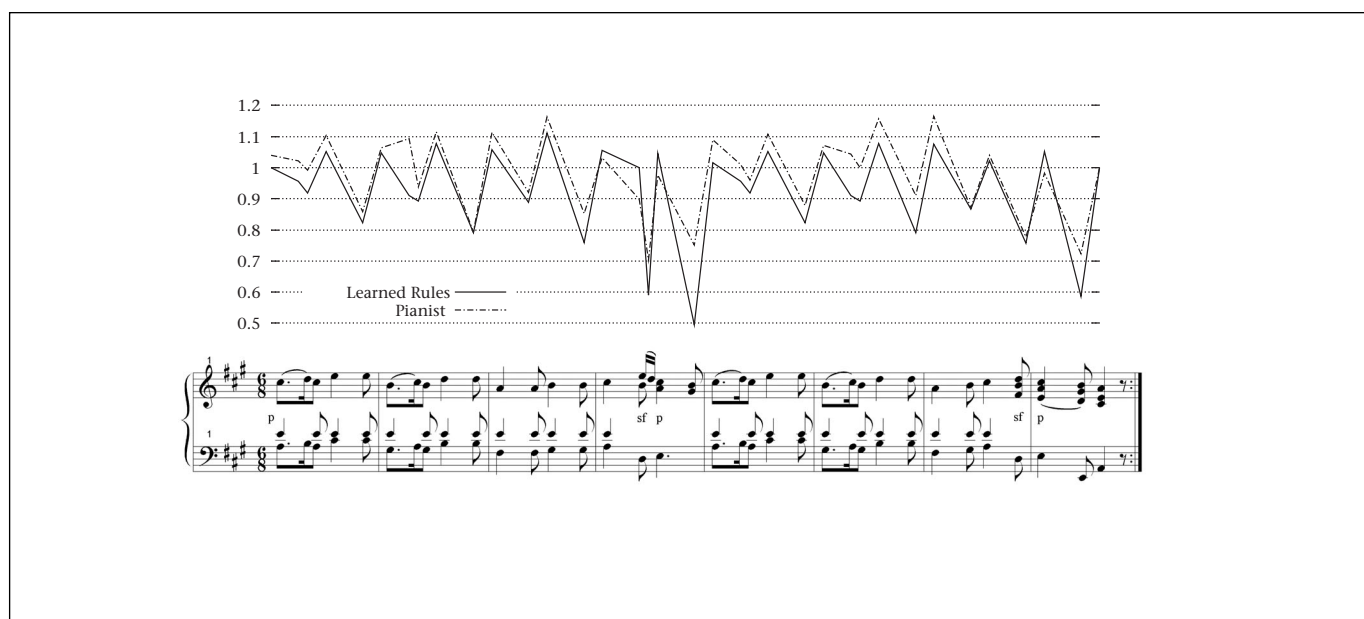
Figure 2. Mozart Sonata K.331, First Movement, First Part, as Played by Pianist and Learner.

The curve plots the relative tempo at each note; notes above the 1.0 line are shortened relative to the tempo of the piece, and notes below 1.0 are lengthened. A perfectly regular performance with no timing deviations would correspond to a straight line at $y = 1.0$.

tion). In fact, it is just two simple rules (one for note lengthening, one for shortening) that produce the system's timing curve.[4]

Experiments also showed that most of these rules are highly general and robust: They carry over to other performers and even music of different styles with virtually no loss of coverage and precision. In fact, when the rules were tested on performances of quite different music (Chopin), they exhibited significantly higher coverage and prediction accuracy than on the original (Mozart) data they had been learned from. What the machine has discovered here really seem to be fundamental performance principles.

A detailed discussion of the rules, as well as a quantitative evaluation of their coverage and precision, can be found in Widmer (2002b); the learning algorithm PLCG is described and analyzed in Widmer (2003).

Multilevel Learning of Performance Strategies

As already mentioned, not all a performer's decisions regarding tempo or dynamics can be predicted on a local, note-to-note basis. Musicians understand the music in terms of a multitude of more abstract patterns and structures (for example, motifs, groups, phrases), and they use tempo and dynamics to "shape" these structures, for example, by applying a gradual crescendo (growing louder) or decrescendo (growing softer) to entire passages. Music per-

formance is a multilevel phenomenon, with musical structures and performance patterns at various levels embedded in each other.

Accordingly, the set of note-level performance rules described earlier is currently being augmented with a multilevel learning strategy where the computer learns to predict elementary tempo and dynamics "shapes" (like a gradual crescendo-decrescendo) at different levels of the hierarchical musical phrase structure and combines these predictions with local timing and dynamics predicted by learned note-level models. Preliminary experiments, again with performances of Mozart sonatas, yielded very promising results (Widmer and Tobudic 2003). Just to give an idea, figure 3 shows the predictions of the integrated learning algorithm on part of a test piece after learning from other Mozart sonatas. As can be seen in the lower part of the figure, the system manages to predict not only local patterns but also higher-level trends (for example, gradual increases of overall loudness) quite well.

The curve shown in figure 3 is from a computer-generated performance of the Mozart piano sonata K.280 in F major. A recording of this performance was submitted to an International Computer Piano Performance Rendering Contest (RENCON'02) in Tokyo in September 2002,[5] where it won second prize behind a rule-based rendering system that had carefully been tuned by hand. The rating was done by a jury of human listeners. Although this result in
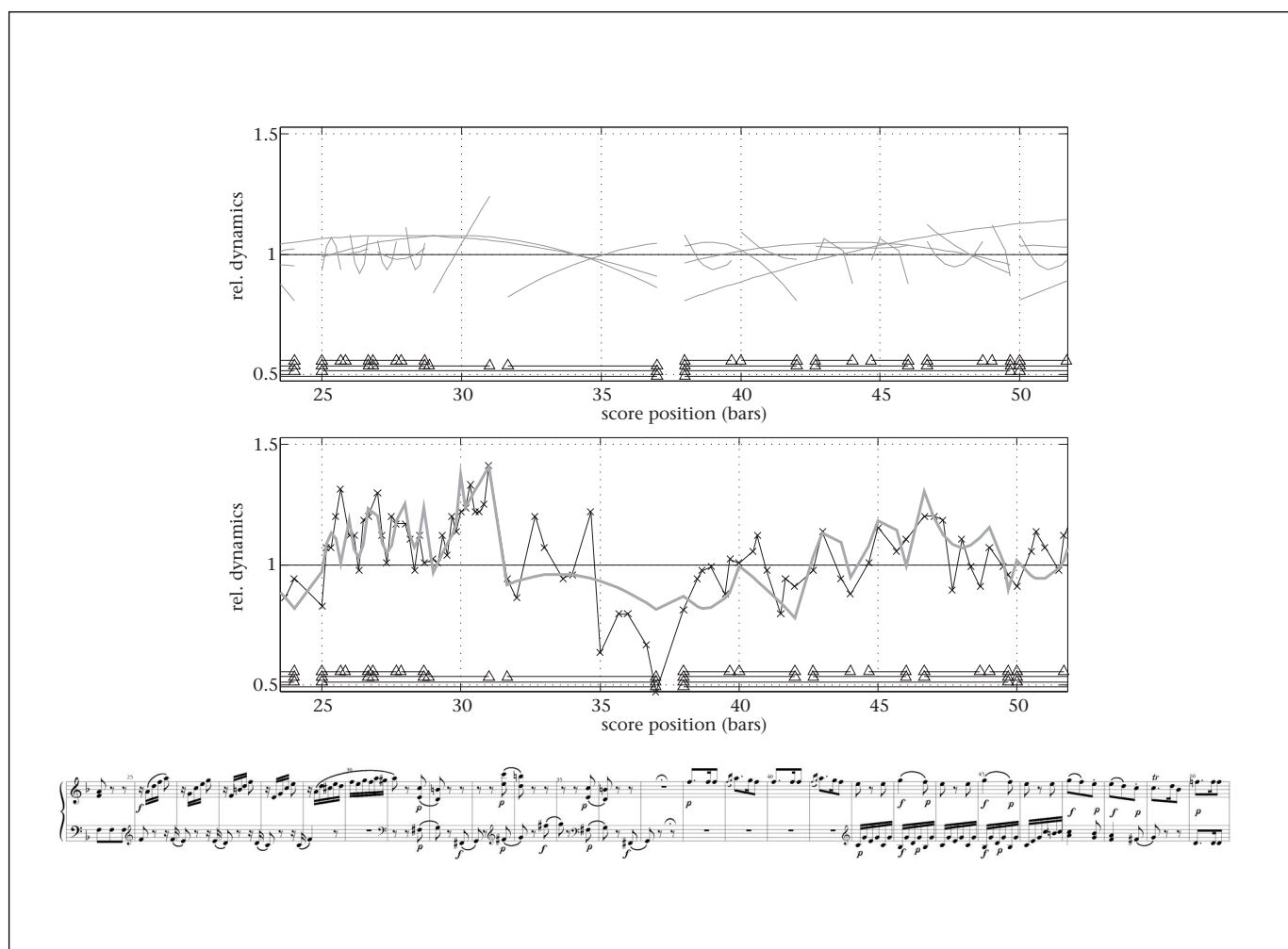
Figure 3. Learner's Predictions for the Dynamics Curve of Mozart Sonata K.280, Third Movement, mm. 25-50.

Top: dynamics shapes predicted for phrases at four levels. Bottom: composite predicted dynamics curve resulting from phrase-level shapes and note-level predictions (gray) versus pianist's actual dynamics (black). Line segments at the bottom of each plot indicate hierarchical phrase structure.

no way implies that a machine will ever be able to learn to play music like a human artist, we do consider it a nice success for a machine learning system.

Studying Differences: Trying to Characterize Individual Artistic Style

The second set of questions guiding our research concerns the differences between individual artists. Can one characterize formally what is special about the style of a particular pianist? Contrary to the research on common principles described earlier, where we mainly used performances by local (though highly skilled) pianists, here we are explicitly interest-

ed in studying famous artists. Can we find the "Horowitz Factor"?

This question might be the more intriguing one for the general audience because it involves famous artists. However, the reader must be warned that the question is difficult. The following is work in progress, and the examples given should be taken as indications of the kinds of things that we hope to discover rather than as truly significant, established discovery results.

Data Acquisition: Measuring Expressivity in Audio Recordings

The first major difficulty is again data acquisition. With famous pianists, the only source of data are audio recordings, that is, records and music CDs (we cannot very well invite them all

to Vienna to perform on the Bösendorfer SE290 piano). Unfortunately, it is impossible, with current signal-processing methods, to extract precise performance information (start and end times, loudness, and so on) about each individual note directly from audio data. Thus, it will not be possible to perform studies at the same level of detail as those based on MIDI data. In particular, we cannot study how individual notes are played.

What is currently possible is to extract tempo and dynamics at the level of the beat.[6] That is, we extract these time points from the audio recordings that correspond to beat locations. From the (varying) time intervals between these points, the beat-level tempo and its changes can be computed. Beat-level dynamics is also computed from the audio signal as the overall loudness (amplitude) of the signal at the beat times.

The hard problem here is automatically detecting and tracking the beat in audio recordings. Indeed, this is an open research problem that forced us to develop a novel beat-tracking algorithm called BEATROOT (Dixon 2001c; Dixon and Cambouropoulos 2000) (see page 127). Beat tracking, in a sense, is what human listeners do when they listen to a piece and tap their foot in time with the music. As with many other perception and cognition tasks, what seems easy and natural for a human turns out to be extremely difficult for a machine.

The main problems to be solved are (1) detecting the onset times of musical events (notes, chords, and so on) in the audio signal, (2) deciding which of these events carry the beat (that includes determining the basic tempo, that is, the basic rate at which beats are expected to occur), and (3) tracking the beat through tempo changes. Tracking the beat is extremely difficult in classical music, where the performer can change the tempo drastically—a slowing down by 50 percent within 1 second is nothing unusual. It is difficult for a machine to decide whether an extreme change in interbeat intervals is because of the performer's expressive timing or whether it indicates that the algorithm's beat hypothesis was wrong.

After dealing with the onset-detection problem with rather straightforward signal-processing methods, BEATROOT models the perception of beat by two interacting processes: The first finds the rate of the beats (*tempo induction*), and the second synchronizes a pulse sequence with the music (*beat tracking*). At any time, multiple hypotheses can exist regarding each of these processes; these are modeled by a multiple-agent architecture in which agents representing each hypothesis compete and cooperate to find the best solution (Dixon 2001c).

Experimental evaluations showed that the BEATROOT algorithm is probably one of the best beat tracking methods currently available (Dixon 2001a). In systematic experiments with expressive performances of 13 complete piano sonatas by W. A. Mozart played by a Viennese concert pianist, the algorithm achieved a correct detection rate of more than 90 percent. However, for our investigations, we needed a tracking accuracy of 100 percent, so we opted for a semiautomatic, interactive procedure. The beat-tracking algorithm was integrated into an interactive computer program that takes a piece of music (a sound file); tries to track the beat;[7] displays its beat hypotheses visually on the screen (figure 4); allows the user to listen to selected parts of the tracked piece and modify the beat hypothesis by adding, deleting, or moving beat indicators; and then attempts to retrack the piece based on the updated information. This process is still laborious, but it is much more efficient than manual beat tracking.

After a recording has been processed in this way, tempo and dynamics at the beat level can easily be computed. The resulting series of tempo and dynamics values is the input data to the next processing step.

Data Visualization: The Performance Worm

An important first step in the analysis of complex data is data visualization. Here we draw on an original idea and method developed by the musicologist Jörg Langner, who proposed to represent the joint development of tempo and dynamics over time as a trajectory in a two-dimensional tempo-loudness space (Langner and Goebl 2002). To provide for a visually appealing display, smoothing is applied to the originally measured series of data points by sliding a Gaussian window across the series of measurements. Various degrees of smoothing can highlight regularities or performance strategies at different structural levels (Langner and Goebl 2003). Of course, smoothing can also introduce artifacts that have to be taken into account when interpreting the results.

We have developed a visualization program called the PERFORMANCE WORM (Dixon, Goebl, and Widmer 2002) that displays animated tempo-loudness trajectories in synchrony with the music. A movement to the right signifies an increase in tempo, a crescendo causes the trajectory to move upward, and so on. The trajectories are computed from the beat-level tempo and dynamics measurements we make with the help of BEATROOT; they can be stored to file
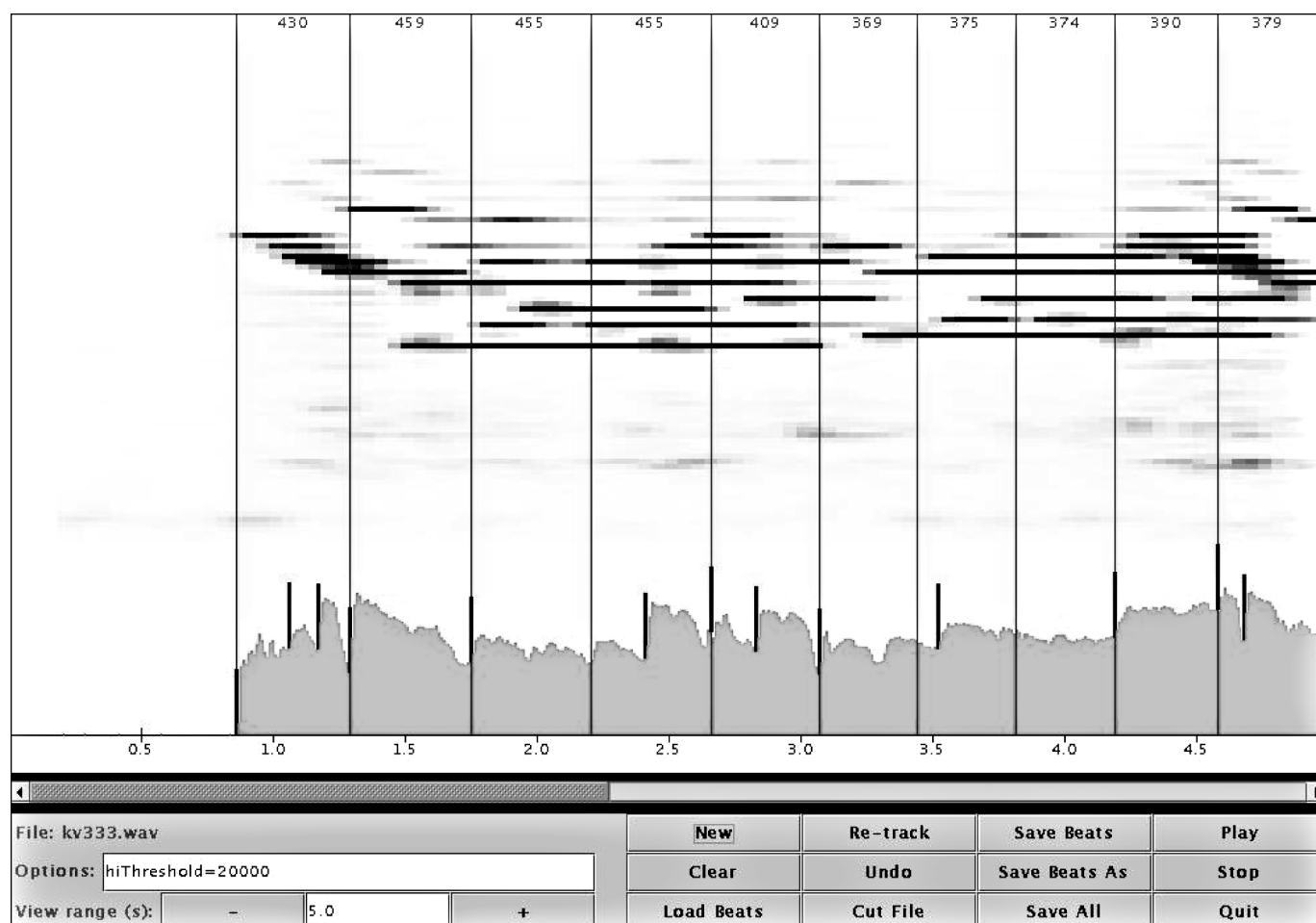
Figure 4. Screen Shot of the Interactive Beat-Tracking System BEATROOT Processing the First Five Seconds of a Mozart Piano Sonata.

Shown are the audio wave form (bottom), the spectrogram derived from it (black "clouds"), the detected note onsets (short vertical lines), the system's current beat hypothesis (long vertical lines), and the interbeat intervals in milliseconds (top).

and used for systematic quantitative analysis (see discussion later).

To give a simple example, Figure 5 shows snapshots of the WORM as it displays performances of the same piece of music (the first eight bars of Robert Schumann's Von fremden Ländern und Menschen, from Kinderszenen, op.15) by three different pianists.[8] Considerable smoothing was applied here to highlight the higher-level developments within this extended phrase.

It is immediately obvious from figure 5 that Horowitz and Kempff have chosen a similar interpretation. Both essentially divide the phrase into two four-bar parts, where the first part is played more or less with an accelerando (the worm moves to the right) and the second part with a ritardando, interrupted by a local speeding up in bar 6 (more pronounced in the Kempff performance). Their dynamics strategies are highly similar too: a general crescendo (interweaved, in Horowitz's case, with two local

reductions in loudness) to the loudness climax around the end of bar five, followed by a decrescendo toward the end of the phrase. Martha Argerich's trajectory, however, betrays a different strategy. In addition to a narrower dynamics range and a slower global tempo, what sets her apart from the others is that, relatively speaking, she starts fast and then plays the entire phrase with one extended ritardando, interrupted only by a short speeding up between bars six and seven.[9] Also, there is no really noticeable loudness climax in her interpretation.

The point here is not to discuss the musical or artistic quality of these three performances—to do this, one would have to see (and hear) the phrase in the context of the entire piece—but simply to indicate the kinds of things one can see in such a visualization. Many more details can be seen when less smoothing is applied to the measured data.

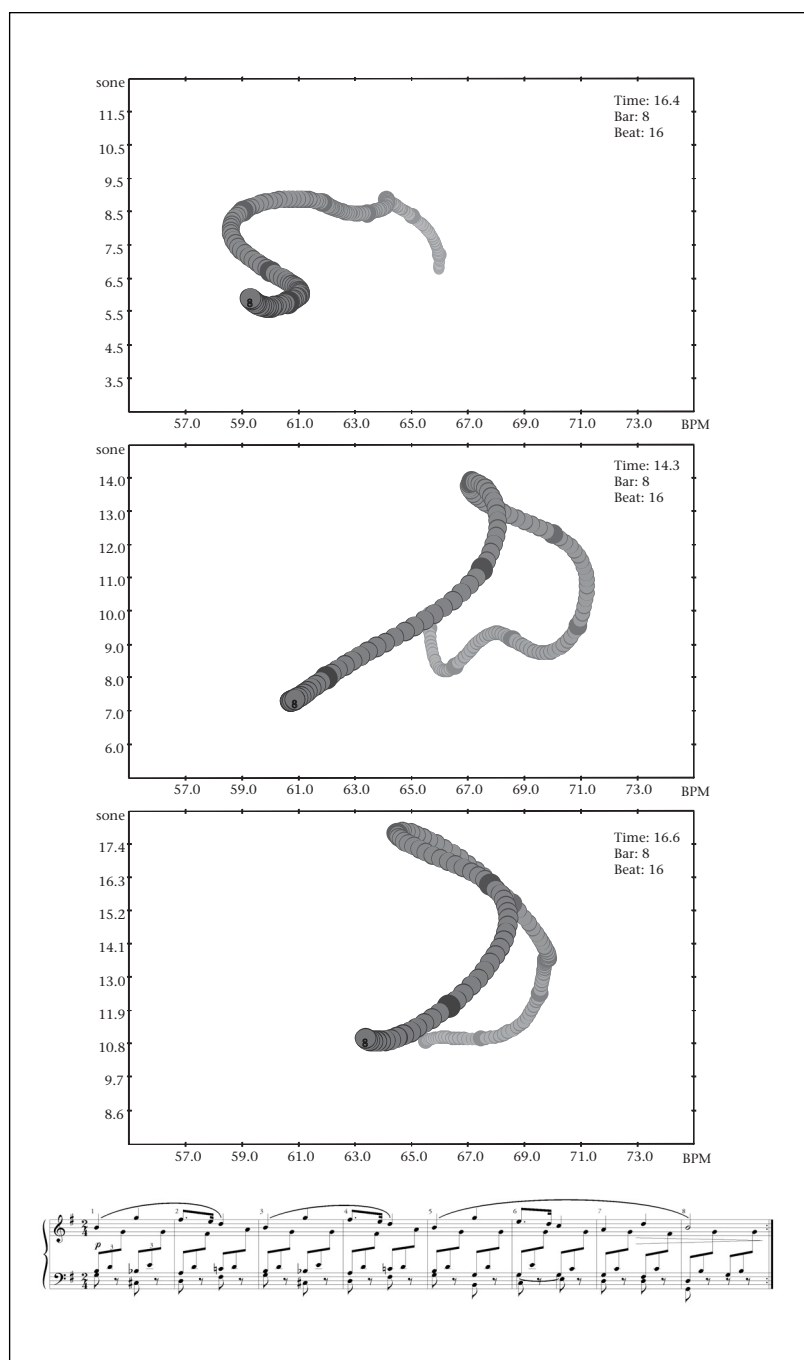We are currently investing large efforts into measuring tempo and dynamics in recordings

Figure 5. Performance Trajectories over the First Eight Bars of Von fremden Ländern und Menschen (from Kinderszenen, op.15, by Robert Schumann), as Played by Martha Argerich (top), Vladimir Horowitz (center), and Wilhelm Kempff (bottom).

Horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in sone (Zwicker and Fastl 2001). The largest point represents the current instant; instants further in the past appear smaller and fainter. Black circles mark the beginnings of bars. Movement to the upper right indicates a speeding up (accelerando) and loudness increase (crescendo) and so on. Note that the *y* axes are scaled differently. The musical score of this excerpt is shown at the bottom.

of different pieces by different famous pianists, using the interactive BEATROOT system. Our current collection (as of January 2003) amounts to more than 500 recordings by pianists such as Martha Argerich, Vladimir Horowitz, Artur Rubinstein, Maurizio Pollini, Sviatoslav Richter, and Glenn Gould, playing music by composers such as Mozart, Beethoven, Schubert, Chopin, Schumann, or Rachmaninov. Precisely measuring these took some 12 person-months of hard work.

Figure 6 shows a complete tempo-loudness trajectory representing a performance of a Chopin Ballade by Artur Rubinstein. The subsequent analysis steps will be based on this kind of representation.

Transforming the Problem: Segmentation, Clustering, Visualization

The smoothed tempo-loudness trajectories provide intuitive insights, which might make visualization programs such as the WORM a useful tool for music education. However, the goal of our research is to go beyond informal observations and derive objective, quantitative conclusions from the data.

Instead of analyzing the raw tempo-loudness trajectories—which, mathematically speaking, are bivariate time series—directly, we chose to pursue an alternative route, namely, to transform the data representation and, thus, the entire discovery problem into a form that is accessible to common inductive machine learning and data-mining algorithms. To this end, the performance trajectories are cut into short segments of fixed length, for example, two beats. The segments are optionally subjected to various normalization operations (for example, mean and/or variance normalization to abstract away from absolute tempo and loudness and/or absolute pattern size, respectively). The resulting segments are then grouped into classes of similar patterns using clustering. For each of the resulting clusters, a prototype is computed. These prototypes represent a set of typical elementary tempo-loudness patterns that can be used to approximately reconstruct a "full" trajectory (that is, a complete performance). In this sense, they can be seen as a simple alphabet of performance, restricted to tempo and dynamics. Figure 7 displays a set of prototypical patterns computed from a set of Mozart sonata recordings by different artists.

The particular clustering shown in figure 7 was generated by a self-organizing map (SOM) algorithm (Kohonen 2001). A SOM produces a geometric layout of the clusters on a two-dimensional grid or map, attempting to place
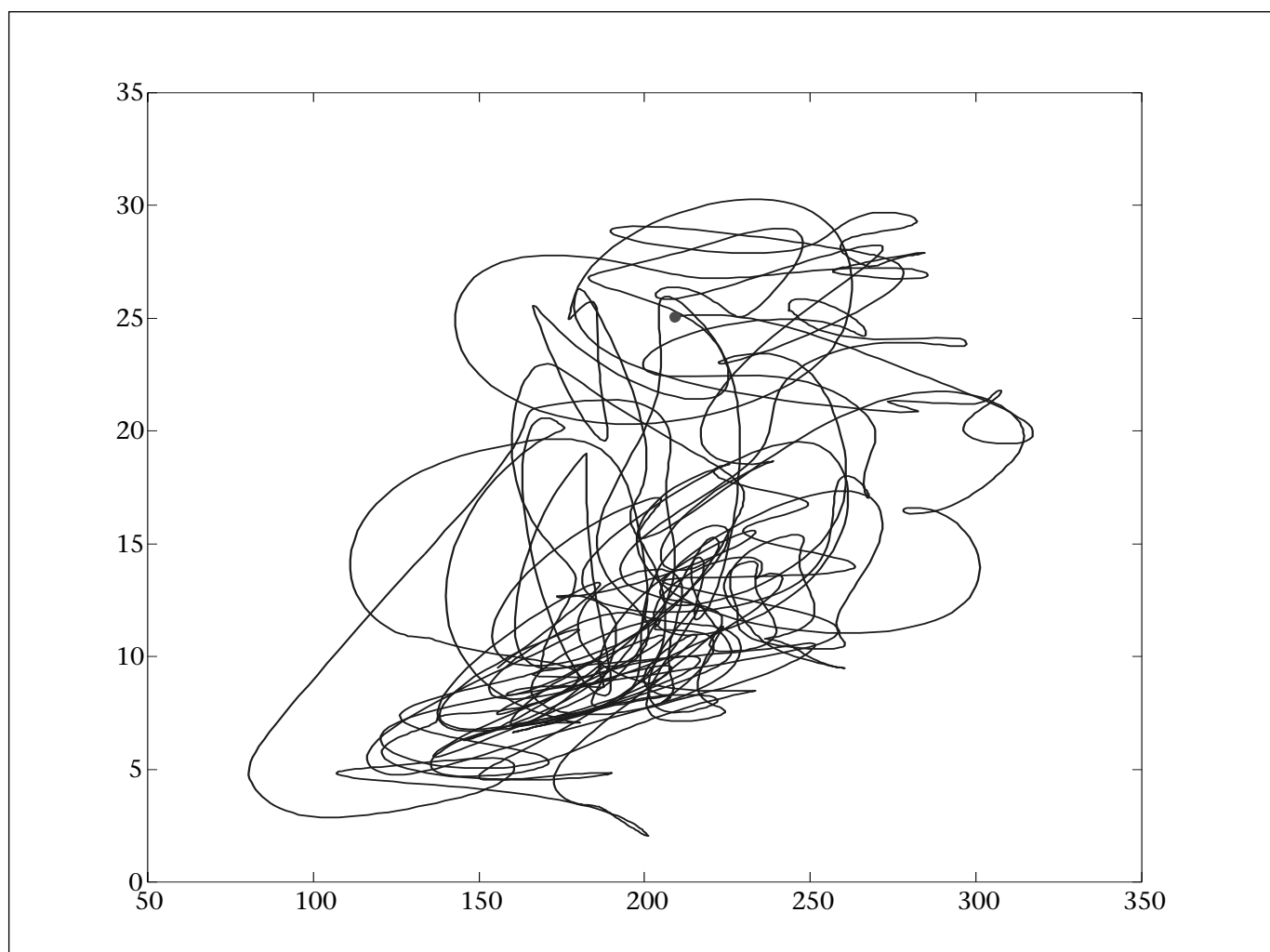
Figure 6. A Complete WORM: *Smoothed Tempo-Loudness Trajectory Representing a Performance of Frédéric Chopin's Ballade op.47 in A Flat Major by Artur Rubinstein.*

Horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in sone (Zwicker and Fastl 2001).

similar clusters close to each other. This property, which is quite evident in figure 7, facilitates a simple, intuitive visualization method. The basic idea, named smoothed data histograms (SDHs), is to visualize the distribution of cluster members in a given data set by estimating the probability density of the high-dimensional data on the map (see Pampalk, Rauber, and Merkl [2002] for details). Figure 8 shows how SDHs can be used to visualize the frequencies with which certain pianists use elementary expressive patterns (trajectory segments) from the various clusters.

Looking at these SDHs with the corresponding cluster map (figure 7) in mind gives us an impression of which types of patterns are preferably used by different pianists. Note that generally, the overall distributions of pattern usage are quite similar (figure 8, left). Obviously, there are strong commonalities, basic principles that all artists adhere to. Most obvious are the common bright areas in the upper right and lower left corners; these correspond to a coupling of tempo and dynamics increases and decreases, respectively (see figure 7). That is, a speeding up often goes along with an increase in loudness, and vice versa. This performance principle is well known and is well in accordance with current performance models in musicology.

The differences between pianists emerge when we filter out the commonalities. The right half of figure 8 shows the same pianists after the commonalities (the joint SDH of all pianists) have been subtracted from each individual SDH. Now we can see clear stylistic differences. Witness, for example, the next-to-rightmost cluster in the bottom row, which represents a slowing down associated with an increase in loudness (a movement of the trajec-
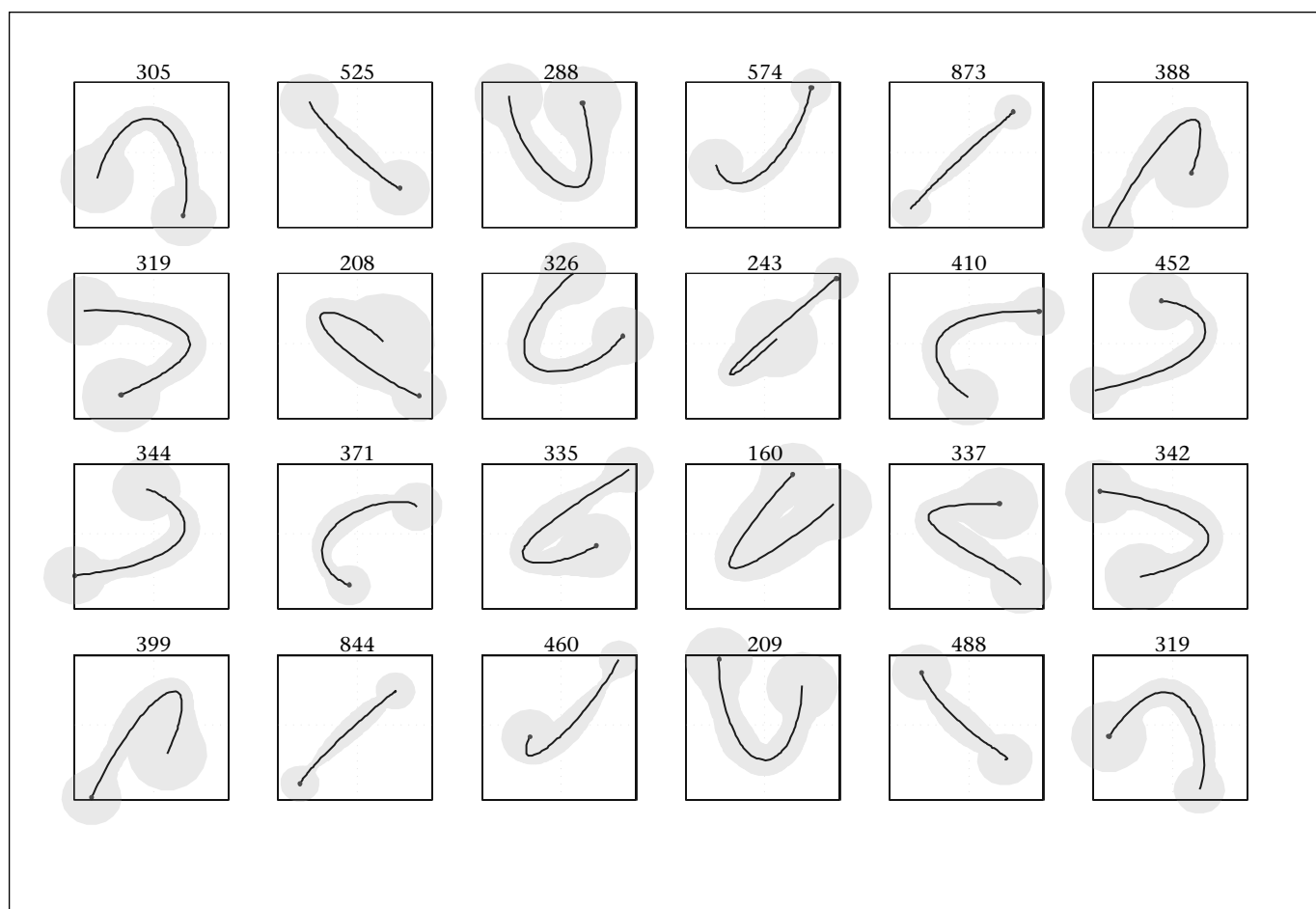
Figure 7. A Mozart Performance Alphabet (cluster prototypes) Computed by Segmentation, Mean, and Variance Normalization and Clustering from performances of Mozart Piano Sonatas by Six Pianists (Daniel Barenboim, Roland Batik, Vladimir Horowitz, Maria João Pires, András Schiff, and Mitsuko Uchida).

To indicate directionality, dots mark the end points of segments. Shaded regions indicate the variance within a cluster.

tory toward the upper left). This pattern is found quite frequently in performances by Horowitz, Schiff, and Batik but much less so in performances by Barenboim, Uchida, and (especially) Pires. An analogous observation can be made in the next-to-leftmost cluster in the top row, which also represents a decoupling of tempo and dynamics (speeding up but growing softer). Again, Pires is the pianist who does this much less frequently than the other pianists. Overall, Maria João Pires and András Schiff appear to be particularly different from each other, and this impression is confirmed when we listen to the Mozart recordings of these two pianists.

Structure Discovery in Musical Strings

The SDH cluster visualization method gives some insight into very global aspects of performance style—the relative frequency with which different artists tend to use certain styl-

istic patterns. It does show that there are systematic differences between the pianists, but we want to get more detailed insight into characteristic patterns and performance strategies. To this end, another (trivial) transformation is applied to the data.

We can take the notion of an alphabet literally and associate each prototypical elementary tempo-dynamics shape (that is, each cluster prototype) with a letter. For example, the prototypes in figure 7 could be named A, B, C, and so on. A *full performance*—a complete trajectory in tempo-dynamics space—can be approximated by a sequence of elementary prototypes and thus be represented as a sequence of letters, that is, a string. Figure 9 shows a part of a performance of a Mozart sonata movement coded in terms of such an alphabet.

This final transformation step, trivial as it might be, makes it evident that our original musical problem has now been transferred into
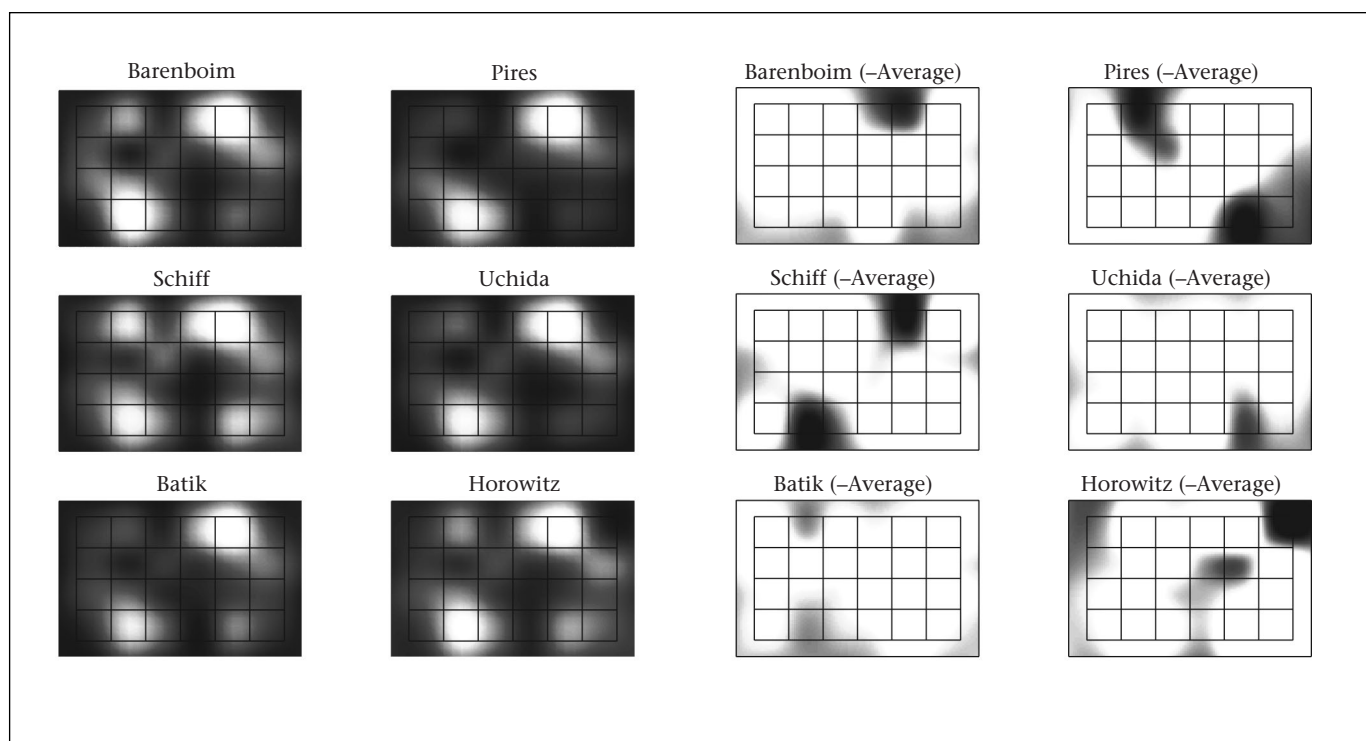
Figure 8. A Smoothed Data Histogram (SDH) Visualization of the Mozart Performances,
Grouped by Pianist (see figure 7 for the corresponding cluster map).

Left: "Plain" SDH showing the relative frequency of pattern use by the individual pianists. Right: To emphasize the differences between the artists, the joint SDH of all pianists was subtracted from each individual SDH. Bright areas indicate high frequency of pattern use.

```
CDSRWHGSNMBDSOMEXOQVWOQQHHSRQVPHJFATGFFUVPLDTPNMECDOVTOMECDSPFXP
OFAVHDTPNFEVHHDXTPMARIFFUHHGIEEARWTTLJEEEARQDNIBDSQIETPPMCDTOMAW
OFVTNMHHDNRRVPHHDUQIFEUTPLXTORQIEBXTORQIECDHFVTOFARBDXPKFURMHDTT
PDTPJARRQWLGFCTPNMEURQIIBDJCGRQIEFFEDTTOMEIFFAVTTPKIFARRTPPPNRRM
IEECHDSRRQEVTTTPORMCGAIEVLGFWLHGARRVLXTOQRWPRRLJFUTPPLSRQIFFAQIF
ARRLHDSOQIEBGAWTOMEFETTPKECTPNIETPOIIFAVLGIECDRQFAVTPHSTPGFEJAWP
ORRQICHDDDTPJFEEDTTPJFAVTOQBHJRQIBDNIFUTPPLDHXOEEAIEFFECXTPRQIFE
CPOVLFAVPTTPPPKIEEFRWPNNIFEEDTTPJFAXTPQIBDNQIECOIEWTPPGCHXOEEUIE
FFICDSOQMIFEEBDTPJURVTTPPNMFEARVTTFFFFRVTLAMFFARQBXSRWPHGFBDTTOU
.............................
```

Figure 9. Beginning of a Performance by Daniel Barenboim (W. A. Mozart piano sonata K.279 in C major)
Coded in Terms of a 24-Letter Performance Alphabet Derived from a Clustering of Performance Trajectory Segments.

a quite different world: the world of string analysis. The fields of pattern recognition, machine learning, data mining, and so on, have developed a rich set of methods that can find structure in strings and that could now profitably be applied to our musical data.[10]

There are a multitude of questions one might want to ask of these musical strings. For example, one might search for letter sequences (substrings) that are characteristic of a particular pianist (see later discussion). One might search for general, frequently occurring substrings that are typical components of performances—stylistic clichés, so to speak. Using such frequent patterns as building blocks, one might try to use machine learning algorithms to induce—at least partial—grammars of musical performance style (for example, Nevill-Manning and Witten 1997). One might also investigate whether a machine can learn to identify performers based on characteristics of their performance trajectories.

We are working along several of these lines. For example, we are currently experimenting with classification algorithms that learn to recognize famous pianists based on aspects of their performance trajectories, both at the level of the raw trajectories (numeric values) and at the level of performance strings. First experimental results are quite encouraging (Stamatatos and Widmer 2002; Zanon and Widmer 2003): there seem to be artist-specific patterns in the performances that a machine can identify.

In the following discussion, we look at a more direct attempt at discovering stylistic patterns typical of different artists. Let's take the previous performance strings and ask the following question: Are there substrings in these strings that occur much more frequently in performances by a particular pianist than in others?

Such a question is a data-mining one. We do not want to bore the reader with a detailed mathematical and algorithmic account of the problem. Essentially, what we are looking for are letter sequences with a certain minimum frequency that occur only in one class of strings (that is, in performances by one pianist). In data-mining terms, these could be called *discriminative frequent sequences*. In reality, patterns that perfectly single out one pianist from the others will be highly unlikely, so instead of requiring uniqueness of a pattern to a particular pianist, we will be searching for patterns that exhibit a certain level of discriminatory power.

Data mining has developed a multitude of methods for discovering frequent subsets or sequences in huge sequences of events or items (for example, Agrawal and Srikant [1994] and Mannila, Toivonen, and Verkamo [1997]). We extended one of these basic methods—the levelwise search algorithm for finding frequent item sets (Agrawal and Srikant 1994)—toward being able to find frequent subsequences that are also discriminative, where discriminatory potential is related to the level of certainty with which one can predict the pianist after having observed a particular pattern. Technically, this method involves computing the entropies of the distribution of the pattern occurrences across the pianists and selecting patterns with low entropy (Widmer 2002a).

In a first experiment with recordings of Mozart piano sonatas—5 sonatas, 54 sections, 6 pianists (Daniel Barenboim, Roland Batik, Glenn Gould, Maria João Pires, András Schiff, and Mitsuko Uchida)—a number of sequences were discovered that are discriminative according to our definition and also look like they might be musically interesting. For example, in one particular alphabet, the sequence FAVT came up as a typical Barenboim pattern, with seven occurrences in Barenboim's Mozart performances, two in Pires, one in Uchida, and none in the other pianists (see also figure 9).

Now what is FAVT? To find out whether a letter sequence codes any musically interesting or interpretable behavior, we can go back to the original data (the tempo-loudness trajectories) and identify the corresponding trajectory segments in the recordings that are coded by the various occurrences of the sequence. As the left part of figure 10 shows, what is coded by the letter sequence FAVT in Daniel Barenboim's performances of Mozart is an increase in loudness (a crescendo), followed by a slight tempo increase (accelerando), followed by a decrease in loudness (decrescendo) with more or less constant tempo. This pattern is, indeed, rather unusual. In our experience to date, it is quite rare to see a pianist speed up during a loudness maximum. Much more common in such situations are slowings down (ritardandi), which gives a characteristic counterclockwise movement of the WORM; for example, the right half of figure 10 shows instantiations of a pattern that seems characteristic of the style of Mitsuko Uchida (8 occurrences versus 0 in all the other pianists). This "Barenboim pattern" might, thus, really be an interesting discovery that deserves more focused study.

What about Vladimir Horowitz, the reader might ask. Where is the Horowitz factor? This is a fair question, given the rather grandiose title of this article. Figure 11 shows a pattern that was discovered as potentially typical of

Horowitz in an experiment with Chopin recordings by various famous pianists, including Horowitz, Rubinstein, and Richter. The tempo-loudness trajectory corresponding to the pattern describes a slowing down with a decrescendo—a movement from the upper right to the lower left—followed by a little speedup—the loop—followed again by a slowing down, now with a crescendo, an increase in loudness. If nothing else, this pattern certainly looks nice. Instantiations of the pattern, distorted in various forms, were found in Horowitz recordings of music by Mozart, Chopin, and even Beethoven (see figure 12).

Is this a characteristic Horowitz performance pattern, a graphic illustration of Horowitz's individuality?

A closer analysis shows that the answer is no.[11] When we actually listen to the corresponding sections from the recordings, we find that most of these patterns are not really perceivable in such detail. In particular, the little interspersed accelerando in the bottom left corner that makes the pattern look so interesting is so small in most cases—it is essentially the result of a temporal displacement of a single melody note—that we do not hear it as a speedup. Moreover, it turns out that some of the occurrences of the pattern are artifacts, caused by the transition from the end of one sonata section to the start of the next, for example.

One must be careful not to be carried away by the apparent elegance of such discoveries. The current data situation is still too limited to draw serious conclusions. The absolute numbers (8 or 10 occurrences of a supposedly typical pattern in recordings by a pianist) are too small to support claims regarding statistical significance. Also, we cannot say with certainty that similar patterns do not occur in the performances by the other pianists just because they do not show up as substrings—they might be coded by a slightly different character sequence! Moreover, many alternative performance alphabets could be computed; we currently have no objective criteria for choosing the optimal one in any sense. Finally, even if we can show that some of these patterns are statistically significant, we will still have to establish their musical relevance, as the Horowitz example clearly shows. The ultimate test is listening, and that is a very time-consuming activity.

Thus, this section appropriately ends with a word of caution. The reader should not take any of the patterns shown here too literally. They are only indicative of the kinds of things we hope to discover with our methods.
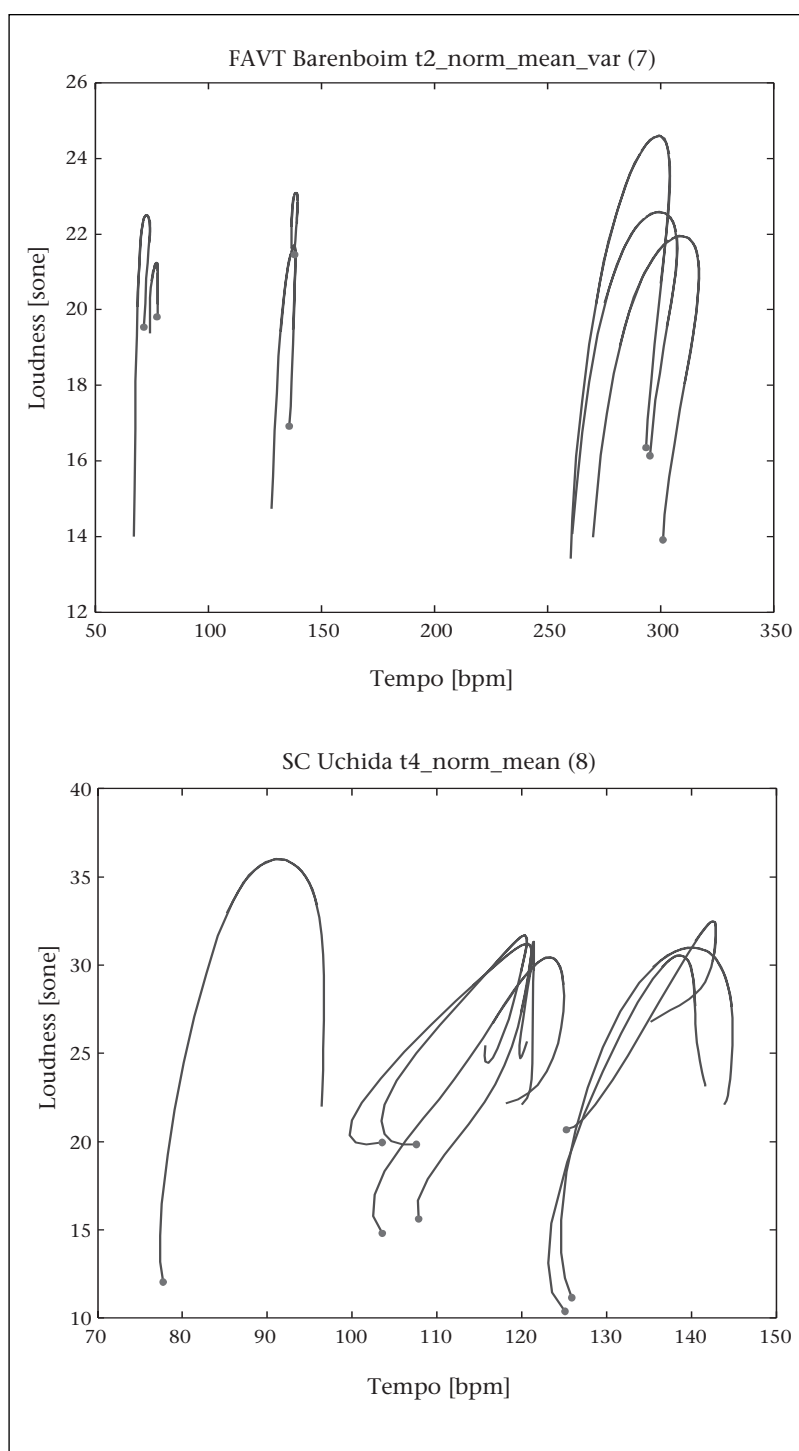

Figure 10. Two Sets of (instantiations of) Performance Patterns: FAVT Sequence Typical of Daniel Barenboim (tpp) and SC pattern (in a different alphabet) from Mitsuko Uchida (bottom).

To indicate directionality, a dot marks the end point of a segment.
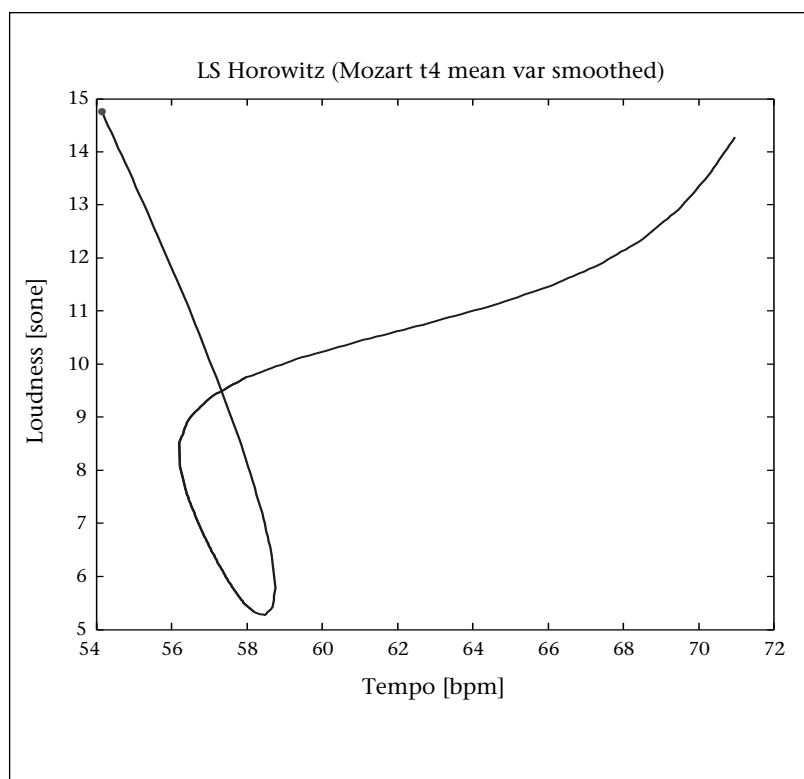
Figure 11. A Typical Horowitz Pattern?

Whether these findings will indeed be musically relevant and artistically interesting can only be hoped for at the moment.

Conclusion

It should have become clear by now that expressive music performance is a complex phenomenon, especially at the level of world-class artists, and the "Horowitz Factor," if there is such a thing, will most likely not be explained in an explicit model, AI-based or otherwise, in the near future. What we have discovered to date are only tiny parts of a big mosaic. Still, we do feel that the project has produced a number of results that are interesting and justify this computer-based discovery approach.

On the musical side, the main results to date were a rule-based model of note-level timing, dynamics, and articulation with surprising generality and predictivity (Widmer 2002b); a model of multilevel phrasing that even won a prize in a computer music performance contest (Widmer and Tobudic 2003); completely new

ways of looking at high-level aspects of performance and visualizing differences in performance style (Dixon, Goebl, and Widmer 2002); and a first set of performance patterns that look like they might be characteristic of particular artists and that might deserve more detailed study.

Along the way, a number of novel methods and tools of potentially general benefit were developed: the beat-tracking algorithm and the interactive tempo-tracking system (Dixon 2001c), the PERFORMANCE WORM (Dixon et al. 2002) (with possible applications in music education and analysis), and the PLCG rule learning algorithm (Widmer 2003), to name a few.

However, the list of limitations and open problems is much longer, and it seems to keep growing with every step forward. Here, we discuss only two main problems related to the level of the analysis:

First, the investigations should look at much more detailed levels of expressive performance, which is currently precluded by fundamental measuring problems. At the moment, it is only possible to extract rather crude and global information from audio recordings; we cannot get at details such as timing, dynamics, and articulation of individual voices or individual notes. It is precisely at this level—in the minute details of voicing, intervoice timing, and so on—that many of the secrets of what music connoisseurs refer to as a pianist's unique "sound" are hidden. Making these effects measurable is a challenge for audio analysis and signal processing, one that is currently outside our own area of expertise.

Second, this research must be taken to higher structural levels. It is in the global organization of a performance, the grand dramatic structure, the shaping of an entire piece, that great artists express their interpretation and understanding of the music, and the differences between artists can be dramatic. These high-level "master plans'"do not reveal themselves at the level of local patterns we studied earlier. Important structural aspects and global performance strategies will only become visible at higher abstraction levels.

These questions promise to be a rich source of challenges for sequence analysis and pattern discovery. First, we may have to develop appropriate high-level pattern languages.

In view of these and many other problems, this project will most likely never be finished. However, much of the beauty of research is in the process, not in the final results, and we do hope that our (current and future) sponsors share this view and will keep supporting what we believe is an exciting research adventure.
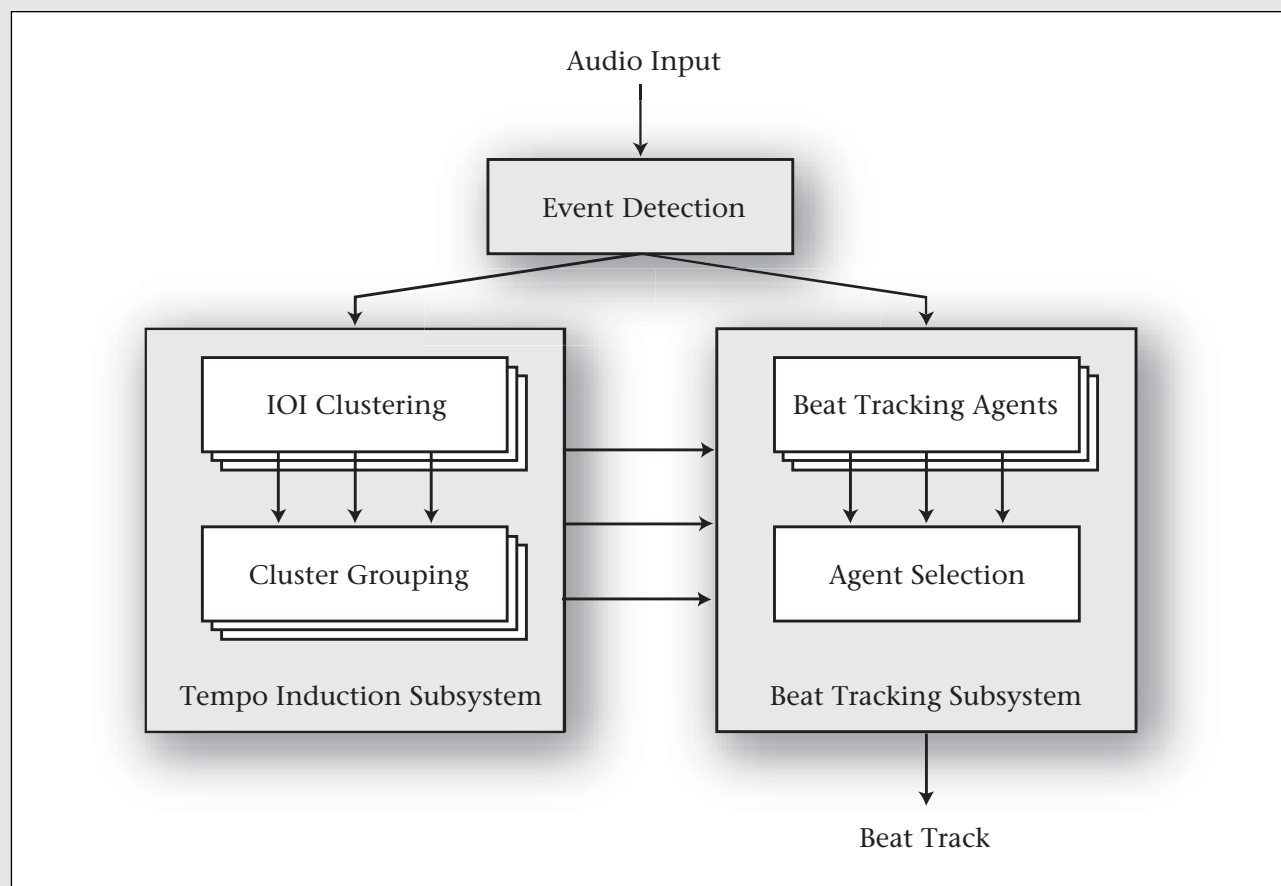
Finding the Beat
with BEATROOT

Figure A. BEATROOT Architecture.

Beat tracking involves identifying the basic rhythmic pulse of a piece of music and determining the sequence of times at which the beats occur. The BEATROOT (Dixon 2001b, 2001c) system architecture is illustrated in figure A. Audio or MIDI data are processed to detect the onsets of notes, and the timing of these onsets is analyzed in the tempo induction subsystem to generate hypotheses of the tempo at various metric levels. Based on these tempo hypotheses, the beat-tracking subsystem performs a multiple hypothesis search that finds the sequence of beat times fitting best to the onset times.

For audio input, the onsets of events are found using a time-domain method that detects local peaks in the slope of the amplitude envelope in various frequency bands of the signal. The tempo-induction subsystem then proceeds by calculating the interonset intervals (IOIs) between pairs of (possibly nonadjacent) onsets, clustering the intervals to find common durations, and then ranking the clusters according to the number of intervals

they contain and the relationships between different clusters to produce a ranked list of basic tempo hypotheses.

These hypotheses are the starting point for the beat-tracking algorithm, which uses a multiple-agent architecture to test the different tempo and phase hypotheses simultaneously and finds the agent whose predicted beat times most closely match those implied by the data. Each hypothesis is handled by a beat-tracking agent, characterized by its state and history. The *state* is the agent's current hypothesis of the beat frequency and phase, and the *history* is the sequence of beat times selected to date by the agent. Based on their current state, agents predict beat times and match them to the detected note onsets, using deviations from predictions to adjust the hypothesized current beat rate and phase or create a new agent when there is more than one reasonable path of action. The agents assess their performance by evaluating the continuity, regularity, and salience of the matched beats.
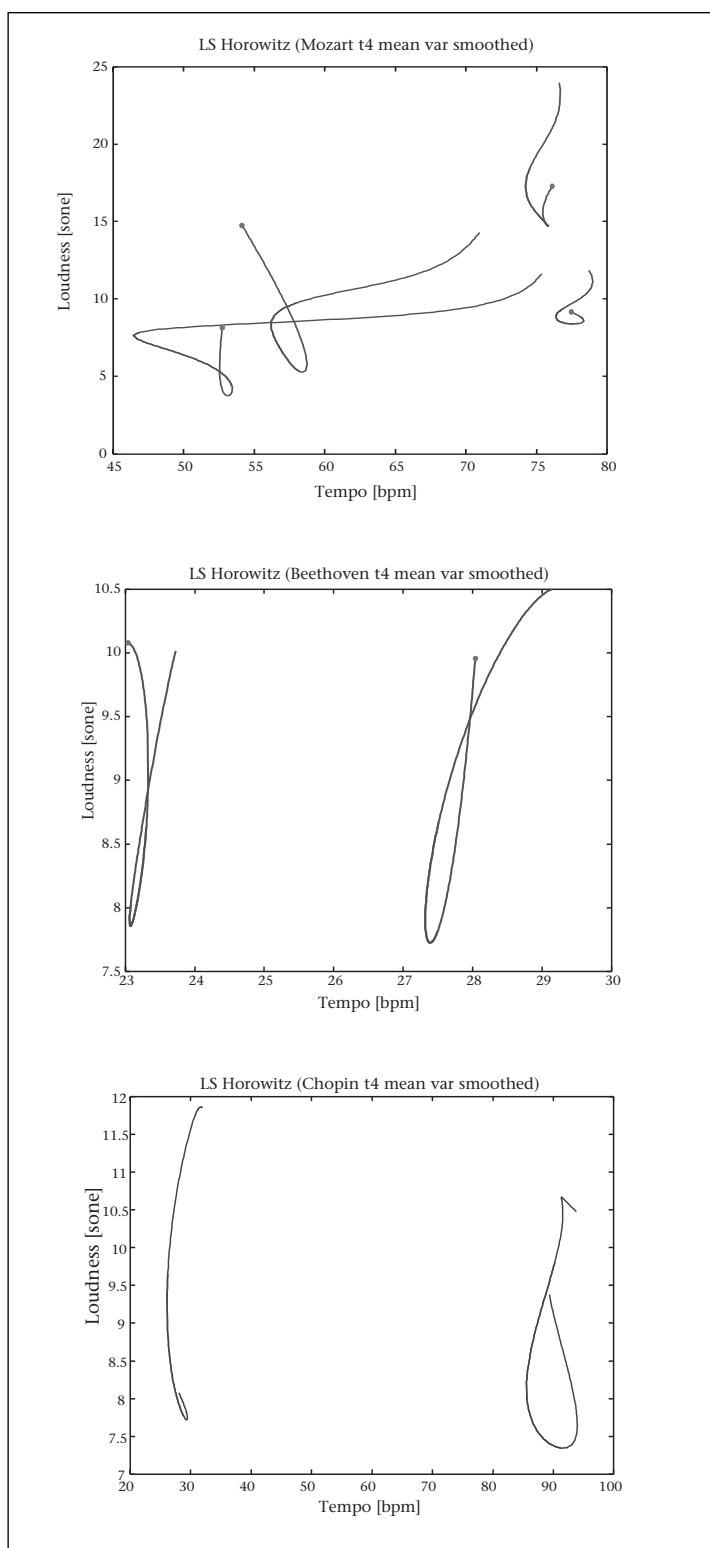
Figure 12. The Alleged Horowitz Pattern in Horowitz Recordings of Music by Mozart (top), Chopin (center), and Beethoven (bottom). (W. A. Mozart: Piano Sonata K281, B^b major, 1st movement, recorded 1989; F. Chopin: Etude op.10 no. 3, E major and Ballade no.4, op.52, F minor, recorded 1952; L. van Beethoven: Piano Sonata op.13 in C minor ["Pathéthique"], 2nd movement, recorded 1963).

Acknowledgments

Notes

1. See also the project web page at www.oefai.at/music.

2. At the moment, we restrict ourselves to classical tonal music and the piano.

3. It should be clear that a coverage of close to 100 percent is totally impossible, not only because expressive music performance is not a perfectly deterministic, predictable phenomenon but also because the level of individual notes is clearly insufficient as a basis for a complete model of performance; musicians think not (only) in terms of single notes but also in terms of higher-level musical units such as motifs and phrases—see the subsection entitled Multilevel Learning of Performance Strategies.

4. To be more precise, the rules predict whether a note should be lengthened or shortened; the precise numeric amount of lengthening or shortening is predicted by a *k*–nearest-neighbor algorithm (with $k = 3$) that uses only instances for prediction that are covered by the matching rule, as proposed in Weiss and Indurkhya (1995) and Widmer (1993).

5. Yes, there is such a thing….

6. The *beat* is an abstract concept related to the metrical structure of the music; it corresponds to a kind of quasiregular pulse that is perceived as such and that structures the music. Essentially, the beat is the time points where listeners would tap their foot along with the music. *Tempo*, then, is the rate or frequency of the beat and is usually specified in terms of beats per minute.

7. The program has been made publicly available and can be downloaded at www.oefai.at/~simon.

8. Martha Argerich, Deutsche Grammophon, 410 653-2, 1984; Vladimir Horowitz, CBS Records (Masterworks), MK 42409, 1962; Wilhelm Kempff, DGG 459 383-2, 1973.

9. Was this unintentional? A look at the complete trajectory over the entire piece reveals that quite in contrast to the other two pianists, she never returns

to the starting tempo again. Could it be that she started the piece faster than she wanted to?

10. However, it is also clear that through this long sequence of transformation steps—smoothing, segmentation, normalization, replacing individual elementary patterns by a prototype—a lot of information has been lost. It is not clear at this point whether this reduced data representation still permits truly significant discoveries. In any case, whatever kinds of patterns might be found in this representation will have to be tested for musical significance in the original data.

11. That would have been too easy, wouldn't it?

Bibliography

Agrawal, R., and Srikant, R. 1994. Fast Algorithms for Mining Association Rules. Paper presented at the Twentieth International Conference on Very Large Data Bases (VLDB), 12–15 September, Santiago, Chile.

Arcos, J. L., and López de Mántaras, R. 2001. An Interactive CBR Approach for Generating Expressive Music. *Journal of Applied Intelligence* 14(1): 115–129.

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24:123–140.

Cohen, W. 1995. Fast Effective Rule Induction. Paper presented at the Twelfth International Conference on Machine Learning, 9–12 July, Tahoe City, California.

Dietterich, T. G. 2000. Ensemble Methods in Machine Learning. In *First International Workshop on Multiple Classifier Systems,* eds. J. Kittler and F. Roli, 1–15. New York: Springer.

Dixon, S. 2001a. An Empirical Comparison of Tempo Trackers. Paper presented at the VIII Brazilian Symposium on Computer Music (SBCM'01), 31 July–3 August, Fortaleza, Brazil.

Dixon, S. 2001b. An Interactive Beat Tracking and Visualization System. In Proceedings of the International Computer Music Conference (ICMC'2001), La Habana, Cuba.

Dixon, S. 2001c. Automatic Extraction of Tempo and Beat from Expressive Performances. *Journal of New Music Research* 30(1): 39–58.

Dixon, S., and Cambouropoulos, E. 2000. Beat Tracking with Musical Knowledge. Paper presented at the Fourteenth European Conference on Artificial Intelligence (ECAI-2000), 20–25 August, Berlin.

Dixon, S.; Goebl, W.; and Widmer, G. 2002. The PERFORMANCE WORM: Real-Time Visualization of Expression Based on Langner's Tempo-Loudness Animation. Paper presented at the International Computer Music Conference (ICMC'2002), 16–21 September, Göteborg, Sweden.

Freund, Y., and Shapire, R. E. 1996. Experiments with a New Boosting Algorithm. Paper presented at the Thirteenth International Machine Learning Conference, 3–6 July, Bari, Italy.

Friberg, A. 1995. A Quantitative Rule System for Musical Performance. Ph.D. dissertation, Royal Institute of Technology (KTH).

Fürnkranz, J. 1999. Separate-and-Conquer Rule Learning. *Artificial Intelligence Review* 13(1): 3–54.

Gabrielsson, A. 1999. The Performance of Music. In *The Psychology of Music,* ed. D. Deutsch, 501–602. 2d ed. San Diego: Academic Press.

Hunter, L., ed. 1993. *Artificial Intelligence and Molecular Biology.* Menlo Park, Calif.: AAAI Press.

King, R. D.; Muggleton, S.; Lewis, R. A.; and Sternberg, M .J. E. 1992. Drug Design by Machine Learning: The Use of Inductive Logic Programming to Model the Structure-Activity Relationship of Trimethoprim Analogues Binding to Dihydrofolate Reductase. In *Proceedings of the National Academy of Sciences 89,* 11322–11326. Washington, D.D.: National Academy Press.

Kohonen, T. 2001. *Self-Organizing Maps.* 3d ed. Berlin: Springer Verlag.

Langner, J., and Goebl, W. 2003. Visualizing Expressive Performance in Tempo-Loudness Space. *Computer Music Journal.* Forthcoming.

Langner, J., and Goebl, W. 2002. Representing Expressive Performance in Tempo-Loudness Space. Paper presented at the ESCOM Conference on Musical Creativity, 5–8 April, Liège, Belgium.

López de Mántaras, R., and Arcos, J. L. 2002. AI and Music: From Composition to Expressive Performances. *AI Magazine* 23(3): 43–57.

Mannila, H.; Toivonen, H.; and Verkamo, I. 1997. Discovery of Frequent Episodes in Event Sequences. *Data Mining and Knowledge Discovery* 1(3): 259–289.

Muggleton, S.; King, R. D.; and Sternberg, M. J. E. 1992. Protein Secondary Structure Prediction Using Logic-Based Machine Learning. *Protein Engineering* 5(7): 647–657.

Nevill-Manning, C. G., and Witten, I. H. 1997. Identifying Hierarchical Structure in Sequences: A Linear-Time Algorithm. *Journal of Artificial Intelligence Research* 7:67–82.

Palmer, C. 1988. Timing in Skilled Piano Performance. Ph.D. dissertation, Cornell University.

Pampalk, E.; Rauber, A.; and Merkl, D. 2002. Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps. Paper presented at the International Conference on Artificial Neural Networks (ICANN'2002), 27–30 August, Madrid, Spain.

Quinlan, J. R. 1990. Learning Logical Definitions from Relations. *Machine Learning* 5:239–266.

Repp, B. H. 1999. A Microcosm of Musical Expression: II. Quantitative Analysis of Pianists' Dynamics in the Initial Measures of Chopin's Etude in E Major. *Journal of the Acoustical Society of America* 105(3): 1972–1988.

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. 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.

Shavlik, J. W.; Towell, G.; and Noordewier, M. 1992. Using Neural Networks to Refine Biological Knowl-

edge. *International Journal of Genome Research* 1(1): 81–107.

Stamatatos, E., and Widmer, G. 2002. Music Performer Recognition Using an Ensemble of Simple Classifiers. Paper presented at the Fifteenth European Conference on Artificial Intelligence (ECAI'2002), 21–26 July, Lyon, France.

Sundberg, J. 1993. How Can Music Be Expressive? *Speech Communication* 13: 239–253.

Todd, N. 1992. The Dynamics of Dynamics: A Model of Musical Expression. *Journal of the Acoustical Society of America* 91(6): 3540–3550.

Todd, N. 1989. Towards a Cognitive Theory of Expression: The Performance and Perception of Rubato. *Contemporary Music Review* 4:405–416.

Valdés-Pérez, R. E. 1999. Principles of Human-Computer Collaboration for Knowledge Discovery in Science. *Artificial Intelligence* 107(2): 335–346.

Valdés-Pérez, R. E. 1996. A New Theorem in Particle Physics Enabled by Machine Discovery. *Artificial Intelligence* 82(1–2): 331–339.

Valdés-Pérez, R. E. 1995. Machine Discovery in Chemistry: New Results. *Artificial Intelligence* 74(1): 191–201.

Weiss, S., and Indurkhya, N. 1995. Rule-Based Machine Learning Methods for Functional Prediction. *Journal of Artificial Intelligence Research* 3:383-403.

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. 2002a. In Search of the Horowitz Factor: Interim Report on a Musical Discovery Project. In Proceedings of the Fifth International Conference on Discovery Science (DS'02), 13–32. Berlin: Springer Verlag.

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

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. 1998. Applications of Machine Learning to Music Research: Empirical Investigations into the Phenomenon of Musical Expression. In *Machine Learning, Data Mining, and Knowledge Discovery: Methods and Applications,* eds. R. S. Michalski, I. Bratko, and M. Kubat, 269–293. Chichester, United Kingdom: Wiley.

Widmer, G. 1995. Modeling the Rational Basis of Musical Expression. *Computer Music Journal* 19(2): 76–96.

Widmer, G. 1993. Combining Knowledge-Based and Instance-Based Learning to Exploit Qualitative Knowledge. *Informatica* 17(4): 371–385.

Widmer, G. and Tobudic, A. 2003. Playing Mozart by Analogy: Learning Multi-Level Timing and Dynamics Strategies. *Journal of New Music Research* 32(3). Forthcoming.

Windsor, L., and Clarke, E. 1997. Expressive Timing and Dynamics in Real and Artificial Musical Performances: Using an Algorithm as an Analytical Tool. *Music Perception* 15(2): 127–152.

Wolpert, D. H. 1992. Stacked Generalization. *Neural Networks* 5(2): 24–-260.

Zanon, P., and Widmer, G. 2003. Learning to Recognize Famous Pianists with Machine Learning Techniques. Paper presented at the Third Decennial Stockholm Music Acoustics Conference (SMAC'03), 6–9 August, Stockholm, Sweden.

Zwicker, E., and Fastl, H. 2001. *Psychoacoustics: Facts and Models.* Berlin: Springer Verlag.

Gerhard Widmer is an associate professor in the Department of Medical Cyternetics and Artificial Intelligence at the University of Vienna and head of the Machine Learning and Data Mining Research Group at the Austrian Research Institute for Artificial Intelligence, Vienna, Austria. He holds M.Sc. degrees from the University of Technology, Vienna, and the University of Madison at Wisconsin and a Ph.D. in computer science from the University of Technology, Vienna. His research interests and current work are in the fields of machine learning, data mining, and intelligent music processing. In 1998, he was awarded one of Austria's highest research prizes, the START Prize, for his work on AI and music. His e-mail address is gerhard @ai.univie. ac.at.

Simon Dixon is a research scientist at the Austrian Research Institute for Artificial Intelligence, with research interests in beat tracking, automatic transcription, and analysis of musical expression. He obtained BSc (1989) and Ph.D. (1994) degrees from the University of Sydney in computer science and an LMusA (1988) in classical guitar. From 1994 to 1999, he worked as a lecturer in computer science at Flinders University of South Australia. His e-mail address is simon@oefai.at.

Werner Goebl is member of the Music Group at the Austrian Research Institute for Artificial Intelligence, Vienna, Austria. He is a Ph.D. student at the Institute for Musicology at Graz University. With his research on expressive music performance, he combines both experience as a performing classical pianist and insight from musicology. His e-mail address is werner.goebl@oefai.at.

Elias Pampalk is a Ph.D. student at the Vienna University of Technology. As a member of the music group of the Austrian Research Institute for Artificial Intelligence, he is developing and applying data-mining techniques for music analysis and music information retrieval. His e-mail address is elias@oefai.at.

Asmir Tobudic obtained an M.Sc. in electrical engineering from the University of Technology, Vienna. He joined the Music Group at the Austrian Research Institute for Artificial Intelligence in September 2001, where he currently works as a Ph.D. student. His research interests concentrate on developing machine learning techniques for building quantitative models of musical expression. His e-mail address is asmir@oefai.at.