

Studying a Creative Act with Computers: Music Performance Studies with Automated Discovery Methods

Gerhard Widmer

Department of Computational Perception, Johannes Kepler University Linz, Austria,
and

Austrian Research Institute for Artificial Intelligence, Vienna

`gerhard.widmer@jku.at`

Abstract

The purpose of this article is to demonstrate how advanced computer methods may be able to provide new insights into a complex creative activity such as music performance. The context is an inter-disciplinary research project in which Artificial Intelligence (AI) methods are used to analyse patterns in performances by human artists. In asking how the computer can take us closer to an understanding of creativity in music performance, we identify two pertinent research strategies within our project: the use of machine learning algorithms that try to discover common performance principles and thus help separate the ‘rationally explainable’ aspects of performance from the more genuinely ‘creative’ ones, and the use of data mining methods that can discover, visualise and describe performance patterns that seem to be characteristic of the style of particular artists and thus may be more directly related to their individual creativity. Some preliminary results are briefly presented that are indicative of the kinds of discoveries these algorithms can make. Some general issues regarding (musical) creativity and its relation to Artificial Intelligence are also briefly discussed.

1 Introduction

Imagine a computer that listens to musicians (classical pianists, say) playing and from this learns to play music ‘expressively’, or at least with musically ‘sensible’ timing, dynamics, articulation, etc. Is that possible? And if so, what would that tell us about the creativity of the human artists? And about the creativity of the learning machine?

We are indeed working on such computer programs. In a large, interdisciplinary research project, we use Artificial Intelligence (AI) technology to study fundamental principles of music performance. We collect large numbers of performances by concert pianists, precisely measure expressive aspects such as timing and dynamics in these, and then apply *learning programs* that analyse the resulting data and look for interesting patterns and regularities. Some of these discovered regularities can be expressed as quantitative predictive rules, which can then also be applied to new pieces to produce computer-generated performances.

The general goal of this research is to contribute new methods and results to the field of music performance research. Our approach is heavily data-driven. It is meant to complement other researchers’ work in the area of musical performance research (see, e.g., [9]), which has produced a wealth of detailed hypotheses and insights, but has often been based on rather

limited sets of performance data. We aim at — and probably are — performing the most data-intensive investigations ever done in musical performance research.

In this article, we will look at this research from the point of view of musical creativity. Although explaining creativity is decidedly not a goal of our research (that is beyond our domain of competence, and beyond what our methods can be expected to produce), it does make sense to ask whether this kind of quantitative, empirical approach may eventually permit us to learn something about the creative aspects of music performance. We will broadly identify two pertinent research strategies within our project: one that tries to separate the ‘rationally explainable’ aspects of performance from the more genuinely ‘creative’ ones (by using machine learning algorithms that try to discover common performance principles and formulate them as rules), and one that is targeted more directly at aspects of individual artistic creativity, by searching for, visualising, and describing performance patterns that seem to be characteristic of the style of particular artists. The former research line has been pursued for quite some time now and has already produced some interesting results, which will briefly be reviewed here. The latter approach is currently being developed; some illustrative examples will be shown, but no claims as to their musical significance can be made at this point. Still, we hope to convince the reader that this is a fruitful avenue to follow.

The structure of the article is as follows. Section 2 starts with some introductory remarks on the class of Artificial Intelligence methods that play a central role in this research. The two research strategies identified above are then presented in sections 3 and 4, respectively, along with some preliminary results and a description of currently ongoing research. In section 5, we come back to the notion of (musical) creativity and how it relates to the Artificial Intelligence approach that is at the core of our research methodology. While scientific caution and philosophical disposition prohibit us from directly attributing creativity (musical or otherwise) to a machine, it seems safe to say that AI methods and models can provide new insight into creative phenomena like expressive music performance, at least indirectly.

2 The methods: Machine Learning and Automated Discovery

As most of the readers will not be familiar with Artificial Intelligence, a few words on the basic methodology of machine learning and automated discovery seem in order. What does it mean for a machine to learn? Generally, learning is a very complex phenomenon, with many different manifestations. In the context of our project, we are interested in programs that can learn general rules from a finite set of example data. The field of *Machine Learning* [13] has developed a wide range of learning algorithms that induce knowledge in a variety of forms, like classification and prediction rules [8], decision trees [15], or logic programs [12].

The fundamental principle underlying all these methods is what is known in logic as *inductive generalisation*: learning, both for living beings and for machines, means drawing general conclusions from specific observations or experiences. For a learning machine, that involves constructing hypotheses that are consistent with the given observations (and possibly satisfy some additional constraints, like simplicity, generality, etc.). Given example data of some phenomenon in some domain and a formal language in which to express its hypotheses, a typical machine learning program searches through the space of all hypotheses (prediction rules, classification rules, or whatever) that it can potentially describe in its representation language, for those hypotheses that best agree with the given data. Usually, these hypotheses do not only describe the data they were derived from, but are *generalisations* of these and

thus can also be used to make predictions in new, unknown situations.

With the help of such inductive learning methods, the computer can analyse large databases of empirical data and can discover new, hitherto unknown and possibly useful regularities. This is used extensively in commercial applications (where the respective application scenario is known as *data mining* [31]), but researchers have shown that machine learning methods can also make interesting and substantial scientific discoveries, mainly in the natural sciences and in mathematics. The work presented here attempts to show that musicology may also gain something from the application of intelligent discovery methods.

3 Direction 1: Separating the ‘Rationally Explainable’ from the (Possibly) ‘Creative’

The main focus of our research so far has been on the machine-assisted search for common patterns in large corpora of performances by skilled pianists. The goal of this is computational, quantitatively testable models of what may be the basic principles or unwritten ‘rules’ that are followed by more or less all skilled musicians. In other words, we want to find and formalise those aspects of performance that are indispensable for the music to sound musical to us. In a sense, we are first trying to characterise the ‘non-creative’ aspects of performance, the things that are *common* to most ‘reasonable’ performances, before trying to explain the artistic individuality of famous musicians.

To that end, large numbers of performances by skilled pianists are collected, aspects of performance (expressive timing, dynamics, articulation) are precisely measured, and machine learning programs are applied that try to find common patterns in these measurements that seem to point to consistent strategies. The learning programs express their findings in terms of explicit prediction rules, which can be directly interpreted by humans and whose accuracy can be tested on new pieces.

3.1 Discovering note-level performance rules

After investing substantial efforts in the acquisition, preparation, and cleaning of a large corpus of expert pianist performances, which in fact forced us to develop a range of new intelligent music processing algorithms of general utility [24], first large-scale learning experiments were started in 2000. To reduce the enormous complexity of the problem, it was decided to first restrict the problem to the level of single notes, i.e., to learn models of the note-to-note expressive changes applied by a pianist (e.g., the lengthening or shortening of a note relative to the local tempo). It is clear that the note level is insufficient as a basis for a complete model of performance, but we first wanted to see how much of observed expressive variations could be described and ‘explained’ at this low level.

The *models* to be constructed are sets of *performance rules*, that is, rules that predict a performer’s expressive choices from properties of the music being played. More precisely, the task for the learning program is to find rules that predict local, note-level timing differences (essentially, lengthening or shortening of notes relative to the nominal duration prescribed by the score), dynamics (whether a note will be played louder or softer than its predecessor), and staccato vs. legato.

The *training data* used in the experiments were 13 complete Mozart piano sonatas, played by the Viennese concert pianist Roland Batik on a Bösendorfer SE290 computer-monitored

grand piano. That is probably one of the largest corpora of precisely measured expert performances ever collected and analysed in empirical performance research; the 13 sonatas represent some 4 hours of music and some 106.000 notes. Every detail of timing, dynamics, articulation, and pedalling can be reconstructed from the data stored by the Bösendorfer SE.

In a first study with state-of-the-art learning algorithms [23], it was shown that there is indeed structure in the data that can be uncovered by machine learning algorithms. However, the resulting sets of performance rules were extremely complex and did not exhibit an adequate level of predictive accuracy when used to predict a pianist’s actual choices. In response to that, a new learning algorithm was developed that is specifically geared towards finding simple partial rule models in complex data [25], and this algorithm has indeed made a number of surprising discoveries.

For instance, when applied to our corpus of Mozart sonata performances, it discovered an extremely small set (17 rules) of very simple rules that correctly predict (or ‘explain’) a large number of the pianist’s expressive or articulatory note-level actions [27]. A detailed discussion of the rules and their musical interpretation is beyond the scope of this article. Generally, it turned out that certain kinds of note (IOI) lengthening, *staccato*, and to a lesser extent the local stresses in dynamics, are surprisingly well predictable, and with extremely few and simple rules. The following rule gives an impression of the simplicity and generality of the rules that were discovered:

```
abstract_duration_context = equal-longer
& metr_strength ≤ 1
⇒ lengthen
```

“Given two notes of equal duration followed by a longer note, lengthen the note 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: the rule correctly predicts 1,894 cases of local note lengthening, which is 14.12% of all the instances of significant lengthening observed in the training data. The number of incorrect predictions is 588 (2.86% 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 a complex corpus such as Mozart sonatas.

A number of other interesting discoveries were made; they are described in detail in [27]. Some of the rules discovered by the machine bear a strong resemblance to expression principles discovered or hypothesised by other performance researchers (e.g., rules proposed in [7]).

The remarkable result is that, at least for timing and articulation, such a high proportion of the observed low-level variations can be predicted by so few (and simple) rules. A striking example of this is shown in Figure 1, which compares the pianist’s timing in the first section of the A major sonata K.331 to the tempo curve predicted by the learned rules. Here, essentially *two* simple rules are sufficient to predict and reproduce the pianist’s interpretation extremely well.¹

¹Admittedly, this is an extreme example. There are also pieces where the rules do not succeed in predicting the pianist well, or make no predictions at all.

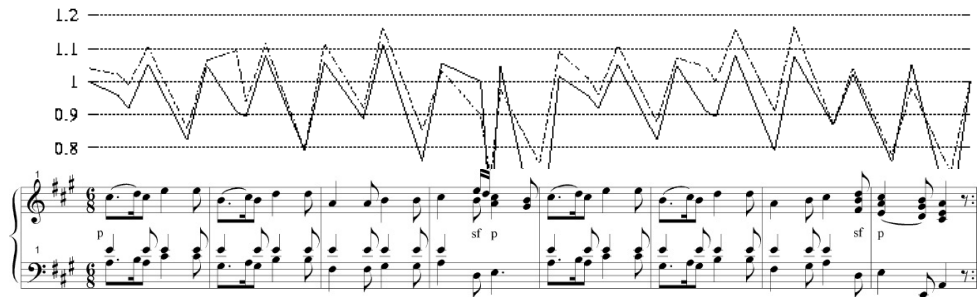


Figure 1: Mozart Sonata K.331, 1st movement, 1st part, as played by pianist and learner. The curve plots the relative tempo at each note — notes above the 1.0 line are shortened relative to the tempo of the piece, notes below 1.0 are lengthened. A perfectly regular performance with no timing deviations would correspond to a straight line at $y = 1.0$

When the rules were tested on Mozart performances by a different artist (the renowned conductor and pianist Philippe Entremont), again recorded on a Bösendorfer SE290, it turned out that they carry over to the new performer with virtually no loss in predictive accuracy, which testifies to the generality of the discovered expression principles, at least with respect to different performers.

Another experiment tested the generality of the discovered rules with respect to *musical style*. They were applied to two pieces by Frédéric Chopin — the Etude Op.10, No.3 in E major (first 20 bars) and the Ballade Op.38, F major (first 45 bars) — and the results were compared to performances of these pieces by 22 Viennese pianists. The results of this experiment were even more surprising. While the dynamics rules failed to predict well on the Chopin performances, the rules for timing and articulation performed extremely well, better in fact than on the original (Mozart) data which the rules had been learned from! That indicates that the rules discovered by the machine represent truly general performance principles. More experimental details are given in [27].

3.2 Extending prediction to phrase-level shapes

Clearly, only a fraction of a performer’s decisions regarding timing 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 (e.g., motifs, groups, phrases), and they use tempo and dynamics to ‘shape’ these structures, e.g., by applying a gradual *crescendo* or *decrescendo* to entire passages. In fact, music performance is a *multi-level* phenomenon, with musical structures and performance patterns at various levels embedded in each other.

Accordingly, the set of note-level performance rules described above is currently being augmented with a multi-level learning strategy where the computer learns to predict elementary tempo and dynamics ‘shapes’ (represented as quadratic functions) at different levels of the hierarchical phrase structure, and combines these predictions with local timing and dynamics modifications predicted by learned note-level models. Preliminary experiments, again with performances of Mozart sonatas, yielded very promising results [29]. Both quantitative evaluations and listening tests show that the multi-level computational performance model can produce quite musical performances.

Unfortunately, the phrase-level models learned by our current learning algorithms are

rather opaque. Our current work in this area focuses on alternative learning algorithms that output an explicit model description that can be put in relation to other phrasing models proposed in the literature (e.g., [19, 20]; see also [30]).

4 Direction 2: Trying to Describe Artistic Individuality

The work described in the previous section tries to identify very basic principles of performance and thus focuses on *commonalities* between performances and performers. In a more recent line of inquiry we have now also started to look at characteristic *differences* between performers. The goal is to find and describe aspects of the individual artistic style of great artists. Our goal here is explicitly to study famous musicians (concert pianists).

4.1 The data: Measuring timing and dynamics at the beat level

The first major difficulty is data acquisition. With famous pianists, the only source of data is *audio recordings*, i.e., records and music CDs (we cannot very well invite them all to Vienna to perform on the Bösendorfer 290SE piano). Unfortunately, it is impossible, with current signal processing methods, to extract precise performance information (start and end times, loudness, etc.) 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 pianos or the Bösendorfer SE. 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*. This could be done manually (as in [16]), which is extremely time-consuming and limits the number of recordings that can be studied with reasonable effort, or (semi-)automatically, with the help of intelligent audio processing methods. We have opted for the machine-assisted approach and developed an interactive *beat tracking system* [3, 4] that estimates the rate and times of musical beats from audio recordings. The system facilitates an iterative, semi-automated approach to beat-level tempo tracking by virtue of a graphical interface that allows a human user to modify the system’s beat hypotheses and retrack arbitrary sections, based on new information.²

From the varying time intervals between the measured beat 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 resulting data consists in pairs of curves that represent changes of tempo and overall loudness at the beat level. Clearly, this is only a fraction of the information that is contained in and makes a complex performance. For instance, important dimensions like articulation or the relative timing and dynamics of individual voices cannot currently be measured. Whether this limited data material still permits useful insights into aspects of creative personal style is a topic of our ongoing research, which is described in the following.

4.2 Making individual performance style visible

An important first step in the analysis of complex data is data visualisation. If properly displayed, the tempo and loudness curves derived from recordings can facilitate intuitive insights into high-level aspects of performance style. Here we draw on an original idea and

²The program has been made publicly available and can be downloaded from <http://www.oefai.at/~simon/beatroot/>.

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 [10]. To provide for a visually appealing display, smoothing is applied to the originally measured series of data points. Various degrees of smoothing can highlight regularities or performance strategies at different structural levels [11].

We have developed a visualisation program called the *Performance Worm* [5, 6] that displays smoothed tempo-loudness trajectories (derived from tempo and dynamics curves measured as described above) in synchrony with the music. A movement to the right signifies an increase in tempo, a *crescendo* causes the trajectory to move upward, etc. The trajectories can be stored to file and be used for systematic quantitative analysis (see below).

To give a simple example, Figure 2 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.³ Considerable smoothing was applied here, in order to highlight the higher-level developments within this extended phrase.

It is immediately obvious from Figure 2 that Horowitz and Kempff have chosen a very similar interpretation. Both essentially divide the phrase into two four-bar parts, where the first is played more or less with an *accelerando* (the worm moves to the right), and the second 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) up to the loudness climax around the end of bar five, followed by a *decrescendo* towards the end of the phrase. Martha Argerich’s trajectory, on the other hand, 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.⁴ 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 — in order to do that, one would have to see (and hear) the phrase in the context of the entire piece — but just to indicate the kinds of things one can see in such a visualisation. Many more details can be seen when less smoothing is applied to the measured data.

4.3 A surprising transformation step

The (smoothed) tempo-loudness trajectories provide intuitive insights, which might make visualisation programs like the Worm a useful tool for music education. But the goal of our research is to go beyond informal observations and derive objective, quantitative conclusions from the data.

The first step towards that is a transformation of the given data, to make them accessible to machine learning and automated discovery methods. The tempo-loudness trajectories computed by the Worm are cut into segments of a specified length (e.g., 2 or 4 beats),

³Martha Argerich, Deutsche Grammophon, 410 653-2, 1984;
Vladimir Horowitz, CBS Records (Masterworks), MK 42409, 1962;
Wilhelm Kempff, DGG 459 383-2, 1973.

⁴Or was this unintentional? A look of 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?

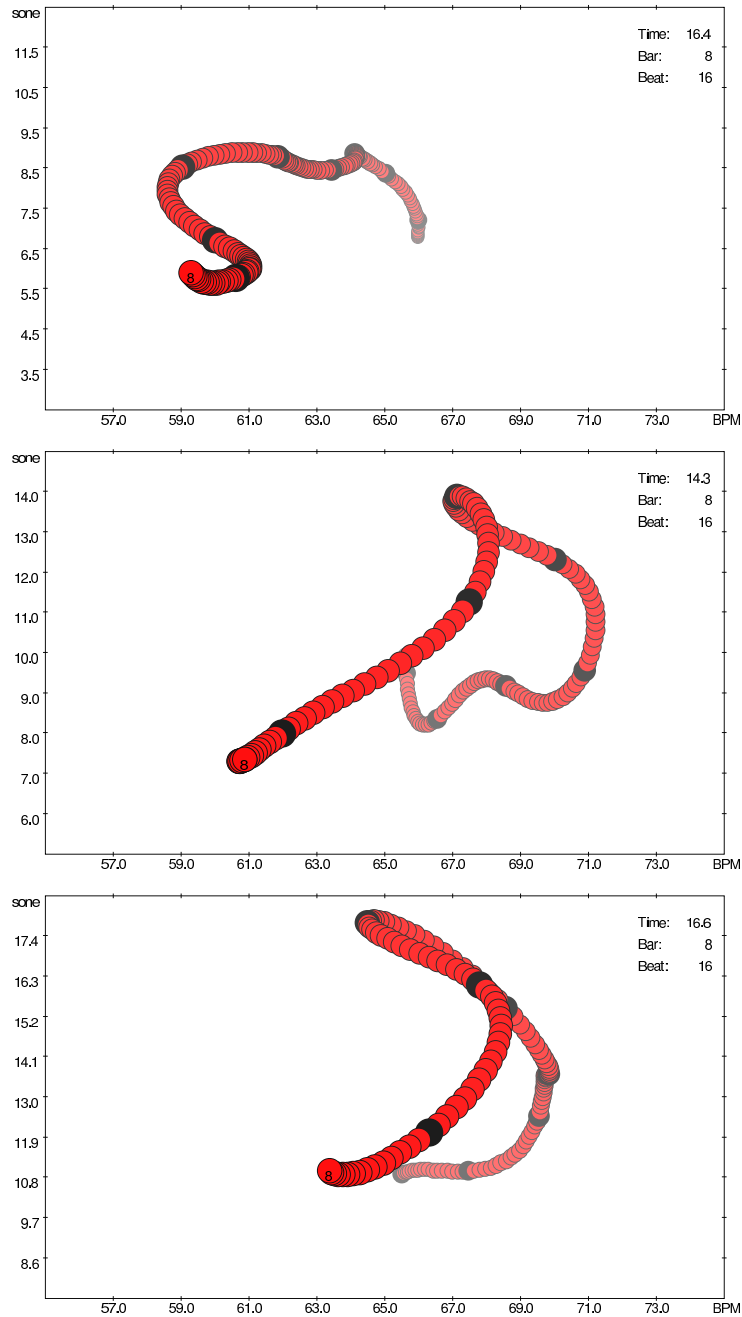


Figure 2: Performance trajectories over the first eight bars of *Von fremden Ländern und Menschen* (from *Kinderszenen*, op.15, by Robert Schumann), as played by Martha Argerich (top), Vladimir Horowitz (center), and Wilhelm Kempff (bottom). Horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in *sone* [32]. The largest point represents the current instant, while 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*) etc. Note that the *y* axes are scaled differently.

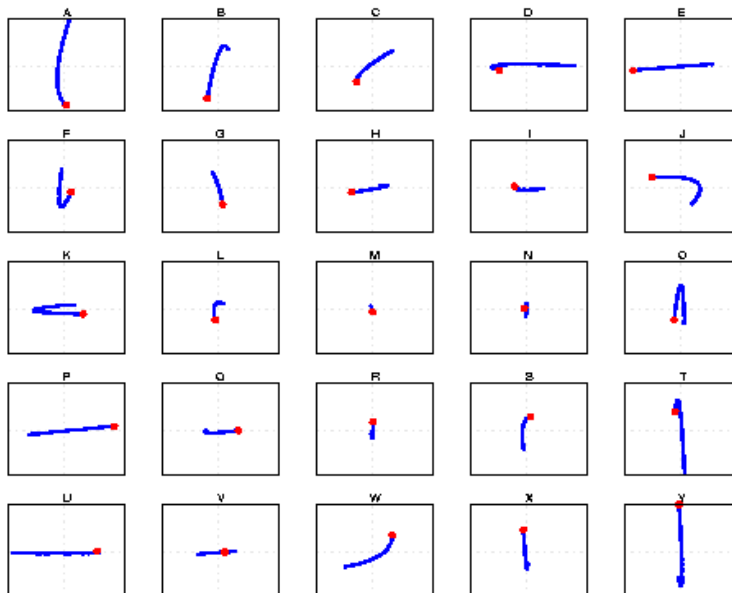


Figure 3: A typical ‘Chopin performance alphabet’ (set of prototypical performance trajectory segments) computed by segmentation and clustering from Chopin recordings by 13 pianists (among them Arthur Rubinstein, Vladimir Horowitz, and Sviatoslav Richter). To indicate directionality, dots mark the end points of segments.

and these segments are optionally subjected to various normalization operations (e.g., 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 via a *clustering* algorithm. For each of the resulting clusters, a *prototype* — the most typical pattern within its class — is computed. These 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 3 shows one such alphabet that was computed from a sizeable number of Chopin performances by a total of 13 different famous pianists. We can take the notion of ‘alphabet’ quite literally and associate each of these typical tempo-loudness patterns with a letter. A complete performance, represented as a tempo-loudness trajectory over an entire piece, can then be approximated by a sequence of these prototypical short patterns. Replacing these by letters yields a new representation of our data: a performance is a sequence of letters, i.e., a *string*, in computer science terminology. Figure 4 shows a part of a Horowitz Chopin performance in this representation.

This final transformation step, trivial as it may be, makes it evident that the original musical problem has now been transferred into a quite different world: the world of *string analysis*. The fields of pattern recognition, machine learning, data mining, etc. have developed a rich set of methods that can find structure in strings and that could now profitably be applied to our musical data.⁵

⁵However, 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

RNRNRXSRLFBCQVIMRYTSBAGCGRNQJBFXOQNLHCKWOCQVSNRNMWRMLLMRQJIIQHC
KVIQVLLRXYOGFLFSOCQVRSQCQHCQVLQVRSJBFMHKHIQWOFNRSHCFNHPWLGHCQVI
QVHINXSWOFLRSLRSBCPVCGBCKVLQWNXORYOAFOFYOCKVMRHBVFCKVIQVRSNLGLH
KMIQWNRNRJKXSLCPWIGBCKVISVLMLMLQMRNMRLMMNMMMRNRNRNHRNNGHCQQVNMN
QHLRNKVINRNRXOFHCPWLCCCPWIVYSVCRVCKJBFNXOXTKVXSBARXBFNLBPCPWO
KWXSSBFJIGGCIBAGFJKVIQWHCXYSWNBXWJINBFWOCWOAFJOQTLKVRSJBFBFWTL
HBFWHDPHCQHXYTLMNLRNBABXWSSSHISLRNEIPVIMLFNMXBAQWGQVMQCKHCPUM
QMIVGMMLLPVKVSVLINLQVHIXJINLGHIMMEDQVJKVHIXJKUHKVHKVE.....
.....

Figure 4: Beginning of a performance by Vladimir Horowitz of Frédéric Chopin’s Ballade op.52 in F minor, coded in terms of the 25 letter ‘performance alphabet’ shown in Figure 3.

4.4 Discovery of characteristic performance patterns

There is a multitude of questions one might want to ask of these musical strings. For instance, one might search for local patterns (substrings) that are characteristic of a particular pianist (that will be demonstrated below). 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 (e.g., [1]) to induce (partial) ‘*grammars*’ of musical performance style. A broad range of methods is available. Whether their application produces results that make musical sense remains to be investigated.

As a first example, let us consider the following question: are there substrings in these musical performance strings that exhibit a certain minimum frequency and that are characteristic of a particular pianist? Such substrings would correspond to typical tempo-loudness patterns applied by a pianist. We are currently developing algorithms for finding such *discriminative frequent subsequences* in large collections of strings [28].

First preliminary experiments with a substantial corpus of measured performances (Mozart piano sonatas, various Chopin Preludes, Etudes, and Ballades) by various pianists are currently being carried out. Our first impression is that the pattern discovery algorithms are indeed capable of finding performance patterns (i.e., letter sequences) that might be characteristic of particular artists. To find out whether such a sequence codes any musically interesting or interpretable behaviour, we can go back to the original data (the tempo-loudness trajectories) and identify the segments of the trajectories coded by the various occurrences of the sequence.

To illustrate, Figure 5 shows two families of performance patterns that were detected that might be typical of the Mozart style of Daniel Barenboim (left) and Mitsuko Uchida (right). The pattern on the left, which was found several times in Mozart piano sonata performances by Daniel Barenboim and almost never in other pianists’ recordings, is characterised by 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. That might indeed be a rather unusual pattern. In our experience so far, it is quite rare to see a pianist speed

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 may be found in this representation will have to be tested for musical significance in the original data.

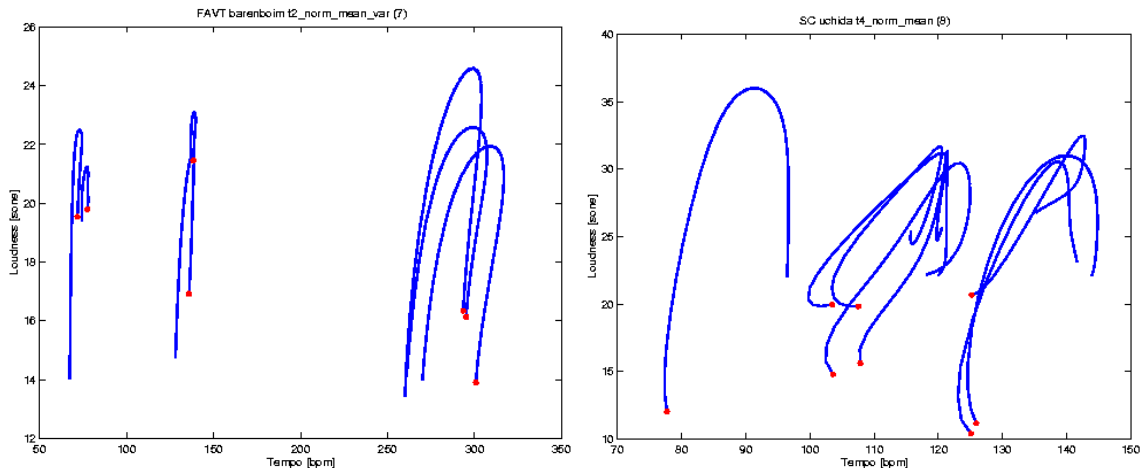


Figure 5: Two sets of (instantiations of) performance patterns: trajectory segments typical of Daniel Barenboim (left) and Mitsuko Uchida (right). To indicate directionality, a dot marks the end point of a segment.

up during a loudness maximum. Much more common in such situations are *slowings down* (*ritardandi*), which gives a characteristic counterclockwise movement of the Worm (as, e.g., in the right half of Figure 5, which shows instantiations of a pattern that seems characteristic of the style of Mitsuko Uchida).

However, the reader must be warned that this is a very difficult data analysis problem, and what is described here is work in progress. The current data situation is too limited to draw serious conclusions, and determining the musical significance of such patterns will be a complex problem. The absolute numbers (7 or 8 occurrences of a supposedly typical pattern in a pianist) are too small to support claims regarding statistical significance. And even if statistical significance can be ascertained, the patterns must be checked for artifacts and for musical and perceptual relevance — after all, a pattern that is due to some technical quirks or that we cannot *hear* in a performance is hardly of interest. So that patterns shown here should not be taken too literally. They are only indicative of the *kinds of things* we hope to discover with our methods in the (near) future. Whether these findings will indeed be musically relevant can only be hoped for at the moment.

4.5 Automatic performer classification

In another attempt to get a grip on the quantitative differences between individual performers, we have recently started to investigate the possibility of *automatic performer classification*. Machine learning algorithms are trained on performances by different pianists and try to learn to identify the pianist in unknown recordings. First experiments with performance data based on MIDI recordings (played by a large number of pianists from the Vienna University of Music) yielded surprisingly good results [17, 18]. For instance, in a ten-way classification task (identifying the correct pianist in ten performances, each played by a different pianist), the computer managed to correctly identify 7 out of 10 pianists, which is significantly better than the 10% recognition rate that would result from random guessing [18].

Current research tries to extend this approach to the problem of identifying world-class pianists, based on properties of the tempo-loudness trajectories extracted from audio record-

ings. These trajectories provide far less information than what can be extracted from MIDI recordings. Very preliminary (and inconclusive) experimental results indicate that it may still be possible to identify certain artists with some degree of certainty. That would indicate that there are indeed style-specific patterns hidden in these performance trajectories. Future work will then aim at pinpointing those features of the performances that are discriminative and predictive (along the lines of [17]).

5 AI and Creativity?

Let us now take a step back and look at the more general issue of (musical) creativity and the role of Artificial Intelligence. Can AI be used to model musical creativity, and are AI programs creative? Let us restrict the question to the particular context of the research project described above. Here, the issue of creativity comes up in at least three different places or levels:

1. **In the phenomenon we are studying.** Surely, expressive performance is a creative act. Beyond mere technical virtuosity, it is the ingenuity with which great artists portray a piece in a new way and illuminate aspects we have never noticed before that commands our admiration. Can the creative aspects of expressive performance be captured in a formal model? Certainly, we cannot expect to formalise an artist's decisions behind expressive performance to the point where they would become completely predictable. That is precluded by the sheer complexity of the phenomenon (if not by deeper reasons). However, Artificial Intelligence can make at least two kinds of contributions, as we hope to have shown here.

One is to find models that as much as possible capture those aspects of musical expression that have a 'rational' basis, i.e., that are constrained by structural properties of the music, cognitive dispositions of human listeners, and cultural norms, including performance tradition. The goal here is not to model the genuinely creative aspects, but to investigate how far one can push the frontier of 'rational explainability', as it were.

In the realm of 'true' (i.e., essentially unpredictable) creativity, we cannot offer any explanatory models, but AI can still make useful contributions by providing tools that allow us to visualise and characterise, and thus possibly enjoy (if not understand) even better, aspects of truly creative musical behaviours.

2. **In the machine that makes discoveries.** Is a learning machine creative? The learning programs have made a number of interesting and surprising discoveries in our project. Is that a manifestation of creativity? It is beyond doubt that a machine can exhibit unexpected (though 'rational') behaviour and do surprising new things, things that were not foreseen by the programmer, but still follow logically from the way the machine was programmed. Seemingly creative behaviour can emerge as a consequence of high *complexity* of the machine's environment, input, input-output mapping, etc.⁶ That is what happens with the machine learning algorithms in our music analysis project. They can find rules that are truly novel in the sense that no human has thought of them (and the machines were not programmed to do/discover explicitly *this*). And that then

⁶One might argue that the same holds for the creativity we attribute to human beings.

appears to us as creative behaviour. Creativity is in the eye of the beholder, one might say.

- 3. In the scientific approach** towards studying expressive performance. Undoubtedly, scientific research is a highly creative activity. To what extent can this activity be modelled in AI systems, and which aspects of the scientific discovery process can be taken over by machines? The idea of an intelligent computer extending the creative abilities of human scientists is appealing. Formalising and automating scientific discovery is an active research area within Artificial Intelligence, with applications predominantly in the formal and natural sciences, and there have been some discoveries that can be said to have been made by the computer (e.g., [2, 14, 21, 22]). Most of these discoveries, however, heavily rely on the human scientist to prepare the right kind of context for the discovery program. The most creative aspects of scientific research — posing the right kinds of questions, setting up appropriate experiments to decide these questions, and interpreting the results — are currently beyond the reach of machines. They require, among other things, a very broad knowledge both of science and of the real world, and the ability to make connections between things that are seemingly far apart (which is probably an important aspect of creativity). What computers can currently help the scientist with are narrower, more well-defined tasks, such as the structuring and condensing of empirical observations, and the search for explanations of these observations. This can still be immensely useful, because it permits us to analyse much larger sets of empirical data than could ever be tackled by a human alone.

To summarise, Artificial Intelligence may help us study creative (musical) behaviour — or more precisely: artifacts that may be manifestations of creative behaviour — in new ways. We cannot and do not claim to study or formalise creativity itself, its sources, inner workings, and effects. We hope to produce empirical material that may be useful to other researchers who are more competent in the study of creativity proper.

Whether machines themselves can be credited with creativity, and whether musical creativity in particular can be captured in formal models is a question that is beyond the scope of the present article. Here, we have settled on the pragmatic view that machines can exhibit behaviour that appears creative to us, and they can produce results and make discoveries that are surprising and useful for musicological research. Whether we attribute creativity to the machines seems secondary in this context.

That is not to say that attempts to explain and formalise creativity are not worthwhile or important. They may not only lead us to ever more intelligent and useful computer programs, but also teach us something about ourselves, give us a deeper understanding of what to me is one of the most elementary and satisfying experiences: to feel that one has created something new.

Acknowledgements

The project is made possible 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). Additional support for our research on machine learning 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. I would like to thank my colleagues Emiliós Cambouropoulos, Simon Dixon, Werner Goebel, Elias Pampalk, and Asmir Tobudic for their cooperation and helpful comments, and the pianists Roland Batik and Philippe Entremont for allowing us to use their performances for our investigations. Thanks also to the L. Bösendorfer Company, Vienna, and in particular Fritz Lachnit for generous help with the Bösendorfer SE290.

References

- [1] Chen, S.F. (1995). Bayesian Grammar Induction for Language Modeling. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pp.228–235.
- [2] Colton, S. (1999). Refactorable Numbers: A Machine Invention. *Journal of Integer Sequences*, vol. 2 (article 99.1.2).
- [3] Dixon, S. (2001). Automatic Extraction of Tempo and Beat from Expressive Performances. *Journal of New Music Research* 30(1), 39–58.
- [4] Dixon, S. (2003). Analysis of Musical Content in Digital Audio. In J. di Marco (ed.), *Computer Graphics and Multimedia: Applications, Problems, and Solutions* (to appear).
- [5] Dixon, S., Goebel, W., and Widmer, G. (2002). The Performance Worm: Real Time Visualisation of Expression Based on Langner’s Tempo-Loudness Animation. In *Proceedings of the International Computer Music Conference (ICMC’2002)*, Göteborg, Sweden.
- [6] Dixon, S., Goebel, W., & Widmer, G. (2002). Real Time Tracking and Visualisation of Musical Expression. In *Proceedings of the 2nd International Conference on Music and Artificial Intelligence (ICMAI’2002)*, Edinburgh, Scotland.
- [7] Friberg, A. (1995). *A Quantitative Rule System for Musical Performance*. Doctoral Dissertation, Royal Institute of Technology (KTH), Stockholm, Sweden.
- [8] Fürnkranz, J. (1999). Separate-and-Conquer Rule Learning. *Artificial Intelligence Review* 13(1), 3–54.
- [9] Gabrielsson, A. (1999). The Performance of Music. In D. Deutsch (Ed.), *The Psychology of Music (2nd ed.)*, pp. 501–602. San Diego, CA: Academic Press.
- [10] Langner, J. and Goebel, W. (2002). Representing Expressive Performance in Tempo-Loudness Space. In *Proceedings of the ESCOM Conference on Musical Creativity*, Liège, April 2002.
- [11] Langner, J. and Goebel, W. (2002). *Visualizing Expressive Performance in Tempo-Loudness Space*. Submitted.
- [12] Lavrac N. and Džeroski S. (1994). *Inductive Logic Programming*. Ellis Horwood, Chichester, UK.
- [13] Mitchell, T.M. (1997). *Machine Learning*. New York, NY: McGraw-Hill.

- [14] 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.
- [15] Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.
- [16] 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.
- [17] Stamatatos, E. (2002). Quantifying the Differences between Music Performers: Score vs. Norm. In *Proceedings of the International Computer Music Conference (ICMC’2002)*, Göteborg, Sweden.
- [18] Stamatatos, E. and Widmer, G. (2002). Music Performer Recognition Using an Ensemble of Simple Classifiers. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI’2002)*, Lyon, France.
- [19] Todd, N. (1989). Towards a Cognitive Theory of Expression: The Performance and Perception of Rubato. *Contemporary Music Review*, vol. 4, pp. 405–416.
- [20] Todd, N. (1992). The Dynamics of Dynamics: A Model of Musical Expression. *Journal of the Acoustical Society of America* 91, pp.3540–3550.
- [21] Valdés-Pérez, R.E. (1995). Machine Discovery in Chemistry: New Results. *Artificial Intelligence* 74(1), 191–201.
- [22] Valdés-Pérez, R.E. (1996). A New Theorem in Particle Physics Enabled by Machine Discovery. *Artificial Intelligence* 82(1-2), 331–339.
- [23] Widmer, G. (2000). Large-scale Induction of Expressive Performance Rules: First Quantitative Results. In *Proceedings of the International Computer Music Conference (ICMC’2000)*. San Francisco, CA: International Computer Music Association.
- [24] Widmer, G. (2001). Using AI and Machine Learning to Study Expressive Music Performance: Project Survey and First Report. *AI Communications* 14(3), 149-162.
- [25] Widmer, G. (2001). Discovering Strong Principles of Expressive Music Performance with the PLCG Rule Learning Strategy. In *Proceedings of the 11th European Conference on Machine Learning (ECML’01)*, Freiburg. Berlin: Springer Verlag.
- [26] Widmer, G. (2001). Inductive Learning of General and Robust Local Expression Principles. In *Proceedings of the International Computer Music Conference (ICMC’2001)*. San Francisco, CA: International Computer Music Association.
- [27] Widmer, G. (2002). Machine Discoveries: Some Simple, Robust Local Expression Principles. *Journal of New Music Research* 31(1), 37–50.
- [28] Widmer, G. (2002). In Search of the Horowitz Factor: Interim Report on a Musical Discovery Project. In *Proceedings of the 5th International Conference on Discovery Science (DS’02)*, Lübeck, Germany. Berlin: Springer Verlag.

- [29] Widmer, G. and Tobudic, A. (2002). Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies. In *Proceedings of the ICAD Workshop on Performance Rendering Systems, 8th International Conference on Auditory Display (ICAD'2002)*, Kyoto, Japan.
- [30] 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(2), 127–152.
- [31] Witten, I.H. & Frank, E. (1999). *Data Mining*. San Francisco, CA: Morgan Kaufmann.
- [32] Zwicker, E. and Fastl, H. (2001). *Psychoacoustics. Facts and Models*. Springer Series in Information Sciences, Vol.22. Berlin: Springer Verlag.