

YQX Plays Chopin

*Gerhard Widmer, Sebastian Flossmann,
and Maarten Grachten*

■ This article is about AI research in the context of a complex artistic behavior: expressive music performance. A computer program is presented that learns to play piano with “expression” and that even won an international computer piano performance contest. A superficial analysis of an expressive performance generated by the system seems to suggest creative musical abilities. After a critical discussion of the processes underlying this behavior, we abandon the question of whether the system is really creative and turn to the true motivation that drives this research: to use AI methods to investigate and better understand music performance as a human creative behavior. A number of recent and current results from our research are briefly presented that indicate that machines can give us interesting insights into such a complex creative behavior, even if they may not be creative themselves.

Consider the following challenge:

A computer program is to play two piano pieces that it has never seen before in an “expressive” way (that is, by shaping tempo, timing, dynamics, and articulation in such a way that the performances sound “musical” or “human”). The two pieces were composed specifically for the occasion by Professor N.N. One is in the style of nocturnes by Frederic Chopin, the other is in a style resembling Mozart piano pieces. The scores of the compositions will contain notes and some expression marks (*f*, *p*, *crescendo*, *andante*, slurs, and so on) but no indications of phrase structure, chords, and so on. The expressive interpretations generated by the program must be saved in the form of MIDI files, which will then be played back on a computer-controlled grand piano. Participants are given 60 minutes to set up their computer, read in the scores of the new compositions (in MusicXML format), and run their performance computation algorithms. They are not permitted to hand-edit performances during the performance-rendering process nor even to listen to the resulting MIDI files before they are publicly performed on the computer grand piano. The computer performances will be graded by a human audience and by the composer of the test pieces, taking into account the degree of “naturalness” and expressiveness. An award will be presented to the system that rendered the best performance.

Now watch (and listen to) the two videos provided at www.cp.jku.at/projects/yqx. The videos were recorded on the occasion of the Seventh Performance Rendering Contest (Rencon 2008),¹ which was held as a special event at the 10th International Conference on Music Perception and Cognition (ICMPC 2008) in Sapporo, Japan, in August 2008. Rencon² is a series of international scientific competitions where computer systems capable of generating “expressive” renditions of music pieces are compared and rated by human listeners. The previous task description is in fact a paraphrase of the task description

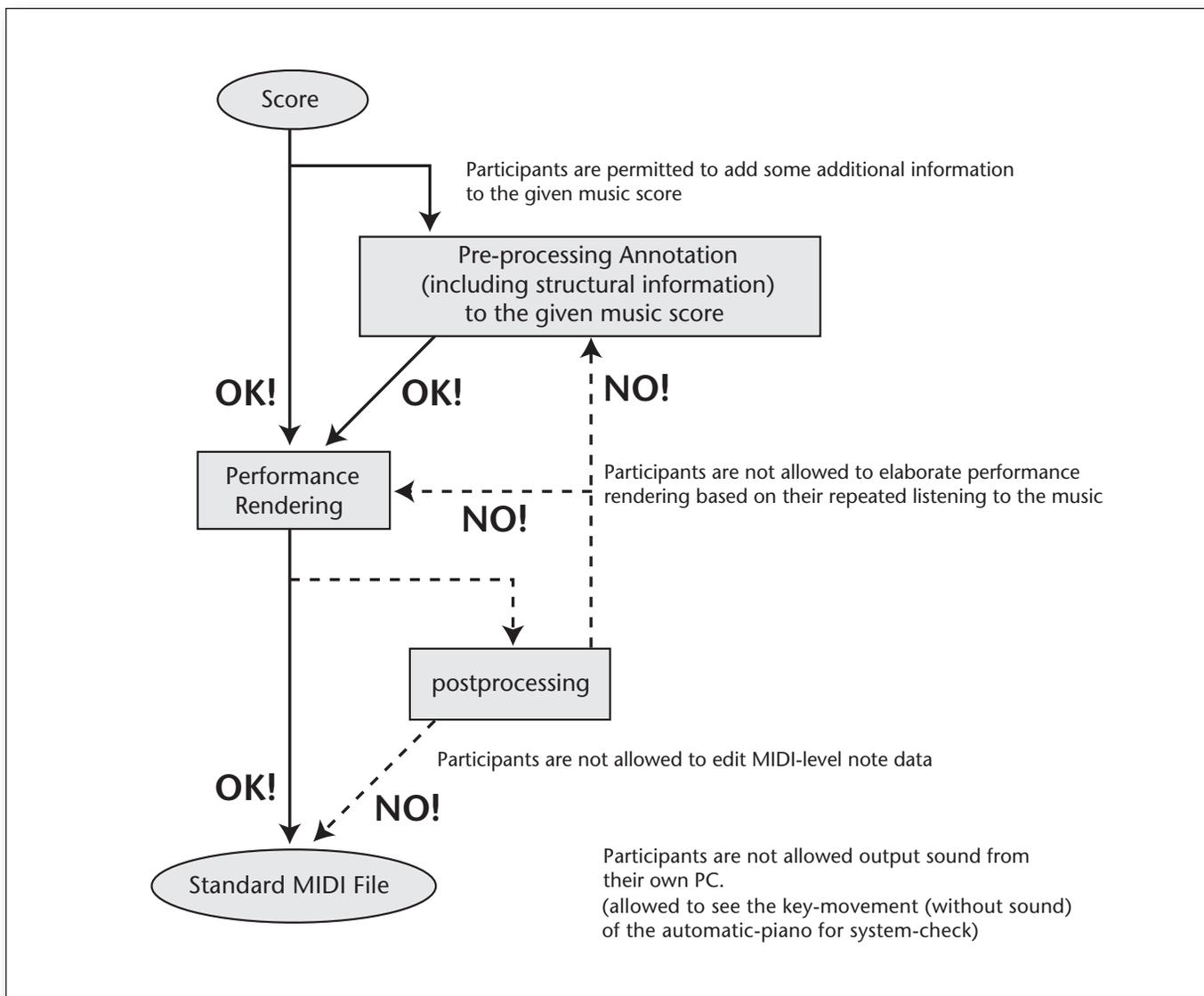


Figure 1. Graphical Summary of Instructions Given to Rencon 2008 Participants.

Based on a figure from www.renconmusic.org. Reproduced with permission.

that was given to the participants at the Rencon 2008 contest. A graphical summary of the instructions (provided by the contest organizers) can be seen in figure 1. What the videos at the website show is our computer program YQX—which was developed by Sebastian Flossmann and Maarten Grachten—performing the two previously mentioned test pieces before the contest audience. Three prizes were to be awarded:

“The Rencon Award will be given to the winner by the votes of ICMPC participants. The Rencon Technical Award will be given to the entrant whose system is judged most interesting from a technical point of view. The Rencon Murao Award will be given to the entrant that affects Prof. Murao (set piece composer) most.”³

YQX won all three of them.

This article is about how one approaches such a task—and why one approaches it in the first place, why winning Rencon Awards is not the main goal of this research, and how AI methods can help us learn more about a creative human behavior.

The Art of Expressive Music Performance

Expressive music performance is the art of shaping a musical piece by continuously varying important parameters like tempo, dynamics, and so on. Human musicians, especially in classical music, 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 *articulation* (the way successive notes are connected). Most of this is *not* specified in the written score, but at the same time it is 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, what distinguishes different performers, 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 contemporary musicology (see Gabriellson [2003], and Widmer and Goebel [2004] for comprehensive overviews of pertinent research, from different angles).

Clearly, expressive music performance is a highly creative activity. Through their playing, artists can make a piece of music appear in a totally different light, and as the history of recorded performance and the lineup of famous pianists of the 20th century show, there are literally endless possibilities of variation and new ways of looking at a masterpiece.

On the other hand, this artistic freedom is not unlimited. It is constrained by specific performance traditions and expectations of audiences (both of which may and do change over time) and also by limitations or biases of our perceptual system. Moreover, it is generally agreed that a central function of expressive playing is to clarify, emphasize, or disambiguate the structure of a piece of music, to make the audience hear a particular reading of the piece. One cannot easily play “against the grain” of the music (even though some performers sometimes seem to try to do that—consider, for example, Glenn Gould’s recordings of Mozart’s piano sonatas).

There is thus an interesting tension between creative freedom and various cultural and musical norms and psychological constraints. Exploring this space at the boundaries of creative freedom is an exciting scientific endeavor that we would like to contribute to with our research.

How YQX Works

To satisfy the reader’s curiosity as to how the Rencon 2008 contest was won, let us first give a brief explanation of how the performance rendering system YQX works. The central performance decision component in YQX (read: *Why QX?*⁴) is based on machine learning. At the heart of YQX is a simple Bayesian model that is trained on a corpus of human piano performances. Its task is to learn to predict three expressive (numerical) dimensions: timing—the ratio of the played as compared to the notated time between two successive notes in the score, which indicates either acceleration or slowing down; dynamics—the relative loudness to be applied to the current note; and articulation—the ratio of how long a note is held as compared to the note duration as prescribed by the score.

The Bayesian network models the dependency of the expressive dimensions on the local score context. The context of a particular melody note is described in terms of features like pitch interval, rhythmic context, and features based on an implication-realization (IR) analysis (based on musicologist Eugene Narmour’s [1990] theory) of the melody. A brief description of the basic YQX model and the features used is given in *The Core of YQX: A Simple Bayesian Model* elsewhere in this article.

The complete expressive rendering strategy of YQX then combines this machine-learning model with several other parts. First, a reference tempo and a loudness curve are constructed, taking into account hints in the score that concern changes in tempo and loudness (for example, loudness indications like *p* (*piano*) versus *f* (*forte*), tempo change indications like *ritardando*, and so on). Second, YQX applies the Bayesian network to predict the precise local timing, loudness, and articulation of notes from their score contexts. To this end, a simple soprano voice extraction method—we simply pick the highest notes in the upper staff of the score⁵—and an implication-realization parser (Grachten, Arcos, and Lopez de Mantaras 2004) are used. The final rendering of a piece is obtained by combining the reference tempo and loudness curves and the note-wise predictions of the network. Finally, instantaneous effects such as accents and fermatas are applied where marked in the score.

In the case of Rencon 2008, YQX was trained on two separate corpora, to be able to deal with pseudo-Chopin and pseudo-Mozart: 13 complete Mozart Piano Sonatas performed by R. Batik (about 51,000 melody notes) and a selection of 39 pieces by Chopin (including various genres such as waltzes, nocturnes, and ballades) performed by N. Magaloff (about 24,000 melody notes).

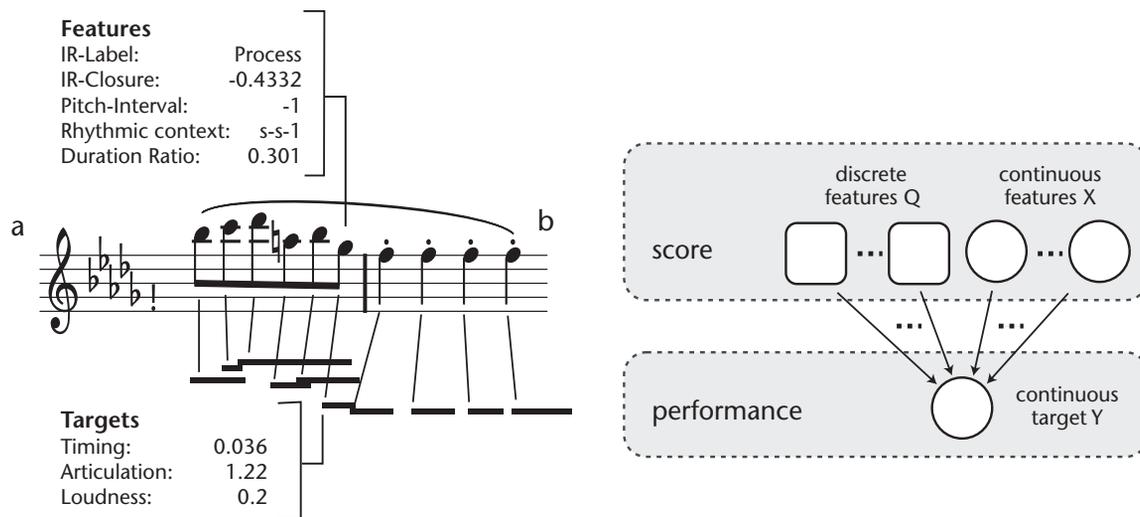


Figure 2. An Example of the Features and Target Values that Are Computed for a Given Note and the Structure of the Bayesian Model.

The Core of YQX: A Simple Bayesian Model

The statistical learning algorithm at the heart of YQX is based on a simple Bayesian model that describes the relationship between the *score context* (described through a set of simple features) and the three *target variables* (the expressive parameters timing, dynamics, articulation) to be predicted. Learning is performed on the individual notes of the *melody* (usually the top voice) of a piece. Likewise, predictions are applied to the melody notes of new pieces; the other voices are then synchronized to the expressively deformed melody.

For each melody note, several properties are calculated from the score (features) and the performance (targets). The features include the *rhythmic context*, an abstract description of the duration of a note in relation to successor and predecessor (for example, short-short-long); the *duration ratio*, the numerical ratio of durations of two successive notes; and the *pitch interval* to the following note in semitones. The implication-realization analysis (Narmour 1990) categorizes the melodic context into typical

categories of melodic contours (IR-label). Based on this a measure of musical closure (IR-closure), an abstract concept from musicology, is estimated. The corresponding targets *IOI-ratio* (tempo), *loudness*, and *articulation* are directly computed from the way the note was played in the example performance.

In figure 2 we show an example of the features and target values that are computed for a given note (figure 2a), and the structure of the Bayesian model (figure 2b). The melody segment shown is from Chopin's Nocturne Op. 9, No. 1, performed by N. Magaloff. Below the printed melody notes is a piano roll display that indicates the performance (the vertical position of the bars indicates pitch, the horizontal dimension time). Shown are the features and targets calculated for the sixth note (G_b) and the corresponding performance note. The IR features are calculated from the note itself and its two neighbors: a linearly proceeding interval sequence ("Process") and a low IR-closure value, which indicates a mid-

phrase note; the note is played legato (articulation ≥ 1) and slightly slower than indicated (IOI-Ratio ≥ 1).

The Bayes net is a simple conditional Gaussian model (Murphy 2002) as shown previously. The features are divided into sets of continuous (X) and discrete (Q) features. The continuous features are modeled as Gaussian distributions $p(x_i)$, the discrete features through simple probability tables $P(q_i)$. The dependency of the target variables Y on the score features X and Q is given by conditional probability distributions $p(y_i | Q, X)$.

The model is trained by estimating, separately for each target variable, multinomial distributions representing the joint probabilities $p(y_i, X)$; the dependency on the discrete variables Q is modeled by computing a separate model for each possible combination of values of the discrete values. (This is feasible because we have a very large amount of training data.) The actual predictions y'_i are approximated through linear regression, as is commonly done (Murphy 2002).

ICMPC-Rencon Set pieces I

My Nocturne

Tadahiro Muraio

Muraio & Rencon © 2008-2011

Figure 3. Bars 1–18 from “My Nocturne,” by Tadahiro Muraio (Aichi University of Education, Japan).
Reproduced with permission.

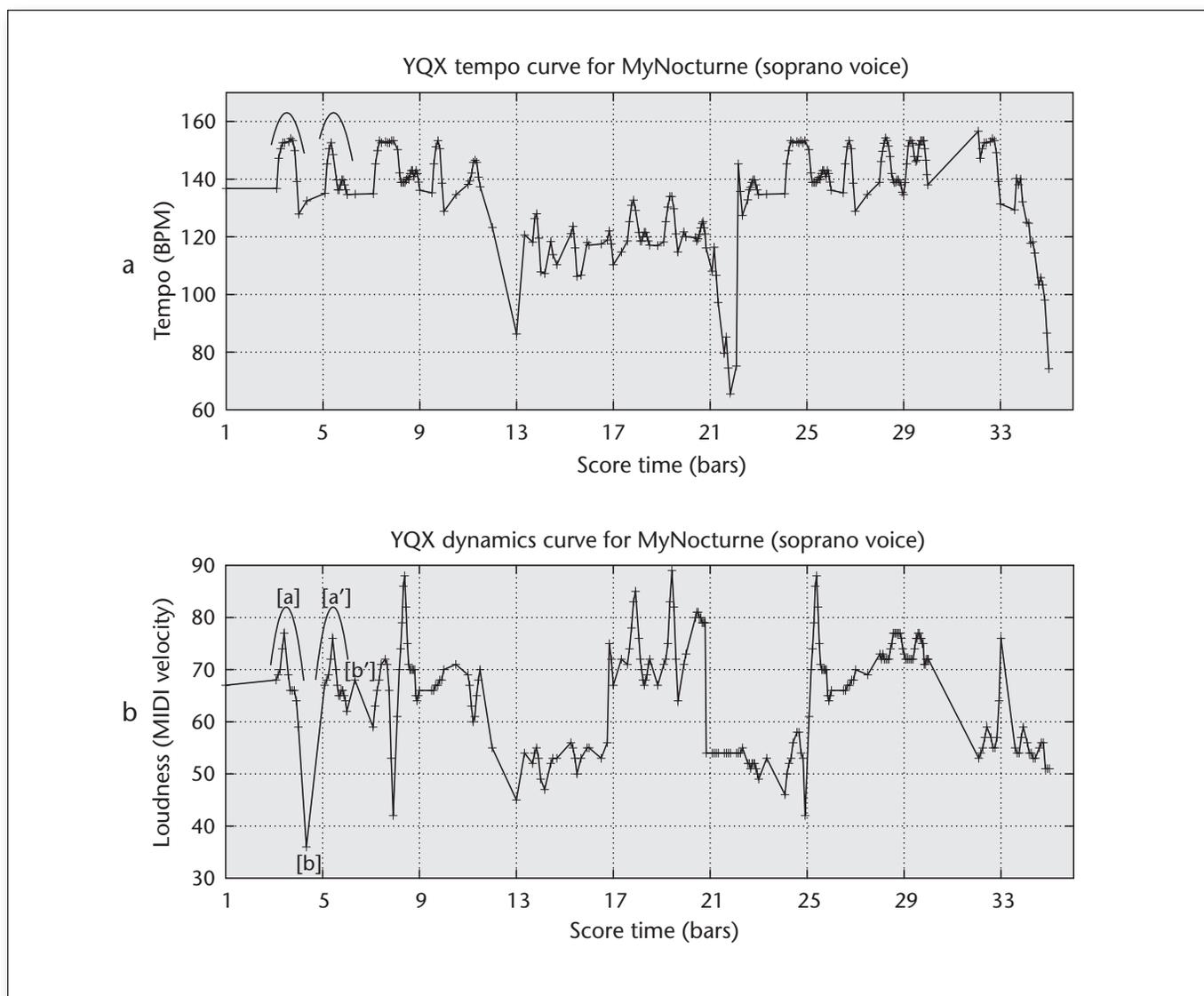


Figure 4. “My Nocturne” as Performed by YQX.

A. Tempo. B. Dynamics. The horizontal axes represent the “logical” score time, that is, the position in the piece, counted in terms of bars from the beginning. In the tempo plot, the vertical axis gives the local tempo YQX played at that point in time (in terms of beats [eighth notes] per minute). In the dynamics plot, the vertical axis gives the loudness of the current melody note (in terms of MIDI velocity). Only notes of the melody of the piece (the highest part in the right hand) are indicated. See the text for an explanation of the specially marked passages.

A Brief Look at YQX Playing a “Chopin” Piece

In order to get an appreciation of what YQX does to a piece of music, let us have a brief look at some details of YQX’s interpretation of the quasi-Chopin piece “My Nocturne,” composed by Tadahiro Murao specifically for the Rencon 2008 contest.¹ Figure 3 shows the first 18 bars of the piece. In figure 4, we show the tempo and dynamics fluctuations (relating to the melody in the right hand) of YQX’s performance over the course of the entire piece. At first sight, the tempo curve (figure 4a)

betrays the three-part structure of the piece: the middle section (bars 13–21) is played more slowly than the two outer sections, and the endings of all three sections are clearly articulated with a strong *ritardando* (an extreme slowing down). This is musically very sensible (as can also be heard in the video), but also not very surprising, as the tempo change for the middle section is given in the score, and also the *ritardandi* are prescribed (*rit.*).

More interesting is the internal structure that YQX chose to give to the individual bars and phrases. Note, for instance, the clear up-down (faster-slower) tempo contour associated with the

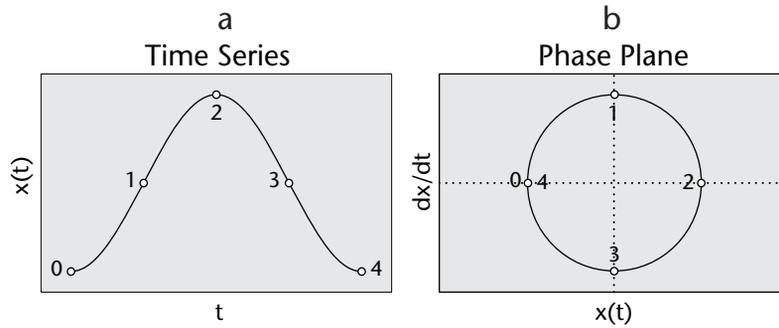


Figure 5. Schematic Relation between the Time Series Representation and the Phase-Plane Representation.

Visualizing Expressive Timing in the Phase Plane

In the phase-plane representation, tempo information measured from a musical performance is displayed in a two-dimensional plane, where the horizontal axis corresponds to tempo, and the vertical axis to the derivative of tempo. The essential difference from the common tempo-versus-time plots is that the time dimension is implicit rather than explicit in the phase plane. Instead, the change of tempo from one time point to the other (the derivative of tempo) is represented explicitly as a dimension. The abundance of names for different types of tempo changes in music (*accelerando* and *ritardando* being the most common denotations of speeding up and slowing down, respectively), proves the importance of the notion of tempo change in music. This makes the phase-plane representation

particularly suitable for visualizing expressive timing.

Figure 5 illustrates the relation between the time series representation and the phase-plane representation schematically. Figure 5a shows one oscillation of a sine function plotted against time. As can be seen in figure 5b, the corresponding phase-plane trajectory is a full circle, starting and ending on the far left side (numbers express the correspondence between points in the time series and phase-plane representations).

This observation gives us the basic understanding to interpret phase-plane representations of expressive timing. Expressive gestures, that is, patterns of timing variations used by the musician to demarcate musical units, typically manifest in the phase plane as (approximately) elliptic

curves. A performance then becomes a concatenation and nesting of such forms, the nesting being a consequence of the hierarchical nature of the musical structure. In figure 6 we show a fragment of an actual tempo curve and the corresponding phase-plane trajectory. The trajectory is obtained by approximating the measured tempo values (dots in figure 6a) by a spline function (solid curve). This function and its derivative can be evaluated at arbitrary time points, yielding the horizontal and vertical coordinates of the phase-plane trajectory. Depending on the degree of smoothing applied in the spline approximation, the phase-plane trajectory reveals either finer details or more global trends of the tempo curve.

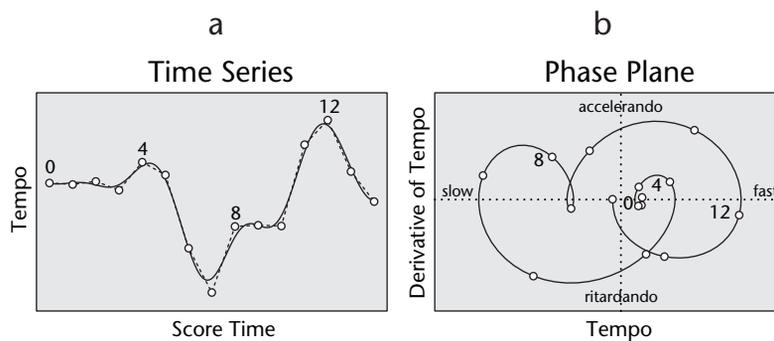


Figure 6. A Fragment of an Actual Tempo Curve and the Corresponding Phase-Plane Trajectory.

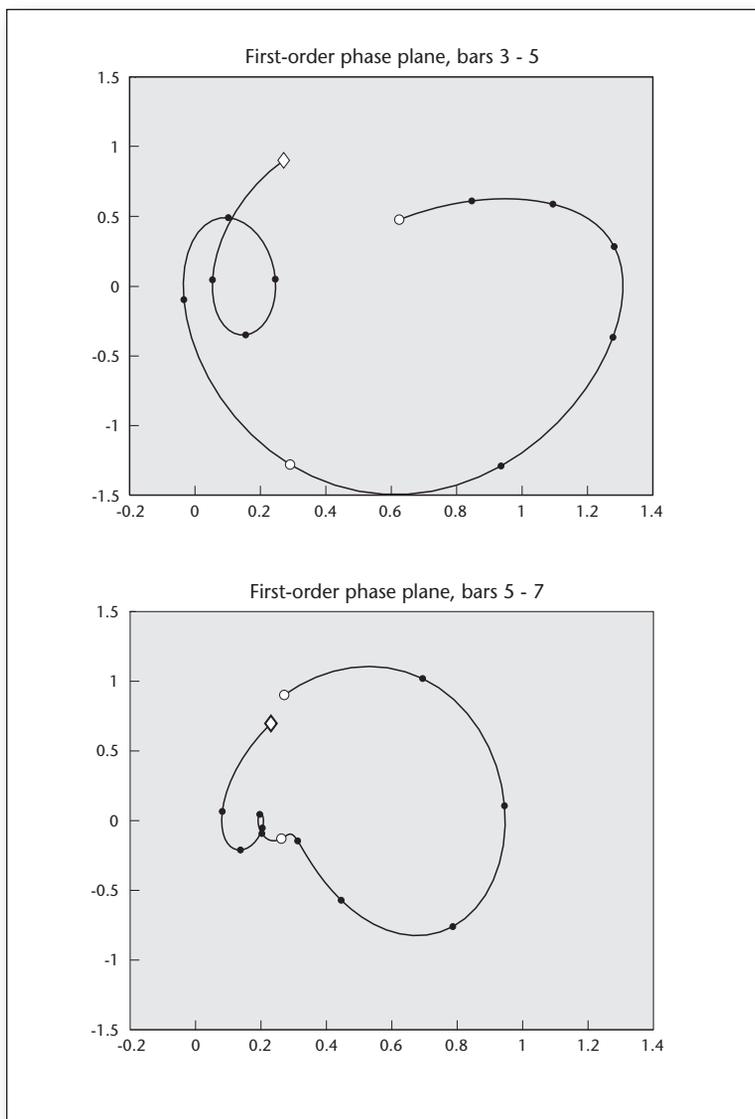


Figure 7. YQX's Expressive Timing (Tempo Changes) in the Phase Plane.

Left: Bars 3–4 of “My Nocturne.” Right: Bars 5–6. Horizontal axis: (smoothed) tempo in beats per minute (bpm). Vertical axis: First derivative of tempo (that is, amount of *accelerando* or *ritardando*). Empty circles indicate the beginnings of new bars; black dots mark beat times (each of the 6 eighth note beats within a bar); the end of a curve is marked by a diamond (to indicate the direction). The tempo and acceleration values have been normalized to give a $\mu = 0.0$ and a $\sigma = 1.0$ over the entire piece; the absolute values thus do not have a direct musical interpretation.

sixteenth-note figure in bar 3, which YQX accompanies with a *crescendo-decrescendo* in the dynamics domain (see indications [a] in figure 4). Note also the extreme softening of the last note of this phrase (the last note of bar 4), which gives a very natural ending to the phrase [b]. Both of these choices are musically very sensible. Note also how the following passage (bar 5—see [a']) is played with an analogous *crescendo-decrescendo*, though it is not identical in note content to the parallel pas-

sage [a]. Unfortunately, there is no corresponding softening at the end [b']: YQX chose to play the closing B^b in bar 6 louder and, worse, with a staccato, as can be clearly heard in the video—a rather clear musical mistake.

The tempo strategy YQX chose for the entire first section (from bar 1 to the beginning of the middle section in bar 13) deserves to be analyzed in a bit more detail. To facilitate an intuitive visual understanding of the high-level trends, we will present this analysis in the *phase plane* (see figures 7 through 9). Phase-plane plots, known from dynamical systems theory, display the behavior of a system by plotting variables describing the state of the system against each other, as opposed to plotting each variable against time. It is common in phase-plane plots that one axis represents the *derivative* of the other axis. For example, dynamical systems that describe physical motion of objects are typically visualized by plotting velocity against displacement. In the context of expressive performance, phase-plane representations have been proposed for the analysis of expressive timing (Grachten et al. 2008), because they explicitly visualize the dynamic aspects of performance. Expressive “gestures,” manifested as patterns of variation in timing, tend to be more clearly revealed in this way. The sidebar Visualizing Expressive Timing in the Phase Plane gives some basic information about the computation and meaning of smoothed timing trajectories in the phase plane. A more detailed description of the computation process and the underlying assumptions can be found in Grachten et al. (2009). We will also encounter phase-plane plots later in this article, when we compare the timing behavior of famous pianists.

Figure 7 shows how YQX chose to play two very similar passages at the beginning of the piece: the first appearance of the main theme in bars 3–4, and the immediately following variation of the theme (bars 5–6); the two phrases are marked as A1 and A2 in the score in figure 3. Overall, we see the same general tempo pattern in both cases: a big arc—a speedup followed by a slow-down (*accelerando-ritardando*)—that characterizes the first part (bar), followed by a further *ritardando* with an intermediate little local speedup (the loop in the phase plot; note that the final segment of the curve (leading to the final diamond) already pertains to the first beat of the next bar). Viewing the two corresponding passages side by side, we clearly see that they were played with analogous shapes, but in a less pronounced or dramatic way the second time around: the second shape is a kind of smaller replica of the first.

A similar picture emerges when we analyze the next pair of corresponding (though again not identical) musical phrases: phrase B1 starting in the middle of bar 9, and its immediate variation B2

starting with bar 11 (see figure 8 for YQX's renderings). Here, we see a speedup over the first part of the phrase (the agitated "gesture" consisting of six sixteenth notes in phrase B1, and the corresponding septuplet in B2), followed by a slowing down over the course of the trill (*tr*), followed by a speedup towards the beginning of the next phrase. The essential difference here is that, in repeat 2, YQX introduces a strong (and very natural-sounding) *ritardando* towards the end, because of the approaching end of the first section of the piece. To make this slowing down towards the end appear even more dramatic,⁶ the preceding speedup is proportionately stronger than in the previous phrase. Overall, we again see YQX treating similar phrases in analogous ways, with the second rendition being less pronounced than the first, and with some variation to account for a different musical context. This can be seen even more clearly in figure 9, which shows the entire passage (bars 9.5–12.5) in one plot. One might be tempted to say that YQX is seeking to produce a sense of "unity in variation."

Of course, YQX's performance of the complete piece is far from musically satisfying. Listening to the entire performance, we hear passages that make good musical sense and may even sound natural or "humanlike," but we also observe quite a number of blatant mistakes or unmusical behaviors, things that no human pianist would ever think of doing.⁷ But overall, the quality is quite surprising, given that all of this was learned by the machine.

Is YQX Creative?

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. While YQX's performances may not exactly command our admiration, they are certainly not unmusical (for the most part), they seem to betray some musical understanding, and they are sometimes surprising. Can YQX be said to be creative, then?

In our view (and certainly in the view of the general public), creativity has something to do with intentionality, with a conscious awareness of form, structure, aesthetics, with imagination, with skill, and with the ability of self-evaluation (see also Boden 1998, Colton 2008, Wiggins 2006). The previous analysis of the Chopin performance seems to suggest that YQX has a sense of unity and coherence, variation, and so on. The fact is, however, that YQX has no notion whatsoever of abstract musical concepts like structure, form, or repetition and parallelism; it is not even aware of the phrase structure of the piece. What seems to be purpose-

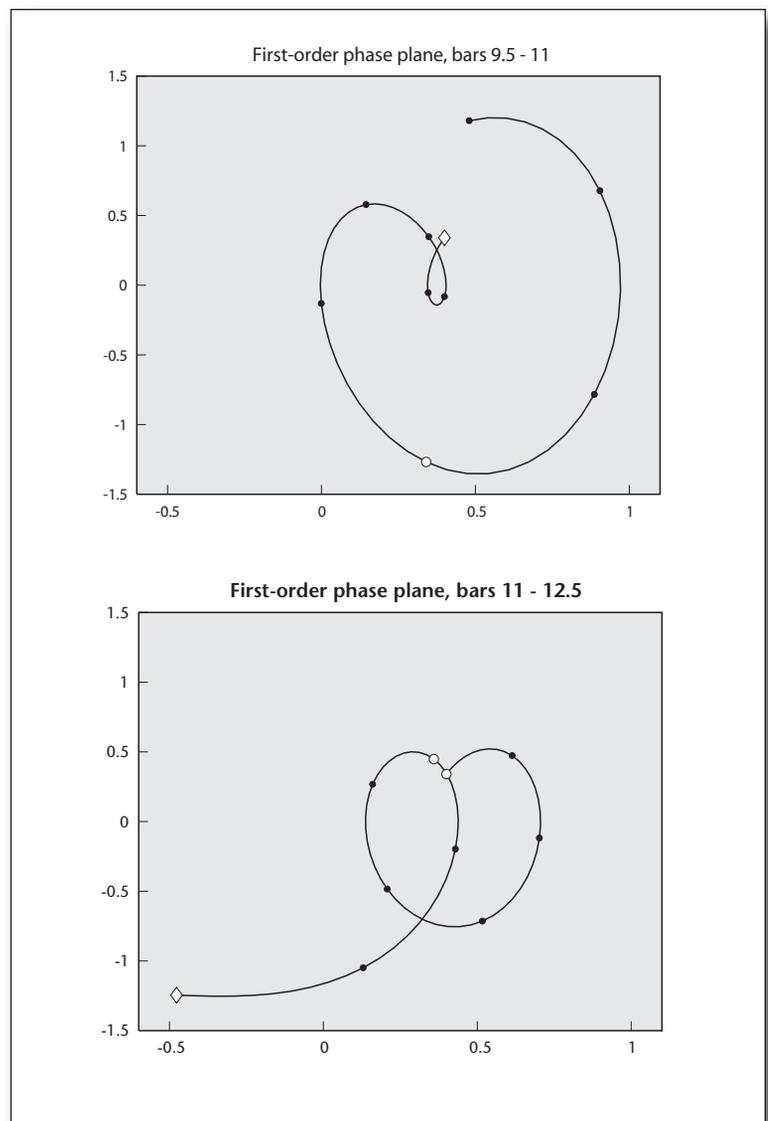


Figure 8. YQX's Expressive Timing (Tempo Changes) over Bars 9.5–11 (Left) and 11–12.5 (Right), Respectively.

ful, planned behavior that even results in an elegant artifact, is "solely" an epiphenomenon emerging from low-level, local decisions—decisions on what tempo and loudness to apply at a given point in the piece, without much regard to the larger musical context.

With YQX, we have no idea why the system chose to do what it did. Of course, it would be possible in principle to trace back the reasons for these decisions, but they would not be very revealing—they would essentially point to the existence of musical situations in the training corpus that are somehow similar to the musical passage currently rendered and that led to the Bayesian model parameters being fit in a particular way.

Other aspects often related to creativity are nov-

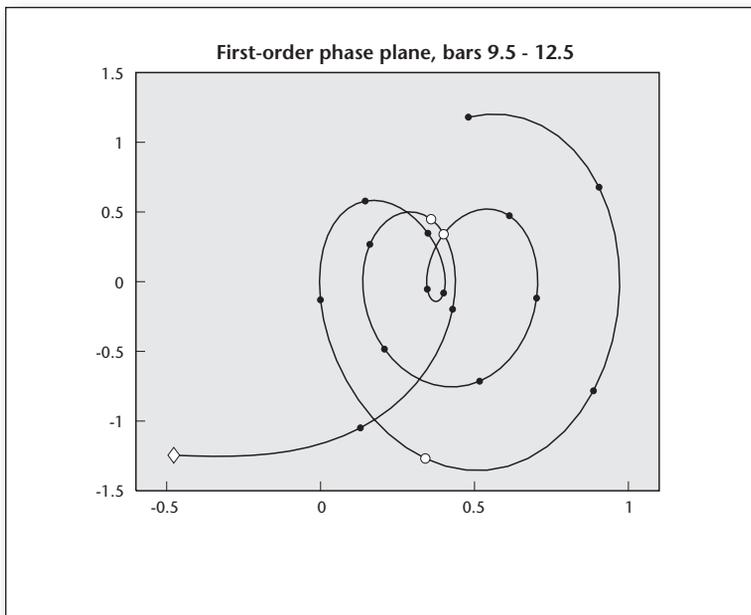


Figure 9. Overall View of YQX's Expressive Timing (Tempo Changes) over Bars 9.5–12.5.

elty, surprisingness, and complexity (for example, Boden 1998, Bundy 1994). YQX does a number of things that are not prescribed in the score, that could not be predicted by the authors, but are nevertheless musically sensible and may even be ascribed a certain aesthetic quality. YQX embodies a combination of hard-coded performance strategies and inductive machine learning based on a large corpus of measured performance data. Its decisions are indeed unpredictable (which also produced quite some nervous tension at the Rencon contest, because we were not permitted to listen to YQX's result before it was publicly performed before the audience). Clearly, a machine can exhibit unexpected (though "rational") behavior 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 behavior may emerge as a consequence of high complexity of the machine's environment, inputs, and input-output mapping.⁸ That is certainly the case with YQX.

The previous considerations lead us to settle, for the purposes of this article, on the pragmatic view that *creativity is in the eye of the beholder*. We do not claim to be experts on formal models of creativity, and the goal of our research is not to directly model the processes underlying creativity. Rather, we are interested in studying creative (human) behaviors, and the artifacts they produce, with the help of intelligent computer methods. YQX was developed not as a dedicated tool, with the purpose of

winning Rencon contests, but as part of a much larger and broader basic research effort that aims at elucidating the elusive art of expressive music performance. An early account of that research was given in Widmer et al. (2003). In the following, we sketch some of our more recent results and some current research directions that should take us a step or two closer to understanding this complex behavior.

Studying a Creative Behavior with AI Methods

As explained previously, the art of expressive music performance takes place in a field of tension demarcated by creative freedom on the one end and various musical and cultural norms and perceptual constraints on the other end. As a consequence, we may expect to find both significant *commonalities* between performances of different artists—in contexts where they all have to succumb to common constraints—and substantial *differences* in other places, where the artists explore various ways of displaying music, and where they can define their own personal artistic style.

Our research agenda is to shed some new light into this field of tension by performing focused computational investigations. The approach is data intensive in nature. We compile large amounts of empirical data (measurements of expressive timing, dynamics, and so on in real piano performances) and analyze these with data analysis and machine-learning methods.

Previous work has shown that both commonalities (that is, fundamental performance principles) and differences (personal style) seem to be sufficiently systematic and pronounced to be captured in computational models. For instance, in Widmer (2002) we presented a small set of low-level performance rules that were discovered by a specialized learning algorithm (Widmer 2003). Quantitative evaluations showed that the rules are surprisingly predictive and also general, carrying over from one pianist to another and even to music of a different style. In Widmer and Tobudic (2003), it was shown that modeling expression curves at multiple levels of the musical phrase hierarchy leads to significantly better predictions, which provides indirect evidence for the hypothesis that expressive timing and dynamics are multi-level behaviors. The results could be further improved by employing a relational learning algorithm (Tobudic and Widmer 2006) that modeled also some simple temporal context. Unfortunately, being entirely case based, these latter models produced little interpretable insight into the underlying principles.

Regarding the quest for individual style, we could show, in various experiments, that personal

style seems sufficiently stable, both within and across pieces, that it can be picked up by machine. Machine-learning algorithms learned to identify artists solely from their way of playing, both at the level of advanced piano students (Stamatatos and Widmer 2005) and famous pianists (Widmer and Zanon 2004). Regarding the identification of famous pianists, the best results were achieved with support vector machines based on a string kernel representation (Saunders et al. 2008), with mean recognition rates of up to 81.9 percent in pairwise discrimination tasks. In Madsen and Widmer (2006) we presented a method to quantify the stylistic intrapiece consistency of famous artists.

And finally, in Tobudic and Widmer (2005), we investigated to what extent learning algorithms could actually learn to replicate the style of famous pianists. The experiments showed that expression curves produced (on new pieces) by the computer after learning from a particular pianist are consistently more strongly correlated to that pianist's curves than to those of other artists. This is another indication that certain aspects of personal style are identifiable and even learnable. Of course, the resulting computer performances are a far cry from sounding anything like, for example, a real Barenboim or Rubinstein performance. Expressive performance is an extremely delicate and complex art that takes a whole lifetime of experience, training, and intellectual engagement to be perfected. We do not expect—to answer a popular question that is often posed—that computers will be able to play music like human masters in the near (or even far) future.

But we may expect to gain more insight into the workings of this complex behavior. To that end, we are currently extending both our probabilistic modeling approach and the work on expressive timing analysis in the phase plane.

Towards a Structured Probabilistic Model

Generally, the strategy is to extend YQX step by step, so that we can quantify exactly the relevance of the various components and layers of the model, by measuring the improvement in the model's prediction power relative to real performance data. While the exact model parameters learned from some training corpus may not be directly interpretable, the structure of the model, the chosen input features, and the way each additional component changes the model's behavior will provide musicologically valuable insights.

Apart from experimenting with various additional score features, two specific directions are currently pursued. The first is to extend YQX with a notion of *local temporal contexts*. To that end, we are converting the simple static model into a dynamic network where the predictions are influenced by the predictions made for the previous notes. In this way, YQX should be able to produce

much more consistent and stable results that lead to musically more fluent performances.

The second major direction for extending YQX is towards *multiple structural levels*. Western music is a highly structured artifact, with levels of groups and phrases hierarchically embedded within each other. It is plausible to assume that this complex structure is somehow reflected in the performances by skilled performers, especially since musicology tells us that one function of expressive playing (and timing, in particular) is to disambiguate the structure of a piece to the audience (for example, Clarke [1991], Palmer [1997]). Indirect evidence for this was provided by our multilevel learning experiments. In Grachten and Widmer (2007) we demonstrated in a more direct way that aspects of the phrase structure are reflected in the tempo and dynamics curves we measure in performances. Our previous multilevel model was a case-based one. We now plan to do multilevel modeling in the probabilistic framework of YQX, which we consider more appropriate for music performance (see also Grindlay and Helmbold 2006).

Towards a Computational Model of an Accomplished Artist

We are in the midst of preparing what will be the largest and most precise corpus of performance measurements ever collected in empirical musicology. Shortly before his death, the Russian pianist Nikita Magaloff (1912–1992) recorded essentially the entire solo piano works by Frédéric Chopin on the Bösendorfer SE290 computer-controlled grand piano in Vienna, and we obtained exclusive permission to use this data set for scientific studies. This is an enormous data set—many hours of music, hundreds of thousands of played notes, each of which is precisely documented (the Bösendorfer computer-controlled piano measures each individual key and pedal movement with utmost precision). It will present us with unprecedented research opportunities, and also formidable data processing challenges. The ultimate challenge would be to build a structured, probabilistic model of the interpretation style of a great pianist. It remains to be seen how far machine learning can take us toward that goal.

Toward Characterizing Interpretation Styles

Representing expressive timing in the phase plane is a new idea that needs further investigation. In a strict sense, the representation is redundant, because the derivatives (the second dimension of the plane) are fully determined by the tempo curve (the first dimension). Still, as one can demonstrate with qualitative examples (Grachten et al. 2008), the representation has certain advantages when used for visual analysis. It permits the human ana-

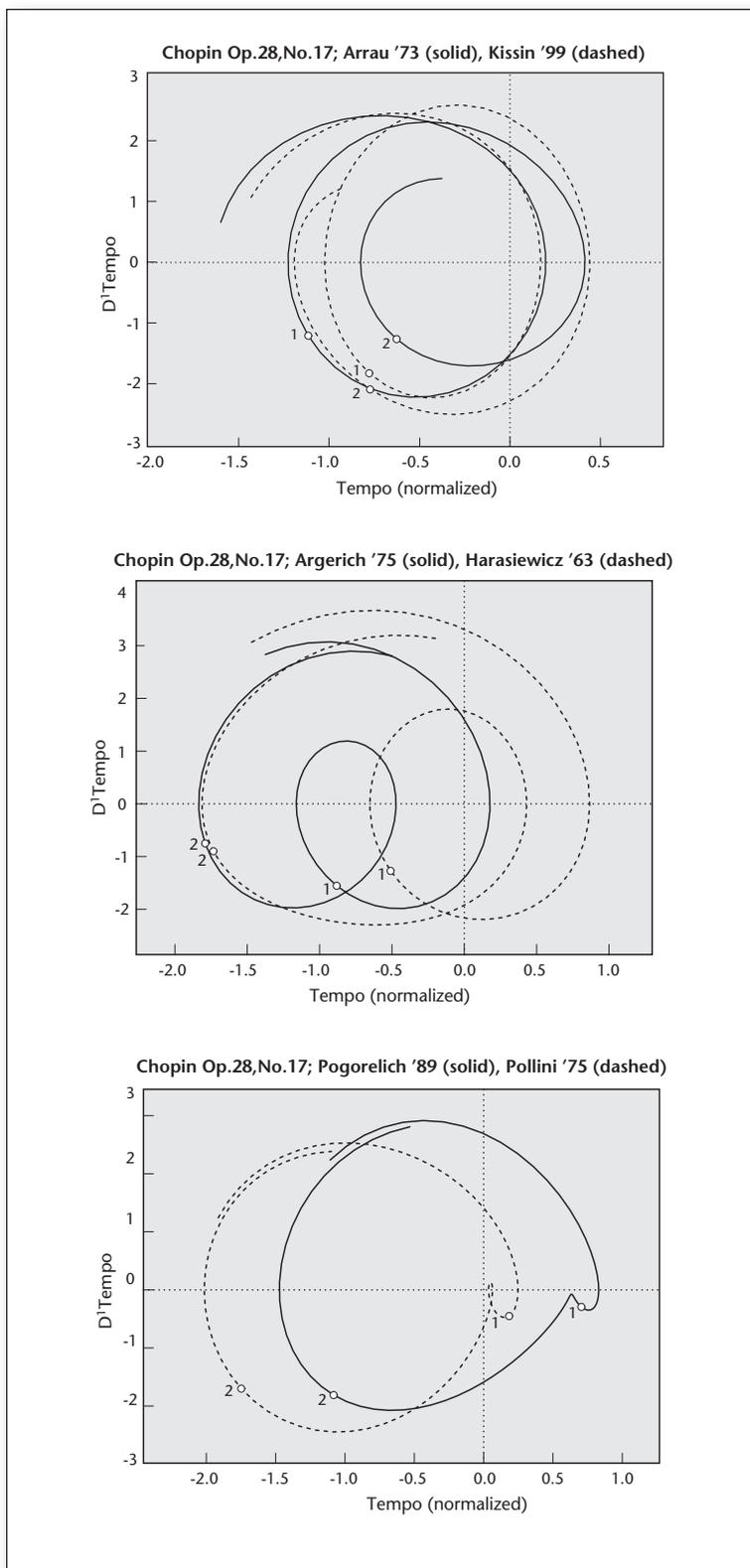


Figure 10. Three Distinct Strategies for Shaping the Tempo in Chopin's *Prélude Op. 28, No. 17*, Bars 73–74, Discovered from Phase-Plane Data.

Claudio Arrau, 1973 and Evgeny Kissin, 1999 (top left); Martha Argerich, 1975 and Adam Harasiewicz, 1963 (top right); Ivo Pogorelich, 1989 and Maurizio Pollini, 1975 (bottom).

lyst to perceive expressive “gestures” and “gestalts,” and variations of these, much more easily and directly than in plain tempo curves. Moreover, even if the second dimension is redundant, it makes information explicit that would otherwise be concealed. This is of particular importance when we use data analysis and machine learning methods for pattern discovery. In Grachten et al. (2009), a set of extensive experiments is presented that evaluate a number of alternative phase plane representations (obtained with different degrees of smoothing, different variants of data normalization, first- or second-order derivatives) against a set of well-defined pattern-classification tasks. The results show that adding derivative information almost always improves the performance of machine-learning methods. Moreover, the results suggest that different kinds of phase-plane parameterizations are appropriate for different identification tasks (for example, performer recognition versus phrase context recognition). Such experiments give us a basis for choosing the appropriate representation for a given analysis task in a principled way.

As a little example, figure 10 shows how a particular level of smoothing can reveal distinct performance strategies that indicate divergent readings of a musical passage. The figure shows three ways of playing the same musical material—bars 73–74 in Chopin's *Prélude Op. 28, No. 17*. They were discovered through automatic phase-plane clustering of recordings by famous concert pianists. For each cluster, we show two prototypical pianists. The interested reader can easily see that the pianists in the first cluster play the passage with two *accelerando-ritardando* “loops,” whereas in the second cluster, the dominant pattern is one large-scale *accelerando-ritardando* with an intermediate speedup-slowdown pattern interspersed, of which, in the third cluster, only a short lingering is left. In essence, pianists in the first cluster clearly divide the passage into two units, shaping each individual bar as a rhythmic chunk in its own right, while pianists in the third read and play it as one musical gesture. This tiny example is just meant to exemplify how an intuitive representation coupled with pattern-detection algorithms can reveal interesting commonalities and differences in expressive strategies. Systematic, large-scale pattern-discovery experiments based on different phase-plane representations will, we hope, bring us closer to being able to characterize individual or collective performance styles.

Conclusion

To summarize, AI may help us study creative behaviors like expressive music performance—or more precisely: artifacts that result from such creative

behaviors—in new ways. That concerns both the “rational,” “rulelike,” or “norm-based” aspects, and the spaces of artistic freedom where artists can develop their very personal style. State-of-the-art systems like YQX can even produce expressive performances themselves that, while neither truly high-class nor arguably very creative, could pass as the products of a mediocre music student. We take this result primarily as evidence that our models capture musically relevant principles. 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. We will continue to test our models in scientific contests such as Rencon, and leave it to the audience to decide whether they consider the results creative.

Acknowledgments

This research is supported by the Austrian Research Fund (FWF) under grant number P19349-N15. Many thanks to Professor Tadahiro Murao for permitting us to reproduce the score of his composition “My Nocturne” and to the organizers of the Rencon 2008 contest for giving us all their prizes.

Notes

1. See www.renconmusic.org/icmpc2008.
2. See www.renconmusic.org.
3. From www.renconmusic.org/icmpc2008/autonomous.htm.
4. Actually, the name derives from the three basic sets of variables used in the system’s Bayesian model: the target (expression) variables Y to be predicted, the discrete score features Q , and the continuous score features X . See The Core of YQX: A Simple Bayesian Model in this article for an explanation of these.
5. Automatic melody identification is a hard problem for which various complex methods have been proposed (for example, Madsen and Widmer [2007]), which still do not reach human levels of performance. Given the specific style of music we faced at the Rencon contest, simply picking the highest notes seemed robust enough. The improvement achievable by better melody line identification is yet to be investigated.
6. We are insinuating purposeful behavior by YQX here, which we will be quick to dismiss in the following section. The point we want to make is that it is all too easy to read too much into the “intentions” behind the decisions of a computer program—particularly so when one is using very suggestive visualizations, as we are doing here.
7. Viewing the videos, the reader will have noticed that both in the Chopin and the Mozart, YQX makes a couple of annoying octave transposition errors (in the octava (*8va*) passages). This is not an indication of creative liberties taken by the system, but rather a consequence of a last-minute programming mistake made by one of the authors during the Rencon contest.
8. One might argue that the same holds for the creativity we attribute to human beings.

References

- Boden, M. 1998. Creativity and Artificial Intelligence. *Artificial Intelligence* 103(1–2): 347–356.
- Bundy, A. 1994. What Is the Difference between Real Creativity and Mere Novelty? *Behavioural and Brain Sciences* 17(3): 533–534.
- Clarke, E. F. 1991. Expression and Communication in Musical Performance. In *Music Language, Speech and Brain*, ed. J. Sundberg, L. Nord, and R. Carlson, 184–193. London: MacMillan.
- Colton, S. 2008. Creativity Versus the Perception of Creativity in Computational Systems. In *Creative Intelligent Systems: Papers from the AAAI 2008 Spring Symposium*, 14–20. Technical Report SS-08-03. Menlo Park, CA: AAAI Press.
- Gabrielsson, A. 2003. Music Performance Research at the Millennium. *Psychology of Music* 31(3): 221–272.
- Grachten, M.; Arcos, J. L.; and Lopez de Mantaras, R. 2004. Melodic Similarity: Looking for a Good Abstraction Level. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, 210–215, Barcelona, Spain. Victoria, Canada: International Society for Music Information Retrieval.
- Grachten, M.; Goebel, W.; Flossmann, S.; and Widmer, G. 2009. Phase-Plane Representation and Visualization of Gestural Structure in Expressive Timing. *Journal of New Music Research*. In press.
- Grachten, M.; Goebel, W.; Flossmann, S.; and Widmer, G. 2008. Intuitive Visualization of Gestures in Expressive Timing: A Case Study on the Final Ritard. Paper presented at the 10th International Conference on Music Perception and Cognition (ICMPC10), Sapporo, Japan, 25–29 August.
- Grachten, M., and Widmer, G. 2007. Towards Phrase Structure Reconstruction from Expressive Performance Data. In *Proceedings of the International Conference on Music Communication Science (ICOMCS)*, 56–59. Sydney, Australia: Australian Music and Psychology Society.
- Grindlay, G., and Helmbold, D. 2006. Modeling, Analyzing, and Synthesizing Expressive Performance with Graphical Models. *Machine Learning* 65(2–3): 361–387.
- Madsen, S. T., and Widmer, G. 2007. Towards a Computational Model of Melody Identification in Polyphonic Music. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*, 495–464, Hyderabad, India. Menlo Park, CA: AAAI Press.
- Madsen, S. T., and Widmer, G. 2006. Exploring Pianist Performance Styles with Evolutionary String Matching. *International Journal of Artificial Intelligence Tools* 15(4): 495–514.
- Murphy, K. 2002. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. dissertation, Department of Computer Science, University of California, Berkeley.
- Narmour, E. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.
- Palmer, C. 1997. Music Performance. *Annual Review of Psychology* 48(1): 115–138.
- Repp, B. 1992. Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schu-



ICWSM-10 to be Held in the Washington, DC Area!

The Fourth International AAAI Conference on Weblogs and Social Media will be held in the Washington, DC area in May 2010. This interdisciplinary conference brings together researchers and industry leaders interested in creating and analyzing social media. Past conferences have included technical papers from areas such as computer science, linguistics, psychology, statistics, sociology, multimedia and semantic web technologies. A full Call for Papers will be available this fall at www.icwsm.org, and papers will be due in mid-January 2010. As in previous conferences, collections of social-media data will be provided by ICWSM-10 organizers to potential participants to encourage experimentation on common problems and datasets. For more information, please write to icwsm10@aaai.org.

mann's Träumerei. *Journal of the Acoustical Society of America* 92(5): 2546–2568.

Saunders, C.; Haroon, D.; Shawe-Taylor, J.; and Widmer, G. 2008. Using String Kernels to Identify Famous Performers from Their Playing Style. *Intelligent Data Analysis* 12(4): 425–450.

Stamatatos, E., and Widmer, G. 2005. Automatic Identification of Music Performers with Learning Ensembles. *Artificial Intelligence* 165(1): 37–56.

Tobudic, A., and Widmer, G. 2006. Relational IBL in Classical Music. *Machine Learning* 64(1–3): 5–24.

Tobudic, A., and Widmer, G. 2005. Learning to Play like the Great Pianists. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05)*, 871–876, Edinburgh, Scotland. Menlo Park, CA: International Joint Conferences on Artificial Intelligence.

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

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

Widmer, G., and Goebel, W. 2004. Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research* 33(3): 203–216.

Widmer, G., and Tobudic, A. 2003. Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies. *Journal of New Music Research* 32(3): 259–268.

Widmer, G., and Zanon, P. 2004. Automatic Recognition of Famous Artists by Machine. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'2004)*, 1109–1110. Amsterdam: IOS Press.

Wiggins, G. 2006. A Preliminary Framework for Description, Analysis and Comparison of Creative Systems. *Journal of Knowledge-Based Systems* 19(7): 449–458.

Gerhard Widmer (www.cp.jku.at/people/widmer) is professor and head of the Department of Computational Perception at the Johannes Kepler University, Linz, Austria, and head of the Intelligent Music Processing and Machine Learning 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 Wisconsin, Madison, and a Ph.D. in computer science from the University of Technology, Vienna. His research interests are in AI, machine learning, pattern recognition, 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.

Sebastian Flossmann obtained an M.Sc in computer science from the University of Passau, Germany. He currently works as a Ph.D. student at the Department for Computational Perception at Johannes Kepler University, Linz, Austria, where he develops statistical models and machine-learning algorithms for musical expression. His e-mail address is sebastian.flossmann@jku.at.

Maarten Grachten obtained an M.Sc. in artificial intelligence from the University of Groningen, The Netherlands (2001), and a Ph.D. in computer science and digital communication from Pompeu Fabra University, Spain (2006). He is presently active as a postdoctoral researcher at the Department of Computational Perception of Johannes Kepler University, Austria. His research activities are centered on computational analysis of expression in musical performances. His e-mail address is maarten.grachten@jku.at.