Learning to Play Like the Great Pianists

Asmir Tobudic
Austrian Research Institute for
Artificial Intelligence, Vienna
asmir.tobudic@ofai.at

Gerhard Widmer
Department of Computational Perception,
Johannes Kepler University, Linz
Austrian Research Institute for
Artificial Intelligence, Vienna
gerhard.widmer@jku.at

Abstract

An application of relational instance-based learning to the complex task of expressive music performance is presented. We investigate to what extent a machine can automatically build 'expressive profiles' of famous pianists using only minimal performance information extracted from audio CD recordings by pianists and the printed score of the played music. It turns out that the machine-generated expressive performances on unseen pieces are substantially closer to the real performances of the 'trainer' pianist than those of all others. Two other interesting applications of the work are discussed: recognizing pianists from their style of playing, and automatic style replication.

1 Introduction

Relational instance-based learning is a machine learning paradigm that tries to transfer the successful nearest-neighbor or instance-based learning (IBL) framework to richer first-order logic (FOL) representations [Emde and Wettschereck, 1996; Tobudic and Widmer, 2005]. As such it is a part of inductive logic programming (ILP), the field of research which – after the euphoria in the nineties – suffered a certain flattening of interest in recent years, the main reason being the difficulties of effectively constraining the extremely large hypothesis spaces. Nevertheless, some ILP systems have recently been shown to achieve performance levels competitive to those of human experts in very complex domains (e.g. [King *et al.*, 2004]).

A successful application of relational IBL to a real-world task from classical music has been presented in previous work ([Tobudic and Widmer, 2005]). A system that predicts expressive timing and dynamics patterns for new pieces by analogy to similar musical passages in a training set has been shown to learn to play piano music 'expressively' with substantial musical quality.

Here we investigate an even more interesting question: can a relational learner learn models that are 'personal' enough to capture some aspects of the playing styles of truly great pianists? A system is presented that builds performance models of six famous pianists using only crude information related to expressive timing and dynamics obtained from the pianist's CD recordings and the printed score of the music. Experiments show that the system indeed captures some aspect of the pianists' playing style: the machine's performances of unseen pieces are substantially closer to the real performances of the 'training' pianist than those of all other pianists in our data set. An interesting by-product of the pianists' 'expressive models' is demonstrated: the automatic identification of pianists based on their style of playing. And finally, the question of automatic style replication is briefly discussed.

The rest of the paper is laid out as follows. After a short introduction to the notion of expressive music performance (Section 2), Section 3 describes the data and its representation in FOL. We also discuss how the complex task of learning expressive curves can be decomposed into a well-defined instance-based learning task and shortly recapitulate the details of the relational instance-based learner DISTALL. Experimental results are presented in Section 4.

2 Expressive Music Performance

Expressive music performance is the art of shaping a musical piece by continuously varying important parameters like tempo, loudness, etc. while playing [Widmer *et al.*, 2003]. Human musicians do not play a piece of music mechanically, with constant tempo or loudness, exactly as written in the printed music score. Rather, skilled performers speed up at some places, slow down at others, stress certain notes or passages etc. The most important parameter dimensions available to a performer are *timing* (tempo variations) and *dynamics* (loudness variations). The way these parameters 'should be' varied is not specified precisely in the printed score; at the same time it is what makes a piece of music come alive, and what distinguishes great artists from each other.

Tempo and loudness variations can be represented as curves that quantify these parameters throughout a piece relative to some reference value (e.g. the average loudness or tempo of the same piece). Figure 1 shows a *dynamics curve* of a small part of Mozart's Piano Sonata K.280, 1st movement, as played by five famous pianists. Each point gives the relative loudness (relative to the average loudness of the piece by the same performer) at a given point in the piece. A purely mechanical (unexpressive) rendition of the piece would correspond to a flat horizontal line at $y = 1.0$. Tempo variations can be represented in an analogous way. In the next section we discuss how predictive models of such curves can be auto-

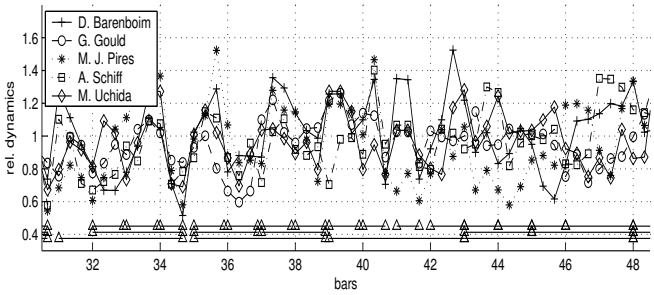

Figure 1: Dynamics curves of performances of five famous pianists for Mozart Sonata K.280, 1^{st} mvt., mm. 31–48. Each point represents the relative loudness at the beat level (relative to the average loudness of the piece by the same performer).

matically learned from information extracted from audio CD recordings and the printed score of the music.

3 Data and Methodology

The data used in this work was obtained from commercial recordings of famous concert pianists. We analyzed the performances of 6 pianists across 15 different sections of piano sonatas by W.A.Mozart. The pieces selected for analysis are complex, different in character, and represent different tempi and time signatures. Tables 1 and 2 summarize the pieces, pianists and recordings selected for analysis.

For learning tempo and dynamics 'profiles' of the pianists in our data set we extract time points from the audio recordings that correspond to *beat* [1] locations. From the (varying) time intervals between these points, the beat-level tempo and its changes can be computed. Beat-level dynamics is computed from the audio signal as the overall loudness of the signal at the beat times as a very crude representation of the dynamics applied by the pianists. Extracting such information from the CD recordings was an extremely laborious task, even with the help of an intelligent interactive beat tracking system [Dixon, 2001]. From these measurements, computing pianists' dynamics and tempo performance curves as shown in Figure 1 – which are the raw material for our experiments – is rather straightforward.

An examination of the dynamics curves in Figure 1 reveals certain trends common for all pianists (e.g. up-down, *crescendo-decrescendo* tendencies). These trends reflect certain aspects of the underlying *structure* of the piece: A piece of music commonly consists of *phrases* – segments that are heard and interpreted as coherent units. Phrases are organized hierarchically: smaller phrases are grouped into higher-level phrases, which are in turn grouped together, constituting a musical context at a higher level of abstraction etc. In Figure 1, the hierarchical, three-level phrase structure of this passage is indicated by three levels of brackets at the bottom. In this work we aim at learning expressive patterns at different levels of such a phrase structure, which roughly corresponds to various levels of musical abstraction.

[1] The beat is the time points where listeners would tap their foot along with the music.

Table 1: Mozart sonata sections selected for analysis. Section ID should be read as <sonataName> : <movement> : <section>. The total numbers of phrases are also shown.

Section ID	Tempo descr.	#phrases
kv279:1:1	fast 4/4	460
kv279:1:2	fast 4/4	753
kv280:1:1	fast 3/4	467
kv280:1:2	fast 3/4	689
kv280:2:1	slow 6/8	129
kv280:2:2	slow 6/8	209
kv280:3:1	fast 3/8	324
kv280:3:2	fast 3/8	448
kv282:1:1	slow 4/4	199
kv282:1:2	slow 4/4	254
kv283:1:1	fast 3/4	455
kv283:1:2	fast 3/4	519
kv283:3:1	fast 3/8	408
kv283:3:2	fast 3/8	683
kv332:2	slow 4/4	549

Table 2: Pianists and recordings.

ID	Pianist name	Recording
DB	Daniel Barenboim	EMI Classics CDZ 7 67295 2, 1984
RB	Roland Batik	Gramola 98701-705, 1990
GG	Glenn Gould	Sony Classical SM4K 52627, 1967
MP	Maria João Pires	DGG 431 761-2, 1991
AS	András Schiff	ADD (Decca) 443 720-2, 1980
MU	Mitsuko Uchida	Philips Classics 464 856-2, 1987

3.1 Phrase Representation in FOL

Phrases and relations between them can be naturally represented in first-order logic. In our collection of pieces, phrases are organized at three hierarchical levels, based on a manual phrase structure analysis. The musical content of each phrase is encoded in the predicate *phrCont/18*. It has the form *phrCont(Id,A1,A2,...)*, where *Id* is the phrase identifier and *A1,A2,...* are attributes that describe very basic phrase properties. The first seven of these are numeric: the length of a phrase, the relative position of the highest melodic point (the 'apex'), the melodic intervals between starting note and apex, and between apex and ending note, metrical strengths of starting note, apex, and ending note. The next three attributes are discrete: the harmonic progression between start, apex, and end, and two boolean attributes that state whether the phrase ends with a 'cumulative rhythm', and whether it ends with a cadential chord sequence. The remaining attributes describe – in addition to some simple information about global tempo and the presence of trills – global characteristics of the phrases in statistical terms: mean and variance of the durations of the melody notes within the phrase (as rough indicators of the general 'speed' of events and of durational variability), and mean and variance of the sizes of the melodic intervals between the melody notes (as measures of the 'jumpiness' of the melodic line).

This propositional representation ignores an essential aspect of the music: its temporal nature. The temporal re-

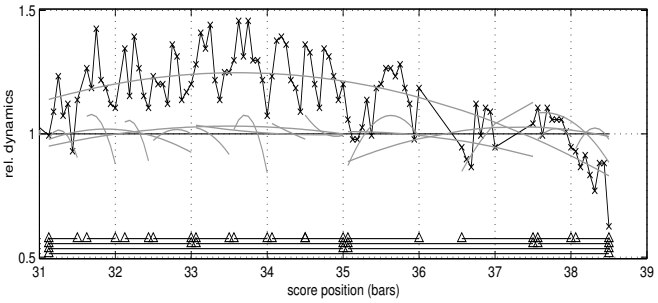

Figure 2: Multilevel decomposition of dynamics curve of performance of Mozart Sonata K.279:1:1, mm.31-38: original dynamics curve plus the second-order polynomial shapes giving the best fit at four levels of phrase structure.

lationships between successive phrases can be naturally expressed in FOL, as a relational predicate *succeeds(Id2,Id1)*, which simply states that the phrase *Id2* succeeds the same-level phrase *Id1*. Supplying the same information in a propositional representation would be very difficult.

What is still needed in order to learn are the training examples, i.e. for each phrase in the training set, we need to know how it was played by a musician. This information is given in the predicate *phrShape(Id,Coeffs)*, where *Coeffs* encode information about the way the phrase was played by a pianist. This is computed from the tempo and dynamics curves, as described in the following section.

3.2 Deriving the Training Instances: Multilevel Decomposition of Performance Curves

Given a complex tempo or dynamics curve and the underlying phrase structure (see Figure 1), we need to calculate the most likely contribution of each phrase to the overall observed expression curve, i.e., we need to decompose the complex curve into basic expressive phrase 'shapes'. As approximation functions to represent these shapes we decided to use the class of second-degree polynomials (i.e., functions of the form $y = ax^2 + bx + c$), because there is ample evidence from research in musicology that high-level tempo and dynamics are well characterized by quadratic or parabolic functions [Todd, 1992]. Decomposing a given expression curve is an iterative process, where each step deals with a specific level of the phrase structure: for each phrase at a given level, we compute the polynomial that best fits the part of the curve that corresponds to this phrase, and 'subtract' the tempo or dynamics deviations 'explained' by the approximation. The curve that remains after this subtraction is then used in the next level of the process. We start with the highest given level of phrasing and move to the lowest. As tempo and dynamics curves are lists of multiplicative factors (relative to a default tempo), 'subtracting' the effects predicted by a fitted curve from an existing curve simply means dividing the y values on the curve by the respective values of the approximation curve.

Figure 2 illustrates the result of the decomposition process on the last part (mm.31–38) of the Mozart Sonata K.279, 1st movement, 1st section. The four-level phrase structure our music analyst assigned to the piece is indicated by the four levels of brackets at the bottom of the plot. The elementary phrase shapes (at four levels of hierarchy) obtained after decomposition are plotted in gray. We end up with a training example for each phrase in the training set — a predicate $phrShape(Id, Coeff)$, where $Coeff = \{a, b, c\}$ are the coefficients of the polynomial fitted to the part of the performance curve associated with the phrase.

Input to the learning algorithm are the (relational) representation of the musical scores plus the training examples (i.e. timing and dynamics polynomials), for each phrase in the training set. Given a test piece the learner assigns the shape of the most similar phrase from the training set to each phrase in the test piece. In order to produce final tempo and dynamics curves, the shapes predicted for phrases at different levels must be combined. This is simply the inverse of the curve decomposition problem: Given a new piece, the system starts with an initial 'flat' expression curve (i.e., a list of 1.0 values) and then successively multiplies the current values by the multi-level phrase predictions.

3.3 DISTALL, a Relational Instance-based Learner

We approach phrase-shape prediction with a straightforward nearest-neighbour (NN, IBL) method. Standard propositional k-NN is not applicable to our data representation discussed in section 3.1. Instead, we use DISTALL, an algorithm that generalizes propositional k-NN to examples described in first-order logic [Tobudic and Widmer, 2005].

DISTALL is a representative of the line of research first initiated in [Bisson, 1992], where a clustering algorithm together with its similarity measure was presented. This work was later improved in [Emde and Wettschereck, 1996], in the context of the relational instance-based learning algorithm RIBL, which in turn can be regarded as DISTALL's predecessor. We skip technical details here, but the main idea behind DISTALL's similarity measure is that the similarity between two objects depends not only on the similarities of their attributes, but also on the similarities of the objects related to them. The similarities of the related objects depend in turn on their attributes and related objects. For our music learning task it means that the 'shaping' of the current (test) phrase depends not only on its attributes, but also on the preceding and succeeding music (through the relation *succeeds(Id1,Id2)*, see section 3.1), which is – from a musical point of view – a rather intuitive idea. For a more detailed description of DISTALL see [Tobudic and Widmer, 2005].

Experimental results with DISTALL on MIDI-like (i.e., very detailed) performance data produced by a local pianist are reported in [Tobudic and Widmer, 2005]. The new contribution of the current paper is that we have laboriously measured audio recordings by truly famous artists and can show — for the first time — that DISTALL actually succeeds in capturing something of personal artistic performance style.

4 Experiments

4.1 Learning Predictive Performance Models

For each pianist, we conducted a systematic *leave-one-piece-out* cross-validation experiment: each of 15 pieces was once

Table 3: Results of piecewise cross-validation experiment. The table cells list correlations between learned and real curves, where rows indicate the 'training pianist', and columns the pianist whose real performance curves are used for comparison. The correlations are averaged over all pieces, weighted by the relative length of the piece. Each cell is further divided into two rows corresponding to *dynamics* and *tempo* correlations, respectively. The highest correlations in each row are printed in bold.

	compared with					
learned from	DB	RB	GG	MP	AS	MU
DB	**.44**	.21	.26	.34	.38	.28
	.44	.27	.26	.32	.31	.31
RB	.21	**.32**	.09	.19	.19	.17
	.28	**.42**	.20	.22	.30	.27
GG	.25	.09	**.36**	.19	.21	.22
	.25	.18	**.32**	.23	.29	.28
MP	.33	.19	.19	**.39**	.33	.28
	.31	.23	.27	**.38**	.28	.34
AS	.36	.17	.20	.31	**.40**	.26
	.32	.29	.28	.25	**.41**	.32
MU	.27	.18	.21	.28	.26	**.38**
	.34	.30	.32	.36	.37	**.50**

Table 4: Detailed results (tempo dimension) of the cross-validation experiment with Mitsuko Uchida was the 'training' pianist. For each piece, the correlations between predicted curve and actual tempo curves from all pianists are given. The average over all pieces is given in the last row (reproduced from the last row of Table 3)

Piece	DB	RB	GG	MP	AS	MU
kv279:1:1	.36	.31	.22	.36	.29	**.50**
kv279:1:2	.31	.35	.35	.28	.34	**.47**
kv280:1:1	.40	.25	.26	.55	.42	**.78**
kv280:1:2	.42	.37	.42	.37	.47	**.54**
kv280:2:1	.53	.34	.52	.58	.59	**.69**
kv280:2:2	.44	.13	.39	.42	.50	**.58**
kv280:3:1	.72	.66	.62	**.82**	.68	.77
kv280:3:2	.51	**.52**	.44	.49	.43	.51
kv282:1:1	.38	.45	.36	.44	.44	**.72**
kv282:1:2	.23	.36	.33	.42	.46	**.52**
kv283:1:1	.21	.10	.12	.16	.22	**.31**
kv283:1:2	.17	.18	.21	.23	.22	**.32**
kv283:3:1	.30	.28	.42	.42	.42	**.52**
kv283:3:2	.19	.15	.29	.21	.32	**.36**
kv332:2	.32	.17	.14	.22	.16	**.35**
Total	.34	.30	.32	.36	.37	**.50**

left aside as a test piece while the remaining 14 performances (by the same pianist) were used for learning. DISTALL's parameter for the number of nearest neighbors was set to 1, and the parameter for the depth of starting clauses (see [Tobudic and Widmer, 2005]) to 4 (meaning that the distance between two phrases can be influenced by at most 4 preceding and 4 succeeding phrases).

The expressive shapes for each phrase in a test piece are predicted by DISTALL and then combined into a final tempo and dynamics curve, as described in section 3.2. The resulting curves are then compared to the real performance curves of all pianists (for the same test piece). If the curve learned from the performances of one pianist is more similar to the real performance curve of the 'teacher' pianist than to those of all other pianists, we could conclude that the learner succeeded in capturing something of the pianist's specific playing style. The described procedure is repeated for all pieces and all pianists in our data set.

Correlation is chosen as a measure of how well the predicted curve 'follows' the real one. The curves are first normalized so that their autocorrelations are identically 1, giving a correlation estimate between curves as a number in the range [-1,1]. The results of the cross-validation experiment averaged over all pieces (weighted by the relative length of the pieces) are given in Table 3.

Interestingly, the system succeeded in learning curves that are substantially closer to the 'trainer' than all others, for all pianists. Some of the pianists are better 'predictable' than others, e.g. Daniel Barenboim and Mitsuko Uchida, which might indicate that they play Mozart in a more 'consistent' way. While at first sight the correlations may not seem impressive, one should keep in mind that artistic performance is far from predictable, and the numbers in Table 3 are averages over all pieces (about half an hour of concert-level piano music per artist). Moreover, the correlation estimates in Table 3 are somewhat unfair, since we compare the performance curve produced by composing the polynomials predicted by the learner, with the curve corresponding to the pianists' *actual* performances. However, what DISTALL learned from was not the actual performance curves, but an *approximation* curve which is implied by the three levels of quadratic functions that were used as training examples. Correctly predicting these is the best the learner could hope to achieve.

Table 4 shows a more detailed picture of the cross-validation experiment, where the training pianist was Mitsuko Uchida and the numbers refer to correlations in the tempo domain. In 13 out of 15 cases the learner produces tempo curves which are closer to Uchida's playing than to any other pianist, with correlations of .7 and better (e.g., for kv280:3:1 and kv280:1:1). The results are even more interesting if we recall that the learner is given a very crude, beat-level representation of the tempo and dynamics applied by the pianist, without any details about e.g. individual voices or timing details below the level of beat. On the other hand, the piecewise results revealed that some of the pianists (e.g. Gould or Pires) seem to be less 'predictable' with our approach than Uchida (not reported here for lack of space).

Figure 3 shows an example of successful performance style learning. We see a passage from a Mozart piano sonata as 'played' by the computer after learning from recordings of other pieces by Daniel Barenboim (top) and Mitsuko Uchida (bottom), respectively. Also shown are the performance curves corresponding to these two pianists' actual performances of the test piece. In this case it is quite clearly visible that the curves predicted by the computer on the test piece are much more similar to the curves by the respective 'teacher' than to those by the other pianist.

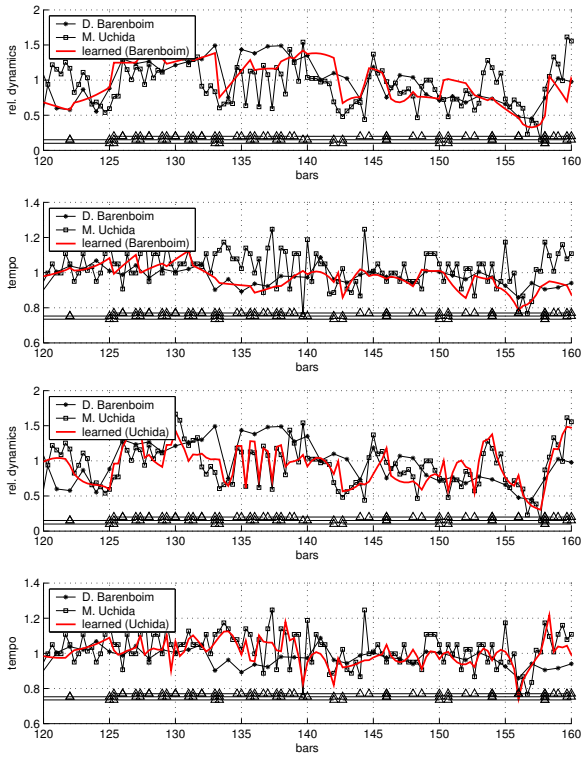

Figure 3: Dynamics and tempo curves produced by DISTALL on test piece (Sonata K.283, 3^{rd} mvt., 2^{nd} section, mm.120–160) after learning from Daniel Barenboim (top) and Mitsuko Uchida (bottom), compared to the artists' real curves as measured from the recordings.

Admittedly, this is a carefully selected example, one of the clearest cases of style replication we could find in our data. The purpose of this example is more to give an indication of the complexity of the curve prediction task and the difference between different artists' interpretations than to suggest that a machine will always be able to achieve this level of prediction performance.

4.2 Identification of Great Pianists

The primary goal of our work is learning predictive models of pianists' expressive performances.[2] But the models can also be used in a straightforward way for recognizing pianists. The problem of identifying famous pianists from information obtained from audio recordings of their playing has been addressed in the recent literature [Saunders *et al.*, 2004; Stamatatos and Widmer, 2002; Widmer and Zanon, 2004]. In [Widmer and Zanon, 2004], a number of low-level scalar features related to expressive timing and dynamics are extracted from the audio CD recordings, and various machine learning algorithms are applied to these. In [Saunders *et al.*, 2004], the sequential nature of music is addressed by representing performances as strings and using string kernels in conjunction with kernel partial least squares and support vector machines. The string kernel approach is shown to achieve better performance than the best results obtained in [Widmer and Zanon, 2004]. A clear result from both works is that identification of pianists from their recordings is an extremely difficult task.

[2]Note that learned tempo and dynamics curves as produced by our system can be used to build truly machine generated expressive performances: using the predicted tempo and dynamics curves (i.e. relative tempo and dynamics for each beat in the piece), it is straightforward to calculate tempo and dynamics for each note in the piece (e.g. by interpolation).

Table 5: Confusion matrix of the pianist classification experiment. Rows correspond to the test performances of each pianist (15 per row), columns to the classifications made by the system. The rightmost column gives the accuracy achieved for all performances of the respective. The *baseline accuracy* in this 6-class problem is 16.67%.

	prediction						
pianist	DB	RB	GG	MP	AS	MU	Acc.[%]
DB	11	0	0	2	2	0	73.3
RB	1	12	1	0	0	1	80.0
GG	1	1	10	0	0	3	66.7
MP	0	0	1	12	0	2	80.0
AS	1	0	2	0	10	2	66.7
MU	0	0	1	0	0	14	93.3
Total	-	-	-	-	-	-	**76.7**

The pianists studied in the present paper are identical to those in [Widmer and Zanon, 2004] and [Saunders *et al.*, 2004]; unfortunately, the sets of recordings differ considerably (because manual phrase structure analyses, which are needed in our approach, were available only for certain pieces), so a direct comparison of the results is impossible. Still, to illustrate what can be achieved with a relational representation and learning algorithm, we briefly describe a classification experiment with DISTALL.

Each of the 15 pieces is set aside once. The 6 performances of that piece (one by each pianist) are used as test instances. A model of each pianist is built from his/her performances of the remaining 14 pieces. The result is two predicted curves per pianist for the test piece (for tempo and dynamics), which we call model curves. The final classification of a pianist, represented by his/her tempo and dynamics curves t_t and t_d on the test piece, is then determined as

$$c(t_t, t_d) = argmax_{p \in P}\left(\frac{corr(t_t, m_{pt}) + corr(t_d, m_{pd})}{2}\right) \quad (1)$$

where P is set of all pianists and m_{pt} and m_{pd} are the pianists' model tempo and dynamics curves. In other words, the performance is classified as belonging to the pianist whose model curves exhibit the highest correlation (averaged over tempo and dynamics) with the test curves. For each pianist, DISTALL is tested on the 15 test pieces, which gives a total number of 90 test performances. The *baseline* accuracy – the success rate of pure guessing – is 15, or 16.67%. The confusion matrix of the experiment is given in Table 5.

Again, it turns out that the artists are identifiable to varying degrees, but the recognition accuracies are all clearly above the baseline. In particular, note that the system correctly iden-

tifies performances by Uchida in all but one case. Obviously, the learner succeeds in reproducing something of the artists' styles in its model curves. While these figures seem to compare very favourably to the accuracies reported in [Widmer and Zanon, 2004] and [Saunders *et al.*, 2004], they cannot be compared directly, because different recordings were used and, more importantly, the level of granularity of the training and test examples are different (movements in [Widmer and Zanon, 2004; Saunders *et al.*, 2004] vs. sections in this paper), which probably makes our learning task easier.

4.3 Replicating Great Pianists?

Looking at Figure 3, one might be tempted to consider the possibility of automatic style replication: wouldn't it be interesting to supply the computer with the score of a new piece and have it perform it 'in the style of', say, Vladimir Horowitz or Arthur Rubinstein? This question is invariably asked when we present this kind of research to the public. Unfortunately (?), the answer is: while it might be interesting, it is not currently feasible.

For one thing, despite the huge effort we invested in measuring expressive timing and dynamics in recordings, the amount of available training data is still vastly insufficient vis-a-vis the enormous complexity of the behaviour to be learned. And secondly, the sort of crude beat-level variations in tempo and general loudness capture only a very small part of what makes an expressive interpretation; essential details like articulation (e.g., staccato vs. legato), pedalling, the shaping of individual voices, etc. are missing (and will be very hard to measure from audio recordings at all). A computer performance based only on applying these beat-level tempo and loudness changes will not sound anything like a human performance, as can be readily verified experimentally. Thus we have to admit that the title we chose for this paper is a bit pretentious: the computer cannot be expected to play like the great pianists – at least not given the current methods and available training data. It can, however, extract aspects of personal style from recordings by great pianists, as has been shown in the our experiments.

5 Conclusion

An application of relational instance-based learning to a difficult task from the domain of classical music was presented: we showed how the problem of learning models of expressive piano performance can be reduced to applying simple expressive phrase-patterns by analogy to the most similar phrases in the training set. In particular, it was shown that a relational learner like DISTALL succeeds in learning performance strategies that obviously capture aspects individual artistic style, which was demonstrated in learning and artist classification experiments.

The ultimate goal of this kind of research is not automatic style replication or the creation of artificial performers, but to use computers to teach us more about the elusive artistic activity of expressive music performance. While it is satisfying to see that the computer is indeed capable of extracting information from performance measurements that seems to capture aspects of individual style, this can only be a first step. In order to get real insight, we will need learning algorithms that, unlike nearest-neighbor methods, produce interpretable models. That is a natural next step in our future research.

Acknowledgments

This research is supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung (FWF)* (project no. Y99-INF). The Austrian Research Institute for Artificial Intelligence acknowledges basic financial support by the Austrian Federal Ministry for Education, Science, and Culture, and the Federal Ministry of Transport, Innovation, and Technology. Thanks to Werner Goebl for performing the harmonic and phrase structure analysis of the Mozart sonatas.

References

[Bisson, 1992] Gilles Bisson. Learning in FOL with a similarity measure. In *Proceedings of the 10th AAAI*, 1992.

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

[Emde and Wettschereck, 1996] Werner Emde and Dietrich Wettschereck. Relational instance-based learning. In *Proceedings of the 13th International Conference on Machine Learning*, 1996.

[King *et al.*, 2004] R.D. King, K.E. Whelan, F.M. Jones, P.G.K. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, and S.G. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971):247–252, 15 January 2004.

[Saunders *et al.*, 2004] Craig Saunders, David R. Hardoon, John Shawe-Taylor, and Gerhard Widmer. Using string kernels to identify famous performers from their playing style. In *Proceedings of the 15th European Conference on Machine Learning*, 2004.

[Stamatatos and Widmer, 2002] Efstathios Stamatatos and Gerhard Widmer. Music performer recognition using an ensemble of simple classifiers. In *Proceedings of the 15th European Conference on Artificial Intelligence*, 2002.

[Tobudic and Widmer, 2005] Asmir Tobudic and Gerhard Widmer. Relational IBL in classical music. *Machine Learning*, to appear, 2005.

[Todd, 1992] McAngus N. Todd. The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91(6):3540–50, 1992.

[Widmer and Zanon, 2004] Gerhard Widmer and Patrick Zanon. Automatic recognition of famous artists by machine. In *Proceedings of the 17th European Conference on Artificial Intelligence*, 2004.

[Widmer *et al.*, 2003] Gerhard Widmer, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic. In search of the Horowitz factor. *AI Magazine*, 24(3):111–130, 2003.