A Machine Learning Analysis of Expressive Timing in Pianists' Performances of Schumann's "Träumerei"

Gerhard Widmer

Department of Medical Cybernetics and Artificial Intelligence, University of Vienna, and
Austrian Research Institute for Artificial Intelligence,
Schottengasse 3, A−1010 Vienna, Austria
gerhard@ai.univie.ac.at

Abstract

The paper describes a recent attempt at reconstructing, by means of machine learning techniques, expressive performance skills from examples of real musical performances. An inductive machine learning algorithm is used to analyze the expressive timing (*rubato*) patterns in some actual performances by various famous pianists of Robert Schumann's "Träumerei" (from "Kinderszenen", op.15). Two approaches based on the same learning algorithm, but using different vocabularies for describing example performances and formulating the rules are described. The experimental results are quite interesting and instructive, but they also point to some rather serious limitations of the available data collection.

1 Introduction

The paper describes a recent attempt at automatically reconstructing, by means of machine learning techniques, expressive performance skills from examples of real musical performances. The work represents another step in the "analysis by resynthesis" research programme to the study of musical expression (Widmer, 1994b). "Analysis by resynthesis" in this context means that we develop computer programs that analyze examples of human performances, learn general expression rules from these, and test the learned rules by applying them to new pieces to produce expressive performances. In addition to learning about the general learnability of the skill (or art) of musical expression, one may also hope to gain new musical insight through an analysis of the explicit expression rules generated by the inductive learning programs. This may be seen as complementary to Sundberg et al.'s (1983) "analysis by synthesis" approach, where a set of expression rules are postulated (based on musical intuition, experience, and various insights from music theory), and their behaviour and adequacy is then tested by applying them to new pieces, and where cycles of testing and analysis can lead to refinement of the rules and tuning of their parameters.

The focus of the experiments to be reported is Robert Schumann's romantic piano piece "Träumerei" (from "Kinderszenen", op. 15). We will present a machine learning analysis of the expressive timing (*rubato*) patterns in some real performances of this piece by various pianists. Precise measurements of these performances were collected by Repp (1992). Repp himself used the data for an extensive statistical analysis that attempted to identify commonalities and differ-

ences between the pianists' performances and styles. The goal of our machine learning analysis, on the other hand, is to discover music-structural criteria that might govern or 'explain' regularities in expressive timing and that, if formulated explicitly in the form of rules, would allow the computer to 'predict' or determine an appropriate rubato structure for new pieces. Thus, one could say that while statistical studies like Repp's are *descriptive* in nature, machine learning studies like the one described here aim at *constructive* or *explanatory* results.

More precisely, we will present two sets of experiments based on the same inductive learning algorithm, but using different vocabularies for describing example performances and formulating the rules. Empirical tests of the learned rules will produce quite interesting musical results, but eventually we will find that the general results of this preliminary study are rather inconclusive. While the rules extracted from the example performances do allow the system to produce sensible performances of an unseen test piece, the rules themselves do not seem to capture principles of expressive timing that are truly general. This result seems due mainly to the specific limitations of the available data collection. Sections 4 and 5 will discuss this in more detail.

2 The data

Bruno Repp (1992) has assembled a sizeable collection of empirical performance data relating to Robert Schumann's *Träumerei*. He measured the exact timing (note onset times) in 28 performances by 24 well-known pianists, down to a resolution of about 2 milliseconds. A list of the pianists and the recordings used, as well as the score and a structural (melodic/rhythmic) analysis of the piece can be found in (Repp, 1992). The measurements are in the form of lists of *interonset intervals (IOIs)* that specify the absolute time in milliseconds between the onset of the major melody tones (usually the notes of the 'soprano'). The other voices were largely ignored.

This data set was used as the basis for our experiments. Note that Repp's data only capture the dimension of *expressive timing (rubato)*, dynamics was not taken into account. Dynamics data could have been handled in an analogous way by our system, had they been available. Also, while Repp's measurements reflect not only the relative timing deviations, but also the *absolute tempo* of the 28 performances (which differed wildly between individual pianists), the dimension of absolute tempo was disregarded in our experiments. All the tempo curves in the following sections relate to a hypothetical 'average' global tempo.

From a machine learning point of view, it would have been more desirable to have performances of several different pieces rather than different performers playing the same piece. The data are well suited to the kind of statistical similarity/difference investigations performed by Repp. But the nature of our learning task, which is to discover rules that make timing decisions based on structural aspects of musical situations, requires that a large and diverse collection of musical situations be available for learning, if the resulting rules are to be general and reliable. The set of distinct musical patterns contained in just one (rather short) piece of music is very limited, so the rules that can be extracted from these examples will tend to be highly specific to the particular training piece. We will return to this problem later.

3 Experimental setting

3.1 The machine learning scenario

The general scenario in our approach is as follows: Expressive performances by musicians are collected, represented in the computer in some symbolic form, and submitted to an inductive

machine learning algorithm as examples of 'correct' or 'sensible' interpretations. More precisely, input to the learning algorithm are the notes of pieces (currently only the melodies) as given in the score, along with tempo curves representing the expressive timing deviations applied by a performer. Each note is described in terms of various intrinsic properties (such as name, pitch, duration), simple relations between the note and its predecessor and successor notes (e.g., interval, direction of interval), and some higher-level descriptors that describe the roles that the note plays in various structural dimensions (e.g., metrical strength, relative position in the grouping or phrase structure, etc.).

The learner's task is to extract from these examples a set of general rubato or timing rules that specify general conditions for when to apply, say, an accelerando or a ritardando. In addition, the algorithm must learn to determine the precise numeric degree of accelerando or ritardando to be applied. The learned rules can be used to compute tempo curves for new pieces. These can then be analyzed graphically, played and listened to, and the rules themselves are also open to inspection and analysis.

3.2 The learning algorithm

A new learning algorithm by the name of IBL-Smart had to be developed for this type of learning scenario. The algorithm basically integrates a symbolic and a numeric generalization strategy. The symbolic component learns explicit rules that determine the appropriate classification of some note or unit in some piece of music (e.g., whether a particular note should be played longer or shorter than notated), and the numeric part is an instance-based learning algorithm (Aha et al., 1991) that in effect builds up numeric interpolation tables for each learned symbolic rule to predict precise numeric values. The details of the algorithm cannot be discussed here, the reader is referred to (Widmer, 1993) for a detailed presentation.

Output of the learning algorithm is then a set of symbolic decision rules, each associated with numeric interpolation tables that determine the exact expression values for each specific situation.

For the present experiments, the inductive learning algorithm FOIL (Quinlan, 1990) was used for the symbolic learning part. The rules produced by FOIL are standard PROLOG clauses and can be directly applied to new pieces by the problem solving component.

3.3 Training and test data

Common machine learning practice dictates that the available data be split into a training set, which is given to the inductive learning algorithm as the basis for learning, and an independent test set, on which the quality and accuracy of the learned concepts or rules is then evaluated. At the highest level, the *Träumerei* is composed of two parts of length 8 and 16 bars, respectively, where the first part is obligatorily repeated. In the experiments, we used various pianists' performances of the *second part* for learning. The *first part* of the piece was then used for testing: the learned rules were applied to it to produce an expressive interpretation.

In each experiment, we only used a small number of selected performances for the training phase, rather than all 28. Given the nature of our learning task, increasing the number of example performances would not increase the number of different musical patterns that the learner can look at, as the music is the same for every performance.

4 Approach 1: Learning at the structure level

4.1 Structure-level learning

The first approach tested was the one that we had already used in previous experiments with other types of music, notably, Chopin waltzes (Widmer, 1994a). Learning proceeds not at the level of individual notes (e.g., by learning rules that would determine whether a particular note should be played longer or shorter than notated), but rather at the level of *musical structures*. The melody of a training piece is first subjected to a rough structural analysis, which identifies various structural units, such as groups, phrases, and musical "surface patterns" like linearly ascending or descending melodic lines, arpeggiated chords (*triadic melodic continuations*, in the terminology of (Narmour, 1997)), and other types of melodic or rhythmic structures that tend to be heard as distinct units by listeners. Most of these structures were derived (in a very loose way) from Narmour's (1977) *Implication−Realization Model*. The tgiven empo curve associated with the piece is then analyzed to find rough prototypical *expressive shapes* that can be associated with each of the structures found. Currently, the repertoire of shapes is limited to only five types made up of straight lines: `even_level` (no recognizable rising or falling tendency of the curve in the time span covered by the structure), `ascending` (an ascending tendency from the beginning to the end of the time span), `descending`, `asc_desc` (first ascending, then descending), and `desc_asc`. The system selects those shapes that minimize the deviation between the actual curve and an idealized shape defined by straight lines. The result of this analysis step are pairs <musical structure, expressive shape> that are passed to the learner as training examples.

The output of the learner is a set of rules that specify conditions under which a certain type of expressive shape should be applied to a specific musical structure in a piece. The rules can then directly be applied to new pieces to produce expressive performances. Where musical structures overlap or are contained within each other, the respective shapes suggested by the rules are combined by simple averaging to produce the final expression curve. This strategy had given quite good results in experiments with performances of Chopin waltzes. See (Widmer, 1994a) for more details.

4.2 Experiment 1: Arrau, Ashkenazy, Brendel

For the first experiment, we chose three pianists from the top of Repp's list, namely, Claudio Arrau, Vladimir Ashkenazy, and Alfred Brendel. Their performances of the second part of the *Träumerei* were given to the learner as training examples.[1)]

Figure 1 shows the tempo curves of the three pianists performing the second part of the piece (ms. 9−24). The labels on the x axis indicate the absolute distance from the beginning of the piece in terms of quarter notes ("score time"). The curves plot the *relative tempo* at each point as the ratio of played vs. notated duration (relative to the average tempo of the entire performance, which would be a straight line at y = 1.0); that is, the higher the curve, the faster the local tempo. Also, the grouping structure, as explicitly given to the system as part of its structur-

[1)] Actually, these pianists are numbers 2 to 4 on Repp's list. Number one, Martha Argerich, was excluded because her performance is somewhat unusual. In fact, Bruno Repp said it struck him as "mannered" and "eccentric and distorted", a characterization which was also partially supported by his statistical analyses.

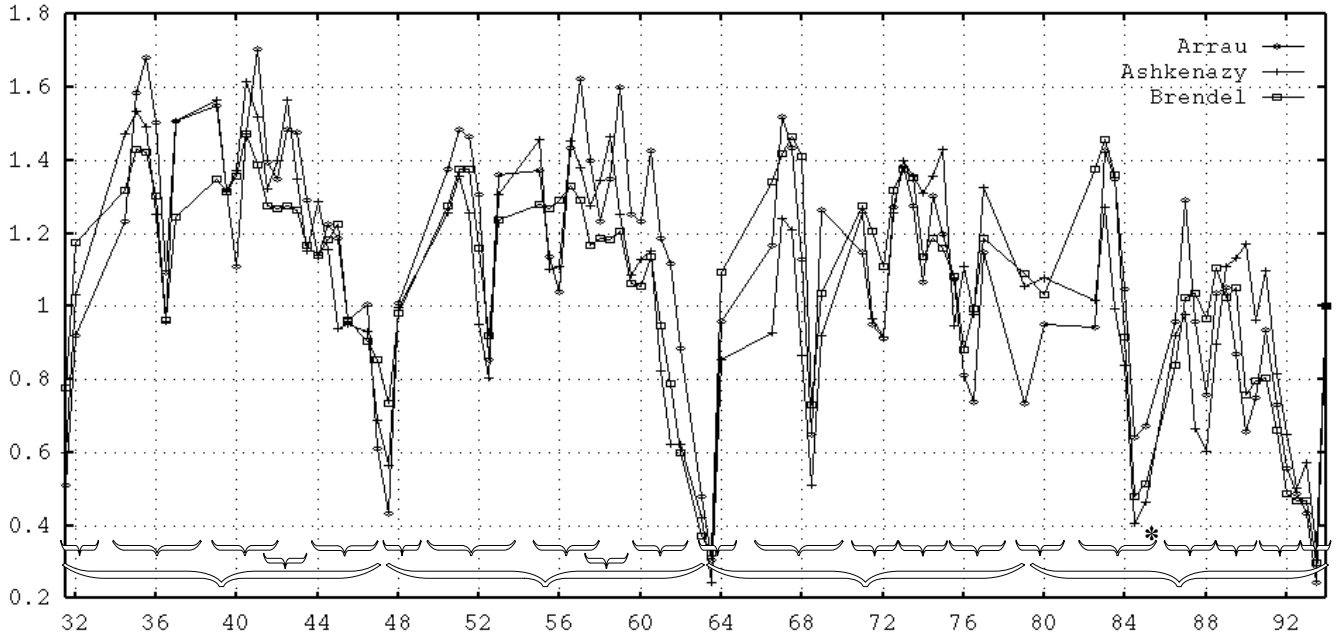

Figure 1: Second part of the Träumerei, as played by three pianists (tempo curves)

al information, is indicated by curly brackets below the plot. The finer level of grouping corresponds more or less directly to Repp's (1992, p.2549) structuring into 'melodic gestures'. The second level makes explicit the obvious high-level phrase structure of the piece.

It is quite evident from the plot that there is significant agreement between the performances at a global level, but also a lot of differences in the fine details. All three pianists observed the major *ritardandi* dictated by important structural boundaries – e.g., major phrase endings – and/or prescribed by expression markings in the score. The extreme *ritardando* in the third to last bar, marked by an asterisk in figure 1, is due to a *fermata* in the score.

Ideally, one would expect the learning algorithm to correctly extract the major common trends, and to learn some average strategy for those situations where the pianists' performances diverge. However, it must be made very clear at this point that by its very design, our learning algorithm (like any other standard inductive machine learning method) is not prepared to distort the training examples given to it, for instance by averaging over them. It searches for descriptions (generalizations) that cleanly separate examples of one class from examples of another. Thus, if the same musical passage is played with an *accelerando* by one pianist, but with a *ritardando* by another, that will be interpreted as a conflicting situation, and nothing will be learned from it. That is another reason for using only a few performances as training examples, rather than all 28.

With that in mind, we now take a look at figure 2, which shows how the system performed the test piece (the first part of the *Träumerei*) after learning from the three example performances. Below the plot, the figure sketches four of the expressive shapes the system decided to apply, along with the musical structures by which they were motivated. For instance, a descending shape (a *ritardando* from beginning to end) was applied to the 'triadic melodic continuation' in measure 1 (the arpeggiated chord F-A-C-F). The shape was suggested by the following learned rule:

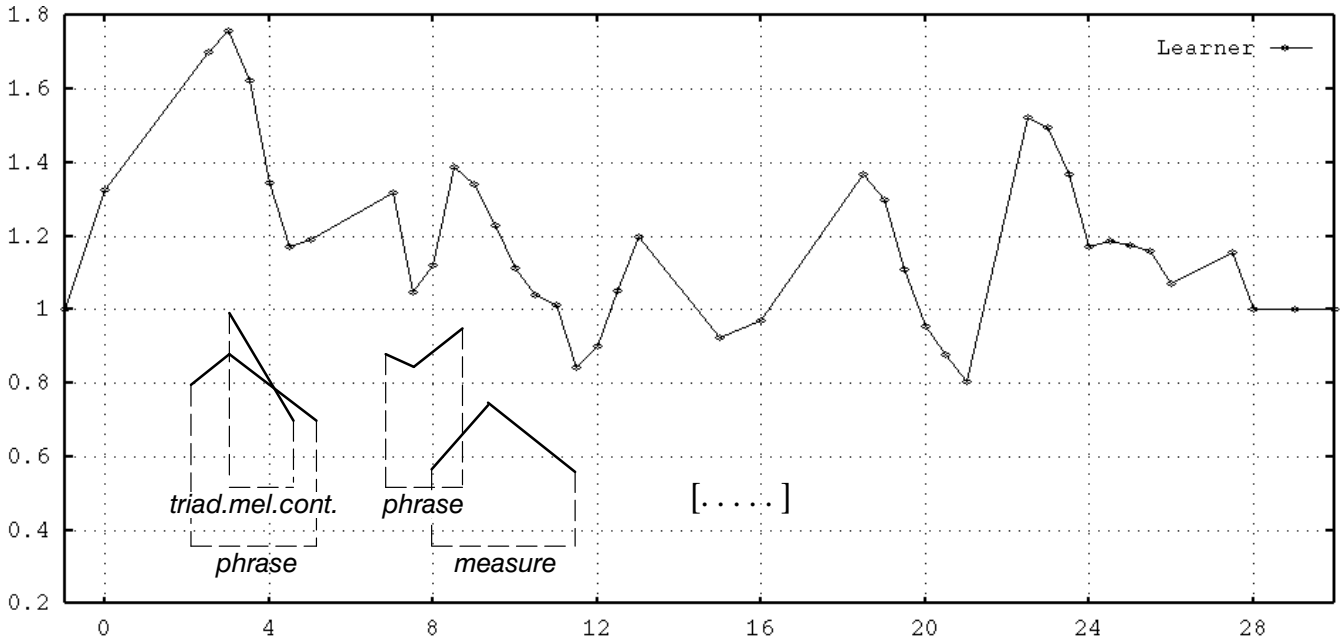

Figure 2: First part of the Träumerei, as played by the system after learning

```
descending( triadic_melodic_continuation( FIRST, LAST)) :–
        rel_position_in_phrase( FIRST, Pos),
        Pos  < 0.25,
        in_process( FIRST, rhythmic_gap_fill).
```

(“Apply a descending shape to a ‘triadic melodic continuation’ (identified by its first and last notes) if the first note is relatively early in the current phrase (its relative position within the phrase is < 0.25) and if the note also occurs in a ‘rhythmic gap fill’ figure (a certain type of rhythmic pattern).”)[2)]

The final shape of the tempo curve was computed by starting from a straight line at y = 1.0 and applying the shapes suggested by the rules one by one, always integrating new shapes into the existing curve by averaging the two.

The quality of the result can be more easily judged if we compare it to a musician’s performance. Figure 3 compares the system’s interpretation to one of its teachers’ (Brendel’s) performances of the same piece. The plot shows considerable agreement in the overall, high-level trends: a pronounced *accelerando–ritardando* shape over the main melodic gesture of the piece, the ascending sequence E-F-A-C-F-F (1); speeding up again towards the end of measure 2, which starts a descending chain of three four-note groups (2); these groups are played with a general *ritardando* tendency (3) in measure 3, followed by a pronouced speeding up towards the end of the sequence (4), which is marked by the half-note G in measure 4. The variation of the opening gesture in measure 5 is again associated with a clear *accelerando–ritardando* shape (5), though the system does not replicate Brendel’s tendency to play the penultimate eighth-note A slower than the final dotted-quarter A, but rather decides to end the *ritardando* on the last note of the group (*). Brendel’s phrasing of the rest of the first part is not well replicated by the

[2)] It is doubtful whether this particular rule reflects any general rubato principle; it is probably highly specific to the particular piece it was learned from. Section 5 has more to say on this general problem.

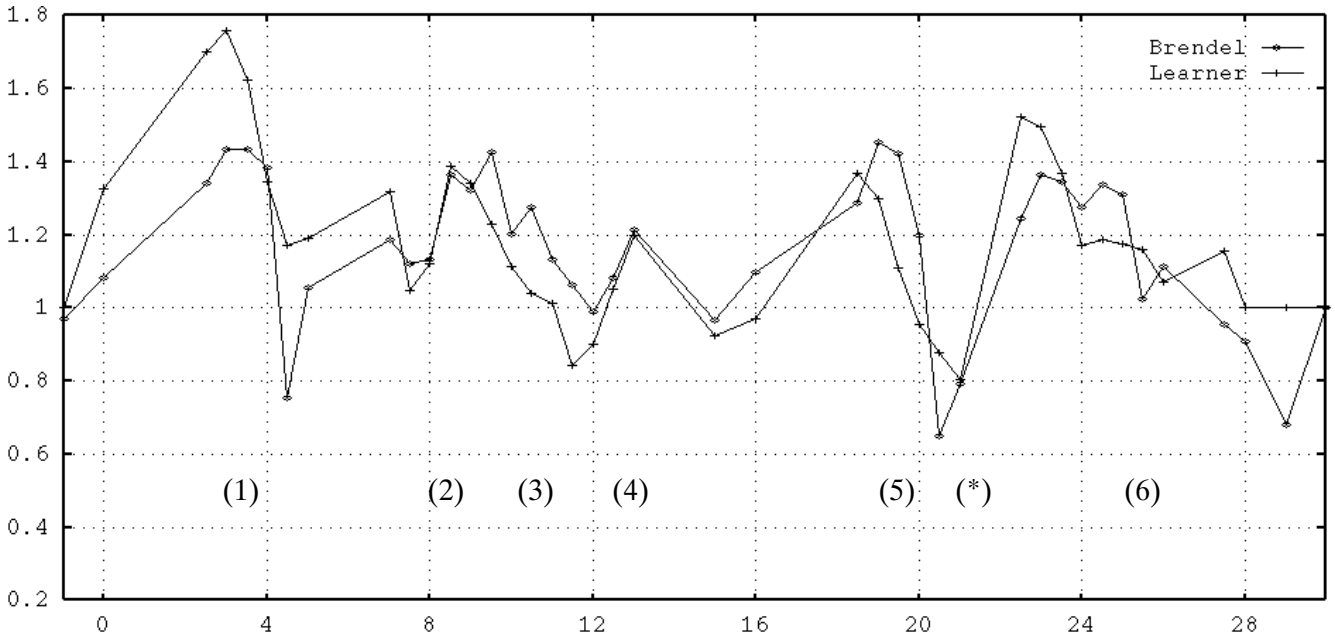

Figure 3: Comparison learner − Brendel on test piece

system, though the general trend (starting fast, gradually slowing down towards the end) is still quite noticeable (6). The finer phrasing of this passage, however, is rather different from Brendel's (though it does not sound too bad when played).

There are some problems with the system's performance that are evident from the plotted curve:

- The absolute tempo of the *accelerando−ritardando* shape at the beginning of the piece (1) is too high (at least it is significantly faster than Brendel's tempo). That is not the fault of the general shapes suggested by the rules, but rather an effect of the numeric inter- and extrapolation strategy employed by the learner to determine specific values for the height of the curve. Two possible explanations suggest themselves, which together might account for the effect: (1) one of the 'teachers' (Arrau) also applied an extreme *accelerando* to (variants of) this melodic gesture, both in relative and absolute terms (see figure 1), and (2) the performance curves of the second part of the piece, which were used for learning, are dominated by four or five very extreme *ritardandi*, which may have had a distorting effect on the numeric learning process.

- The system does not do a very good job in replicating the finer phrasing structure in measures 3 and 7. A preliminary analysis of the learning process revealed that this is due, at least in part, to the limited set of abstract *expressive shapes* (see above) that the learner can identify in a given performance curve. Patterns in the curve that are more complex than, say, an up-down shape, will be grossly approximated to fit one of the available linear prototypes. We are planning to introduce more complex abstract patterns into the learner's shape vocabulary (though that would probably not help in the current study, because the introduction of more complex target concepts increases the amount of training data needed to obtain stable learning results).

- There is no final *ritardando* in the last measure. Here, the system simply did not know what to do, it had not learned any rule that applied to this passage. Again, that is primarily an effect of the limited training data available.

Generally, however, we consider the result to be quite satisfactory, especially given that the performances of the three 'teachers', though fairly similar at a high level, are quite different in some of the finer details.

4.3 Experiment 2: Horowitz

Repp's data collection also includes three performances by Vladimir Horowitz. His statistical analyses revealed quite clearly that Horowitz's performance style is strikingly different from that of most of the other pianists. In a second experiment, the three performances by Horowitz (again only of the second part of the piece) were used as training examples, in order to see how well his style could be captured and replicated by the learner.

Figure 4 shows the system's performance of the test piece (the first part of the *Träumerei*) after learning from the three Horowitz examples. For comparison, we also plot the timing curve of one of Horowitz's performances on the test piece.

First of all, we note that Horowitz's performance is indeed very different from, say, Brendel's (cf. figure 3). The most striking differences are the shortened upbeat at the beginning of the piece (the quarter note C is reduced almost to a dotted eighth), the extreme tempo changes at the end of the first and the beginning of the second major melodic motif (positions 5 and 7 on the x axis), and the phrasing of the next to last measure. The learner did pick up the extreme way of playing the material of measure 2 − in fact, its rendition of this passage is even more extreme than Horowitz's; that is a parallel of the effect we had already seen in the first experiment − but it failed to replicate the shortening of the upbeat, and it failed rather miserably on the musical material of measures 3 and 7.

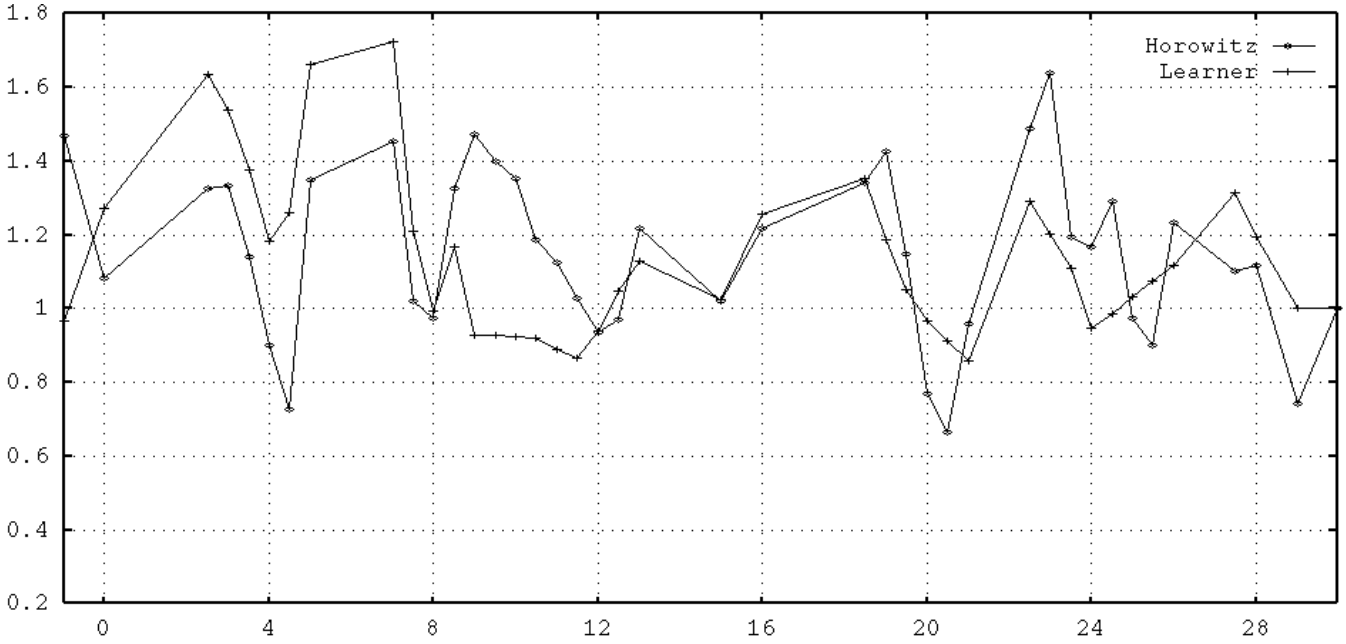

Figure 4: Comparison learner − Horowitz on test piece

The upbeat mismatch is easily explained: of the four similar upbeats that appear in part two of the piece (of which two are eighth-note upbeats, one is notated as a grace note only, and the last one is a quarter note), *all* are lengthened by Horowitz (in all three performances), so this particular way of playing the very beginning of the piece could not be predicted from looking at the second part.

As for the serious discrepancies in measures 3 and 7, they are again partly explained by the above-mentioned limitation of abstract expressive shapes that the learner can recognize and apply (that is a problem especially with measure 7). In addition, an analysis of the Horowitz performances reveals that he played comparable sections in the second part of the piece quite differently than in the first part (though he is extremely consistent across the three performances).

In summary, this case study is too limited to allow us to draw general conclusions. It does indicate that the learning system may be able to acquire performer- or style-specific rules, but it also reveals serious limitations of our current representation scheme that prevent the system from learning more refined expression principles.

5 Approach 2: Note-level learning with the KTH vocabulary

5.1 Target concepts and representation language

An alternative approach was tested in a second set of experiments. Given the current limitations of the structure-based approach, we wanted to compare it to a learning system that learns expression rules directly at the note level. That is, each individual note in a performance is interpreted as an example of *accelerando* (if it was played faster than its predecessor) and *ritardando*. In addition, each example is associated with a numeric value that represents the precise degree of *accelerando* or *ritardando* applied. The desired output of the learning system would be a set of rules for note-level *accelerando* and *ritardando*, respectively. That is the same level on which most of the KTH expression rules are formulated (see, e.g., Friberg, 1991). For the learning algorithm proper, the abstract task is the same as above − learn symbolic decision rules plus numeric interpolation tables for deciding the precise degree of tempo change − so the same learning algorithm was used as in the previous experiments.

The *vocabulary*, i.e., the set of descriptors or features used to describe each individual note and its role in the musical context, consisted of intrinsic note attributes (name, pitch, duration), information about the immediate context, i.e., the immediate predecessor and successor notes (duration of the neighboring notes and intervals between them and the current note), a few attributes expressing the position of the note in some obvious structural dimensions (like the metrical strength of the note and its relative position within the current phrase), and finally of two concepts that were adopted from the vocabulary of the KTH rule set: the *melodic* and *harmonic charge*, as defined in (Friberg, 1991) and (Sundberg, 1993), with additional information about the difference in harmonic charge (and the distance) between the harmony underlying the current note and the previous and following harmonies. With respect to the phrase attributes, we defined two levels of grouping, in analogy to (Sundberg et al., 1991), namely, subphrases and phrases, where the subphrases in the *Träumerei* correspond to Repp's *melodic gestures*, while the phrases are the four-measure groups that naturally describe the high-level structure of the piece (cf. figure 1).

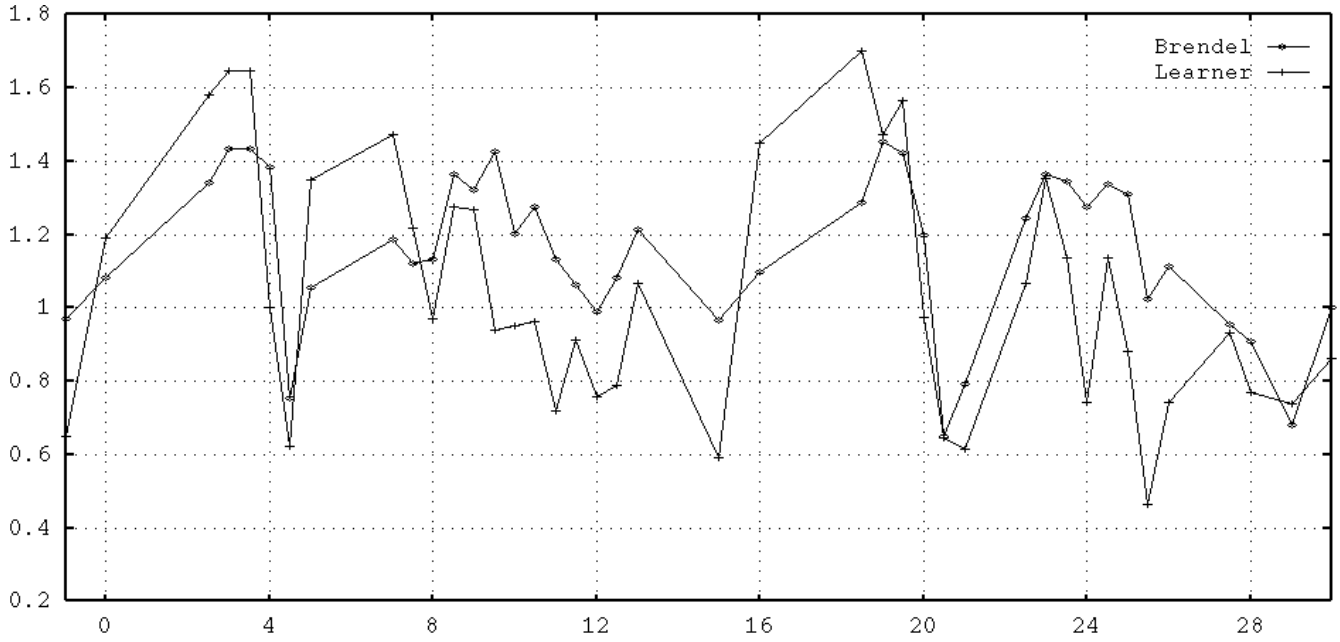

Figure 5: Performance of test piece after note-level learning with KTH vocabulary

Given this particular vocabulary and the fact that learned rules would be formulated at the note level, we were also interested in whether the learning algorithm would actually rediscover some of the KTH rules, or at least some similar general principles.[3)]

5.2 Experiment: Arrau, Ashkenazy, Brendel

Learning experiments with various sets of pianists were performed. Here, we show one typical result, derived from the same data as in section 4.2 above: the performances by Arrau, Ashkenazy, and Brendel of the second part of the *Träumerei* were used as training examples, and the rules learned from that were then tested on the first part.

Figure 5 shows the result and again compares it to Brendel's performance. A comparison with the result of the structure-based learner (see figure 3) reveals significant differences and suggests that the structure-based learner is superior, at least on this task: the differences between the system's and Brendel's performance are markedly larger in the note-level learning case, and the system's variations generally tend to be rather extreme. However, these are preliminary results that should be taken with a grain of salt. For one thing, the training material that was available for the experiments is too limited to permit us to draw conclusions that go beyond this particular piece of music. And secondly, a quantitative analysis of the results − we computed the degree of agreement between the system's solution and the three pianists' performances − showed that for the note-level learner the error in terms of the absolute difference between the curves is indeed larger, but that the number of "correct" classification decisions (i.e., the number of notes where both a pianist and the system made the same categorical decision: *accelerando* or *ritardando*) is in fact a bit higher for the note-level learner than for the structure-based

[3)] We had already observed that effect in previous experiments with other types of music, where our learning system had discovered rules that turned out to be variants of some of the expression principles postulated by Sundberg and co-workers (see Widmer, 1995).

one. Experiments with larger and more diverse example sets and much more detailed analyses will be needed to establish with reasonable confidence which of the two approaches is better, and in which situations.

As for the interpretability of the learned rules, our hopes that the system might discover general principles akin to the KTH rules were not really fulfilled. A few of the learned rules are indeed interesting and have, at least in part, a relatively straightforward musical interpretation. Here are two examples:

```
ritardando( Note) :–
        rel_position_in_phrase( Note, Pos),
        Pos > 0.5183871,
        harmonic_charge_diff_next( Note, NHCDiff),
        NHCDiff > 3.
```

("Play the current note slower if the note is in the second half of the current phrase (its relative position in the phrase is > 0.51) and the difference in harmonic charge between the current harmony and the following one is larger than 3.0 (i.e., the next chord has higher harmonic charge.")
That is reminiscent of Friberg's principle of increasing the duration of notes when a chord of higher harmonic charge is approaching.

```
ritardando( Note) :–
        rel_position_in_phrase( Note, Pos),
        Pos > 0.75875,
        int_prev( Note, PInt),
        at_most( PInt, min3).
```

("Slow down if the current note is toward the end of the current phrase (within the last quarter of the phrase's duration) and the interval between the note and its predecessor is not larger than a minor third.")

However, many of the rules produced by the generalization algorithm were rather complex and do not appeal to our musical intuition, like the following:

```
accelerando( Note) :–
        metrical_strength( Note, MS),
        MS <= 3,
        duration_of_phrase( Note, PhrDur),
        PhrDur <= 16,
        rel_position_in_phrase( Note, PhrPos),
        PhrPos <= 0.612903,
        dur_next( Note, NDur),
        NDur <= 0.5,
        int_next( Note, NInt),
        int_prev( Note, PInt),
        wider_interval( NInt, PInt).
```

Such rules describe musical situations and expression patterns that do appear in this particular piece, but they are not likely to be very general. We conclude from this that the data sample available for learning was too limited, not in terms of the number of example performances, but in terms of the diversity of the musical material. Examples from other pieces would be needed to be able to learn general principles that apply to an entire genre and abstract away from the peculiarities of a single piece or the style of a particular pianist.

There is a second, independent phenomenon that contributes to this effect. Experiments with larger numbers of example performances have shown that the number and complexity of rules

produced by the learner increases with the amount of training data. That is a clear sign of what is known in the machine learning literature as *overfitting*: the learning algorithm attempts to find a set of rules that explain every single observed tempo deviation in each of the performances, even those that are very rare and idiosyncratic. The result is poor generalization. Machine learning has developed a wealth of techniques that go under the name of *pruning*. Their common goal is to enable the learner to find simpler rule sets by distinguishing between strong, relevant regularities and coincidental patterns or even errors (*noise*) in the data. We plan to repeat the Schumann experiments with the inductive learning algorithm FOSSIL, developed at our institute (Fürnkranz, 1994). FOSSIL provides a variety of mechanisms for explicit pruning control, which might allow us to discover more general and robust expression patterns.

A final problem with the rediscovery of KTH-type expression rules is that the KTH rules are *additive* in nature, whereas the rules that a standard machine learning algorithm looks for are *exclusive*: each observed effect is to be completely explained by *one* rule. It is not at all obvious at this point how the machine learning approach could be modified to discover additive, or even partially conflicting, influence patterns.

6 Conclusions

To summarize, the paper has described two approaches to learning rules of expressive performance from examples of human performances. Experiments with expressive timing data from performances by various pianists of Robert Schumann's *Träumerei* were presented. Though some of the results are quite interesting and encouraging, the experiments point to a number of open problems that need to be addressed.

Apart from the limitations of the data available for our experiments, which were discussed in detail in various sections of this paper, the learning system itself suffers from a number of shortcomings. The most obvious limitations are the limited repertoire of *expressive shapes* that can be recognized in expression curves and the fact that the structural analysis of given pieces is done in a rather crude way (and only for the melody). Extending these components would increase the expressiveness of the system's representation language and would make it possible to at least describe (and hopefully also find) more refined expression principles.

General questions that are of great importance to the eventual success of the general approach concern mainly the appropriate *level of modelling*. For instance, if we adopt the structure-level approach to learning (see section 4), what are the structural units in the music that are relevant to expression, that is, how do we segment the observed performance curves into meaningful chunks that can be explained? What is the appropriate vocabulary to describe musical situations and structures? And would it be fruitful to pursue a combination of structure-level and note-level learning? And finally, in our experiments we ignored the aspect of absolute tempo, because that could not be learned from the examples, but one must also be aware of the effect that factors like global tempo have on the details of a performance (see, e.g., Desain and Honing, 1991). More global factors will have to be taken into account in a full model of expression.

Specific efforts in the immediate future will be devoted to assembling a large collection of real data (measurements of recordings of a diverse set of pieces) and extensive experiments with different music-structural vocabularies; experience in machine learning shows that the choice of vocabulary has an enormous impact on the results achievable by a particular learning algorithm and also on the general learnability of the concepts of a domain.

Considering all this, there is ample room for further research within the "analysis by resynthesis" approach. Despite the mixed results of the present experiments, we have reasons to believe in the general utility of the approach, and we hope that experiments with richer data sets will lead to general results that may be of interest to the study of musical expression, both in a practical and in a theoretical sense.

Acknowledgments

I would like to thank Bruno Repp for allowing me to use his dataset. The Austrian Research Institute for Artificial Intelligence gratefully acknowledges financial support from the Austrian Federal Ministry of Science, Technology, and the Arts.

References

Aha, D.W., Kibler, D. and Albert, M. (1991). Instance-Based Learning Algorithms. *Machine Learning* 6(1), pp. 37−66.

Desain, P., and Honing, H. (1991). Tempo Curves Considered Harmful. A Critical Review of the Representation of Timing in Computer Music. In *Proceedings of the International Computer Music Conference (ICMC-91)*. San Francisco: International Computer Music Association.

Friberg, A. (1991). Generative Rules for Music Performance: A Formal Description of a Rule System. *Computer Music Journal* 15(2), pp. 56−71.

Fürnkranz, J. (1994). FOSSIL: A Robust Relational Learner. In *Proceedings of the Seventh European Conference on Machine Learning (ECML-94)*. Berlin: Springer Verlag.

Narmour, E. 1977. *Beyond Schenkerism*. Chicago: University of Chicago Press.

Quinlan, J.R. (1990). Learning Logical Definitions from Relations. *Machine Learning* 5(3), pp. 239−266.

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), pp. 2546−2568.

Sundberg, J., Askenfelt, A. and Frydén, L. (1983). Musical Performance: A Synthesis-by-rule Approach. *Computer Music Journal* 7(1), pp. 37−43.

Sundberg, J., Friberg, A. and Frydén, L. (1991). Common Secrets of Musicians and Listeners: An Analysis-by-Synthesis Study of Musical Performance. In P. Howell, R. West, and I. Cross (eds.), *Representing Musical Structure*. London: Academic Press.

Sundberg, J. (1993). How Can Music Be Expressive? *Speech Communication* 13, pp. 239−253.

Widmer, G. (1993). Combining Knowledge-Based and Instance-Based Learning to Exploit Qualitative Knowledge. *Informatica* 17, special issue on Multistrategy Learning, pp. 371−385.

Widmer, G. (1994a). Learning Expression at Multiple Structural Levels. In *Proceedings of the International Computer Music Conference (ICMC−94)*, Aarhus, Denmark.

Widmer, G. (1994b). Studying Musical Expression with AI and Machine Learning: "Analysis by Resynthesis". In *Proceedings of the Aarhus Symposium on Generative Grammars for Music Performance*, KTH, Stockholm.

Widmer, G. (1995). Modelling the Rational Basis of Musical Expression. *Computer Music Journal* 19(2) (in press).