Applications of Machine Learning to Music Research: Empirical Investigations into the Phenomenon of Musical Expression

Gerhard Widmer

Dept. 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
http://www.ai.univie.ac.at/~gerhard

Abstract

This chapter describes an application of machine learning techniques to the study of a fundamental phenomenon in tonal music. Learning algorithms are described that induce general rules of *expressive music performance* from example of real performances by musicians. Motivated by the insight that general knowledge about music plays an essential role in the way humans learn this task, we present two alternative approaches to knowledge-based learning. In both cases, the domain knowledge provided to the learner is based on established theories of tonal music. Experimental results show that both approaches lead to a significant improvement of the learning results, compared to purely inductive learning.

However, this project is more than basic machine learning research. Due to its thorough grounding in music theory, the project can also be viewed as a contribution to the scientific field of music research or musicology; it has produced results that have found their way also into the literature of that scientific discipline. These will also be touched on in this chapter.

1 Introduction

This chapter describes an application of machine learning that may at first sight seem somewhat unusual or even esoteric: learning algorithms are applied to problems of tonal music. In a project that has evolved over several years, we have used machine learning methods to study the foundations of a fundamental musical skill that lies at the heart of music as an art form, namely, *expressive music performance*. Several learning systems have been developed that try to learn general rules of expressive performance from examples of performances by human musicians.

The project started as basic machine learning research in the area of knowledge-based learning. The initial aim was to investigate various ways of introducing domain knowledge into the learning process and to study the general nature and impact of such knowledge. Music was selected as a test domain because it provided a set of difficult learning tasks (and, admittedly, for reasons of personal interest). As the domain analysis progressed and more and more emphasis was put on a principled and musically plausible modelling of domain knowledge, the project gradually turned into a truly interdisciplinary endeavor. It began to produce results of interest to musicology that have in the meantime found their way also into the literature of that scientific discipline (see, e.g., Widmer, 1993a, 1995a, 1995b).

What this chapter presents, then, is a genuine application of machine learning — not to a "practical" (e.g., industrial) problem, but to another branch of *science*. The potential of machine learning as a contributing technique for other scientific domains — notably, biochemistry and molecular biology — has been demonstrated by a number of researchers (e.g., Hunter, 1993; King et al., 1992; Muggleton et al., 1992; Shavlik et al., 1992). This chapter attempts to show that also more 'informal' domains like music can benefit from machine learning experiments.

As an interdisciplinary project, our work was guided by questions from, and has produced results of interest to both fields involved. From a machine learning perspective, our objective was to study various types of weak (i.e., imprecise and incomplete) domain knowledge, and ways of using it to bias a learner towards better hypotheses. The results that will be presented here are two alternative approaches to knowledge-based learning: in the first approach, an inductive learning algorithm takes advantage of explicitly represented qualitative domain knowledge to guide its search for generalizations (section 4). Section 5 describes an alternative strategy. Domain knowledge is used to transform the entire learning task to a higher abstraction level where relevant regularities become more readily apparent. Experimental results show that both approaches lead to an improvement of the learning results.

From the viewpoint of musicology, the central problem to be investigated was the notion of musical knowledge. Relevant questions included: What kind of general musical knowledge do music listeners possess? How can it be formalized? What is the relation between this knowledge and expressive performance? What structural aspects of music pieces determine or influence the acceptability of performances? It was our belief that machine learning can shed new light into these matters, and the results of our experiments are indeed informative. A comprehensive presentation and analysis of the experiments from a music-theoretic point of view is, of course, beyond the scope of this article. In sections 6 and 7, we will at least try to hint at some of the most interesting results.

This chapter can only give a broad overview of our projects, the algorithms and results, but we do hope to give the reader an appreciation of the promise that machine learning holds for scientific fields like musicology and, not least, we hope to convey some of the fascination of AI-based music research.

2 The object of study: expressive music performance

When played exactly as written, most pieces of music would sound utterly mechanical and lifeless. *Expressive performance* (or *interpretation*) is the art of 'shaping' a piece of music by playing it not exactly as given in the written score, but continuously varying certain musical parameters during a performance, e.g., speeding up or slowing down, growing louder or softer, placing micro-pauses between events, etc. There are numerous parameter dimensions that can be affected by a performer, some of which are limited to particular instruments (e.g., *vibrato*). In this project, we concentrate on the two most important expression dimensions, *dynamics* (variations of loudness) and *rubato* or *expressive timing* (variations of local tempo). The relevant musical terms are *crescendo* vs. *diminuendo* (increase vs. decrease in loudness) and *accelerando* vs. *ritardando* (speeding up vs. slowing down), respectively. Our programs will be shown the melodies of pieces as written and recordings of these melodies as played expressively by a human pianist. From that they will have to learn general principles of expressive interpretation (in the form of rules), which should enable them to play new pieces

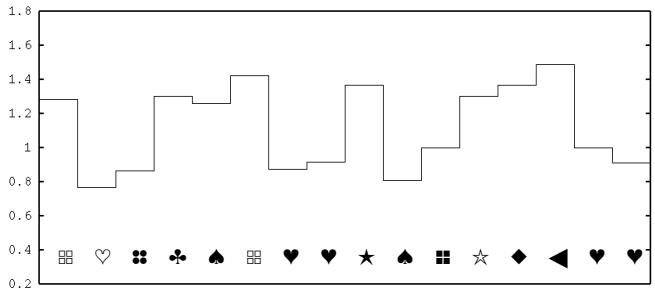

Figure 1: An abstract training example.

more or less expressively.

Sometimes composers place explicit expression marks in the score (e.g., the command *cresc* underneath a musical passage), but more often than not expressive form is left implicit, to be decided by the performer on the basis of his/her musical understanding. Our systems will be shown only the notes of a piece, with no explicit expression marks.

To be more precise, input to the learner are *melodies* of musical pieces (i.e., sequences of notes), where each note in a melody is associated with two numeric values: the exact *loudness* with which the note was played by a performer (the *dynamics* dimension) and the precise *tempo* (i.e., the ratio of the duration as actually played vs. the duration as prescribed by the score). The learner's task is to induce rules that allow it to determine exactly how loud and how fast any note in a new given pieces should be played. The problem is thus a *numeric prediction task*.

Section 4.3 below will present a knowledge-based algorithm for this type of induction problem, but first we take a closer look at what music theory can tell us about the problem.

3 The nature and importance of background knowledge

Consider again the abstract learning task: training examples are melodies, i.e., sequences of notes, where each note is associated with a numeric value that represents the precise degree of loudness or local tempo that has been applied to the note by a musician in a particular performance. These numeric values can be viewed as defining a curve (a *performance curve*, in music-technical terms) above the melody. The task is to learn to 'draw' 'correct' or at least 'sensible' curves above new melodies, i.e., new sequences of notes.

Figure 1 tries to give the reader an intuition of what the problem looks like to a learner without any knowledge of music: to a naive learner, the individual notes are simply generic *symbols* with various intrinsic characteristics or features. This abstract representation illustrates the difficulty of the learning task. It is quite evident that one of the main problems is that of *context*: one symbol alone does not uniquely determine the numeric value (the height of the curve) associated with it. It is not at all clear, however, what the relevant context is, whether there are only local or also nonlocal context influences.

It is a fact that humans (e.g., music students) learn general principles of expressive performances quite effectively, from rather few examples. The reason is, of course, that we as humans possess additional *knowledge* about the *meaning* of the symbols. To us, this is music, and that gives us an *interpretation framework* for the symbols. Listeners do not perceive a presented piece as a simple sequence of unrelated symbols or events, but they immediately and

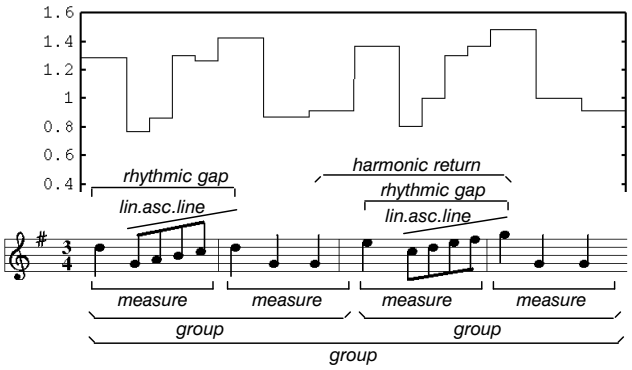

Figure 2: The problem as perceived by a human learner.

automatically interpret it in structural terms. For instance, they segment the flow of events into 'chunks' (motives, groups, phrases, etc.); they intuitively hear the *metrical structure* of the music, i.e., identify a regular alternation of strong and weak beats and know where to tap their foot. Linearly ascending or descending melodic lines are often heard as one group, and so are typical rhythmic figures and other combinations of notes. To a human learner, then, the above training example would look more like Figure 2.

Many more dimensions of musical structure can be identified, and it has been shown that acculturated listeners extract these structures in a highly consistent manner, and mostly without being aware of it. All these structures and patterns are related to the rise and fall of loudness or tempo that are observed (heard) in a performance. That is the (unconscious) musical 'knowledge' that listeners and musicians automatically bring to bear when listening to or playing a piece.

Results from musicology support this hypothesis. Numerous recent studies tell us that expression is not arbitrary, but highly correlated with the structure of music as it is perceived by performers and listeners. In fact, expression is a means for the performer to emphasize certain structures and maybe de-emphasize others, thus conducing the listener to 'hearing' the piece as the performer understands it.[1]

The following sections describe two different approaches towards providing the learning algorithm with general background knowledge about musical structure and its possible relation to expressive performance. The knowledge itself will be based in both cases on two well-known theories of tonal music — Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music* and Eugene Narmour's (1977) *Implication-Realization Model*. Both theories postulate certain types of structures that are claimed to be perceivable by human listeners.

That is also where the project becomes interesting to musicology: experimental results with these learning systems can provide empirical evidence for or against the relevance of various parts of the underlying music theories and will generally help us identify those structural

[1] A clarifying remark to readers who feel that we are trivializing the artistic phenomenon of expressive performance by reducing it to a function of structural patterns in the music: We are not talking here about the highly artistic details that distinguish a great pianist or other performer, and that derive in part from his/her deep understanding of music history, experience with styles, social circumstances, and artistic intentions. What is being investigated here is the "rational" component of expression, the types of musical behavior and understanding that are more or less common and agreed upon among musicians — in other words, what a music student must learn in order to produce acceptable performances. Even a virtuoso's performance is constrained by and large by such common norms of interpretation.

dimensions of music that seem to have the most 'explanatory power' with respect to given expressive performances.

4 Approach I: Learning at the note level with explicit qualitative background knowledge

The first approach we pursued was very much in the tradition of what is generally known as *knowledge-based* or *knowledge-intensive learning*: knowledge about musical structure perception was formulated in an explicit (albeit abstract, incomplete, and partly inconsistent) *domain theory* (Mitchell et al., 1986). A learning algorithm by the name of IBL-SMART was developed that can use the knowledge to advantage.

4.1 The target concepts

Learning proceeds at the level of notes. Each individual note is a training example, and the induced rules refer to individual notes as well. The goal is to learn rules that determine the precise degrees of loudness and tempo to be applied to each note in a piece. We have thus two separate (numeric) learning tasks — dynamics and tempo — and accordingly, the system will learn two sets of rules.

In order to make the problem accessible to a symbolic, knowledge-based induction algorithm, we split it into a symbolic classification task and a numeric prediction task. In both expression dimensions, we distinguish two classes of notes: those that are associated with a *rise* of the performance curve (relative to the previous note), and those that witness a *fall* of the curve. In the dynamics dimension, the relevant musical terms for the two classes are *crescendo* (an increase in loudness) and *decrescendo* (a decrease), and in the tempo dimension, *accelerando* (an increase in tempo, i.e., speeding up) and *ritardando* (slowing down). These are common musical concepts.

The learner induces classification rules that distinguish between instances of the two classes. In addition, for each of these rules it learns a scheme to predict precise numeric values, i.e., by *how much* the dynamics or tempo curve should rise or fall at a particular point. The learning algorithm IBL-SMART, which was developed for this class of problems, is described in section 4.3 below.

4.2 The qualitative domain theory

The musical examples are initially described only through intrinsic features of the individual notes (e.g., pitch (tone height), duration, relative position in the piece) and some simple relations between pairs of adjacent notes (the interval between two notes, and the direction of the interval). As we have tried to show in section 3 above, this is hardly sufficient for effective learning. Knowledge about relevant musical structure is needed. We have devised a structured symbolic *domain theory* that represents what we consider general musical intuitions that ordinary human listeners possess. Most of this knowledge is only approximate and uncertain, and that is made explicit in the formulation of the theory. Figure 3 sketches the general structure of the theory. A more detailed discussion can be found in (Widmer, 1993a, 1995a).

The model consists of two major components. The lower part, named *model of structural hearing* in figure 3, is basically a set of programs that perform a structural analysis of a given melody and explicitly annotate the melody with various musical structures that we believe

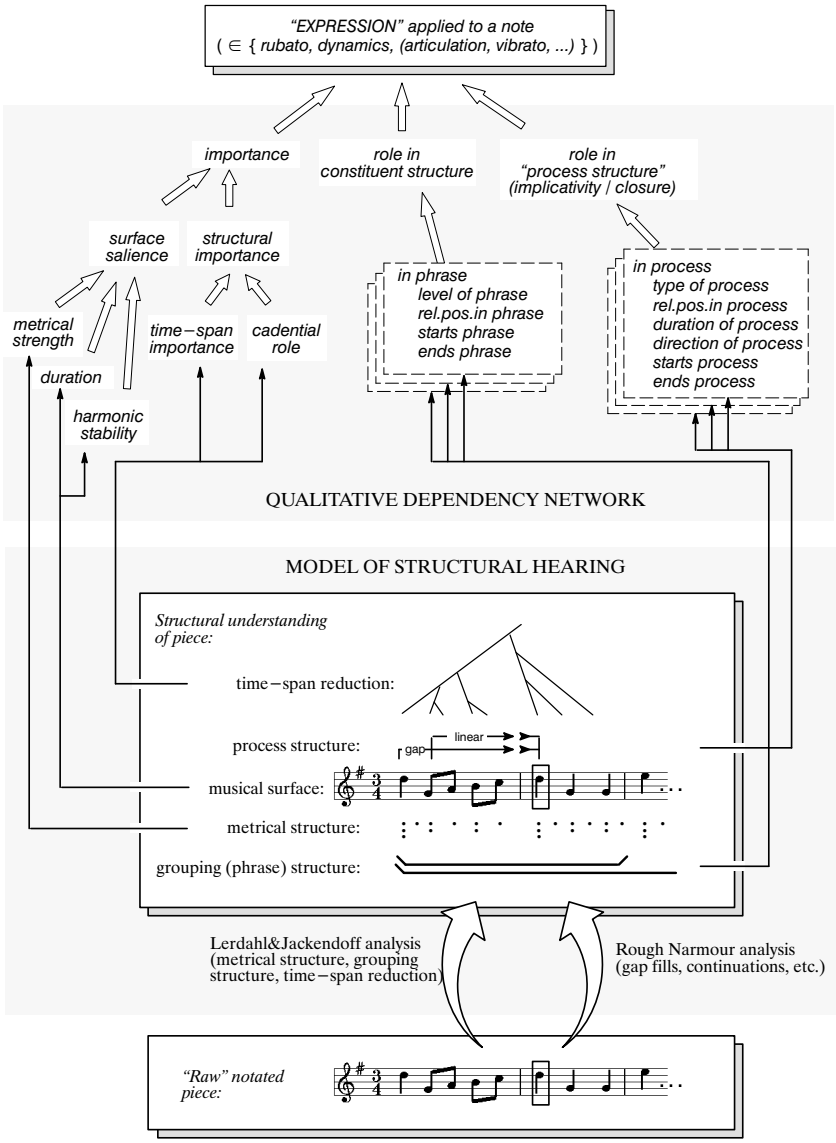

Figure 3: Structure of the qualitative background model.

are perceived by human listeners. This part of the model is based on the well-known music theories by Lerdahl and Jackendoff (1983) and Narmour (1977). Essentially, the purpose is to construct meaningful higher-level descriptors that capture aspects of musical context. These can then be referred to in the induction process.

The upper part of the theory (the *qualitative dependency network*) expresses our intuitions concerning possible relations between structural aspects of the music and appropriate expressive performance decisions (i.e., the symbolic target concepts). It is similar in structure to the 'classical' domain theories as used in *Explanation-Based Learning (EBL)* (Mitchell et al., 1986). It is a hierarchy of statements relating non-operational predicates (including the target concepts) to more operational, specific conditions. However, these statements may describe relations of various strength and specificity:

Strict (deductive) rules: As in EBL, the domain theory contains some strict deductive rules of the form `Q` $:-$ `P`$_1$, `P`$_2$, ... that specify sufficient conditions (`P`$_1$, `P`$_2$, ...) for some (non-operational) predicate `Q` to be true.

Directed qualitative dependencies: A statement of the form `q+(A,B)` can be paraphrased as "the values of attributes `A` and `B` are positively proportionally related" or "high (or low) values of `A` tend to produce high (or low) values of `B`, all other things being equal". Negative dependency `q-(A,B)` is defined analogously. Obviously, this type of knowledge is less precise and logically weaker than strict rules. It does not permit deductive reasoning. Similar types of knowledge items have been proposed in (Michalski, 1983) and (Collins and Michalski, 1989).

Undirected qualitative dependencies: A statement `depends_on`(`Q`, [`P`$_1$, `P`$_2$, ...]) denotes an unspecific, undirected relation between the set of predicates `P`$_i$ and the (non-operational) predicate `Q`. Basically, it says that the value (or truth value) of `Q` depends somehow on the values (or truth values) of the `P`$_i$, but we do not know the exact function that defines this dependency. Similar types of general knowledge items have been described in (Russell, 1989) and (Bergadano et al., 1989). They are used to focus the learner on sets of relevant predicates or attributes in the search for rule refinements.

Most of the arrows in figure 3 represent qualitative dependencies. For instance, the following statement at the top level of the theory relates the phenomenon of loudness variations to some abstract musical notions:

```
depends_on( crescendo(Note,X),
   [stability(Note,S), goal_directedness(Note,G), closure(Note,C)]).
```

"Whether crescendo should be applied to a note (and if so, the exact amount X) depends, among other things, on the musical stability S of the note, on its degree G of melodic 'goal-directedness', and on its degree of melodic 'closure'.".

Abstract notions like `stability`, `goal_directedness`, and `closure` are then again related to lower-level musical effects, all the way down to some surface features of training instances, for example:

```
q+( metrical_strength(Note,X),  stability(Note,Y)).
q+( harmonic_stability(Note,X), stability(Note,Y)).
```

"The perceived degree of stability Y of a note is positively proportionally related to (among other things) the metrical strength X of the note" etc.

where `metrical_strength` is a numeric and `harmonic_stability` is a symbolic attribute (with a discrete, ordered domain of qualitative values). Both are defined as operational and are computed by the lower part of the domain theory — the *model of structural hearing.*

4.3 The learning algorithm IBL-SMART

A knowledge-based learning algorithm by the name of IBL-SMART was developed for the purpose of this project. According to the two-part structure of the learning task as defined

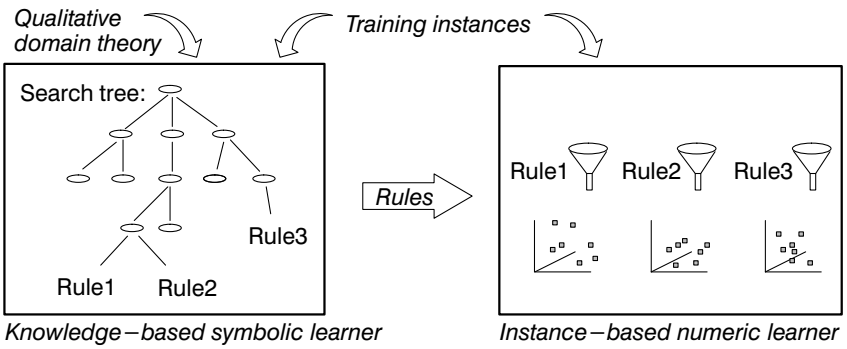

Figure 4: Integration of symbolic learning and numeric learning in IBL-SMART.

in section 4.1 above, IBL-SMART is composed of two major components (see figure 4)[2]: a *symbolic learning component* that learns to distinguish between the symbolic target concepts (e.g., *crescendo* and *decrescendo*) and can utilize domain knowledge in the form of a qualitative model, and an *instance-based* component that stores the instances with their precise numeric attribute values and can predict the target value for some new note by numeric interpolation over known instances. The connection between these two components is as follows: each rule (conjunctive hypothesis) learned by the symbolic component describes a subset of the instances; these are assumed to represent a subtype of the target concept (e.g., some particular type of *crescendo* situations). All the instances covered by a rule are given to the instance-based learner to be stored together in a separate instance space. Predicting the target value for some new note in a new piece then involves matching the note against the symbolic rules and using only those numeric instance spaces (interpolation tables) for prediction whose associated rules are satisfied by the note.

IBL-SMART's symbolic component is a non-incremental discrimination algorithm that learns classification rules in disjunctive normal form (DNF). It has been specifically designed to be able to use imprecise, qualitative background knowledge as contained in our domain theory. The algorithm starts with a nonoperational definition of the target concept (e.g., *crescendo*) and performs stepwise top-down operationalization (specialization) by growing a heuristic best-first search tree. Expressions (nodes of the tree) are refined by operationalizing non-operational predicates or by inductively adding new conditions that discriminate between positive and negative examples. A node becomes a leaf when it covers only positive training instances; it then represents one conjunct (rule) in the final DNF hypothesis. The search terminates when a certain percentage of the positive examples are covered.

Operationalization steps that reduce a non-operational predicate to more basic ones are based on rules or dependency statements given in the domain theory. In the case of strict rules, this is identical to the method used in Explanation-Based Generalization (Mitchell et al., 1986). In the case of a qualitative dependency, say, `q+(A,B)`, the operationalization step consists in replacing the non-operational predicate `B` with `A`. The algorithm creates successor nodes by replacing `B(X,.)` with $\texttt{A(X,}a_i\texttt{)}$ for all values a_i appearing in positive instances covered by the current node. Which of these node expansions is most promising and will

[2]The name IBL-SMART reflects the two components: IBL stands for *Instance-Based Learning* (Aha et al., 1991) and characterizes the numeric component, and SMART is a tribute to the ML-SMART algorithm (Bergadano and Giordana, 1988), which provided some of the ideas for the search strategy of the symbolic learner.

Figure 5: Beginnings of three little minuets by J.S.Bach.

most likely be expanded further is then determined by a heuristic *evaluation function*, which guides the search. The function takes into account empirical measures like the 'purity' of the current node, i.e., the ratio of positive / negative instances covered by the expression, but also semantic criteria, like the degree to which the attribute values involved in the operationalization observe the proportionality relation postulated by some dependency statement in the domain theory.

By taking into account both such inference-dependent plausibility measures and information about the numbers of instances covered, the search heuristic combines weak, imprecise background knowledge with empirical information from the training data, producing hypotheses that tend to correspond to the background knowledge as much as the data permits and overriding the background knowledge if the data is in conflict with the knowledge. A more detailed description of the search strategy can be found in (Widmer, 1993b).

4.4 An experiment

The system has been tested with pieces from various musical epochs and styles (Bach minuets, Chopin waltzes, even jazz standards). Here we present two typical results.

Figure 5 shows the beginnings of three well-known minuets from J.S.Bach's *Notenbüchlein für Anna Magdalena Bach*. All three pieces consist of two parts. The second parts of the pieces were used for training: they were played on an electronic piano by the author, and recorded through a MIDI interface. After learning, the system was tested on the first parts of the same pieces. In this way, we combined some variation in the training data (three different pieces) with some uniformity in style (three pieces from the same period and with similar characteristics; test data from the same pieces as training data, though different).

The training input consisted of 212 examples (notes), of which 79 were examples of crescendo, and 120 were examples of decrescendo (the rest were played in a neutral way). The system learned 14 rules and, correspondingly, 14 interpolation tables characterizing crescendo situations, and 15 rules for decrescendo. Quite a number of instances were covered by more than one rule.

Applying these rules to new pieces produces expressive performances. The quality of these

Figure 6: Beginning of a training piece as played by teacher (*dynamics* curve).

Figure 7: Beginning of a test piece as played by learner after learning (*dynamics* curve).

is not easy to measure, as there is no precise criterion to decide whether some performance is right or wrong. Judging the correctness is a matter of listening. Unfortunately, we cannot attach a recording to this article so that the reader can appreciate the results. Instead, figure 6 depicts a part of one of the training pieces (the second part of the first minuet in G major as played by the author), and figure 7 shows the performance generated by the system for a test piece (the first part of the same minuet) after learning. The figures plot the relative loudness with which the individual notes were played; a level of 1.0 represents average loudness.

The reader familiar with standard music notation may appreciate that there are strong similarities in the way similar types of phrases are played by the human teacher and the learner. Note, for instance, the crescendo in lines rising by stepwise motion, and the decrescendo patterns in measures with three quarter notes. Note also the consistent pattern of accents (loud notes) at the beginnings of measures. Given the limited amount of training data, the degree of generalization achieved is quite remarkable. In addition, an inspection of

Figure 8: Beginning of test piece as played after learning without domain theory.

the symbolic rules learned in this experiment reveals that the system had re-discovered some expression principles that had been formulated years ago by music theorists (see section 7).

When we perform the same experiment *without* the domain theory, we get an impression of the importance of the musical background knowledge. Without the domain model, IBL-SMART is reduced to a purely empirical discrimination algorithm.

Figure 8 shows the system's performance of the same test piece after learning from the Bach minuets in this way. There is a marked deterioration in the resulting performance from learning *with* knowledge (figure 7) to learning *without* knowledge (figure 8). The variations applied by the restricted system are of rather mixed quality. In some cases (e.g., the decrescendo patterns in measures 4 and 5), they do make sense, in others (e.g., the stress on the last notes in measures 1, 3, and 6) the system's decisions run counter to musical intuition. Obviously, the domain theory contributes significantly to successful learning, especially when the number of available training examples is rather small, as in the current case.

Apart from such qualitative evaluations, we have also performed some quantitative measurements to establish beyond doubt the benefits of the knowledge-based approach. Section 7.1 has more to say on that.

5 Approach II: Learning at the structure level via knowledge-based abstraction

Despite some encouraging results with the first approach, it became clear eventually that the note level is not really appropriate from a musical point of view. For one thing, though the performances produced by the system were in large part musically sensible, they lacked a certain smoothness and a sense of both local and global form. Second, it is psychologically implausible that performers think and decide on a purely local level in terms of single notes; rather, they tend to comprehend music in terms of higher-level abstract forms like phrases etc. And finally, as observed by Sloboda (1985), expression is a *multi-level* phenomenon: expressive shapes, like musical structures, appear at multiple levels. Local expression patterns may be embedded within larger patterns (e.g., shaping of ornaments within an overall crescendo). A

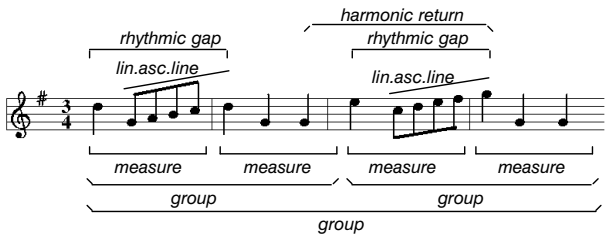

Figure 9: Structural interpretation of part of Bach minuet.

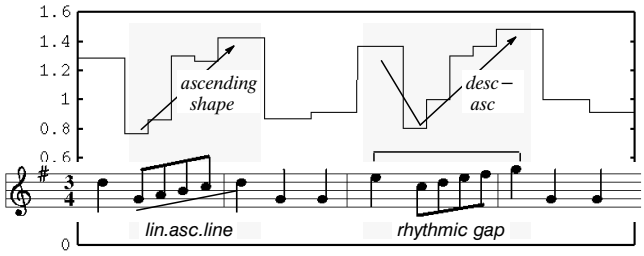

Figure 10: Two of the expressive shapes found in Bach recording.

sensible formalization of musical expression should reflect that.

Consequently, we have developed an alternative approach that abandons the note level and tries to learn expression rules directly at the level of musical structures. The essence of the approach is a *knowledge-based abstraction strategy* that transforms the training examples and the entire learning problem to a musically plausible abstraction level. The induced expression rules will then also relate to that abstraction level.

The problem transformation proceeds in two stages. The system first performs a musical analysis of the given melody. Analysis routines, based again on selected parts of the theories by Lerdahl and Jackendoff (1983) and Narmour (1977), identify various structures in the melody that might be heard as units or 'chunks' by a listener or musician. The result is a rich annotation of the melody with identified structures. Figure 9 exemplifies the result of this step with an excerpt from a simple Bach minuet. Among the perceptual chunks identified here are four *measures* heard as rhythmic units, three *groups* heard as melodic units or "phrases" on two different levels, two *linearly ascending melodic lines*, two rhythmic patterns called *rhythmic gap fills* (a concept derived from Narmour's theory), and several others. Note that these musical structures can be of widely varying scope — some consist of two or three notes only, others may span several measures. As training examples will be defined by such structures, the system will learn to recognize and apply expression at multiple levels.

In the second step, the abstract *target concepts* for the learner are identified. The system tries to find prototypical *shapes* in the given expression (dynamics and tempo) curves that can be associated with these structures. Prototypical shapes are rough trends that can be identified in the curve. The system distinguishes five kinds of shapes: `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 up to a certain point, then descending), and `desc_asc` (first descending, then ascending). The system selects those shapes that minimize the deviation between the actual curve and an idealized shape defined by straight lines.

Figure 10 illustrates this step for the dynamics curve associated with the Bach example (derived from a performance by the author). We take a look at two of the structures found in figure 9: the ascending melodic line in measures 1–2 has been associated with the shape `ascending`, as the curve shows a clear ascending (*crescendo*) tendency in this part of the recording. And the 'rhythmic gap fill' pattern in measures 3–4 has been played with a `desc_asc` (*decrescendo – crescendo*) shape.

The results of the transformation phase are passed on to IBL-SMART.[3] Each pair <*musical structure, expressive shape*> is a training example. Each such example is further described by a quantitative characterization of the shape (the precise loudness/tempo values, relative to the average loudness and tempo of the piece, of the curve at the extreme points of the shape) and a description, in terms of music-theoretic features, of the structure and the notes contained in it (e.g., note duration, harmonic function, metrical strength, ...).

The *output* of *IBL-Smart* is a set of general rules that decide, given the description of a musical structure, what kind of expressive shape should be applied to it, and exactly *how much* crescendo, accelerando, etc. should be applied.

Applying learned rules to new problems is then rather straightforward: given the score of a new piece (melody) to play expressively, the system again first transforms it to the abstract structural level by performing its musical analysis. For each of the musical structures found, the learned rules are consulted to suggest an appropriate expressive shape (for dynamics and rubato). The interpolation tables associated with the matching rules are used to compute the precise numeric details of the shape. Starting from an even shape for the entire piece (i.e., equal loudness and tempo for all notes), expressive shapes are applied to the piece in sorted order, from shortest to longest. Expressive shapes are overlayed over already applied ones by averaging the respective dynamics and rubato values. The result is an expressive interpretation of the piece that pays equal regard to local and global expression patterns, thus combining micro- and macro-structures.

5.1 An experiment

Here are some results of an experiment with waltzes by Frédéric Chopin. The training pieces were five rather short excerpts (about 20 measures on average) from the three waltzes Op.64 no.2, Op.69 no.2, and Op.70 no.3, played by the author on an electronic piano and recorded via MIDI. The results of learning were then tested by having the system play other excerpts from Chopin waltzes.

As an example, figure 11 shows the system's performance, in terms of both loudness and tempo variations, of the beginning of the waltz Op.18 after learning from the five training pieces. Again, values of 1.0 mean average loudness or tempo, higher values mean that a note has been played louder or faster, respectively. The arrows have been added by the author to indicate various structural regularities in the performance. Note that while the written musical score contains some explicit expression marks added by the composer (e.g., commands like *cresc*, *sf* or *p* and graphical symbols calling for large-scale crescendo and decrescendo), the system was not aware of these; it was given the notes only.

In a qualitative analysis, the results look and sound musically convincing. The graphs suggest a clear understanding of musical structure and a musically sensible shaping of these structures, both at micro and macro levels. At the macro level (arrows above the graphs),

[3]Since there is no explicit domain theory any more in this approach, we have used FOIL (Quinlan, 1990) as the symbolic learning component of IBL-SMART in all the experiments described below.

Figure 11: Waltz op.18, E♭ major, as played by learner: *dynamics* (top) and *tempo* (bottom).

for instance, both the dynamics and the tempo curve mirror the four-phrase structure of the piece. In the dynamics dimension, the first and third phrase are played with a recognizable crescendo culminating at the end point of the phrases (the B♭ at the beginning of the fourth and twelfth measures — see positions (beats) 9 and 33 in the plot). In the tempo dimension, phrases (at least the first three) are shaped by giving them a roughly parabolic shape — speeding up at the beginning, slowing down towards the end. That agrees well with theories of rubato published in the music literature (e.g., Todd, 1989).

At lower structural levels, the most obvious phenomenon is the phrasing of the individual measures, which creates the distinct waltz 'feel': in the dynamics dimension, the first and metrically strongest note of each measure is emphasized in almost all cases by playing it louder than the rest of the measure, and additional melodic considerations (like rising or

falling melodic lines) determine the fine structure of each measure. In the tempo dimension, measures are shaped by playing the first note slightly longer than the following ones and then again slowing down towards the end of the measure.

The most striking aspect is the close correspondence between the system's variations and Chopin's (or the score editor's) explicit expression marks (which were not visible to the system!). The reader trained in reading music notation may appreciate how the system's dynamics curve closely parallels the various crescendo and decrescendo markings and also the *p* (*piano*) command in measure 5. Two notes were deemed particularly worthy of stress by Chopin and were explicitly annotated with *sf* (*sforzato*): the B♭'s at the beginning of the fourth and twelfth measures. Elegantly enough, our program came to the same conclusion and emphasized them most extremely by playing them louder and longer than any other note in the piece; the corresponding places are marked by arrows with asterisks in figure 11.

6 A machine learning analysis of real artistic performances

All experiments so far used performances by the author himself as training examples. One might be concerned about a possible bias in these data, intentional or inadvertent. (Though, given the author's far from perfect piano technique and the rather poor keyboard on which the examples were recorded, whatever bias there may be in the data is definitely dominated by involuntary errors and noise).

This section briefly describes experiments performed with *real* data, that is, performances of a complete piece by a number of internationally famous pianists. The results shed some light into significant differences in personal performance styles between different artists. The experiments have also helped us pinpoint a number of weaknesses of the current apprach. Appropriate refinements of the strategy and the music-theoretic vocabulary are currently under way. We cannot present a detailed discussion here — the following is only intended to give the reader an impression of the complexity of the phenomenon and a glimpse of the results we have achieved so far. Further details can be found in (Widmer, 1995b).

The piece in question is Robert Schumann's romantic piano piece *"Träumerei"* (from *"Kinderszenen"*, op. 15). Figure 12 shows the score of the entire piece. Bruno Repp (1992) has measured the tempo deviations in 28 performances of this piece by 24 well-known pianists. This data set was used as the basis for a suite of experiments. Repp's data only capture the dimension of expressive timing (tempo), dynamics was not taken into account.

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.

Three pianists from the top of Repp's list — Claudio Arrau, Vladimir Ashkenazy, and Alfred Brendel — were chosen for the first experiment. Their performances of the second part of the *Träumerei* were used as training examples. The respective tempo curves are shown in figure 13. (To facilitate an easier comparison of several curves, we are using a slightly different plot style here). As before, the labels on the x axis indicate the absolute distance from the beginning of the piece in terms of quarter notes ("score time"). The plot represents the relative tempo variations — the higher the curve, the faster the local tempo.

It is quite evident 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

Figure 12: "Träumerei" by Robert Schumann (from the 'Urtext Edition', W. Boetticher (ed.), G. Henle Verlag, Munich, 1977).

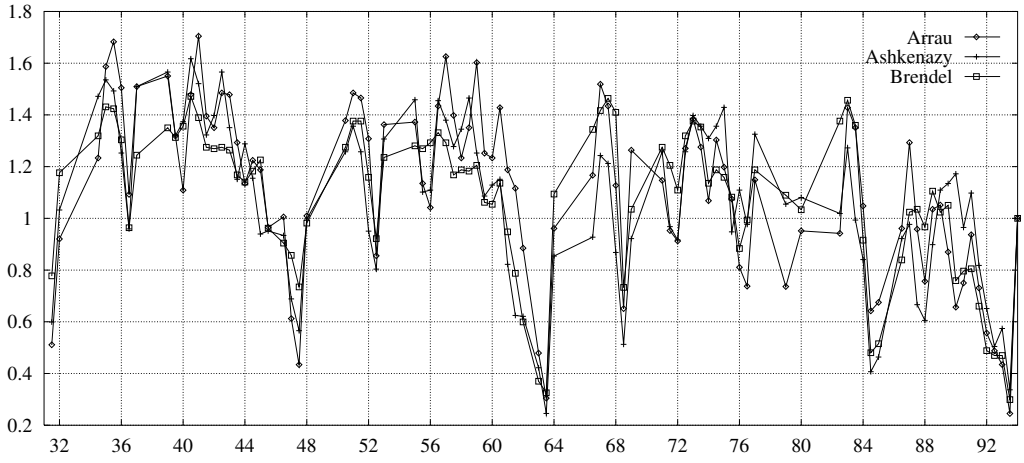

Figure 13: Second part of *Träumerei* as played by three pianists (tempo curves).

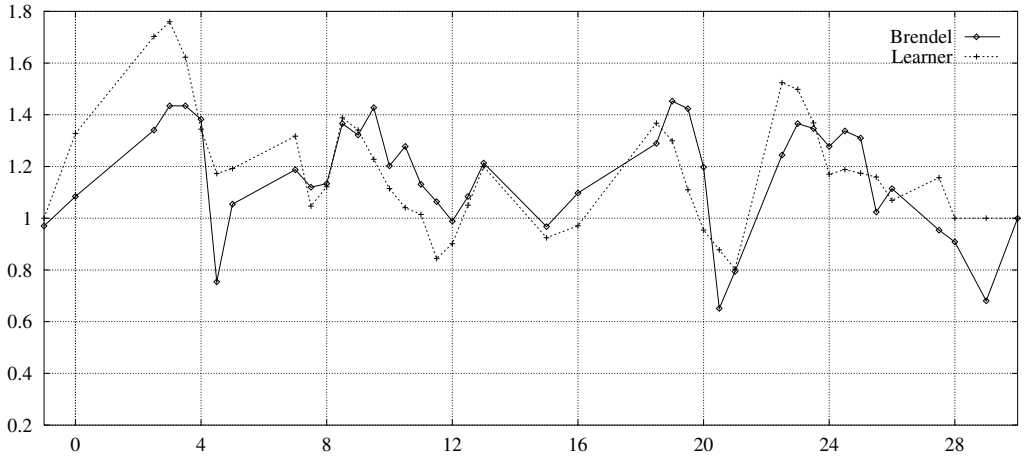

Figure 14: Comparison learner – Brendel on test piece (first part of "Träumerei").

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 is due to a *fermata* in the score.

Figure 14 shows the system's performance of the test piece (the first part of the *Träumerei*) after learning from these three examples and compares it to one of its teachers' (Brendel's) performances of the same piece.

The plot shows considerable agreement in the overall, high-level trends, but also some discrepancies in the finer details (e.g., the finer phrasing structure in measures 3 and 7). Some of these discrepancies point to shortcomings of our current system. For instance, the system fails to replicate Brendel's way of phrasing the small melodic motifs in measures 3 and 7. Deeper analysis revealed that this is due to the limited set of abstract *expressive shapes* (see section 5) that the learner can identify in a given performance curve. We are planning to introduce more complex abstract patterns into the learner's shape vocabulary. Generally, however, we consider the result very satisfactory, especially given that the performances of the three teachers, though fairly similar at a high level, are quite different at lower levels.

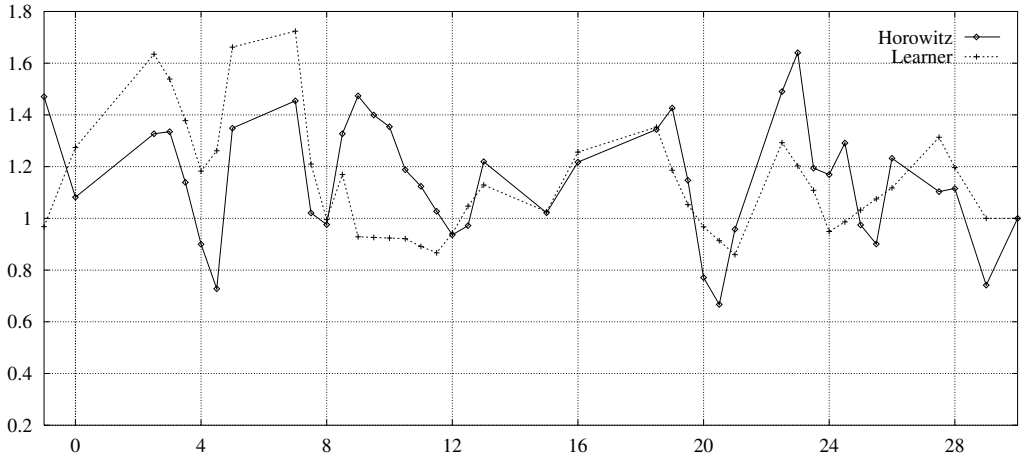

Figure 15: Comparison learner – Horowitz on test piece (first part of *Träumerei*).

Another interesting dimension that can be explored with the help of machine learning is personal style differences between individual artists. Repp's data collection also includes three performances by Vladimir Horowitz, who is known for his very distinctive interpretations. In another experiment, the three performances by Horowitz (again only of the second part of the piece) were used as training examples. Figure 15 shows the system's performance of the test piece after learning from the three Horowitz examples, and compares it to one of Horowitz's performances.

It is quite obvious that Horowitz's performance is indeed very different from, say, Alfred Brendel's (cf. figure 14). The learner does seem to manage to replicate part of the Horowitz style, but not as well as that of more 'standard' interpretation styles such as Brendel's. We cannot give a conclusive explanation at this point, but one may conjecture that Horowitz's style is more idiosyncratic, his performance decisions cannot be so easily related to or 'explained' by obvious structural features of the music. We do expect that further analysis of the learned rules and more detailed experiments will provide insights into specific aspects of performance differences that may be of interest to musicology in general. In any event, we can show experimentally that the two knowledge-based approaches to learning are superior to learning without musical knowledge (see the next section).

7 Discussion of experimental results

In the introduction to this chapter, it was claimed that, as an interdisciplinary project, our work should produce results of interest to both disciplines involved. The example results presented in the previous sections have hinted at some of these. Here, we will look at the results a bit more closely, both from a machine learning and a musicology perspective.

7.1 Quantitative analysis

From the viewpoint of machine learning, the main contribution of this project is the introduction and comparison of two different approaches to knowledge-based learning: the first consists in making incomplete and very imprecise domain knowledge explicit in the form of a qualitative domain theory and devising an inductive learning algorithm that uses the theory

	naive approach (no knowledge)	approach 1 (qual. domain theory)	approach 2 (abstraction)
matches/accelerando	58.46 %	61.54 %	55.38 %
matches/ritardando	50.91 %	54.55 %	78.18 %
Total matches	55.00 %	58.33 %	65.83 %

Table 1: Percentage of agreement between learner and teachers (unweighted).

	naive approach (no knowledge)	approach 1 (qual. domain theory)	approach 2 (abstraction)
matches/accelerando	61.93 %	58.88 %	57.87 %
matches/ritardando	40.83 %	55.03 %	76.92 %
Total matches	52.19 %	57.10 %	66.67 %

Table 2: Percentage of agreement (weighted by metrical strength).

to guide its heuristic search. The alternative approach uses domain knowledge to transform the training examples and the entire learning problem to musically plausible abstraction levels. Our results with Bach minuets, briefly hinted at in section 4.4, weakly indicated that the introduction of additional knowledge (in that case through the first approach) does indeed improve the learning results. However, one would like to obtain quantitative results that clearly prove that hypothesis.

A fundamental problem with our application domain, at least from a machine learning point of view, is that a precise quantitative evaluation of the results is not possible. The musical quality of an expressive performance cannot be quantified. There is no one 'correct' interpretation, aesthetic judgements can depend on many extra-musical factors, and global qualities like the coherence or balance of a performance are very difficult to formalize. Nonetheless, we have performed some simple measurements in order to at least get some weak indications as to the relative merits of our learning approaches.

For instance, we experimentally compared three algorithms on the Schumann learning task: algorithm 0 (the 'naive' algorithm) is IBL-SMART *without* any domain knowledge, thus restricted to purely empirical learning. Algorithm 1 is the same system *with* the qualitative domain theory, learning at the note level as described in section 4, and algorithm 2 is IBL-SMART with knowledge-based abstraction as described in section 5. Each of the three algorithms was trained on the performances of the second piece of the *Träumerei* by the three pianists Claudio Arrau, Vladimir Ashkenazy, and Alfred Brendel. The learned rules were then applied to the first part of the piece, and the resulting performances were compared to the respective performances by the three 'teachers' by counting the number of agreements of categorical decisions (i.e., how often both the pianist and the learner applied a *ritardando* or an *accelerando* to a note). Table 1 summarizes these 'predictive accuracy' measurements, averaged over all three pianists. The reader should keep in mind that an agreement of 100% is strictly impossible, as the three pianists' performances differ in a lot of details.

The summary line (the total percentage of matches) indicates significant advantages of the knowledge-based systems over the learner without domain knowledge. And among the

former, approach 2 (knowledge-based abstraction) clearly outperforms approach 1 (learning at the note level). That confirms our previous qualitative evaluations (by musical analysis and listening tests) and also supports the theoretic hypothesis that structure abstraction is musically more plausible than direct application of knowledge at the level of individual notes.

But such quantitative results should be taken with a grain of salt. Simply counting the number of matching decisions is far too simplistic. Not every note in a piece is equally important, and some errors are far more critical than others. All that depends in a complex way on aspects of the musical context. A musically meaningful comparison should take all the relevant factors into account, but that would presuppose a complete theory of 'correct' interpretation, a thing which obviously does not exist (which is why we started our empirical research in the first place).

As a first approximation to a more elaborate comparison, table 2 lists the results of the same experiment if we apply a simple *weighting scheme* to the counts: each match/mismatch between system and teacher is weighted by the relative *metrical strength* of the underlying note. This is meant to be a *very* rough measure of the relative importance of notes. In the weighted analysis, the differences between the three learners come out even more clearly, with the abstraction-based approach winning by a big margin. Whatever the ultimate musical validity of these measurements, they do provide strong evidence for the utility of the musical background knowledge and the effectiveness of our knowledge-based learners.

7.2 Useful qualitative results for musicology

From the perspective of musicology, the qualitative aspects of our results are more informative. Generally, since the domain knowledge — be it in the form of a domain theory or in the form of abstraction operators — is based on two recent theories of tonal music, the musical quality of our learners' expressive performances (and the superiority over learning without knowledge) provides additional empirical evidence for the relevance of these music theories.

More detailed insights can be gained by directly inspecting the learned expression rules. For instance, an analysis of rules learned from different types of music have revealed different structural dimensions of the music to be relevant (Widmer, 1995a). Also, experiments have shown that while abstraction to the structure level generally provides better results for various types of classical music, for other styles like jazz the note level is more adequate — note level rules perform better and have more explanatory potential.

A very interesting result was that the system in effect re-discovered variations of some expression rules that were postulated by music theorists some years ago (e.g., Sundberg et al., 1983; Friberg, 1991), based mainly on musical intuition and experience. For instance, one of the rules discovered by our learner reads:

```
ritardando( Note, X) :-
   interval_prev( Note, I),
   at_least( I, maj6),
   dir_prev( Note, up).
```

which may be paraphrased as *"Increase the duration (by a certain amount X) of all notes that terminate an upward melodic leap of at least a major sixth."* This is a specialization of rule 4 from (Sundberg et al., 1983), which increases the duration of all notes that terminate a melodic leap (in either direction). Several other variants of Sundberg rules were discovered

through learning. Our experiments have thus produced additional empirical support for the appropriateness of the Sundberg rules. These results are also the starting point for new investigations with various music-theoretic vocabularies that we are currently performing in cooperation with Johan Sundberg and colleagues.

8 Conclusion

This chapter has shown how machine learning can profitably be applied to the study of real problems in the field of tonal music. Compared to 'hard' sciences like physics and chemistry, music is in many ways 'softer' — many aspects are not quantifiable, and that makes it difficult to perform the kinds of precise experiments and analyses that are considered the norm in inductive learning research. Nonetheless, machine learning can make useful qualitative contributions, for instance to the empirical evaluation of existing theories of the domain. Prerquisites for the success of such projects are a thorough analysis of the application domain and existing theories thereof, and a conscious approach to domain modelling. That includes the careful design of vocabulary and representation language, which can contain (and hide) a lot of domain-specific knowledge and implicit assumptions.

Our projects have yielded a number of interesting musical results, and we view our "analysis-by-resynthesis" approach (i.e., having machine learning programs reproduce observed phenomena and analyzing the results) as a viable alternative or addition to more traditional methods in musicology (Widmer, 1994b).

From a machine learning perspective, such interdisciplinary projects can be beneficial as well: new application domains can motivate the development of new learning models and algorithms, which need not at all be domain-specific. Our algorithm IBL-SMART, for instance, is a general inductive learner that may well be useful for other classes of applications.

Future work in this project will concentrate primarily on aspects of domain modelling. Experiments with different music-theoretic vocabularies and different types of music will give us a more detailed insight into the regularities and possible explanations of performance styles. The compilation of a large collection of real performance data turns out to be difficult (mainly for copyright reasons), but it will be essential to the success of this enterprise.

Acknowledgments

The author would like to thank Bruno Repp for the permission to use his collection of Schumann performance data. Financial support for the Austrian Research Institute for Artificial Intelligence is provided by the Austrian Federal Ministry for Science, Research, and Arts.

References

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

Bergadano, F. and Giordana, A. (1988). A Knowledge Intensive Approach to Concept Induction. In *Proceedings of the Fifth International Conference on Machine Learning*, Ann Arbor, MI.

Bergadano, F., Giordana, A., and Ponsero, S. (1989). Deduction in Top-Down Inductive Learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, Ithaca, N.Y.

Collins and Michalski 1989. The Logic of Plausible Reasoning: A Core Theory. *Cognitive Science* 13(1), pp. 1–49.

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

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

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

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Michalski, R.S. (1983). A Theory and Methodology of Inductive Learning. In R.S. Michalski, J.G. Carbonell and T.M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach*, vol. I. Palo Alto, CA: Tioga.

Mitchell, T.M., Keller, R.M., and Kedar-Cabelli, S.T. (1986). Explanation-Based Generalization: A Unifying View. *Machine Learning* 1(1), pp. 47–80.

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

Narmour, E. (1977). *Beyond Schenkerism: The Need for Alternatives in Music Analysis.* Chicago, Ill.: Chicago University 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.

Russell, S.J. (1989). *The Use of Knowledge in Analogy and Induction.* London: Pitman.

Shavlik, J.W., Towell, G., and Noordewier, M. (1992). Using Neural Networks to Refine Biological Knowledge. *International Journal of Genome Research* 1(1), pp. 81–107.

Sloboda, J. (1985). *The Musical Mind: The Cognitive Psychology of Music.* Oxford: Clarendon Press.

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

Todd, N. (1985). A Model of Expressive Timing in Tonal Music. *Music Perception* 3, pp. 33–59.

Widmer, G. (1993a). Understanding and Learning Musical Expression. In *Proceedings of the International Computer Music Conference (ICMC-93)*, Tokyo, Japan.

Widmer, G. (1993b). Combining Knowledge-Based and Instance-Based Learning to Exploit Qualitative Knowledge. *Informatica* 17, Special Issue on Multistrategy Learning, pp. 371–385.

Widmer, G. (1994a). The Synergy of Music Theory and AI: Learning Multi-Level Expressive Interpretation. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, Seattle, WA. Menlo Park, CA: AAAI Press.

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

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

Widmer, G. (1995b). A Machine Learning Analysis of Expressive Timing in Pianists' Performances of Schumann's "Träumerei". In J. Sundberg (ed.), *Proceedings of the Stockholm Symposium on Generative Grammars for Music Performance.* Royal Institute of Technology (KTH), Stockholm, Sweden.