Modeling the Rational Basis of Musical Expression

Gerhard Widmer

Austrian Research Institute for Artificial Intelligence, and
Department of Medical Cybernetics and Artificial Intelligence
University of Vienna,
Freyung 6/2
A−1010 Vienna, Austria
gerhard@ai.univie.ac.at

Expressive interpretation and performance of written music is a phenomenon of central interest to music research. As a manifestation of human musical competence it is a challenge to musicologists and psychologists alike, and as a specifically artistic phenomenon, expressive performance is a permanent topic of artistic debate and music-theoretic analyses. Not surprisingly, this interest has also inspired research into formal models that try to identify the regularities and mechanisms underlying musical expression.

Music theorists have tried to relate expression patterns to musical structure. In most cases, the size and shape of expressive gestures are assumed to be more or less closely linked to the phrase or grouping structure, at various levels, of the piece being performed. Some formal computational models have also been proposed. For instance, Sundberg et al. (1983) have presented a system of rules that express some simple hypotheses about local expression patterns. The general approach was to derive hypotheses by intuition and then to test them by experimentation (”analysis-by-synthesis”). Other authors take a very different approach, relating the subject to extra-musical phenomena, for instance by hypothesizing analogies between expression and general motion in physical space (e.g.,Todd 1990).

Common to all these models is the assumption that there is some rational basis to expressive interpretation—in other words, that expression is not, as public opinion would have it, an intangible, inexplicable, ”artistic” phenomenon that comes out of thin air, but can in part be traced back to structural features of the music performed, and to the performer’s awareness of these structures.

The objective of the research described here is to investigate this hypothesis more thoroughly, with the help of Artificial Intelligence (AI) methods. Our starting point is the following question (a typical AI question, if you will): is there general musical knowledge that enables one to ”understand” and ”explain” expressive variations, and can this knowledge be made explicit and modeled in a computer program? The research methodology employed in this project is similar to the philosophy followed in (Widmer 1992), where a program was developed that could learn to harmonize melodies. First, we will hypothesize what common, general musical knowledge might be relevant to expressive performance and might, at least in part, ”explain” the phenomenon. This knowledge will then be encoded in an explicit formal model, at the appropriate level of abstraction; care must be taken to ensure that the model is as plausible as possible, both musically and psychologically. And finally, the adequacy of this model will be tested empirically by incorporating it into a learning program that uses the knowledge to learn general performance rules from actual performances by human musicians.

In the project described in (Widmer 1992), this methodology turned out to be quite successful. A simple, abstract model describing roughly how listeners perceive harmonized melodies was constructed, and it was then shown that this model enabled a computer program to learn to harmonize new melodies more effectively. In the case of expressive performance, we start from the hypothesis that one function of expression is to give the listener cues as to the intended structural interpretation of a piece. Given that this hypothesis is correct, certain aspects of expressive performance should become explainable once we have a model of musical structure understanding.

In fact, certain aspects of structural hearing and music comprehension have been spelled out in some recent theories of tonal music, specifically, the theories of Lerdahl and Jackendoff (1983) and Narmour (1977). They will be used as the basis for our computational model of musical knowledge. In this way, the project may also provide empirical evidence for the adequacy and relevance of Lerdahl and Jackendoff's and Narmour's theories. If they contribute to more effective and "intelligent" learning, we have indirect evidence that the structures they postulate are indeed relevant to real musical behavior.

Assumptions and Limitations

The phenomenon to be modeled is expressive interpretation of written music. We wish to develop computer models and programs that exhibit musical competence by producing musically sensible, if not really original and artistic, interpretations (performances) of given pieces. Expressive interpretation means variations in tempo, dynamics, relative note duration, etc. The investigation is restricted to the purely musical level; we do not deal with actual physical performance with its attendant physical and physiological aspects.

However, the objective is not just to write a computer program that produces the desired output. We also want to understand what the musical basis is that enables people to produce sensible interpretations, and allows students to learn this skill (to a greater or lesser extent). Thus, musical and psychological plausibility of the program and the knowledge encoded therein is of importance if the results of this study are to have any theoretical significance.

The approach to be presented here is guided by two important assumptions, the first of which (at least) is widely shared among music researchers. The first assumption is that one important role of expressive interpretation is the communication of an understanding of musical structure, from the performer to the listener (see, e.g., Sundberg et al. 1991). Expression serves to emphasize structure. Consequently, we may assume that expression becomes partly explainable if we know the relevant structural dimensions, how they are perceived by human listeners, and how they can be made explicit by expressive variation along various dimensions.

The second assumption touches on the AI aspect of this work. It is our belief that knowledge about the human perception of musical structure can be described explicitly and encoded in a computer program in operational form. Both Lerdahl and Jackendoff's theory and Narmour's Implication–Realization model capture relevant aspects of musical understanding as it is exhibited by most experienced listeners of Western tonal music. Additional knowledge concerning the connection between musical structure and actual expression decisions can be specified at least in qualitative form.

The hypothesis to be tested is whether a computer system, if equipped with such basic musical knowledge, will be able to learn some important principles of expressive interpretation from example performances (and to learn these more effectively than if it did not have this knowledge).

In our current implementation, a number of restrictions have been applied in order to make the task manageable. We deal only with classical Western tonal music; no claim is made that the model of musical knowledge will be relevant for the musics of other cultures. The current system does not handle full polyphony; it deals only with melodies, accompanied by a functional harmony notation. The expression dimensions investigated are rubato and dynamics. Other expression devices like articulation, vibrato or variations in timbre are currently not considered. We do believe, however, that the same or similar methods will be applicable to these other dimensions.

Finally, it is clear that artistic considerations do not play any role in this research. It is the purely "rational" components of the skill that are to be modeled, and the results should also be viewed in this light.

A Qualitative Model of Musical Knowledge Related to Expression

When constructing a model of (musical) knowledge, one has to be judicious about what to include in the model. In a project where one of the goals is to empirically test the validity of a musical hypothesis, it is important that the "knowledge" modeled be psychologically plausible and that it can be assumed to be common, shared by most listeners in our musical culture. The model should also be stated in a general way and should not be tailored specifically to the way it is going to be used (in our case, for learning).

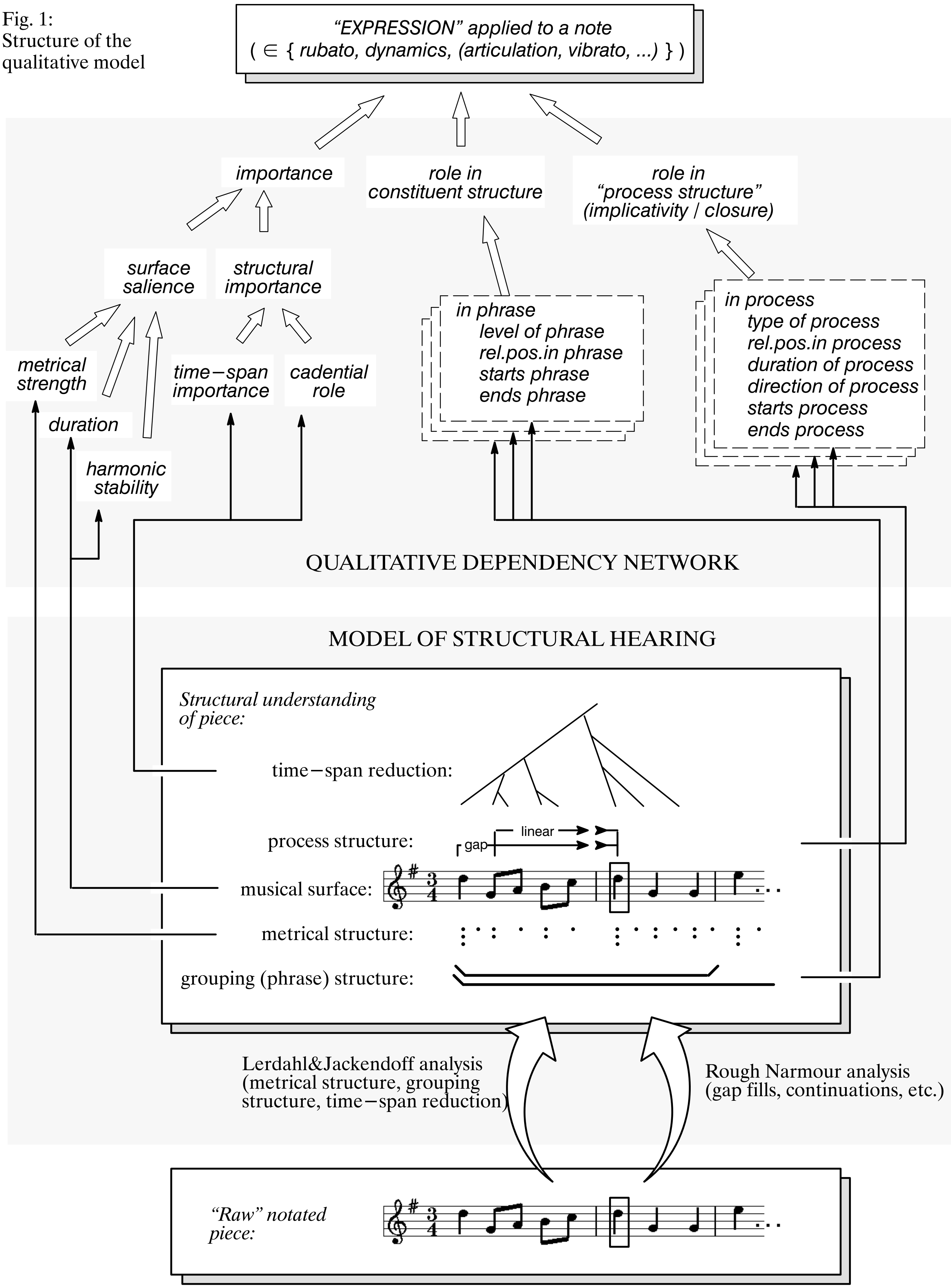
Our model rests in large part upon two well-known theories of music, namely, Lerdahl and Jackendoff's "Generative Theory of Tonal Music" (1983) and Narmour's "Implication–Realization Model" (1977). Both of these can be regarded as plausible theories of musical listening, in that they attempt to formalize the results of structural hearing, if not the actual process. Other parts of the model (the qualitative dependencies) are of a more intuitive origin. They embody our hypothesis that structure relates to expression in a certain way. This is to be tested empirically.

Accordingly, the model consists roughly of two components (figure 1): a model of structural hearing that identifies relevant structures in a piece of music as they may be perceived by a listener (or the performer), and a higher-level dependency network that relates these structural features to the target concept "expression" in various ways. While the structural hearing model is based on established theories of music and is quite precise and crisp, our intuitions as to how musical structures relate to actual expression patterns are rather weak, and therefore the higher-level dependency network can only be stated in a very abstract, qualitative form. That it is still useful will become clear when the learning strategy is presented.

The Model of Structural Hearing

The purpose of the model of structural hearing (sketched in the lower part of figure 1) is to enable the computer to "hear" a piece of music in a way similar to how human listeners would

Fig. 1:
Structure of the
qualitative model



”hear” or ”understand” it. ”Hearing” is used here in a very abstract sense, namely, as the perception of a piece of music at the level of symbolic notes. Structural hearing then, in accordance with Lerdahl and Jackendoff (1983), is the ability to understand music by recognizing complex structures in presented musical surfaces, and to organize heard pieces around these structures. (Of course, there are many other aspects of musical listening—e.g., psychoacoustic constraints, real-time processing considerations, etc.—that certainly do contribute to what structures can and cannot be perceived by listeners. For technical reasons, however, we are forced at the moment to remain at the purely symbolic level).

There are various dimensions along which structure can be perceived. Lerdahl and Jackendoff's *Generative Theory of Tonal Music* models four such dimensions via four distinct sets of rules (grouping or phrase structure, metrical structure, time-span reduction, prolongational reduction). Three of these are used in our model. Lerdahl and Jackendoff's theory stresses the hierarchical nature of musical organization; with very few exceptions, all of the structures it produces are strictly hierarchical. The theory is attractive both for its elegance and for the fact that it is formalized in a set of rules that can almost directly be implemented in a computer program.

Eugene Narmour's *Implication−Realization Model*, on the other hand, argues against strictly hierarchical structures. It suggests that linear connections at various levels and complex interdependencies between such connections give a more truthful picture of the structure of music as actually heard. Narmour's theory is of interest for our listening model because it identifies obvious connections between events in the music surface that would go unnoticed in an analysis according to Lerdahl and Jackendoff's principles.

For instance, in the Bach minuet excerpt shown in figure 1, using only the *Generative Theory of Tonal Music* for an analysis, there would be nothing that indicates a relationship between the G in measure 1 and the D at the beginning of measure 2. In particular, the time-span reduction treats both of them as subordinate to the starting D, with no connection between them. It is the Narmour-style analysis that reveals a significant and perceivable relationship between these two notes: they form the beginning and the end, respectively, of a stepwise ascending line—a ”linear melodic continuation”, in Narmour's terms—that is certainly heard as such by human listeners. The G starts a goal-directed motion, and the D represents the goal and point of closure of this motion. Between them, the two notes define a perceptually relevant group of notes that will usually be associated with a certain type of expression pattern (most likely crescendo, and maybe an accelerando).

The full implication−realization model as described in (Narmour 1977) is very rich; it produces recursive, multi-level analyses, and that is indeed where its power to identify dependencies and interactions between various ”processes” comes from. However, some aspects of the full theory are extremely difficult to operationalize, and we are not aware of any computer implementation of the theory. So for the moment, concepts from Narmour's theory are used in our model only at a very superficial level, to identify common surface patterns in the music that are kind of ”archetypal”, at least in Western tonal music, and that usually receive certain kinds of treatment by a performer.

The current model of structural hearing, then, comprises five levels of structural interpretation. The first level is the *musical surface* itself, i.e., the individual notes and rests as they appear in the notated piece. That includes, for each note, features such as its pitch and duration, its predecessor and successor notes, the underlying harmony, etc. While we expect that many expression decisions will depend on higher structural levels, features of this ”raw” musical surface will also influence details of expression. (This is confirmed by the rules learned by the system—see the appendix.)

The next two levels of analysis are *grouping structure* and *metrical structure*, as defined by Lerdahl and Jackendoff. The *grouping structure* component partitions the musical surface into contiguous segments that are heard as units, based on features of the musical surface, on symmetry principles, etc. The theory provides a set of rules that find plausible grouping boundaries at all hierarchical levels. The perceptually most prominent level of segmentation (and the most meaningful in our context) is probably the level of phrases. Note that in some models of rubato described in the literature (e.g., Todd 1989), grouping or phrase structure is assumed to be the only relevant structural level.

The *metrical structure* component establishes the regular alternation of strong and weak beats that constitutes the metrical interpretation of a piece. Metrical structure is a hierarchical phenomenon, and the metrical strength of an individual event (note) can be calculated as the number of metrical levels in which the event is classified as a strong beat. Metrical strength is one of the factors that determine the stability of individual musical events, and as such is very likely to play an important role in expression decisions.

The fourth level, denoted *process structure* in figure 1, is based very loosely on some concepts from Narmour's *Implication−Realization Model*. The goal is to detect and mark common melodic and rhythmic surface patterns like ascending or descending lines ("linear continuations"), arpeggio-like figures ("triadic continuations"), melodic gap fills of various sorts, harmonic departures and returns, and common rhythmic figures (e.g., "rhythmic gap fills"). For example, in the excerpt in figure 1, the first two notes (D−G) create a melodic gap (a leap) that is consequently filled by a stepwise ascending line G−A−B−C−D, such that the final note D of this pattern creates a certain sense of closure. At the same time, the same notes also constitute what Narmour might call a "rhythmic gap fill" figure—a "countercumulative" duration pattern (quarter note D followed by eighth note G) followed by a sequence of short notes (A−B−C), and finally finished again by a longer note (D). The D in the second measure is also the goal note of a "linear melodic continuation", i.e., a stepwise ascending line.

Our hypothesis is that all these are very typical and common patterns in Western tonal music, that they are readily perceived by a listener, and that they have a direct influence on local stress or emphasis patterns that a performer will apply. In the experiment to be described later, it turns out that patterns like "rhythmic gap fill" do indeed figure very prominently in the expression rules learned by the system. Note again that in the current model this analysis is applied only at the surface level, to find common linear patterns. No attempt is made to take the analysis further, into higher hierarchical levels.

Finally, the fifth level of analysis is Lerdahl and Jackendoff's *Time-span Reduction*. This level of analysis establishes the relative structural importance of the events in the piece in a recursive importance hierarchy. Again, it seems clear that structural importance and stability are criteria that influence interpretation decisions. The result of the analysis is a hierarchical tree structure that mirrors the system's conception of the static structure of the piece. The analysis tree is then translated into a rough qualitative measure of structural importance of each note, and these are the features that are used in higher levels of the model to refer to a note's structural role.

To summarize, the structural hearing part of our musical knowledge model is a set of programs that analyze a given piece (melody) and re-describe each note in terms of its roles in the structural interpretation of the piece. These structural features, as well as features of the musical surface, are then related to the target level—expression—in an organized fashion through a network of qualitative dependencies.

The Qualitative Dependency Network

The upper part of the model in figure 1 is a rough sketch of some intuitions as to how structural and surface features relate to expression dimensions. Instead of just saying that structure relates to expression *somehow*, we can be a bit more precise in defining intermediate concepts and grouping various influences around these intermediate concepts. For instance (follow the leftmost branch in the upper half of figure 1), the relative importance of a note certainly has an influence on how it will be played. In assessing the general importance of a note, we can distinguish its salience at the musical surface, which again depends, among other things, on the metrical strength (derived from metrical structure), the duration (derived from the surface description), and the stability of the note in its harmonic environment (again derived from the surface description of the piece), and the structural importance of the note, which depends on its prominence in the time-span reduction and possibly on whether it plays a role in some cadence; these features are derived from the time-span component of the model.

Similarly, the role that a note plays in the constituent structure of the piece is relevant. Constituent structure being grouping structure in our case, this translates into determining what groups or phrases a note appears in, what its relative position in the phrase is, whether it starts or ends a phrase, etc. These features are derived from the grouping structure component.

Finally, the note's general role in the process structure is relevant. This role depends on which and how many patterns the note appears in, the types, duration, and possibly direction of these patterns, the position of the note within the patterns, etc. Processes or patterns in turn are derived from level four of the structural hearing model.

In other words, what the qualitative dependency network does is to relate surface and structural features of the piece step by step, through intermediate concepts at various levels, to the target concept "expression". Enriching the model with structure in this way not only corresponds to our intuition that there is some structure to the patterns of influence, but also makes it possible for the computer to find more differentiated plausible explanations of expressive gestures.

When trying to formalize these influences (the arrows in the model) in a computer program, one immediately realizes that it is impossible to state the influences precisely, in the form of rules. To begin with, we cannot even precisely quantify the parameters—on what numeric scale should one measure "surface salience", for example? These are very imprecise notions, and thus, in our model, they can only take imprecise, qualitative values that express their relative strength. For instance, the "surface salience" of a note may only be one of {extremely_low, low, medium, high, extremely_high}. Also, the relations between these parameters can only be specified in a qualitative way. We simply don't know exactly how the surface salience of an event relates to how much crescendo, say, will be applied.

The formalization appropriate to the imprecision of our knowledge is in terms of directed or undirected *qualitative dependency statements*. Two comments are necessary for a proper understanding of the following examples. The first is that the statements must be interpreted as probabilistic dependencies. They need not necessarily be correct in all cases. The important thing is that they are usually correct. This is sufficient for the learning component, which relies on plausible explanations. Second, the representation language used for the dependency statements, as for all other parts of the model, is the programming language Prolog (Sterling & Shapiro, 1986). Capitalized names denote variables. Predicates after the ":−" operator represent conditions that must be true in order for the predicate before the ":−" to be valid. Influences are then represented in the form of three types of relations, **q+**, **q−**, and **dep**:

A statement **q+(A, B)** denotes a "qualitative proportionality" between parameters *A* and *B*. Basically, it says that there is a directed dependency between the values of *A* and *B*: the higher the value of *A*, the higher will be the value of *B*. In other words, *A* contributes positively to *B*. Qualitative proportionalities are an important concept in qualitative modelling of physical systems, and the notation has been taken from (Forbus 1984). An example from our model is the relation **q+(metrical_strength(Note,MS), surface_salience(Note,SS))**, which may be paraphrased as "The metrical strength MS of a Note contributes positively to the surface salience SS of the Note—the stronger metrically, the more salient, all other things being equal."

A statement **q−(A, B)** denotes an "inverse qualitative proportionality" between parameters *A* and *B*. The interpretation is analogous to **q+**: *A* contributes negatively to *B*. An example from our model is **q−(degree_of_closure(Note, C), accelerando(Note, A))**, which means "The higher the degree of closure C of a Note (in the sense of Narmour), the lower can we expect the degree of accelerando A to be that is applied by a performer."

A statement **dep(A, B)**, finally, denotes an unspecific "undirected dependency" between *A* and *B*. It says that *A* somehow depends on *B*, among other things. In other words, when trying to find an explanation for *A*, it might be sensible to look at *B*. For instance, the statement
dep(degree_of_implicativity(Note, I), direction_of_process(P, Dir)) :− in_process(Note, P)
in our model is to be interpreted as "If a note appears in a process (surface pattern) P, then the degree of implicativity I of the Note (in the sense of Narmour) depends, among other things, on the direction Dir of that process (if it does have a specified direction). We do believe that Dir influences I, but we do not know exactly how."

All the influences indicated by arrows in the model are formalized as such dependencies. This goes also for the influences on the target category "expression". For instance, the intuition that important events will usually be emphasized by a performer is expressed in the dependency
q+(importance(Note, I), crescendo(Note, X)). ("The higher the importance I of a Note, the more crescendo (X) is likely to be applied to the note.")

In summary, the complete model as sketched in figure 1 takes a raw input piece and relates its musical surface through various analysis and abstraction steps to the target concepts—the various dimensions of expression. The claim is not that the model gives an exhaustive account of the factors that determine how a piece is played (nor even that the best or most sensible interpretation of a piece is determined at all). Other factors may be relevant, and are relevant. In order to predict how a person will play a piece, one would in principle have to know everything, from the performer's history, knowledge, artistic intentions, current mood, etc. all the way to the atmospheric conditions in the concert hall.

The claim is, however, that the factors appearing in our model are among those factors that are relevant to expressive interpretation, and that being aware of these factors enables one to at least partially explain a given performance, and from this to learn basic expression principles much more effectively than if one were reduced simply to the surface structure of presented pieces.

Note again that the model is only qualitative and abstract: it does not specify how some important event will be marked or emphasized in a performance—we don't know enough about that, and performances may vary. The model only specifies the types of situations that may be relevant, the types of features that may be indicators of such situations, and the dimensions along which stress or emphasis may occur (at the moment, variations in tempo and dynamics). As a consequence, the model cannot be used to generate sensible performances; it is simply not precise enough. It can only be used to explain, a posteriori, some aspects of a particular expressive performance, with the help of methods of plausible reasoning (Collins and Michalski 1989; Widmer 1993a). These plausible explanations can then be used by a learning system to find general rules of expressive performance. The explanations may sometimes be partially incorrect and incomplete. In such cases the learning system will nevertheless learn "correct" rules (i.e., rules that are consistent with the examples) empirically by paying more attention to the training data.

Using the Model for Effective Learning

Prior knowledge plays a fundamental role in learning. If it is meaningful and correct, it will enable the learner to converge to the correct hypotheses more quickly, requiring less training. If it is irrelevant or incorrect, it may render the learning task more difficult or even impossible. The effectiveness of learning can be measured along a variety of dimensions, for instance, the time and amount of training data needed, or the accuracy of learned concepts with respect to new test data. That is why learning experiments are a fruitful way to test the relevance of general knowledge to a problem. In our case, the relevance of the musical model was to be tested empirically by giving it to a learning program that would learn, from given performances of known pieces, general rules for expressive performance, and that would do so by using the model as its basic knowledge. The experiment would be considered successful if the results showed a significant improvement of learning with the model over learning without it.

Learning general performance rules from examples involves comparing performed pieces with the written score, determining which situations represent examples of a particular kind of expression type (e.g., crescendo), comparing these situations with situations exemplifying other expression types (e.g., diminuendo), and finding a set of general conditions that characterize all the example situations (all known examples of crescendo) and distinguish them from all other situations (all known examples of diminuendo). To put it in common machine learning terminology, crescendo situations would be considered positive examples, diminuendo and neutral situations would be negative examples. Crescendo would be the target concept, the concept that is to be characterized by general rules.

In complex domains such as the present one, where the learning examples are described by a large number of features, the number of possible consistent generalizations is potentially enormous. Some of these will reflect coincidental similarities in the data, while others will more or less closely approximate the underlying regularities and the hidden target concept. It is in such situations that given knowledge about the domain can make a big difference, if it is relevant and if the learning algorithm can take advantage of it. It may allow the learner to find a solution faster (by concentrating on relevant aspects and not wasting time on irrelevant hypotheses) and to distinguish between plausible and implausible generalizations. Moreover, if the knowledge refers to higher-level concepts that are intuitively understandable to humans, the rules produced by the system will also tend to make sense to humans.

There are some known methods in the machine learning field that can utilize given knowledge to learn more effectively. A very well-known one is *Explanation-Based Generalization (EBG)* (Mitchell et al. 1986), which uses given knowledge (background knowledge, in machne learning terminology) to construct logical explanations of the training examples and learns logically justified rules from these. Applying EBG to the expression problem would require our model of musical knowledge to be complete and correct: it would have to be sufficiently detailed to explain exactly why a performer applied a certain amount of crescendo, say, at a particular point, and why the performer could not have done otherwise. This is obviously impossible, both because of the indeterminacy of the phenomenon of expression and the necessary imprecision of our model. Another complication in our learning problem is that the musical knowledge is symbolic and qualitative, while the target concepts are inherently numeric—after all, we don't just want to learn whether to apply a crescendo or diminuendo, we also need to decide exactly how much.

For these reasons, a new learning algorithm was developed that can learn precise numeric decision rules and at the same time utilize background knowledge that is symbolic, qualitative, incomplete, and maybe partly incorrect. A presentation of the algorithm is beyond the scope of this article, and the details are not really important here. The method has been described in the machine learning literature (Widmer 1993b). Basically, given a set of positive and negative examples of the target concept (e.g., crescendo), it conducts a search for symbolic generalizations that consistently discriminate between positive and negative instances. In the process, it grows an explicit search tree. The result is a set of general symbolic rules that characterize various sub-types of the target concept (e.g., various types of situations that may call for a crescendo). For each of these symbolic rules, it also learns a numeric interpolation table, based on the training examples that are covered by the rule. When deciding how to play a new piece, the system applies all its learned rules and uses the interpolation tables associated with the matching rules to derive specific numeric values for crescendos, accelerandos, etc.

An important characteristic of the learning algorithm is that it tends to find rules that are as much as possible in accordance with the given background knowledge. In particular, it can use directed and undirected dependency statements like the ones contained in our model to guide the search and to distinguish between plausible and implausible generalizations. Generalizations that refer to features that are related to the target concept via dependency statements are preferred, and the degree to which examples are compatible with directed qualitative proportionalities is taken into account when evaluating competing generalizations. In a sense, the search tree can be viewed as a "plausible explanation" of multiple examples; it explains why certain examples are instances of the target concept by relating some of their features to the target concept via the system's background knowledge.

The level of granularity of both the learning process and the system's background knowledge is the level of the individual note: training examples are the individual notes in a piece, along with information about how much crescendo/diminuendo and accelerando/ritardando was applied to them. The structural hearing component of the musical model generates various analyses of a piece and describes each note by the roles it plays in the analyses (e.g., what its importance in time-span reduction is). In the qualitative dependency network, various features of a note are related, through dependency statements, to expectations as to which types of variation are likely to be applied to the note. The learned rules are also formulated at the note level: for a given note, the symbolic rules determine what type of variation (crescendo or diminuendo, acceleran-

Fig. 2: Beginnings of three little minuets by J.S.Bach

do or ritardando) should be applied to it, and the associated numeric interpolation tables are then used to calculate the precise amount of variation.

A technical question is how to define crescendo or ritardando in the training data. While the original Italian terms express processes evolving in time, and crescendo, for instance, is usually understood as an increase in volume relative to the previous events, we found it easier and sufficient for our purposes to define the target concepts at the level of the single note: a note is simply labelled as an instance of crescendo if its loudness level is above the average loudness of the entire piece as played by the performer. The precise degree is the ratio of actual loudness to the average loudness. With ritardando, the problem is a bit more subtle. Musicians who are not extremely well trained have a tendency to gradually and inadvertently speed up or slow down during a performance. In order to filter out this effect and allow the learner to concentrate on the performer's deliberate variations, the system keeps a moving average of the current tempo over the last five notes, and ritardando is then measured as the ratio of the note duration as actually played to the note duration that would correspond to the current averaged tempo. Diminuendo and accelerando are defined analogously.

An Experiment

The model, including the various analysis components, and the learning algorithm have been implemented in Prolog on a Sun SparcStation. Several sets of experiments with comparatively simple piano pieces were performed to determine the effect of the model on the system's ability to learn. This section reports on an experiment with three well-known minuets from J.S. Bach's "Notenbüchlein für Anna Magdalena Bach". The beginnings of the three pieces are shown in figure 2.

All three pieces consist of two parts. The second part of each piece was used for training: they were played on an electronic piano by the author, and recorded through a MIDI interface. These training passages were played with rather exaggerated expressive gesture, to provoke clear responses from the learning system. After learning, the system was tested on the first parts

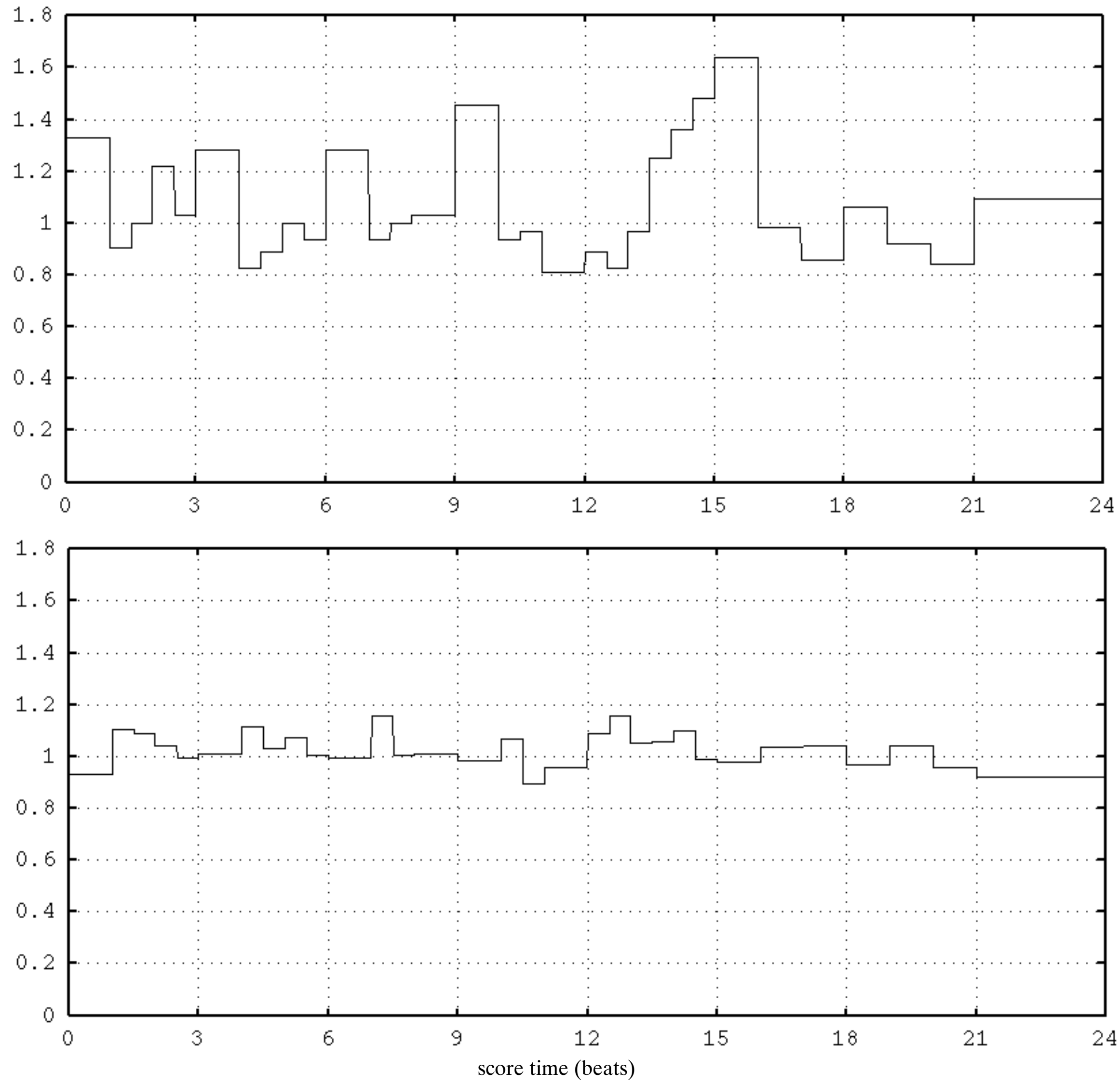


Fig. 3: Part of a training piece as played by teacher: dynamics (top) and rubato (bottom)

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). To illustrate, figure 3 shows the dynamics and the tempo curve of the first training piece (the second half of the first minuet in G major) as played by the author. The dynamics graph (upper half of figure 3) shows the relative loudness with which the individual notes were played. A level of 1.0 would be neutral (with respect to the average level), values above 1.0 represent crescendo (increased loudness), values below 1.0 diminuendo. The tempo curve (lower half) plots the local tempo; values above 1.0 indicate a speeding up (accelerando) relative to the averaged tempo, values below 1.0 represent a slowing down (ritardando).

Each note in the three training pieces was one example. That gives a total of 222 examples. They were used twice, once for learning rules for dynamics, then for rubato. In the dynamics dimension, 79 of the notes were examples of crescendo, 120 of diminuendo (the rest were played in a neutral way). When learning about rubato, 111 of the notes were examples of accelerando, 91 were instances of ritardando. From these 2 x 222 examples, the system learned a total of 58 rules, of which 14 describe situations requiring a crescendo, 15 deal with diminuendo, 13 with ritardando, and 16 with accelerando. A representative sample of rules learned in this experiment is explained and illustrated in the appendix. The search trees constructed by the learning algorithm typically contained a few thousand nodes.

When interpreting the results, there are at least two dimensions that must be taken into account: first, what do the learned rules look like? are they comprehensible, do they make musical sense? and second, how accurate and useful is the learned knowledge? how does it perform on new, unseen pieces?

When looking at the appendix, the reader will notice that some of the learned rules do indeed make intuitive sense; they seem to reflect musically and structurally motivated regularities. On the other hand, some rules are difficult to interpret. They simply capture what we would consider spurious, coincidental patterns in the training data. This is mainly an effect of the limited number of examples that were used for learning. The larger the training set that is fed to the system, the more spurious patterns will disappear from the learned rules, and what is left will approximate the "real" regularities governing musical expression.

As mentioned above, the validity of the learned rules was also evaluated experimentally. The system was presented with some new pieces that it had not seen before, namely, the first parts of the three minuets. It applied the learned rules to the melodies and played the resulting "expressive" interpretations on the electronic piano via MIDI. When several rules applied to a note, the resulting dynamics or rubato factors were averaged. Generally, the results sound surprisingly good. The system has learned, from only three training pieces, to apply sensible variations in most situations. The pieces performed by the system exhibit similar crescendo and accelerando patterns as the author's performances. Of course, the fact that training and test pieces were very similar in style helped. Still, there are considerable surface differences between training and test pieces, and the degree of correct generalization achieved is remarkable.

To illustrate, figure 4 shows the dynamics and tempo curves of a test piece (the first half of the G major minuet) as played by the system after learning. The first thing we note is that the system's performance makes sense in itself: in the dynamics dimension, there is a clear pattern of stresses on the very strong beats (the first beat in each measure), with the first half of the eight bar phrase receiving consistently more stress than the second. There is a pronounced crescendo tendency in ascending lines, and a decrescendo pattern in the bars with three quarter notes. In the tempo dimension, there is a distinct tendency to shorten short notes (which corresponds to Friberg et al.'s (1991) "The shorter the shorter" rule), and also a slowing down at the end of the phrase ("phrase final lengthening").

There are remarkable similarities between the author's and the system's style of expression. A comparison of figures 3 and 4 reveals clear parallels in the dynamics dimension—the patterns associated with eighth note groups and especially rising lines of eighth notes are very similar, despite the fact that in terms of surface details, the melody fragments are rather different. The tempo dimension, though not as pronounced in its range of variations, also exhibits similarities between teacher and learner, most notably a tendency to accelerate towards the middle of a measure, especially with eighth note patterns, and then to decelerate again.

Returning to the learned rules for a moment, it is instructive to look at the concepts that appear most often in the rules. These are the features that the system considered most predictive or

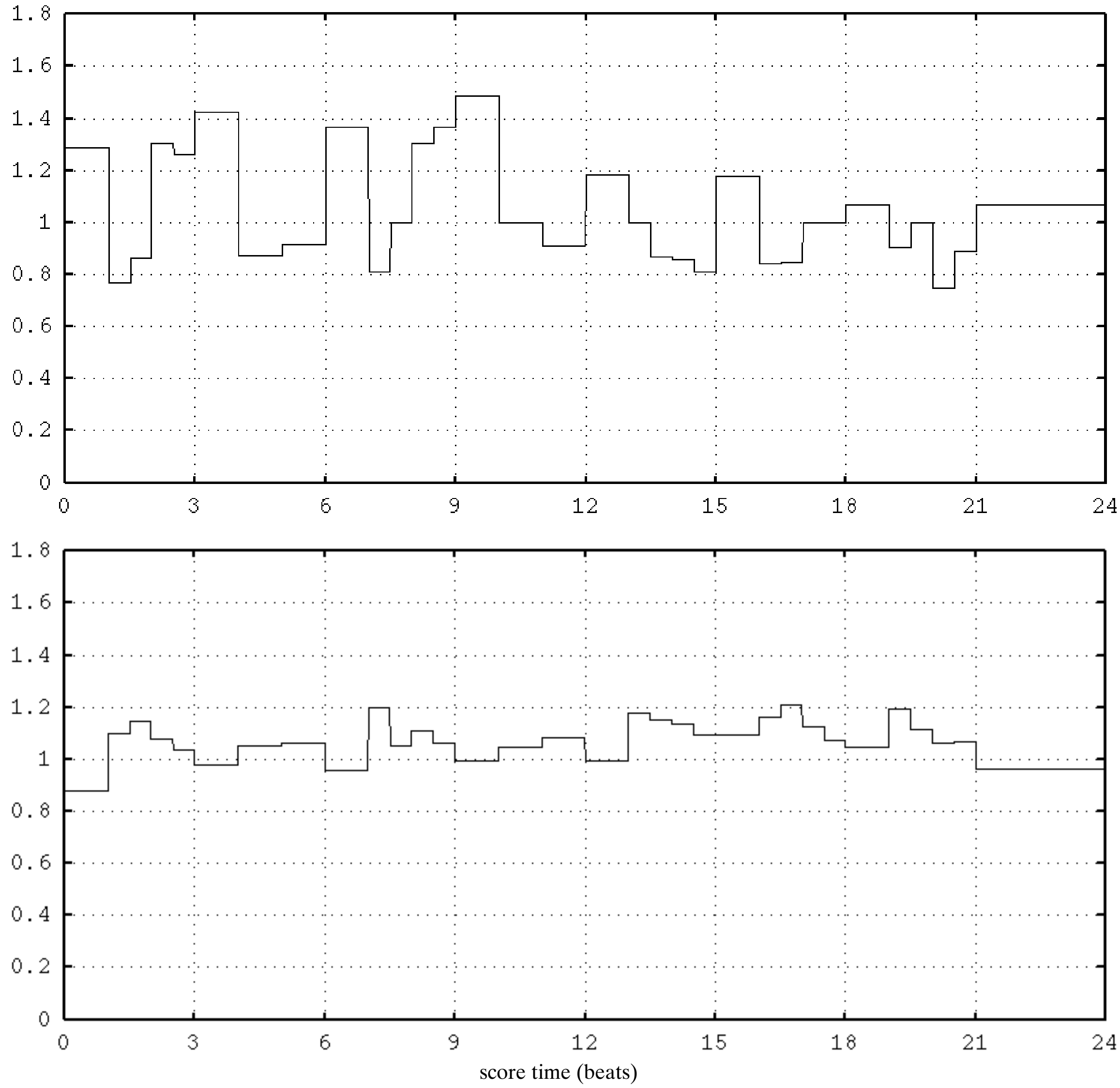


Fig. 4: Part of a test piece as played after learning: dynamics (top) and rubato (bottom)

relevant for making expression decisions. An analysis of the entire set of rules learned brings to light some interesting tendencies:

Rhythmic and metrical features like metrical strength and duration seem to be very important; they occur in a large number of rules both for dynamics and rubato.

Melodic and rhythmic surface patterns derived from Narmour's theory (especially gap fills and linear melodic continuations) appear quite often in the rules. This confirms our intuition that such linear patterns are easily perceivable and may thus play an important role in local expression decisions.

Time-span importance seems to be of some relevance; it appears in quite a number of dynamics rules (crescendo and diminuendo), to a lesser extent in rules concerning rubato.

On the other hand, there are some rules that refer to the underlying grouping or phrase structure (via the predicate *rel_position_in_phrase*), but not too many. This comes as something of a surprise, as many authors have stressed group boundaries and phrase structure as the determining factors, especially for rubato. From our results we may conclude that, at least for many local expression decisions, melodic and rhythmic surface patterns seem to be at least as important as phrase structure.

Generally, the fact that the system makes extensive use of the music-theoretic terms defined in the qualitative model indicates that the information provided by the model is really relevant. The reader should be aware that there is a large number of possible alternative rules that would also be consistent with the given examples. That the system chose the rules it did is due to the fact that these rules are more concise and explain the training data well. In this way, the empirical results indirectly provide evidence for the relevance of both Lerdahl and Jackendoff's and Narmour's theories—at least for the relevance of the structural concepts we borrowed from these theories.

The Importance of the Musical Knowledge

In order to appreciate the importance of the musical knowledge, it is instructive to run the learning program on the same training data, but without the underlying model. That is, the pieces that are input to the program are described only in terms of surface features like duration of the notes, intervals between melody notes, the underlying harmony, and the stability of the notes with respect to the underlying harmony. No other music-theoretic features (which would be the result of structural hearing) are used, and the qualitative depencendy network is omitted. In such a situation, the system can learn only in a purely empirical way, by comparing the various situations and generalizing inductively.

When this experiment was performed, the results were considerably worse. The system learned 48 rules, of which 10 dealt with crescendo, 11 with diminuendo, 15 with accelerando, and 12 with ritardando. Figure 5 shows the result, for the dynamics dimension, produced by the system on the test piece after learning in the restricted mode. When comparing this to figure 4, the deterioration is evident. The variations applied by the restricted system are of rather mixed quality. In some cases (e.g., the decrescendo patterns in measure 4 and in measure 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. Similar results were obtained with the other test pieces.

The reasons for this deterioration are equally evident. The restricted vocabulary makes it impossible for the system to find and formulate rules that refer to high-level structures; expression decisions can only be based on local features of the piece, and that is simply not sufficient in all cases. In addition, the absence of the qualitative dependency part of the model leaves the system without any knowledge about the relevance of different features to the target concept. The effect is that in the search for generalizations that are consistent with the examples, all available descriptors are tried indiscriminately, which increases the chances of finding spurious rules that just happen to fit the data but do not reflect any real regularities.

From these results we may conclude that the musical knowledge encoded in our structured model is not only sensible and relevant (it leads to useful and partly comprehensible learning results), but actually necessary: without this knowledge, learning general rules for musical expression from examples is considerably more difficult. This confirms the hypothesis that motivated the whole project.

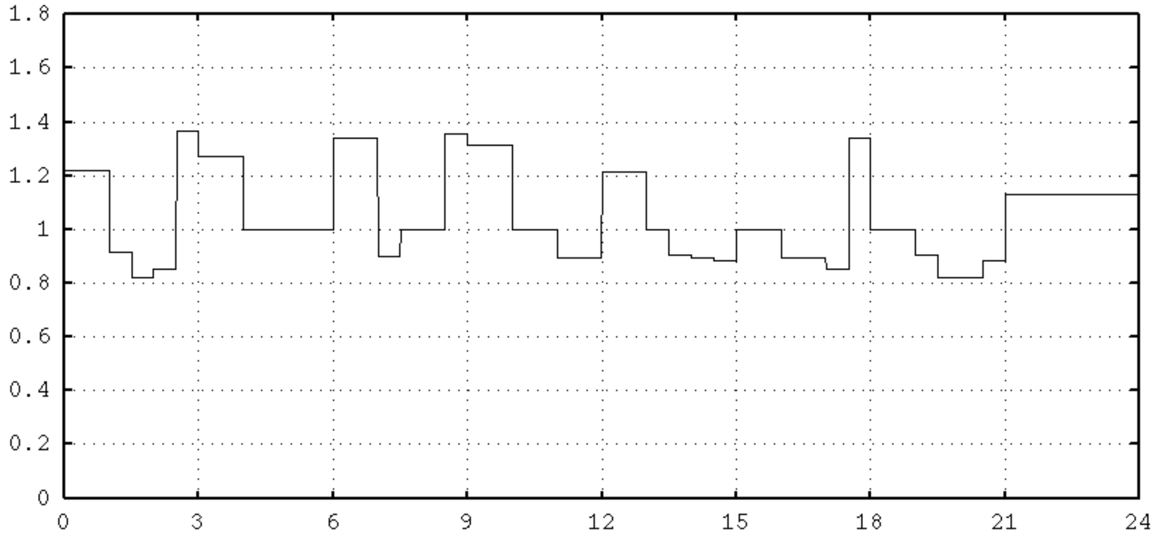


Fig. 5: Test piece as played after learning without model (dynamics)

Discussion and Related Work

As mentioned in the introduction, a number of researchers have developed formal models of various aspects of musical expression. For instance, Sundberg et al. (1983) have proposed seven rules for some simple rubato, dynamics, and articulation effects. A larger set of rules for contemporary keyboard music has been presented in (Friberg et al. 1991). While the original rules were of an extremely local nature, Sundberg and colleagues later also incorporated expression rules operating at higher structural levels (Sundberg et al. 1991).

Interestingly, our system re-discovered some of Sundberg's expression principles. Take, for instance, the following rule learned in the above experiment (see rule 56 and the associated explanation in the appendix):

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

"Increase the duration (by an amount X specified in the corresponding interpolation table) of all notes that terminate an upward melodic leap of at least a major sixth."

This is closely related to Rule 4 in (Sundberg et al. 1983), which increases the duration of all notes that terminate a melodic leap (in either direction). The nice thing is that in our approach, such rules fall out automatically, as a result of learning from specific performances.

Katayose et al. (1990) have presented an operational system that learns interpretation rules from actual performances. The system employs statistical methods (autocorrelation) to extract surface patterns from the music; these patterns are then stored, along with the way they were played. While this system seems to be quite effective, it applies its statistics solely at the surface level and thus sheds little if any explanatory light on the possible cognitive or musical basis of expression. Similar kinds of criticism have been levelled at pattern-directed composition systems such as David Cope's EMI (Cope 1992).

Desain and de Vos (1990) have also used autocorrelation measures to detect expressive timing patterns in actual performances, in the context of the POCO project (Honing 1990). The emphasis there is more on analysis than on automatic synthesis of expressive timing.

Todd (1989) has presented a quantitative model of rubato that simulates primarily ”phrase final lengthening” by a parabolic function that distorts note duration as a function of the (given) phrase structure of the piece. In a later article (Todd 1990), a causal analogy between such measured rubato patterns and the phenomenon of physical motion in space is hypothesized. (Todd 1992) presents a similar ”physical” model for the expressive dimension of dynamics.

What distinguishes our approach from these more mathematical or statistical methods is the emphasis on explicit cognitive modelling and on learning, and generally that we want to study expressive performance as a skill rather than an abstract phenomenon. The goal is to learn more about the musical knowledge and intuitions underlying this skill.

Of course, the current model is far from complete. For instance, Carlson et al. (1989) mention ”predictability” as another plausible source of expressive differentiation. This is related to the communicative aspects of music; they derive their hypothesis from observed parallels between speech and music. Basically, their claim is that exceptional, unpredictable events in speech as in music are especially important to the information content of the message being transmitted, and therefore usually receive special emphasis by the speaker or performer. Predictability could be modeled at least in part by implication denials in the Narmour-based part of our model; the point where an implication is denied would be regarded as a remarkable (unpredictable) event.

Shaffer (1989) rightly points out that there is much more to musical performance than knowledge and control: there are things like emotion, social factors and social conscience, an awareness of personal and cultural history, etc. that would have to be modeled, not to mention the physical involvedness in music-making, which adds an additional important dimension to the act of musical performance. Clearly, these issues are outside the scope of a project such as ours. Modeling emotion or social awareness is way beyond what Artificial Intelligence could hope to achieve at this stage.

There are more obvious problems that will have to be solved before we can begin to speak of a realistic model of musical expression. For instance, the system should be extended to handle free polyphony, and not just single lines. Also, as Sloboda (1985) has noted, expressive variation occurs at many levels at once; long-term variations such as gradual changes in speed and intensity may be related to large-scale aspects of musical structure, while local variations within a bar are more related to the micro-structure of a piece. A full-scale model of expression would have to explicitly deal with multiple interpretation levels.

One must also be aware of the warnings expressed in (Desain and Honing 1991), where it was argued that timing curves cannot be interpreted independently of the rhythmic substrate and the context. For instance, a particular timing curve that gives a perfect rubato performance of a piece at a particular speed may yield an unacceptable performance when the global speed of the piece is increased. Indeed, in our model, there is no explicit relation between the learned performance rules and such contextual factors as global tempo.

These and other limitations notwithstanding, we do believe that the kind of research described here is useful and can produce interesting new insights. The possibility of empirically verifying general music theories with the help of Artificial Intelligence methods provides the science of music with a new methodology. The experimental results we have produced so far (also with other musical styles) seem to confirm the hypothesis stated in the introduction, namely, that many aspects of expression can be explained by referring to the underlying musical structure, and also that the music theories used here capture aspects of structure that are relevant to expressive performance. In particular, this seems true of some of Narmour's surface patterns, such as gap fills of various sorts, which occur over and over again in the learned rules.

There is also a practical side to this project. Instruments that could learn to adapt their style of performance to human or other performers might open up interesting and exciting new possibilities, especially in interactive settings as described, e.g., in (Rowe 1992). Here, more research will be needed to develop robust and efficient learning methods that can operate in real time.

Acknowledgments

This research was sponsored in part by the Austrian "Fonds zur Förderung der Wissenschaftlichen Forschung" under grant P8756−TEC. Financial support for the Austrian Research Institute for Artificial Intelligence is provided by the Austrian Federal Ministry for Science and Research. I would like to thank Robert Trappl for his continuing support of this research. Thanks to Johan Sundberg, Paolo Petta and Erich Prem for helpful comments.

References

Carlson, R., A. Friberg, L. Frydén, B. Granström, and J. Sundberg. 1989. "Speech and Music Performance: Parallels and Contrasts." *Contemporary Music Review* 4: 391−404.

Clarke, E. 1988. "Generative Principles in Music Performance." In J. Sloboda (ed.), *Generative Processes in Music*. Oxford: Clarendon Press.

Collins, A., and R.S. Michalski. 1989. "The Logic of Plausible Reasoning: A Core Theory." *Cognitive Science* 13(1): 1−49.

Cope, D. 1992. "Computer Modeling of Musical Intelligence in EMI." *Computer Music Journal* 16(2): 69−83.

Desain, P., and S. de Vos. 1990. "Autocorrelation and the Study of Musical Expression." In *Proceedings of the International Computer Music Conference (ICMC−90)*. San Francisco: International Computer Music Association.

Desain, P., and H. Honing. 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.

Forbus, K.D. 1984. "Qualitative Process Theory". *Artificial Intelligence* 24(1−3): 85−169.

Friberg, A., L. Frydén, L. Bodin, and J. Sundberg. 1991. "Performance Rules for Computer-Controlled Contemporary Keyboard Music." *Computer Music Journal* 15(2): 49−55.

Honing, H. 1990. "POCO: An Environment for Analysing, Modifying, and Generating Expression in Music." In *Proceedings of the International Computer Music Conference (ICMC−90)*. San Francisco: International Computer Music Association.

Katayose, H., T. Fukuoka, K. Takami, and S. Inokuchi. 1990. "Expression Extraction in Virtuoso Music Performances." In *Proceedings of the Tenth International Conference on Pattern Recognition*, Atlantic City, N.J.

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

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

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

Rowe, R. 1992. "Machine Listening and Composition with Cypher." *Computer Music Journal* 16(1): 43–63.

Shaffer, L.H. 1989. "Cognition and Affect in Musical Performance." *Contemporary Music Review* 4: 381–389.

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

Sterling, L., and E. Shapiro. 1986. *The Art of Prolog*. Cambridge: MIT Press.

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

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

Todd, N. 1989. "Towards a Cognitive Theory of Expression: The Performance and Perception of Rubato." *Contemporary Music Review* 4: 405–416.

Todd, N. 1990. *The Kinematics of Musical Expression*. Manuscript.

Todd, N. 1992. "The Dynamics of Dynamics: A Model of Musical Expression." *Journal of the Acoustical Society of America* 91(6): 3540–3550.

Widmer, G. 1992. "Qualitative Perception Modelling and Intelligent Musical Learning." *Computer Music Journal* 16(2): 51–68.

Widmer, G. 1993a. "Learning with a Qualitative Domain Theory by Means of Plausible Explanations." In R.S. Michalski and G. Tecuci, eds. 1993. *Machine Learning: A Multistrategy Approach, vol. IV.* San Mateo, CA: Morgan Kaufmann.

Widmer, G. 1993b. "Plausible Explanations and Instance-Based Learning in Mixed Symbolic/Numeric Domains." In *Proceedings of the Second International Workshop on Multistrategy Learning (MSL–93)*, Harper's Ferry, W.VA.

Appendix: Some Typical Rules Learned

The following is a sample of representative rules that were learned in the experiment described in the article. Each rule is paraphrased and is accompanied by a passage from one of the test pieces where it would apply. The note to which it would apply is marked by a box.

Here is a short guide to reading the rules. Rules are written in standard Prolog notation (see Sterling and Shapiro 1986). The ":−" functor is to be read as "IF". The expression before the ":−" (the "head" of the rule) specifies the type of expressive variation to be applied to a note. The terms following the ":−" (the "body" of the rule) represent the conditions that the note must satisfy to qualify for this treatment. Conditions are implicitly connected by "AND". Variable names begin with capital letters, constants (i.e., specific values) are written in lowercase. Mnemonic variable names have been substituted by the author (internally, the system uses generic variable names like _393 or _1134).

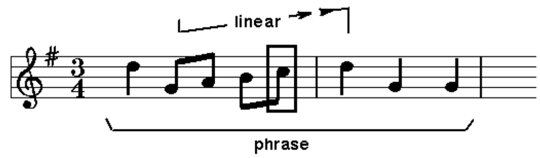
When applying a rule, the system instantiates the variable *Note* in the head with the current note that is to be played. The conditions are then evaluated for this note. If all conditions are satisfied, the interpolation table that was learned along with the rule (not shown in the following) is consulted to compute a specific numeric value for the variable X in the head. The value computed depends on the values determined for the numeric variables in the rule body during condition evaluation. The resulting value X is the precise amount of crescendo etc. that will be applied to the note.

Note durations are measured in terms of quarter notes, i.e., a duration of 0.5 equals an eighth note, 2.0 equals a half note, etc. Metrical strength is based on the metrical structure (computed according to Lerdahl and Jackendoff's rules) and is measured on a scale from 1 to the number of metrical levels found. The relative position of a note within some phrase or pattern is measured on a normalized scale from 0 to 1, where 0.0 indicates the beginning of the phrase, and 1.0 the end.

Some selected rules:

RULE 3:

```
crescendo(Note,X) :–
    rel_position_in_linear_melodic_cont(Note,Pos1),
    Pos1 > 0.42,
    dir_prev( Note, up),
    rel_position_in_phrase( Note, Pos2),
    Pos2 < 0.68,
    dur_prev( Note, PDur),
    PDur < 0.75,
    dir_next( Note, up).
```



”Apply a certain amount X of crescendo to the current note if
the note is inside a linear melodic continuation pattern (Narmour)
and its relative position (Pos1) in this pattern is in the second half (> 0.42)
and the direction of the interval leading to the current note is up
and the relative position of the note in the current phrase (Pos2) is in the first two thirds (<0.68)
and the duration of the previous note (PDur) is < 0.75
and the direction of the interval leading to the next note is up.”

RULE 7 (”emphasize long notes reached by upward motion”):

```
crescendo( Note, X) :–
    duration( Note, Dur),
    Dur > 2.0,
    dir_prev( Note, up).
```

”Apply a certain amount X of crescendo to the current Note if
the Note’s duration is greater than 2.0
and the direction of the interval leading to the note is up.”

RULE 8 (”emphasize important notes reached by upward motion”):

```
crescendo( Note, X) :–
    timespan_importance( Note, extremely_high),
    dir_prev( Note, up).
```

”Apply a certain amount X of crescendo to the current note if
the note is extremely important in the time-span reduction
and the direction of the interval leading to the note is up.”

RULE 11 (spurious?):

```
crescendo( Note, X) :–
    interval_next( Note, p8),
    dir_next( Note, down).
```

”Apply a certain amount X of crescendo to the current note if
the interval leading to the next note is a downward jump of an octave (p8).”

RULE 13 (”emphasize strong, stable events”—last condition spurious?):

```
crescendo( Note, X) :–
    metrical_strength( Note, MS),
    MS > 4.0,
    harmonic_stability( Note, high),
    timespan_importance( Note, medium).
```

”Apply a certain amount X of crescendo to the current note if
the note is metrically strong (> 4.0)
and the note is highly stable w.r.t. the underlying harmony
and the note is of medium importance in the time-span reduction.”

RULE 17 (”deemphasize unimportant, short notes”):

```
diminuendo( Note, X) :–
    timespan_importance( Note, extremely_low),
    dur_next( Note, NDur),
    NDur < 0.75,
    duration( Note, Dur),
    Dur < 0.75,
    metrical_strength( Note, MS),
    MS < 4.0.
```

”Apply a certain amount X of diminuendo to the current note if
the timespan importance of the note is extremely low
and the duration of the following note is < 0.75
and the duration of the current note is < 0.75
and the current note is not too strong metrically (< 4.0).”

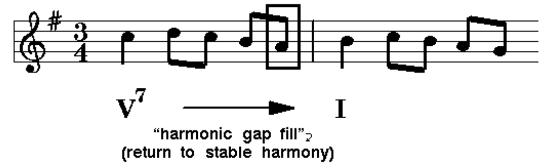
RULE 18 (”deemphasize short notes in melodic gap fills”):

```
diminuendo( Note, X) :–
    rel_position_in_linear_melodic_gap_fill( Note, Pos),
    Pos < 0.63,
    dur_next( Note, NDur),
    NDur < 0.75,
    duration( Note, Dur),
    Dur < 0.75,
    metrical_strength( Note, MS),
    MS < 4.0.
```

”Apply a certain amount X of diminuendo to the current note if
the note is inside a linear melodic gap fill pattern (Narmour)
and its relative position (Pos) in this pattern is in the first two thirds (< 0.63)
and the duration of the following note is < 0.75
and the duration of the current note is < 0.75
and the current note is not too strong metrically (< 4.0).”

RULE 21 (spurious):

```
diminuendo( Note, X) :–
    rel_position_in_harmonic_gap_fill( Note, Pos),
    Pos > 0.89,
    metrical_strength( Note, MS),
    MS < 4.0.
```



”Apply a certain amount X of diminuendo to the current note if the note is inside a harmonic gap fill pattern (Narmour) and its relative position (Pos) in this pattern is close to the end (> 0.89) and the current note is not too strong metrically (< 4.0).”

RULE 25 (obviously spurious):

```
diminuendo( Note, X) :–
    interval_next( Note, p4),
    dir_prev( Note, down).
```

”Apply a certain amount X of diminuendo to the current note if the interval leading to the next note is a perfect 4th and the direction of the interval leading to the current note is down.”

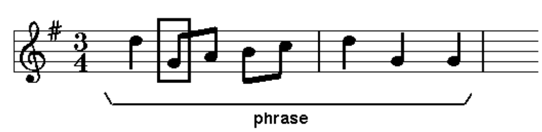
RULE 31 (”accelerate at beginning of rhythmic gap fills”):

```
accelerando( Note, X) :–
    rel_position_in_rhythmic_gap_fill( Note, Pos),
    Pos < 0.27,
    dir_prev( Note, down),
    metrical_strength( Note, MS),
    MS < 4.0.
```

”Apply a certain amount X of accelerando to the current note if the note is inside a rhythmic gap fill pattern (Narmour) and its relative position (Pos) in this pattern is close to the beginning (< 0.27) and the direction of the interval leading to the note is down and the current note is not too strong metrically (< 4.0).”

RULE 37 (”accelerate at beginning of phrase”):

```
accelerando( Note, X) :–
    rel_position_in_phrase( Note, Pos),
    Pos < 0.16,
    metrical_strength( Note, MS),
    MS < 4.0.
```



”Apply a certain amount X of accelerando to the current note if the relative position of the note in the current phrase (Pos) is close to the beginning (< 0.16) and the current note is not too strong metrically (< 4.0).”

RULE 51 (”retard at end of line”):

```
ritardando( Note, X) :–
    rel_position_in_linear_melodic_cont( Note, Pos),
    Pos > 0.92,
    metrical_strength( Note, MS),
    MS > 4.0.
```

”Apply a certain amount X of ritardando to the current note if
the note is inside a linear melodic continuation pattern (Narmour)
and its relative position (Pos) in this pattern is close to the end (> 0.92)
and the note is extremely strong metrically (> 4.0).”

RULE 52 (spurious):

```
ritardando( Note, X) :–
    harmonic_stability( Note, medium),
    metrical_strength( Note, MS),
    MS > 4.0.
```

”Apply a certain amount X of ritardando to the current note if
the harmonic stability around the note is medium
and the note is very strong metrically (> 4.0).”

RULE 56 (”retard at end of melodic leap”):

```
ritardando( Note, X) :–
    interval_prev( Note, PI),
    at_least( PI, maj6),
    dir_prev( Note, up).
```

”Apply a certain amount X of ritardando to the current note if
the interval leading to the current note (PI) is a jump of at least a major 6th
and the direction of this leap is up.”