Overview of the KTH rule system for musical performance

Anders Friberg, Roberto Bresin, and Johan Sundberg

Department of Speech, Music and Hearing, KTH, Stockholm

ABSTRACT

The KTH rule system models performance principles used by musicians when performing a musical score, within the realm of Western classical, jazz and popular music. An overview is given of the major rules involving phrasing, micro-level timing, metrical patterns and grooves, articulation, tonal tension, intonation, ensemble timing, and performance noise. By using selections of rules and rule quantities, semantic descriptions such as emotional expressions can be modeled. A recent real-time implementation provides the means for controlling the expressive character of the music. The communicative purpose and meaning of the resulting performance variations are discussed as well as limitations and future improvements.

INTRODUCTION

A musical performance within Western classical music typically consists of the combined artistic contribution of the composer and the performer. The composer's part, the score, has long been studied and scrutinized in terms of its structural aspects such as harmony, melody, form, instrumentation, as well as of studying the composer's intention or the inherent emotional expression (Hevner, 1937). The performer's expressive contribution has certainly been acknowledged but has until recently been more difficult to study. The contribution from the performer involves the shaping of all audible performance parameters such as tempo, sound level, articulation and vibrato with respect to individual notes, phrases and the whole piece. These parameters can not be studied as readily as the notation in the score. However, they did become easier to study when new acoustic measurement methods were developed (Seashore, 1938). Since then, considerable knowledge has been gained about how performance parameters contribute to structural, motional, and emotional communication (Gabrielsson, 1999; Friberg & Battel, 2002; Juslin & Sloboda, 2001).

With further developments in computer-based tools, it also became possible to formulate and run algorithmic models for musical performance. Thus, for the first time theories of music performance could be verified by listening to model-generated performances. Coinciding with the introduction of personal computers around 1980, a number of such models were developed. One computational model concerned the metrical patterns in the Viennese waltz by Bengtsson and Gabrielsson (1983). They found that the three beats in the measure were performed in a short-long-intermediate pattern. They extended these measurements into a model for the performance of metrical patterns of the Viennese waltz taking into account several metrical levels and performance parameters. Clynes (1983) proposed a rather speculative model suggesting that each composer has a specific "pulse" describing rhythmic and dynamic changes related to the metrical structure (see also Widmer & Goebl, 2004). The large tempo changes reflecting the phrase structure often found in Romantic music were modeled by Todd (1985, 1989). Based on measurements of piano performances, he proposed a general model of tempo as a function of phrase structure. Clarke (1988) suggested a set of nine

Correspondence concerning this article should be addressed to Anders Friberg, dept. of Speech, Music and Hearing, KTH Lindstedtsvägen 24, S-100 44 Stockholm, e-mail: afriberg@csc.kth.se

rules relating musical structure to changes in timing, dynamics and articulation in piano performance.

The development of the KTH rule system started when co-author Sundberg started collaboration with the late Professor Lars Frydén who was a renowned teacher and violin player. The aim at that time was to enhance the recently developed singing synthesis system (Larsson, 1977). Frydén had specific ideas about how music performance should be related to musical structure. These ideas were implemented in a rule system controlling the singing synthesizer. Sundberg et al. (1982, 1983) presented a set of eleven rules constituting the start of the rule system. Modified versions of these original rules are still in use. One source of inspiration for Frydén was the musical analysis suggested by Istvan Ipolyi (1952), with whom he had studied performance, see also Hansen (forthcoming). The rule system has since then been continuously developed within several research projects.

Throughout the development, a general aim has been to contribute to research in the psychology of human communication. New research findings in other areas have been taken into account, in some cases motivating the development of new rules. Progress on the rule system has profited significantly from research and data on music cognition, such as the key profiles quantified by Krumhansl and Kessler (1982; see Sundberg, 1993), motor production (Gilden et al. 1995; Juslin et al. 2002), human movement gestures (Friberg & Sundberg, 1999), speech communication (Carlson et al. 1989), and communication of emotions in music (Juslin, 2001; Bresin & Friberg, 2000).

The first set of melodic rules was implemented on an adapted version of the RULSYS system, a program originally developed for rule-based speech synthesis (Carlson & Granström, 1975). It could only process monophonic melodies and the synthesized sound was played on a singing synthesizer adjusted to an instrumental sound. However, each parameter could be controlled by time envelopes allowing for example accents at the beginning of a note or a crescendo during one note. For handling polyphony, it was necessary to develop a custom-made synthesis system for music performance. The result, the Director Musices (DM) program, was based on the MIDI protocol for input/output and is still the major platform for the rule development. Presently, it is available for most operating systems and can be downloaded at our website (see note 1). Parts of the rule system are implemented in other programs (Bresin, 1993; Bresin & Friberg, 1997, Hellkvist, 2004; Kroiss, 2000). Recently, the possibility of real-time control of music performance

was developed. The expressive deviations are still computed by the rules in DM but the application of the deviations from the score is made within the program pDM (Friberg, 2006). In this way, the performance can be "conducted" in real-time, controlling its expressive character by means of the rules.

The last overview of the rule system appeared in Friberg (1995b) and a number of rules have been introduced since that time. There are new rules for articulation, ensemble swing in jazz, and the modeling of neural motor noise. Several rules have also been improved. The rule for musical punctuation was optimized using a training set of melodies. Rules for final ritardando and phrasing were modified and extended using models of human motion. In the last five years we have focused on macro-level aspects such as using the existing rules to shape a performance according to semantic descriptions including emotional and motional aspects (Bresin, 2000; Bresin & Friberg, 2000). Recently, considerable effort has been devoted to the real-time control of the rule system, which has led to new applications (Friberg, 2005; 2006). A general overview of the communication of music structure in music performance was given in Friberg and Battel (2002).

In this paper, we outline the methods used in the development and refinement of the rules, and we provide an overview of the rules, grouped according to major aspects of performance and macro-level control. Finally, the limitations and future prospects of the rule system are discussed[1].

Method

The original method used to develop the KTH rule system was *analysis-by-synthesis* (Gabrielsson, 1985). First, a tentative model was implemented and applied to different music examples. Next, the effects of each rule were assessed by listening, which typically led to some modifications of the rule. Finally, the process was iterated several times, assessing the effects of the rules on numerous musical examples within the musical style. For more than two decades our musical mentor was Lars Frydén, who served as an inspiration both for implementing new rules and for judging the resulting performances. When the formulation of a set of rules was established, listening experiments were conducted in order to test the generalizability of those rules for a larger population of listeners. This is a rather straightforward procedure, although the effects of rules often depend on the musical situation. Given the almost unlimited number of possible musical
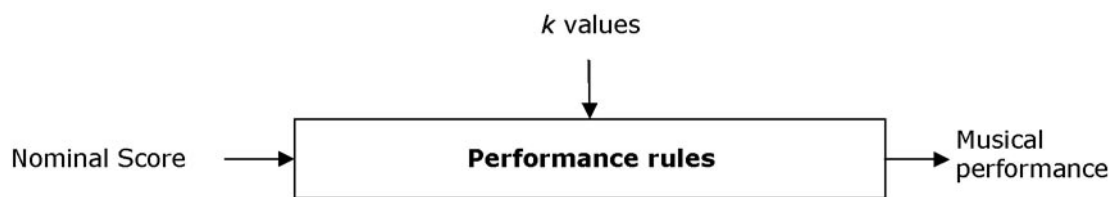
Figure 1.

The performance rules act on the nominal score and generate a musical performance. The effect of each rule is controlled by the k values.

situations, there is little hope of verifying the rules for music examples spanning all possible contexts. Over the course of the project, we have assembled a set of melodic themes, each representing a typical musical context. This set now consists of about 50 melodies and is the main source for testing the rule system. In the development of the more recent models such as the Punctuation rule, this set of annotated melodies was used for automatic training of the rule parameters.

A complementary method for rule development has been *analysis-by-measurement*. Here the rule is formulated using data from real performances. This means that the model can be tested by directly comparing the model outcome with the measurements. However, a further evaluation by listening is still useful. In some cases, we have also used motion data from physical human gestures, on the assumption that the variation of performance parameters may originate from these motor patterns. For the Punctuation rule, an alternative method was used (see section Articulation below). Here the internal rule parameters were optimized using a training set of melodies annotated by a music expert. Unless otherwise noted, the rules below were developed using the analysis-by-synthesis method.

The initial rule formulations were rather simple with respect to context definitions and the resulting change of performance variables. Since then, these rules have been gradually refined, reformulated, and sometimes incorporated into more complex rules. Testing an existing rule in a new musical context often initiated the development of a new rule.

How is a rule-based model of music performance evaluated? Since the reference is the listener's cognition of the music, there is no one "correct" performance. However, listeners tend to agree about many aspects of music performance, reflecting performance conventions within the studied styles. Therefore, the evaluation process begins with perceptual judgments by the development team, and then proceeds to formal listening experiments involving a larger sample of

listeners (e.g. Friberg, 1995b; Thompson et al. 1989; Sundberg et al. 1991; Friberg & Sundberg, 1999).

RULES

General design

As illustrated in Figure 1, the basic rule-scheme is rather simple. The input is the notated music as represented in a score. The rules affect various performance parameters such as timing, sound level and articulation, and ultimately generate a musical performance. The overall magnitude of each rule may be adjusted with the k parameter. Smaller values of k are used when subtle changes are appropriate; larger values of k are used when greater expressive changes are desirable. Within the limitations of the rule system, different combinations of k values can be used for modeling different performer styles, stylistic conventions or different emotional expressions (see below). Thus, there is no "optimal" setting of k values that would be appropriate for any type of music.

The output is a symbolic representation of the resulting musical performance that can be used for controlling a synthesizer.

Most of the rules consist of two parts: a *context* part describing *when* to trigger the rule, and an *execution* part describing *how* to perform this musical situation. The aim of the rule system is to find general principles of music performance. This means that the rules are designed such that the context definition of the rule would trigger only for the intended musical situation. It also means that it should be applicable regardless of style and musical instrument. Most of the rules generate expressive actions based on melodic features (pitch and duration), rather than on the metric structure.

Tempo must be considered in the context definition, because it is important that the rules have the same perceptual effect regardless of the tempo. It has recently been debated whether the perception of a performance expression remains constant across different tempi (Repp,

1994; Desain & Honing, 1994; Honing, 2005a). For most of the rules, our starting point has been to implement expressive deviations relative to the tempo, i.e. analogous to Weber's law[2]. This approach works well within a tempo range when note durations are not too short[3]. The most notable exceptions are the rules Duration contrast and Swing ensemble, described below.

The role of performance marks such as accents or phrase marks in the score may also be considered. These marks are often treated more as a guideline for the performance than mandatory ways of performing the piece. Furthermore, these marks are often inserted by the editor rather than the composer. Therefore, we have in general avoided incorporating such marks in the score, thus, mainly trying to model the performance from the "raw" score. One exception is the set of rules for notated *staccato/legato*.

Certain aspects of musical structure must be provided manually by the user because they are not readily evident from surface properties. For example, automatic extraction of phrase structure and harmonic structure is a difficult process (Temperley, 2001; Ahlbäck, 2004), but these analyses are essential for the phrasing and tonal tension rules respectively. Thus, these structural characteristics must be added to the score manually by the user so that they can trigger the phrasing and tonal tension rules. One exception is the rule for musical punctuation, which automatically finds the melodic grouping structure on a lower level.

Because the rules act upon a range of structural characteristics of music, they provide an expressive interpretation of musical structure. An analogous function is observed in speech prosody, which introduces variation in a range of acoustic features (intensity,

Table 1.
An overview of the rule system

Phrasing	
Phrase arch	Create arch-like tempo and sound level changes over phrases
Final ritardando	Apply a ritardando in the end of the piece
High loud	Increase sound level in proportion to pitch height
Micro-level timing	
Duration contrast	Shorten relatively short notes and lengthen relatively long notes
Faster uphill	Increase tempo in rising pitch sequences
Metrical patterns and grooves	
Double duration	Decrease duration ratio for two notes with a nominal value of 2:1
Inégales	Introduce long-short patterns for equal note values (swing)
Articulation	
Punctuation	Find short melodic fragments and mark them with a final micropause
Score legato/staccato	Articulate legato/staccato when marked in the score
Repetition articulation	Add articulation for repeated notes.
Overall articulation	Add articulation for all notes except very short ones
Tonal tension	
Melodic charge	Emphasize the melodic tension of notes relatively the current chord
Harmonic charge	Emphasize the harmonic tension of chords relatively the key
Chromatic charge	Emphasize regions of small pitch changes
Intonation	
High sharp	Stretch all intervals in proportion to size
Melodic intonation	Intonate according to melodic context
Harmonic intonation	Intonate according to harmonic context
Mixed intonation	Intonate using a combination of melodic and harmonic intonation
Ensemble timing	
Melodic sync	Synchronize using a new voice containing all relevant onsets
Ensemble swing	Introduce metrical timing patterns for the instruments in a jazz ensemble
Performance noise	
Noise control	Simulate inaccuracies in motor

duration, pitch, and timbre) to provide an expressive interpretation of grammatical structure and semantic structure in language (Carlson et al. 1989).

The rules should not be interpreted as prescriptive in the sense that all performances must obey them. Rather, the rules should be considered as a set of principles and parameters from which the user of the rule system can choose freely. Even the application of "inverted" rules may be desirable in some contexts, i.e. using negative *k* values. Thus, the rule system can be viewed as a sort of "musician's toolbox" – a set of basic methods available for shaping a performance. The selection of rules and rule quantities is collected in a *rule palette*, which defines a particular performance style. Such rule palettes have been defined for different styles (jazz, Baroque, Romantic), as well as different emotional expressions (anger, happiness, sadness); see below.

An overview of all the rules discussed below is given in Table 1. The following section provides a brief overview of the rules. Some rules have been presented in several publications. Here, references are limited to the first presentation and the most comprehensive description. For a complete list of publications, see the web site given in note 1. The focus is on the rules that have been clearly verified and are currently used and implemented in the program Director Musices.

Phrasing

Musical phrasing is an important high-level issue, particularly prominent in Western classical music from the Romantic period (Hudson, 1994). A musical phrase is often performed with an arch-like shape applied to tempo and dynamics (Gabrielsson, 1987; Repp, 1992). The phrase is typically slow/soft in the beginning, fast/loud in the middle and ends slow/soft, modeling a *crescendo/accelerando decrescendo/rallentando* pattern. Todd (1985, 1989) proposed an early model for large phrase units using data from a piano equipped with sensors that could register all key-presses. The *Phrase arch* rule extended his model to include shorter phrase structures, and introduced several new parameters in order to account for the variations found in real performances regarding structural level of the phrase, the amount of change in the beginning and end, the position of the maximum of the arch, the duration of the last note in the phrase, and the shape of the curve (Friberg, 1995a). The rule was modeled and tested on the data for Schumann's *Träumerei* measured by Repp (1992). Later, the shape of the arch was modified according to hand gesture data (Juslin et al. 2002). As mentioned, the phrase analysis is currently supplied

manually, i.e., markings indicating the start and the end of a phrase. Typically, several instances of the Phrase arch rule acting on different structural levels are applied simultaneously.

The *Final ritardando* rule provides an alternative phrasing for the end of the piece (Friberg & Sundberg, 1999). It was intended to model the final *ritardandi* often found in Baroque pieces with a motoric and rhythmic character, in which the tempo is basically constant (similar to most jazz, pop and other styles where meter is important) (Sundberg & Verillo, 1980). This type of character may suggest a motion metaphor of a mass moving at constant speed. Following this idea, the rule was modeled from measurements of stopping runners translating velocity to tempo and position to score position. Score position is defined as the nominal time if the music were to be played with a constant tempo. In the physical domain, this model has the simple property that the braking energy is constant during the stopping phase and is thus likely to minimize the physical effort in a perfectly planned halt. The model yielded the best rating of several alternatives in a listening experiment (Friberg & Sundberg, 1999).

The *High loud* rule increases the sound level proportional to pitch (Sundberg et al. 1982; Friberg, 1991). Since the pitch shape often co-varies with the phrase such that the highest pitch is in the middle of the phrase, it will in most cases result in a sound level shape marking the phrase structure, also justifying it as a simple phrasing rule.

Micro-level timing

A closer look at the inter-onset interval (IOI, i.e. the time from one tone's onset to that of the next) values from real performances reveals that each note is played with a small variation relative to its nominal duration given by the score. Certainly, a portion of this variation is due to imprecision in the execution. This is modeled in the Noise rule below. However, many systematic variations can be observed. For example, Widmer (2002) identified 17 rules describing local timing variations when applying machine learning methods to pianist performances. There are also deviations due to the mechanical properties of the instrument (Goebl 2001).

The *Duration contrast* rule increases the difference in IOI between different note values such that long notes are lengthened and short notes are shortened (Sundberg et al. 1982; Friberg, 1991). The rule introduces a fixed amount of shortening/lengthening in milliseconds depending on the original IOI. This formulation acknowledges that

there are perceptual boundaries regarding IOI. Too short or too long intervals should not be further shortened or lengthened, respectively. The lower boundary was evident in measurements of jazz performances. The shorter note of the typical long-short pattern exhibited a floor value of about 100 ms in the tempo range 180-330 BPM (Friberg & Sundström, 2002). Later it was found that the Duration contrast rule was a rather flexible and useful tool in that it could also be used "inverted", i.e. with negative *k* values. This implies that the contrast will be de-emphasized instead (Bresin & Friberg, 2000). Increasing duration contrast appears to have the perceptual effect of increasing the "energy" of the performance, while decreasing the contrast creates the opposite smoothing, "calm" effect. Performers used duration contrast in such a way when expressing different emotions (Gabrielsson & Juslin, 1996, c.f. Bresin & Friberg, 2000).

The *Faster uphill* rule complements the Duration contrast rule in terms of pitch. Notes in an ascending motion are shortened (Frydén et al. 1988; Friberg, 1991). Presumably, this has a similar effect to that of the Duration contrast rule in increasing the perceived "energy".

Metrical patterns and grooves

Music with regular rhythmic patterns results induce a stable perception of beat and higher metrical units such as the measure. These rhythmic patterns often have a corresponding performance pattern in terms of IOI, sound level and articulation.

A common rhythmic pattern found in 3/4 and 6/8 meter is a half note followed by a quarter note. This pattern is often performed with reduced duration contrast (Gabrielsson, 1987; Gabrielsson et al. 1983; Henderson, 1936). This was implemented as the rule *Double duration*, which performs any 2 to 1 duration ratio with reduced duration contrast, keeping the total duration of the two tones the same (Friberg, 1991). The conflict between Double duration and Duration contrast rules is solved by not allowing Duration contrast to trigger when Double duration can been applied.

Another common pattern is the alternating long-short pattern commonly found in a variety of musical styles including Baroque (Hefling, 1993), folk, as well as jazz music. The *Inégales* rule was our first implementation of this pattern (Friberg, 1991). Later, the Ensemble swing rule refined the application in a jazz context; see below. The rules assume that the notes in question are notated with equal note values[4].

There are a multitude of metrical patterns for various types of music. Therefore, the current rule system could easily be extended by implementing more of these known patterns such as that found in the Viennese waltz (Bengtsson & Gabrielsson, 1983).

Articulation

The term articulation is used to describe the amount of *legato/staccato* with which a note is being played. It is defined as the ratio between the note duration (i.e. sounding duration) and the IOI; thus a value of 1 represents legato and a value of around 0.5 represents a staccato. It is an important parameter for changing the overall character of the piece including motional as well as emotional aspects (De Poli et al. 1998; Bresin & Battel, 2000; Juslin, 2001).

The *Punctuation* rule attempts to identify the small melodical fragments and perform them by inserting a micropause after the last note as well as lengthening it (Friberg et al. 1998). The analysis part uses a subset of 13 rules to determine the boundaries of the melodic fragments based on structural discontinuity (Ahlbäck, 2004; Cambouropoulos, 1998). These sub-rules try to find melodic boundaries, for example, by identifying pitch leaps and long notes in a context of five notes. This is similar to the more common term grouping analysis (Lerdahl & Jackendoff, 1983). The internal parameters were automatically optimized and the outcome was analyzed using a set of 52 annotated melodies.

The *Score legato* and *Score staccato* rules are intended to be used where there are corresponding marks in the original score. They are modeled after measurements of piano performances (Bresin, 2000, 2001; Bresin & Widmer, 2000). These rules model the amount of articulation according to the different tempo indications in the score as well as the expressive intentions of the performer.

The *Repetition articulation* rule adds a micropause between repeated notes (Friberg, 1991). The duration of the micropause can be modulated according to the expressive intentions of the performer (Bresin & Battel, 2000; Bresin, 2001).

As a complement, the *Overall articulation* rule can be used for changing the articulation of all notes. This is useful for modeling emotional expressions as well as for real-time control. For *k* = 1, an articulation ratio of 0.75 is applied to all notes longer than 100 ms.

Tonal tension

The concept of tonal tension has been used both in music theory and psychology to explain, for example,

the tension-relaxation pattern perceived in a harmonic cadence. A number of musical features have been suggested to contribute to tonal tension, such as consonant/dissonant chordal structures, unusual melody notes relative to the harmony, or unexpected changes of key (Friberg & Battel, 2002). Two of these are modeled in the rule system relating to music with traditional harmony in which a harmonic analysis can be made. An alternative model is also suggested for atonal music. The *Melodic charge* rule emphasizes tones that are far away from the current root of the chord on the circle-of-fifths (Sundberg et al. 1982; Friberg, 1991). Thus, an F# in the melody during a C major chord will get the largest emphasis while a C note will get no emphasis. The note is performed with increased sound level and duration in proportion to the emphasis. The key profiles by Krumhansl and Kessler (1982) were found to correlate with the melodic charge (r = 0.86, Sundberg, 1993). In a similar way, the *Harmonic charge* rule emphasizes chords that are far away from the current key on the circle-of-fifths (Sundberg et al. 1982; Friberg, 1991). It is performed by gradual *crescendi/rallentandi* towards the emphasized chord regions. As an alternative, the *Chromatic charge* rule was developed for atonal music (Friberg, 1991; Friberg et al. 1991). This emphasizes regions in which the melody consists of relatively small intervals. Possibly, the occurrence of such small intervals is relatively rare in atonal music and therefore creates tension. Expert listeners rated performances of music by Boulez, Webern, Xenakis and random sequences. Almost without exception, they preferred versions in which the Chromatic charge rule was applied compared to dead-pan performances. In the listening experiment, the chromatic charge was applied together with a few other rules and the comparison was made between the complete version and a dead-pan version. However, the Chromatic charge was the only macro-level rule yielding comparatively large deviations. It is, therefore, unlikely that the effect of the Chromatic charge rule was unnoticed by the listeners.

Intonation

Intonation and tuning has fascinated many musicians and music theorists (e.g. Barbour, 1951; 2004). It is apparent that a fixed tuning can not fit current intonation practice. Rather, intonation practice is dependent on the musical context and seems to be a compromise between several partly contradictory considerations. Most researchers agree that a good melodic intonation is dependent on the melodic context (Fyk, 1995;

Gabrielsson, 1999). For example, in a leading tone to tonic progression, the leading tone is often played sharper in pitch than is indicated in equal temperament tuning, resulting in a melodic interval that is smaller than a semitone (i.e., less than 100 cents). This corresponds roughly to a Pythagorean tuning relative to the current tonic. On the other hand, good harmonic intonation is achieved by reducing the beating between partials, a particularly important feature in barbershop singing (Hagerman & Sundberg, 1980). This intonation corresponds roughly to a just tuning relative to the current chord.

The *Melodic intonation* rule determines the intonation of each note depending on the melodic context and its relation to the root of the current chord (Frydén et al. 1988; Friberg, 1991). Roughly, minor seconds are performed more narrowly than equal temperament. This rule was specifically designed for monophonic melodies. The resulting intonation correlated well with measurements of professional violinists (Garbuzov, 1948; Sundberg, 1993). The *Harmonic intonation* rule was designed to minimize audible beats in chords, which implies a just intonation of all simultaneous notes using the root of the current chord as the reference. Interestingly, however, the tuning implied by the Harmonic intonation rule is often opposite to the tuning implied by the Melodic intonation rule. For example, the major third is tuned relatively flat according to the Harmonic intonation rule, but relatively sharp according to the Melodic intonation rule (for an ascending melodic interval). In order to solve this dilemma, the Harmonic and Melodic intonation rules were combined in the rule *Mixed intonation* (Sundberg et al. 1989; Friberg, 1991). Here each new note is first intonated following the melodic rule. Then the intonation is slowly changed towards the harmonic intonation with the root of the current chord as the reference. In this way relatively short notes are intonated melodically while relatively long notes will gradually move towards less audible beats within the current chord.

The *High sharp* rule stretches the scale somewhat so that every interval will be increased in proportion to the interval size (Friberg, 1991). The normal setting of the rule gives a stretching of 3 cents per octave. This stretching is in accordance with the physical features of some instruments. In tones produced by piano strings, the distance between adjacent partials is slightly greater than the fundamental frequency. This deviation from strict harmonicity arises from the physical properties of the piano strings. Thus, in order to minimize beats between the partials of simultaneously sounding tones, octaves in pianos are slightly stretched relative

to strict harmonic octaves. Indeed, listeners tend to prefer stretched octaves in listening tests (Sundberg & Lindquist, 1973). Note that the rule was developed and tested using monophonic melodies.

Ensemble timing

Since many of the rules introduce small independent tempo changes to each part, a polyphonic performance will be unsynchronized. While certainly not modeling the listening-adaptation process in a real ensemble, the *Melodic sync* rule solves any synchronization problem (Sundberg et al. 1989). The first step is to create a new part consisting of all onsets in any of the original parts. In the case of several simultaneous notes, the note with the highest melodic charge will be used for the new part. Thus, this new "melody" will reflect all rhythmic activity in the music but with the notes selected from any of the original parts. All the rules are then applied to this new part and the resulting tempo variations are transferred back to the original parts. Any other strategy tested, such as selecting the lead part and letting the accompaniment follow, led to many practical problems in specific musical situations. This method adopts the idea of "listening to the short notes" in the sense that it is the short notes that decide the duration of any simultaneously sounding long notes.

An interesting open question is what makes the music "swing". In jazz music it is often attributed to the typical long-short pattern applied to consecutive eighth notes. One principle regarding the timing in jazz ensembles was found by Friberg and Sundström (2002). Analyzing commercial recordings of several well-known jazz musicians, they found a common timing pattern such that the soloist was delayed on the downbeats (quarter notes) relative to the drums, while the soloist and the drummer were synchronized on the upbeats (intermediate eighth notes). This implies that the ratio of the long-short pattern (swing ratio) was different for the drummer and the soloist; that is, the drummers' swing ratio was relatively large and the soloists' relatively small. Furthermore, these timing differences diminished gradually with increasing tempo, leading to smaller swing ratios. The *Ensemble swing* rule implements these patterns using the average curves from the measurements.

Another ensemble principle is the concept of "melodic lead" (Rasch, 1979; Palmer, 1996; Vernon, 1936; c.f. Goebl, 2001). It was found that (in particular) pianists tend to play the melody ahead of the accompaniment by, on average, about 20 ms. This is not implemented as a rule but can be achieved in Director Musices using a delay parameter in the score specification.

Performance noise

Musicians often spend considerable time practicing to play consecutive notes with even timing, such as in scale exercises. However, there are limitations both in the motor system and in perceiving small timing variations. Therefore, random variations (i.e., noise) are always present to some degree in any performance by a musician. One problem in modeling these variations is that the noise is difficult to separate from intentional variations in real performances. To address this issue, the modeling of the noise component has been based on psychoacoustic experiments, involving finger tapping tasks and models of human timing proposed by Gilden et al. (1995).

The *Noise* rule consists of two distinct components (Juslin et al. 2002). The first component, motor delay noise, is assumed to originate from the effectuation of each tone gesture. It is modeled using white noise added to each tone onset time and tone sound level. Thus, this component only affects the onsets individually and does not result in any long-term tempo drift. The second component, assumed to originate from an internal time-keeper clock, is modeled using 1/f noise with the overall amount dependent on the IOI. The 1/f noise results in relatively more dominant slow variations usually leading to a small tempo drift. The resulting deviation from the two components closely follows the just noticeable difference (JND) in perception experiments (Juslin et al. 2002). Interestingly, listeners rated performances with the Noise rule applied as more "human" but not more "musical" compared to performances without the rule.

Rule combinations

In rendering a musical performance, several rules are applied to the same score. A specific performance is obtained by selecting rules and the corresponding rule parameters in a rule palette. Typically, only a subset of the rules is used in one rule palette. For example, for overall phrasing, either the Phrase arch rule or the Final ritardando rule is used since they overlap in the end of the piece. When combining several rules, the effect of each rule is in principle added or multiplied together for each performance parameter. In the case where there are several rules acting on the same note and parameter, this may result in undesired side-ef-
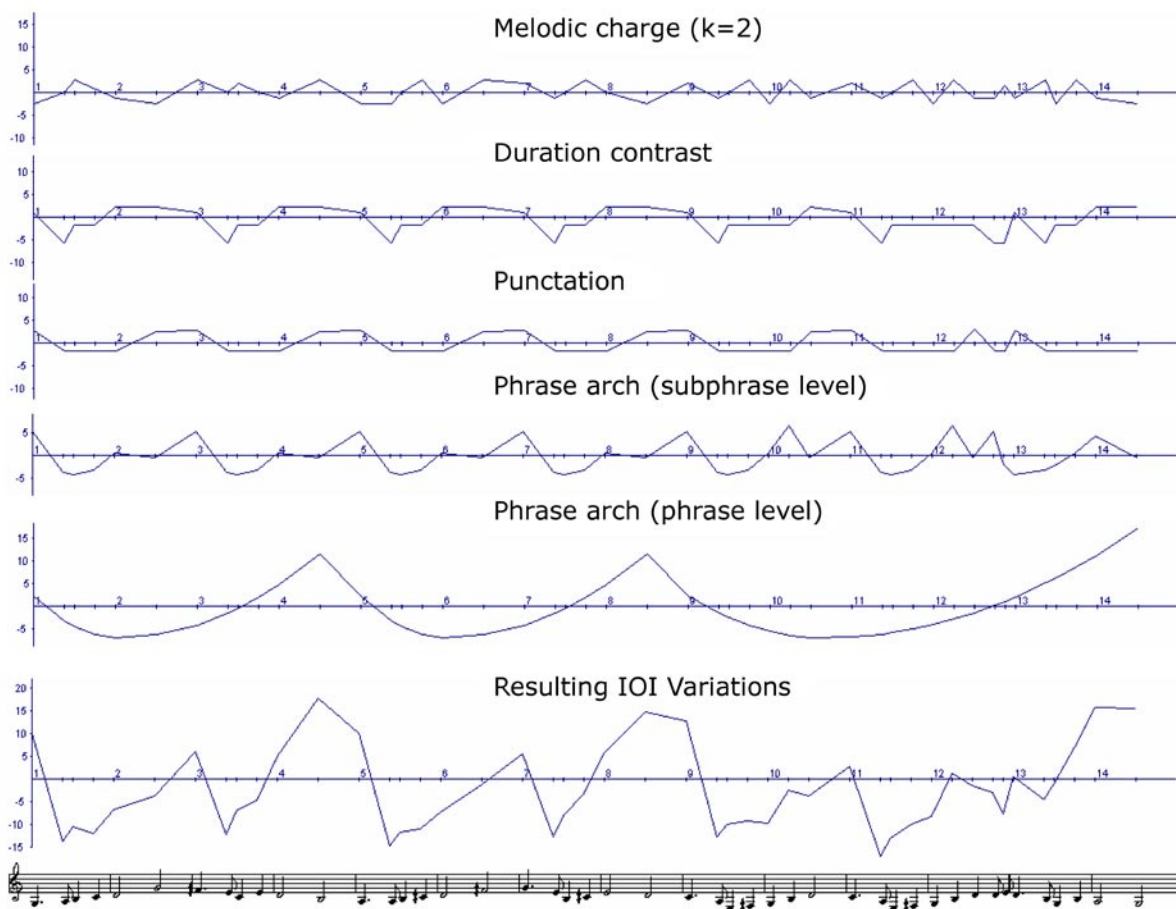
Figure 2.

The resulting IOI deviations by applying Phrase arch, Duration contrast, Melodic charge, and Punctuation to the Swedish nursery tune "Ekorr'n satt i granen". All rules were applied with the rule quantity k=1 except the Melodic charge rule that was applied with k=2.

fects. For example, when several rules that act upon note durations are combined, a note might be lengthened too much. Some of these possible rule conflicts have been solved in the rule context definitions. A micro-timing rule and a phrasing rule, although acting on the same notes, work on different time scales and will not interfere with each other. Figure 2 illustrates the effect of IOI variations resulting from six rules applied to a melody.

How to shape a performance: Using the rules for modeling semantic performance descriptions.

It can be rather difficult to generate a specific performance with the rule system, given the many degrees-of-freedom of the whole rule system with a total of about 30-40 parameters to change. This procedure is greatly simplified if mappings are used that translate descriptions of specific expressive musical characters to corresponding rule parameters. Overall descriptions

of the desired expressive character are often found at the top of the score. They may refer to direct tempo indications (*lento, veloce, adagio*) but also to motional aspects (*andante, corrente, danzando, fermo, con moto*) or emotional aspects (*furioso, con fuoco, giocoso, vivace, tenero*). These semantic descriptions of the expressive character can be modeled by selecting an appropriate set of rules and rule parameters in a rule palette. Research on emotional expression in music performance has shown that there tends to be agreement among Western listeners and performers about how to express certain emotions in terms of performance parameters (Juslin, 2000). Using these results as a starting point, we modeled seven different emotional expressions using the KTH rule system (Bresin & Friberg, 2000; Bresin, 2000). In addition to the basic rule system, we also manipulated overall tempo, sound level and articulation. A listener test confirmed the emotional expression resulting from the defined set of rule parameters (rule palettes) for two different music examples. Table 2 suggests some

Table 2.

Qualitative changes of overall performance parameters and rule quantities for portraying four different emotional expressions

	Happy	Sad	Angry	Tender
Overall changes				
Tempo	somewhat fast	slow	fast	slow
Sound level	medium	low	high	low
Articulation	staccato	legato	somewhat staccato	legato
Rules				
Phrase arch	small	large	negative	small
Final ritardando	small	-	-	small
Punctuation	large	small	medium	small
Duration contrast	large	negative	large	–

qualitative changes of overall performance parameters and rule parameters for expressing different emotions. The list can be expanded to all rules. For example, the overall change of articulation can be obtained by using a combination of all articulation rules. The exact values need to be determined according to aesthetical and stylistic aspects. From our experience, other semantic descriptions, such as those given by De Poli et al. (1998), can easily be modeled using the rule system. Several such descriptions and their corresponding rule parameters are suggested in Friberg (2006).

APPLICATIONS

Since their introduction, computers have been used for playing music both in digital format (i.e. audio files) and in symbolic notation (e.g. MIDI). This has raised a need for tools for expressive performance in order to obtain expressive renderings of computer-played music. The KTH rule system could be applied to any polyphonic music in symbolic notation such as music databases available on the Internet[5].

On the consumer market, there are many devices that emit alert sounds. We live in an acoustic environment that is populated by machine-generated musical events such as music-based mobile phone ringtones, wake-up calls, or alert sounds from household appliances. All these sounds should be designed to raise the attention of the user. The KTH rule system has recently been applied to ringtones. MIDI ringtones often have quantized durations and sound levels, and therefore sound unnatural and dull to our ears. Using the performance rules, it is possible to obtain ringtones that sound more musical and that differ in emotional expressive character (Bresin & Friberg, 2001).

A recent application of the rule system has been in conductor systems for controlling the performance of

a symbolicscore in real-time. Previous systems have largely been limited to the control of overall dynamics and tempo (Mathews, 1989; Borschers et al. 2004). Using the rule system, it is also possible to control micro-level aspects of the performance. The first attempt in this direction combined a static application of the rule system and the Radio Baton conducting system developed by Max Mathews and co-workers (Mathews et al. 2003). Here, the rules were applied such that the global tempo was not affected. The conductor still controlled the overall tempo and dynamics but the micro-level structure of the music was shaped by the rules.

Due to the properties of the rule formulation and implementation issues, the rules are first applied in DM to the whole score and then the resulting performance can be played. By using a two-step procedure, the rules can also be controlled in real-time. The rules are first applied in DM and all the deviations from each rule are individually collected in a file. The music is then played using the newly developed pDM application in which the deviation resulting from each rule can be applied and scaled (Friberg, 2006). Mappers are provided in pDM that interpolate between different semantic descriptions including the activity-valence space commonly used for characterizing different emotions. By analyzing the overall gesture properties using a small video camera, a user can "conduct" the expressivity of a performance in a very flexible manner. This was suggested as a way to enhance the listening experience rather than using it as a professional performance tool (Friberg, 2005).

The KTH rule system could also be used for developing and testing theories of music performance, music cognition, and perception. DM and pDM allow for a systematic variation of the contribution given by a single performance parameter. For example, the rule system could be used to identify and quantify perform-

154

ance parameters that are important for conveying different emotional expressions. This could be achieved by preparing a battery of stimuli and rating them in a listening test, or by asking professional musicians to target a specific emotion by controlling in real-time the performance of a particular score.

ALTERNATIVE MODELS

Recently several alternative computational models have been developed. Gerhard Widmer and co-workers use machine learning techniques for automatic deduction of rules for music performance (Widmer, 1995). They have used this strategy for extracting rules both at note level (Widmer, 2002) and at phrase level (Widmer & Tobudic, 2003). While the primary purpose of this strategy is to understand the underlying principles, it has also been used for computer-generated performances.

Mazzola (1990, 2002) proposed a complex theory of music analysis and performance modeling based on advanced mathematics. The theory is implemented in several computer models but has been verified with empirical methods to a limited extent only (see Widmer & Goebl, 2004).

Bresin (Bresin et al. 1992; Bresin, 1998) proposed an artificial neural network (ANN) that models the performance style of a pianist. ANNs were trained with data from real piano performances and learned the style of one professional pianist. Analyses of the output by the trained ANNs showed that they were clearly replicating the qualitative behavior of some of the KTH rules, such as Duration contrast and Leap articulation. Probably the most interesting behavior of the ANNs was related to their non-linear characteristics. Even if the ANN's output could be partly explained as the effect of KTH rules, its output did not correspond to an additive combination of rules, but to a non-linear combination of rules.

The positive experience with the ANNs, and the difficulty in exactly explaining their functioning, led to the design of a fuzzy rule system (Bresin et al. 1995). This system realized fuzzy versions of the original KTH rules. It allowed the effects of the rules to be combined in a non-additive way and, at the same time, it was possible to describe the behavior of the system exhaustively.

Another model based on an artificial-intelligence technique is that proposed by Arcos and co-workers (Arcos et al. 1998). They developed a case-based reasoning system. The system is trained by using expressive parameters extracted from real performances. These parameters are successively used, in combination with musical knowledge based on Narmour's im-

plication/realization model (Narmour, 1990), as a basis for performing new scores. To our knowledge, the system was applied only to saxophone jazz performances. More extensive overviews of computational modeling of music performance can be found in De Poli (2004) and Widmer and Goebl (2004).

DISCUSSION AND FUTURE IDEAS
Evaluation

Most of the initial KTH rules were developed using the analysis-by-synthesis method using only a few expert judges. However, support for these rules has in many cases been confirmed either from listening tests or from other investigations (Friberg, 1995b). For example, duration contrast was found to be used in patterns of a dotted eighth note followed by a sixteenth note (Gabrielsson, 1987). It was also used as an expressive device when performers were asked to portray different emotions (Gabrielsson & Juslin, 1996). The KTH rule system was partly confirmed by the machine learning system used by Widmer (2002). Looking at the note level, his system predicted a set of 17 rules for the performance of individual notes. Some of these resembled the KTH rules.

Recent rules in the DM system have been based on empirical evaluation. Rules such as the Phrase arch rule were derived from measurements taken from performance data, and the goodness of fit between empirical data and the effects of the rule provided empirical validation for the rule (Friberg, 1995a). This procedure was extended in the development of the Punctuation rule by using two different rule models and two sets of data (Friberg et al. 1998). One data set was used for the rule development and the other for the final testing. During the development of the Final ritardando rule we used measured *ritardandi* both for evaluation and for parameter fitting. In addition, the ritardando model was verified in a listening test where several different models and parameter values were tested (Friberg & Sundberg, 1999).

Recently, the method used for developing the Final ritardando rule was evaluated and compared to an alternative model (Honing, 2005b) according to the model selection criteria (1) goodness-of-fit, (2) model simplicity, and (3) degree of surprise in the model outcome (Honing, in press 2006). Honing came to the conclusion that the final ritard model showed good fit with the data but was quite flexible in that it could predict a great variety of curves, and that the surprise factor was low. His overall conclu-

sion was that the final ritard model fulfilled only one out of his three criteria. However, the comparison method is not entirely unproblematic. Instead of using the more obvious number of parameters as a criterion for determining model simplicity, Honing investigated the total area of tempo variations from the start to the end of the ritard that the model was able to produce. He concluded that this area was larger for the final ritard model than for the alternative model. It seems, however, that this is a somewhat debatable interpretation of the concept of degrees-of-freedom. Since the alternative model suggests an area for permissible ritards and that the final ritard model suggests a specific curve family, it is obvious that the alternative model produces an infinite number of more possible curves as compared to the final ritard model. Another aspect is that the final ritard model was additionally evaluated by a listening test limiting the possible parameter space to a narrow range. This was not considered in Honing's comparison.

One general contribution of modeling is that it results in a wealth of specific parameter data. Many of those values have been determined intuitively during the analysis-by-synthesis loop. Evaluations in form of listening experiments have rather strict limitations both regarding the number of possible parameters and musical examples that can be assessed at the same time. Therefore, it is often impractical to make an objective assessment of the specific effect of each parameter. So far only the overall effect of the each rule or a combination of rules has been assessed in listening experiments.

Most empirical evaluations of the rule system have considered aesthetic effects of the rules (Thompson et al. 1989; Sundberg et al. 1991; Friberg, 1995b). By using a number of rating scales including emotional, motional or expressive aspects (Bresin & Friberg, 2000; Juslin et al. 2002) we will presumably gain a clearer picture of the expressive/communicative function of each rule. By using listener estimations of such a variety of communicative aspects applied to different musical styles, it may be possible to identify general communicative features as well as features attributed to certain stylistic conventions.

Purpose and meaning

The performance rules can be viewed as a set of performance principles mainly used to enhance the structure of the music (Friberg & Battel, 2002). This follows

from the fact that most rules use structural context descriptions as input. Thus, for example, the phrasing rules enhance the division in phrases already apparent in the score. This indicates an interesting limitation for the freedom of expressive control: it is not possible to violate the inherent musical structure. One example would be to introduce *ritardandi* and *accelerandi* in the middle of a phrase. From our experience with the rule system, such a violation will not be perceived as musical.

Looking more closely into the communicative aspects of musical structure, Sundberg (2000) suggested that the rules enhance the inherent structure of the music using two main principles: grouping and differentiation. Both differentiation and grouping are important for enhancing speech communication (Carlson et al. 1989). Differentiation serves to facilitate the perception of categories. For example, by applying the Duration contrast rule, the difference between long and short duration categories will be enhanced. The grouping principle serves the purpose of dividing the musical stream into chunks similar to word boundaries in speech, thus facilitating the understanding of the melodic structure. Examples of grouping rules include the Punctuation and Phrase Arch rules.

As mentioned, the rules also have other expressive functions. Figure 3 identifies the important sources of performance variation. To the left are the musically relevant communicative intentions. They are divided into structural (differentiation, grouping), emotional and motional aspects. To the right are the limitations in terms of neural noise and technical performance aspects. The communication of emotional expression can largely be explained by average measures of performance parameters (Juslin, 2000). For example, a sad performance is characterized by slow tempo, soft dynamics and legato articulation. However, the emotional communication can be improved by also applying specific rule setups (see Table 2). One example is the Phrase arch rule, which was found to be effective for modeling sadness. Therefore, there is an interaction between communication of structure and emotional/motional expression (not shown in the Figure). The communication of different emotional expressions can be very effective if the different parameters, such as the tempo, are exaggerated. However, this would violate the aesthetic quality of the performance. This implies that there are aesthetic boundaries that impose restrictions on expressive communication. In this way, a real performance may be viewed as a compromise between a number of different communicative goals.
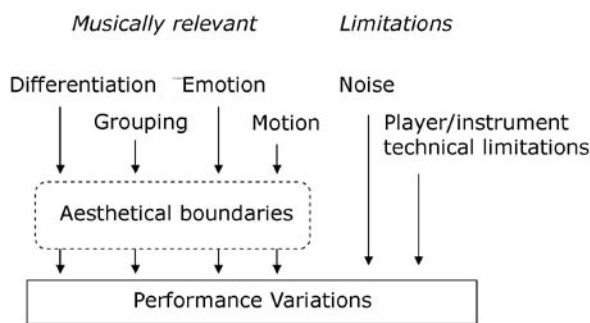
Figure 3.
An overview of the different sources of performance variations.

What's missing?

As a violin player, Frydén hypothesized that musical performance should be shaped according to melodic features only. He therefore assumed that metre and rhythm are consequences of the melodic structure. We know, however, from previous research that the metric structure is in fact an independent perceptual factor that influences the performance (Clarke, 1985). Frydén acknowledged the need for metrical performance patterns in certain cases, e.g. some waltzes. However, these patterns were not properly formulated into a rule. In short, the rule system could easily be extended with rules modeling metrical patterns in metrically oriented music. As a starting point, existing measurements of different rhythms could be used (e.g. Bengtsson & Gabrielsson, 1983).

Due to the technological limitations of the first computer system, all of the early rules were developed using monophonic melodies. These rules often alter the global tempo somewhat. The small, barely perceptible rule quantities used in the beginning made these tempo deviations rather negligible. However, when we tried to increase the rule quantities and extend the rule system to polyphony, these tempo deviations created undesired side-effects. Our experience from testing different music examples is that, although we accept considerable tempo changes in a monophonic melody, when a rhythmic accompaniment is added to the same melody, any micro-level change of tempo (e.g. shortening one note) will sound unmusical. Therefore, more work is needed to formulate micro-level rules such that the global tempo is not affected. One way of doing that is to use "duration stealing"; that is, to subtract the same amount of duration from the neighboring notes as has been added to the target note. This method is used for the Double duration rule.

The intention during the development has been to generalize each rule so that it would be possible to apply it to a variety of musical examples. However,

some of the rules are clearly associated with a certain style. For example, the Phrase arch rule is associated primarily with music from the Romantic era. In fact, Romantic music is possibly the best target for a music performance model since it is common and acceptable to make very large performance variations in this style. Possibly the traditional categorization in styles such as classical (Baroque, Romantic), jazz (bebop, swing), folk music is not an appropriate classification for modeling music performance. We have found that the rhythmic/melodic/harmonic character of the piece is more important for determining an appropriate selection of rules and parameter than musical style. For example, in a slow melodically oriented piece we may accept quite large relative tempo variations[6] while a fast piece with a strong rhythmic accompaniment demands a steady tempo. This difference due to rhythmic character may have a perceptual base. A common theory of time perception suggests that we have an internal clock synchronized with the perceived beat of the music (Povel & Essen, 1985). If the music is highly repetitive rhythmically, this clock percept is likely to be reinforced. Any tempo change in the music must therefore force the internal clock to change. On the other hand if the music has a melodic (rhythmically weak) character, this clock percept is only weakly supported, making a change more acceptable. The only dependence on tonality that we found was the need for a different tonal tension rule for atonal music.

Still, the rules certainly do not cover all possible musical performances. Clear evidence of this is seen in the difficulty experienced in applying the rules in reverse, that is, trying to find a rule set corresponding to a given a performance (Zanon & De Poli, 2003; Sundberg et al. 2003). Since the rules are applied with the same values for the whole piece, a possible extension would be to allow rule parameters to change during the performance. This allows a better fit to a given performance but at the same time substantially increases the number of parameters. A more reasonable model could be to determine an overall expressive character for each section. This would restrict the rule parameter changes so that they follow a mapping model according to the semantic performance descriptions mentioned above.

One way of regarding the rule system is illustrated in Figure 4. All possible performances define a vast space in which each note can be changed independently. The musical performance space defines the subspace of performances that are musically acceptable (in a broad sense)[7]. The rule subspace defines all possible performances by the rule system. We believe that the rule system substantially limits the expressive space

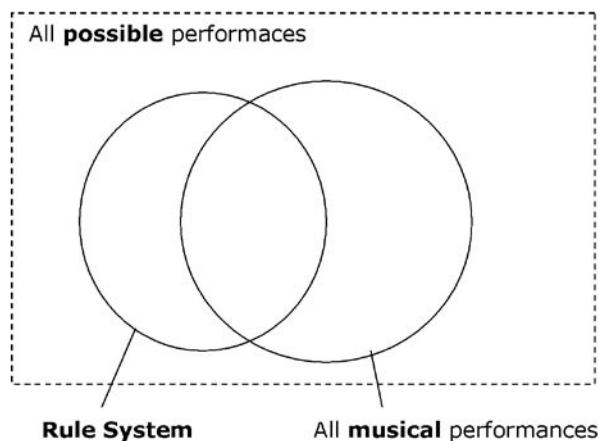**Figure 4.**

A schematic view on the variations that can be produced by the rule system versus all possible variations and all musical variations.

relative to all possible performances and that a good part of all musical performances is covered (the intersection of the circles). However, in order to make the rule system more congruent with the space of all musical performances, two things are needed. (1) The basic rule system needs to be complemented with missing parts, such as metric/rhythmic rules (expanding the rule system space) and (2) the possible combinations of rules need to be restricted, for example, by extending and using a semantic descriptive space, as outlined above (making the rule system space smaller and move towards the space of all musical performances).

ACKNOWLEDGEMENTS

Notes

[1] For additional information about the KTH rule system including sound examples, literature and program downloads see http://www.speech.kth.se/music/performance

[2] Weber's law states that the perceived intensity change of a stimulus is a constant proportion of the original stimulus value, see e.g. http://en.wikipedia.org/wiki/Weber-Fechner_law

[3] For inter-onset intervals (IOI) less than 250 ms, the JND for timing deviations does not follow Weber's law (Friberg and Sundberg, 1995), thus, it is likely that small expressive deviations will not do so either. A lower boundary seems to be about 100 ms where individual events are not clearly perceived (Friberg and Sundström, 2002; London, 2004).

[4] In this case, other notational praxis have been used such as using patterns of a dotted eighth note followed by a sixteenth note.

[5] Kunst der Fuge http://www.kunstderfuge.com/ Classical Music Archives http://www.classicalarchives.com/

[6] That is, we are assuming that the tempo variations follow Weber's law.

[7] The same discussion can be applied if we consider the musical space restricted to a certain style.

References

Ahlbäck, S. (2004). *Melody Beyond Notes. A study in melody cognition*. (Doctoral thesis) Publications from the department of Musicology, Göteborg University.

Arcos, J. L., Lopez de Mantaras, R. & Serra, X. (1998). Generating expressive musical performances with SaxEx. *Journal of New Music Research*, *27*, 194-210.

Barbour, J. M. (1951, 2004). *Tuning and Temperament*. Courier Dover Publications.

Bengtsson, I. & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg, (ed.), *Studies of Music Performance*, Stockholm: Royal Swedish Academy of Music, Publication No. 39, (pp. 27-60).

Borchers, J., Lee, E. & Samminger, W. (2004). Personal Orchestra: A Real-Time Audio/Video System for Interactive Conducting. *Multimedia Systems 9*, 458–465.

Bresin, R. (1993). MELODIA: a program for performance rules testing, teaching, and piano scores performing. In *Proceedings of the X Italian Colloqium on Music Informatics (CIM)*, AIMI, Padova.

Bresin, R. (1998). Artificial neural networks based models for automatic performance of musical scores, *Journal of New Music Research, 27*, 239-270.

Bresin, R. (2000). *Virtual Virtuosity – Studies in automatic music performance*. Doctoral Dissertation, Speech Music and Hearing, Stockholm: KTH, TRITA-TMH 2000:9, ISSN 1104-5787, ISBN 91-7170-643-7.

Bresin, R. (2001). Articulation rules for automatic music performance. In Proceedings of the International Computer Music Conference 2001, Havana, (pp. 294-297).

Bresin, R. & Battel, G. U. (2000). Articulation strategies in expressive piano performance. Analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart's Sonata in G major

(k 545). *Journal of New Music Research, 29*, 211-224.

Bresin, R. & Friberg, A. (1997). A multimedia environment for interactive music performance. In *Proceedings of KANSEI - The Technology of Emotion AIMI International Workshop, Genova, 1997*, (pp. 64-67).

Bresin, R. & Friberg, A. (2000). Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal, 24*, 44-63.

Bresin, R. & Friberg, A. (2001). Expressive musical icons. In *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland*, (pp. 141-143).

Bresin, R. & Widmer, G. (2000). Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results, *TMH-QPSR - Speech Music and Hearing Quarterly Progress and Status Report*, 2000-4, 1-6.

Bresin, R., De Poli, G. & Ghetta, R. (1995). Fuzzy performance rules. In *Proceedings of the KTH Symposium on ''Grammars for music performance'', Stockholm*, (pp. 15-36).

Bresin, R., De Poli, G. & Vidolin A. (1992). Symbolic and sub-symbolic rules system for real time score performance. In *Proceedings of the International Computer Music Conference 1992, San Jose*, (pp. 211-214).

Cambouropoulos, E. (1998). *Towards a General Computational Theory of Musical Structure*. PhD Thesis, Faculty of Music and Department of Artificial Intelligence, University of Edinburgh.

Carlson, R., Friberg, A., Frydén, L., Granström, B. & Sundberg, J. (1989). Speech and music performance: parallels and contrasts. *Contemporary Music Review, 4*, 389-402.

Carlson, R. & Granström, B. (1975). A phonetically oriented programming language for rule description of speech. In G. Fant, (ed.), *Speech communication 2*, Stockholm: Almquist & Wiksell, (pp.245-253).

Clarke, E. F. (1985). Structure and Expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition*, London: Academic press, (pp. 209-236).

Clarke, E. F. (1988). Generative principles in music performance. In J. Sloboda, (ed.), *Generative processes in music*, Oxford: Claredon press, (pp. 1-26).

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg, (ed.), *Studies of Music Performance*, Stockholm: Royal Swedish Academy of Music, Publication No. 39, (pp. 76-181).

De Poli, G. (2004). Methodologies for expressiveness modeling of and for music performance. *Journal of New Music Research 33*, 189-202.

De Poli, G., Rodà, A. & Vidolin, A. (1998). Note-by-Note Analysis of the influence of Expressive Intentions and Musical Structure in Violin performance. *Journal of New Music Research, 27*, 293-321.

Desain, P. & Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research, 56*, 285-292.

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

Friberg, A. (1995a). Matching the rule parameters of Phrase arch to performances of "*Träumerei*": A preliminary study. In A. Friberg & J. Sundberg (eds.), *Proceedings of the KTH symposium on Grammars for music performance May 27, 1995*, (pp. 37-44).

Friberg, A. (1995b). *A Quantitative Rule System for Musical Expression*. Doctoral dissertation, Royal Institute of Technology, Sweden.

Friberg, A. (2005). Home conducting – Control the overall musical expression with gestures. In *Proceedings of the 2005 International Computer Music Conference*, San Francisco: International Computer Music Association. (pp. 479-482).

Friberg, A. (2006). pDM: an expressive sequencer with real-time control of the KTH music performance rules. *Computer Music Journal, 30*, 37-48.

Friberg, A. & Battel, G. U. (2002). Structural Communication. In R. Parncutt & G. E. McPherson (Eds.), *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*. (pp. 199-218) New York: Oxford University Press.

Friberg, A. & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America, 98*, 2524-2531.

Friberg, A. & Sundberg, J. (1999) Does music performance allude to locomotion? A model of final *ritardandi* derived from measurements of stopping runners. *Journal of the Acoustical Society of America, 105*, 1469-1484.

Friberg, A. & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception, 19*, 333-349.

Friberg, A., Bresin, R., Frydén, L., & Sundberg, J. (1998) Musical punctuation on the microlevel: Automatic identification and performance of small melodic units. *Journal of New Music Research, 27*, 271-292.

159

Friberg, A., Frydén, L., Bodin, L.-G., & Sundberg, J. (1991) Performance Rules for Computer-Controlled Contemporary Keyboard Music. *Computer Music Journal, 15*, 49-55.

Frydén, L., Sundberg, J. & Askenfelt, A. (1988). Perception aspects of a rule system for converting melodies from musical notation into sound. *Archives of Acoustics 13*, 269-278.

Fyk, J. (1995) *Melodic Intonation, Psychoacoustics, and the Violin*. Zielona Góra: Organon.

Gabrielson, A. & Juslin, P. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*, 68–91.

Gabrielsson, A. (1985). Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception, 3,* 59-86.

Gabrielsson, A. (1987). Once again: the theme from Mozart's piano sonata in A major (K. 331) A comparison of five performances. In A. Gabrielsson (ed.), *Action and Perception in Rhythm and Music*, Stockholm: Royal Swedish Academy of Music, Publication No. 55, (pp. 81-103).

Gabrielsson, A. (1999). Music Performance. in D. Deutsch (ed.), *The Psychology of Music* (2nd ed.), New York: Academic press. (pp. 501-602).

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology, 24*, 193-213.

Garbuzov, N. (1948). *Zonal nature of a pitch perception*. Moscow: Academy of Science of the USSR, in Russian.

Gilden, D. L., Thornton, T. & Mallon, M. W. (1995). 1/f Noise in Human Cognition. *Science, 267*, 1837-1839.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*, 563-572.

Hagerman, B. & Sundberg, J. (1980). Fundamental frequency adjustment in barbershop singing. *Journal of Research in Singing, 4*, 3-17.

Hansen, K.F. (2006). Musical structure: A translation of István Ipolyi's (1952) Innføring i Musikkspråkets Opprinnelse og Struktur. *Speech, Music and Hearing Quarterly Progress and Status Report, 48*.

Hefling, S. E. (1993). *Rhythmic alteration in Seventeenth- and Eighteenth-Century Music*: Notes inégales *and* overdotting. New York: Schirmer Books.

Hellkvist, A. (2004). Implementation of Performance Rules in Igor Engraver. Master thesis, Department of Information Technology, Uppsala University, Sweden.

Henderson, A. T. (1936). Rhythmic organization in artistic piano performance. In C. E. Seashore (ed.), *Objective analysis of musical performance. Univ. of Iowa Studies in the Psychology of Music, Vol IV*, Iowa City: University of Iowa, (pp. 281-305).

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49*, 621-30.

Honing, H. (2005a). Timing is tempo-specific. *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association. (pp. 359-362).

Honing, H. (2005b). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception, 23*, 79-85.

Honing, H. (in press 2006). Computational modeling of music cognition: A case study on model selection. *Music Perception 23,* 365-376.

Hudson, R. (1994). *Stolen Time - The History of Tempo Rubato*. Oxford: Clarendon press.

Ipolyi, I. (1952). *Innføring i musikspråkets opprinnelse och struktur*. Bergen: J.W. Eides Forlag (in Norwegian).

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1797-1813.

Juslin, P. N. (2001). Communication of emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-337). New York: Oxford University Press.

Juslin, P. N., Friberg, A. & Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae special issue 2001-2002*, 63-122.

Juslin, P. N. & Sloboda J. A. (Eds.) (2001). *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.

Kroiss W. (2000). Parameteroptimierung für ein Modell des musikalischen Ausdrucks mittels genetischer Algorithmen. Master thesis, Department of Medical Cybernetics and Artificial Intelligence, University of Vienna

Krumhansl, C. L. & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in spatial representation of musical keys, *Psychological Review, 89*, 334-368.

Larsson, B. (1977) Music and singing synthesis equipment (musse). STL-QPSR, 18(1):38-40.

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory*

of tonal music. Cambridge, Mass.: The MIT Press.

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press.

Mathews, M. V. (1989). The Conductor Program and the Mechanical Baton. In M. Mathews and J. Pierce, eds. *Current Directions in Computer Music Research*. Cambridge, Massachusetts: MIT Press, (pp. 263–282).

Mathews, M. V., Friberg, A., Bennett, G., Sapp, C. & Sundberg, J. (2003). A marriage of the Director Musices program and the conductor program. In R Bresin (Ed.) *Proceedings of the Stockholm Music Acoustics Conference 2003*, Vol I, (pp. 13-16).

Mazzola, G. (1990). *Geometrie der Töne. Elemente der Mathematischen Musiktheorie*. Basel: Birkhäuser Verlag.

Mazzola, G. (Ed.) (2002). *The topos of music – geometric logic of concepts, theory, and performance*. Basel: Birkhäuser Verlag.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*, Chicago: University of Chicago Press.

Palmer, C. (1996). On the Assignment of Structure in Music Performance. *Music Perception. 14*, 23-56.

Povel, D.-J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411-440.

Rasch, R. A. (1979). Synchronization in Performed Ensemble Music. *Acustica, 43,* 121-131.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America, 92*, 2546-2568.

Repp, B. H. (1994). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56*, 269-284.

Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill. (reprinted 1967 by Dover Publications, New York).

Sundberg, J. (1993). How can music be expressive? *Speech Communication, 13*, 239-253.

Sundberg, J. (2000). Grouping and differentiation. Two main principles in the performance of music. In T. Nakada (Ed.), *Integrated human brain science: Theory, method application (music)*. (p. 299-314). Amsterdam: Elsevier.

Sundberg, J. & Lindquist, J. (1973). Musical octaves and pitch. *Journal of the Acoustical Society of America, 54*, 922-929.

Sundberg, J. & Verrillo, V. (1980). On the anatomy of the ritard: A study of timing in music. *Journal of the Acoustical Society of America, 68*, 772-779.

Sundberg, J., Askenfelt, A. & Frydén, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal, 7*, 37-43.

Sundberg, J., Friberg, A. & Bresin, R. (2003). Attempts to reproduce a pianist's expressive timing with Director Musices performance rules. *Journal of New Music Research, 32*, 317-326.

Sundberg, J., Friberg, A. & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review, 3*, 89-109.

Sundberg, J., Friberg, A. & Frydén, L. (1991). Threshold and preference Quantities of Rules for Music Performance. *Music Perception, 9*, 71-92.

Sundberg, J., Frydén, L. & Askenfelt, A. (1982). What tells you the player is musical? A study of music performance by means of analysis-by-synthesis. *KTH Speech Transmission Laboratory Quarterly Progress and Status Report, 23*, 135-148.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press.

Thompson, W. F., Sundberg, J., Friberg, A. & Frydén, L. (1989). The Use of Rules for Expression in the Performance of Melodies. *Psychology of Music, 17*, 63-82.

Todd, N. P. McA. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33-58.

Todd, N. P. McA. (1989). A computational model of rubato. *Contemporary Music Review, 3*, 69-88.

Vernon, L. N. (1936). Synchronization of chords in artistic piano music. In C.E. Seashore (ed.) *Objective analysis of musical performance. University of Iowa studies in the psychology of music Vol. 5*, (pp 306-345), Iowa city: University of Iowa press.

Widmer, G. (1995). Modeling rational basis for musical expression. *Computer Music Journal, 19*, 76-96.

Widmer, G. (2002). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research, 31*, 37-50.

Widmer, G. & Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research, 33*, 203-216.

Widmer, G. & Tobudic, A. (2003). Playing Mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research, 32*, 259-268.

Zanon, P. & De Poli, G. (2003). Estimation of Time-Varying Parameters in Rule Systems for Music Performance. *Journal of New Music Research, 32*,