

Roberto Bresin and Anders Friberg

Speech Music Hearing Department (TMH)
Royal Institute of Technology (KTH)
Drottning Kristinas väg 31
100 44 Stockholm, Sweden
[roberto, andersf]@speech.kth.se
www.speech.kth.se/music/

When interpreting a musical score, performers introduce deviations in time, sound level, and timbre from the values indicated by composers. These deviations depend on the expressive mechanical and acoustical possibilities offered by the instrument they are playing and can vary in nature. Musical structure, biological motion, timekeeper variance and motor variance, and the performer's expressive intentions are the most common sources of deviation in a performance. In this paper, we focus on the deviations that render different emotional characteristics in music performance as part of expressive intentions.

Many important contributions to research in this area derive from Alf Gabrielsson's group at the University of Uppsala. Much of their work has centered on the so-called *basic emotions* (anger, sadness, happiness, and fear, sometimes complemented with solemnity and tenderness). The group found that all of these emotions, as conveyed by players, could be clearly recognized by an audience containing both musically trained and untrained listeners (Juslin 1997a; Juslin 1997b). Other researchers have shown that more complex emotional states can be successfully communicated in performance, although it is not completely clear how these states are defined—because performers and listeners often use different terms in describing intentions and perceived emotions (Canazza et al. 1997; Battel and Fimbiani 1998; De Poli, Rodà, and Vidolin 1998; Orio and Canazza 1998).

The identification of a communication code in real performances is another essential task in research on emotional aspects of music performance, and researchers have approached this problem in several ways. Gabrielsson and Juslin (1996) asked professional musicians to play a set of simple melodies with different prescribed emotions. From these renditions, the authors identified sets of potential acoustical cues that players of different instruments (violin, flute, and guitar) used to encode

Emotional Coloring of Computer-Controlled Music Performances

emotions in their performances which were similar to the sets listeners used for decoding these emotions (Juslin 1997c). A difficulty with this experimental design is that conflicts may exist between the general character of the composition and the prescribed emotional quality of its performance.

Another important aspect is to what extent the characteristic performance variations contain information that is essential to the identification of emotion. The significance of various performance cues in the identification of emotional qualities of a performance has been tested in synthesis experiments. Automatic performances were obtained by setting certain expressive cues to greater or lesser values, either in custom developed software (Canazza et al. 1998) or on a commercial sequencer (Juslin 1997c), and in formal listening tests listeners were able to recognize and identify the intended emotions. In the computer program developed by Canazza et al., the expressiveness cues were applied to a "neutral" performance as played by a live musician with no intended emotion, as opposed to a computer-generated "deadpan" performance. Juslin manually adjusted the values of some previously identified cues by means of "appropriate settings on a Roland JX1 synthesizer that was MIDI-controlled" by a Synclavier III. No phrase markings were used, and Juslin assumes that this reduced listeners' ratings of overall expressiveness.

The present work attempts to further explore the flexibility of Director Musices (DM), a program for automatic music performance that has been developed in our group (Friberg et al. 2000). More specifically, we have explored the possibilities of using the DM program to produce performances that differ with respect to emotional expression.

Director Musices

The DM program is a LISP language implementation of the KTH performance rule system (e.g., Friberg 1991, 1995a; Sundberg 1993), for automatic

performance of music. It contains about 25 rules that model performers' renderings of, for example, phrasing, accents, or rhythmic patterns. The DM program consists of a set of context-dependent rules that introduce modifications of amplitude, duration, vibrato, and timbre. These modifications are dependent on and thus reflect musical structure as defined by the score, complemented by chord symbols and phrase markers.

The program has a deterministic nature: it always introduces the same modifications given identical musical contexts. This determinism may appear counterintuitive; however, while DM always generates the same performance of a given composition, it is well known that a piece of music can be performed in a number of different but musically acceptable ways, depending on, among other things, the performer.

However, Friberg (1995b) demonstrated that DM can indeed produce different performances of a piece, simply by varying the magnitude of the effects of a single rule that was applied to four hierarchical phrase levels. By these simple means, DM was capable of generating inter-onset interval (IOI) patterns that closely matched those observed in recordings by three famous pianists. This result indicates that most differences between performances can be explained in terms of how players emphasize and demarcate the structure.

Macro-Rules for Emotional Expressive Performance

Gabrielsson (1994, 1995) and Juslin (1997a, 1997c) proposed a list of expressive cues that seemed characteristic of each of the emotions fear, anger, happiness, sadness, solemnity, and tenderness (see Table 1). The cues, described in qualitative terms, concern tempo, sound level, articulation (staccato/legato), tone onsets and decays, timbre, IOI deviations, vibrato, and final *ritardando*. These descriptions were used as a starting point for selecting rules and rule parameters that could model each emotion. The cues were restricted to those possible on a keyboard instrument, therefore eliminating the cues of tone onset and decay, timbre,

and vibrato, although these do belong to the Gabrielsson and Juslin list of characteristic cues. The cues used here are listed in Table 1.

The method used was analysis by synthesis (Gabrielsson 1985). Rules and rule parameters modeling each emotion were selected by a panel consisting of Roberto Bresin, Anders Friberg, Johan Sundberg, and Lars Frydén, the latter an expert musician and principal advisor in designing the rules in DM. After listening to the resulting performance, the parameters were further adjusted, and the whole process was repeated. After trying several musical examples, a consensus was obtained, resulting in a macro-rule ("rule palette" in DM) consisting of a set of rules and parameters for each emotion. Each macro-rule could be applied with the same parameters to each of the musical examples tried.

There is no model for determining the basic tempo from a musical structure. In measurements by Repp (1992, 1998), the basic tempo in commercial recordings of the same music by different pianists differed by up to a factor of two, thus indicating that such a model is not feasible. Therefore, the basic tempo corresponding to a "normal" musical performance had to be selected by the panel for each music example.

The following six of the existing rules were selected for modeling the considered emotions. *High Loud* (Friberg 1991, 1995a) increases the sound pressure level as a function of fundamental frequency. *Duration Contrast Articulation* (Bresin and Friberg 1998) inserts a micro pause between two consecutive notes if the first note has an IOI of between 30 and 600 msec. The *Duration Contrast* rule (Friberg 1991, 1995a) increases the contrast between long and short notes, i.e., comparatively short notes are shortened and softened, while comparatively long notes are lengthened and made louder. *Punctuation* (Friberg et al. 1998) automatically locates small tone groups and marks them with a lengthening of the last note and a following micropause. In the *Phrase Arch* rule (Friberg 1995b) each phrase is performed with an arch-like tempo and sound level curve, starting slow/soft, becoming faster/louder in the middle, followed by a *ritardando/decrescendo* at the end.

Table 1. Cue profiles for each emotion, as outlined by Gabrielsson and Juslin, are compared with the rule setup used for synthesis with Director Musices.

<i>Emotion</i>	<i>Expressive Cue</i>	<i>Gabrielsson and Juslin</i>	<i>Macro-Rule in Director Musices</i>
<i>Fear</i>	<i>Tempo</i>	Irregular	<i>Tone IOI</i> is lengthened by 80%
	<i>Sound Level</i>	Low	<i>Sound Level</i> is decreased by 6 dB
	<i>Articulation</i>	Mostly staccato or non-legato	Duration Contrast Articulation rule
	<i>Time Deviations</i>	Large Structural reorganizations Final acceleration (sometimes)	Duration Contrast rule Punctuation rule Phrase Arch rule applied on phrase level Phrase Arch rule applied on sub-phrase level Final Ritardando
<i>Anger</i>	<i>Tempo</i>	Very rapid	<i>Tone IOI</i> is shortened by 15%
	<i>Sound Level</i>	Loud	<i>Sound Level</i> is increased by 8 dB
	<i>Articulation</i>	Mostly non-legato	Duration Contrast Articulation rule
	<i>Time Deviations</i>	Moderate Structural reorganizations Increased contrast between long and short notes	Duration Contrast rule Punctuation rule Phrase Arch rule applied on phrase level Phrase Arch rule applied on sub-phrase level
<i>Happiness</i>	<i>Tempo</i>	Fast	<i>Tone IOI</i> is shortened by 20%
	<i>Sound Level</i>	Moderate or loud	<i>Sound Level</i> is increased by 3 dB High Loud rule
	<i>Articulation</i> <i>Time Deviations</i>	Airy Moderate	Duration Contrast Articulation rule Duration Contrast rule Punctuation rule Final Ritardando rule
<i>Sadness</i>	<i>Tempo</i>	Slow	<i>Tone IOI</i> is lengthened by 30%
	<i>Sound Level</i>	Moderate or loud	<i>Sound Level</i> is decreased by 6 dB
	<i>Articulation</i> <i>Time Deviations</i> <i>Final Ritardando</i>	Legato Moderate Yes	Duration Contrast rule Phrase Arch rule applied on phrase level Phrase Arch rule applied on sub-phrase level Obtained from the Phrase rule with the next parameter
<i>Solemnity</i>	<i>Tempo</i>	Slow or moderate	<i>Tone IOI</i> is lengthened by 30%
	<i>Sound Level</i>	Moderate or loud	<i>Sound Level</i> is increased by 3 dB High Loud rule
	<i>Articulation</i> <i>Time Deviations</i>	Mostly legato Relatively small	Duration Contrast Articulation rule Punctuation rule
	<i>Final Ritardando</i>	Yes	Final Ritardando rule
<i>Tenderness</i>	<i>Tempo</i>	Slow	<i>Tone IOI</i> is lengthened by 30%
	<i>Sound Level</i>	Mostly low	<i>Sound Level</i> is decreased by 6 dB
	<i>Articulation</i> <i>Time Deviations</i>	Legato Diminished contrast between long and short notes	Duration Contrast rule
	<i>Final Ritardando</i>	Yes	Final Ritardando rule

The rule can be applied to different hierarchical phrase levels. *Final Ritardando* (Friberg and Sundberg 1999) introduces a *ritardando* at the end of the piece similar to a stopping runner.

It was also necessary to introduce two new rules for controlling overall tempo and sound level. The *Tone Duration* rule shortens or lengthens all notes in the score by a constant percentage. The *Sound Level* rule decreases or increases the sound level (SL) of all notes in the score by a constant value in decibels. The performance variables and the input parameters of these eight rules are listed in Table 2.

The resulting groups of rules, i.e., macro-rules, for each intended emotion are listed in the right column of Table 1. The parameters for each of the rules are shown graphically in Figure 1. The rules contained in a macro-rule are automatically applied in sequence, one after the other, to the input music score. The effects produced by each rule are added to the effects produced by previous rules.

As an example, the resulting deviations for the "angry" version of a Swedish nursery rhyme (*Ekorrn satt i granen*, or *The Squirrel Sat on the Fir Tree*, composed by Alice Tegnér and henceforth referred to as *Ekorrn*) are shown in Figure 2. The relative time deviation for each note's IOI varied between -20 and -40%. The negative values imply that the original tempo was slower, in accordance with the observations of Gabrielsson and Juslin. The IOI deviation curve contains small and quick oscillations; this reflects our attempt to reproduce what Gabrielsson and Juslin described as "moderate variations in timing." The large time deviations associated with short notes was produced by the duration contrast rule, thus realizing the "increased contrast between long and short notes" noted by Gabrielsson and Juslin. The graph also shows that a slight *accelerando/crescendo* appears in the end instead of a final *ritardando/descrescendo*.

The off-time duration (DRO) reached a maximum of 110 msec at the endings of phrases and subphrases. The Duration Contrast Articulation rule introduced small articulation pauses after all comparatively short notes, thus producing something equivalent to the "mostly non-legato articulation" observed by Gabrielsson and Juslin. The Punctuation

rule inserted larger articulation pauses at automatically detected structural boundaries. The overall net result was that all DRO values were positive.

At the endings of phrases and subphrases, the sound level increased, an effect caused by the inversion of the Phrase Arch rule. This rule attempted to realize the "structural reorganization" proposed by Gabrielsson and Juslin. The sound level curve shows positive deviations corresponding to the increased sound level also noted by Gabrielsson and Juslin. A more detailed description of the macro-rule used for the "angry" version of *Ekorrn* can be found in a previous work by the authors (Bresin and Friberg 2000).

Listening Experiment

According to Juslin (1997b), forced-choice judgments and free-labeling judgments give similar results when listeners attempt to decode a performer's intended emotional expression. Therefore, it was considered sufficient to make a forced-choice listening test to assess the efficiency of the emotional communication.

Music

Two clearly different pieces of music were used and are given in Figures 3 and 4. One was the melody line of *Ekorrn*, written in a major mode. The other was a computer-generated piece in a minor mode (henceforth called *Mazurka*), written in an attempt to portray the musical style of Frédéric Chopin (Cope 1992). The phrase structure with regard to the levels of sub-phrase (level 6), phrase (level 5) and piece (level 4) was marked in each score (shown in Figures 3 and 4), thus meeting the demands of the Phrase Arch rule.

Performances

Applying the macro-rules listed in Table 1 using the k-values shown in Figure 1, seven different versions of each of the two scores were generated.

Figure 1. Values used for the indicated rule parameters in the various rules for the emotions tenderness (T), sadness (Sa), solemnity (So), happiness (H), anger (A), and fear (F). Each rule is represented by a panel.

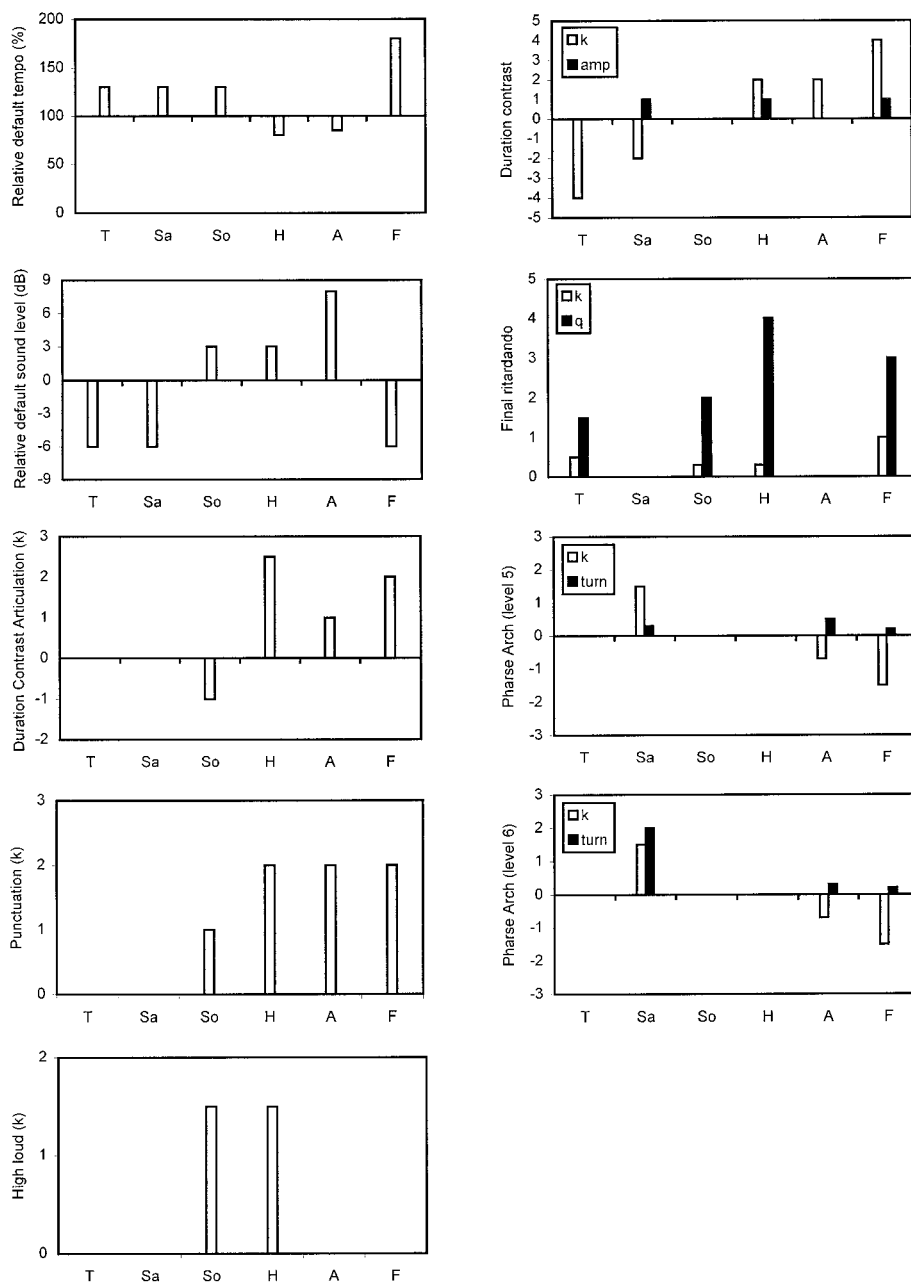


Table 2.

(a) List of the variables affected by the rules and of the parameters used for setting up the rules.

<i>Rule name</i>	<i>Affected variables</i>	<i>Parameters (default values)</i>
<i>High Loud</i>	<i>sl</i>	
<i>Duration Contrast</i>	<i>dr sl</i>	:amp 1 :dur 1
<i>Duration Contrast Articulation</i>	<i>dro</i>	
<i>Punctuation</i>	<i>dr dro</i>	:dur 1 :duroff 1
<i>Phrase Arch</i>	<i>dr sl</i>	:phlevel 7 :power 2 :amp 1 :next 1 :2next 1 :turn 2 :last 1 :acc 1
<i>Final Ritard</i>	<i>dr</i>	:q 3

(b) Description of the affected variables.

<i>dl</i>	Sound level in decibels relative to a constant value defined in DM (default value is $sl = 0$)
<i>dr</i>	Total inter-onset duration in msec
<i>dro</i>	Offtime duration, i.e., offset to next onset

(c) Description of the parameters.

<i>amp</i>	Sets the sound level as a factor multiplied by the default value
<i>dur</i>	Amount of the effect on the duration
<i>duroff</i>	Amount of the effect on the off-time duration
<i>phlevel</i>	Indicates the level of phrasing at which the rule produces the effect
<i>power</i>	Determines the shape of the accelerando and ritardando functions
<i>next</i>	Used to modify the amount of ritardando for phrases that also terminate a phrase at the next higher level
<i>turn</i>	The position of the turning point between the accelerando and the ritardando
<i>last</i>	Changes the duration of the last note of each phrase
<i>acc</i>	Controls the amount of accelerando, expressed in terms of factor multiplied by $k/2$
<i>q</i>	Curvature parameter

Note: The effect of each rule can be amplified or damped by multiplying it with an emphasis parameter called k .

Figure 2. Percent IOI deviation (top), off-time deviation (middle), and sound level deviation (bottom) for the "angry" version of Ekorn.

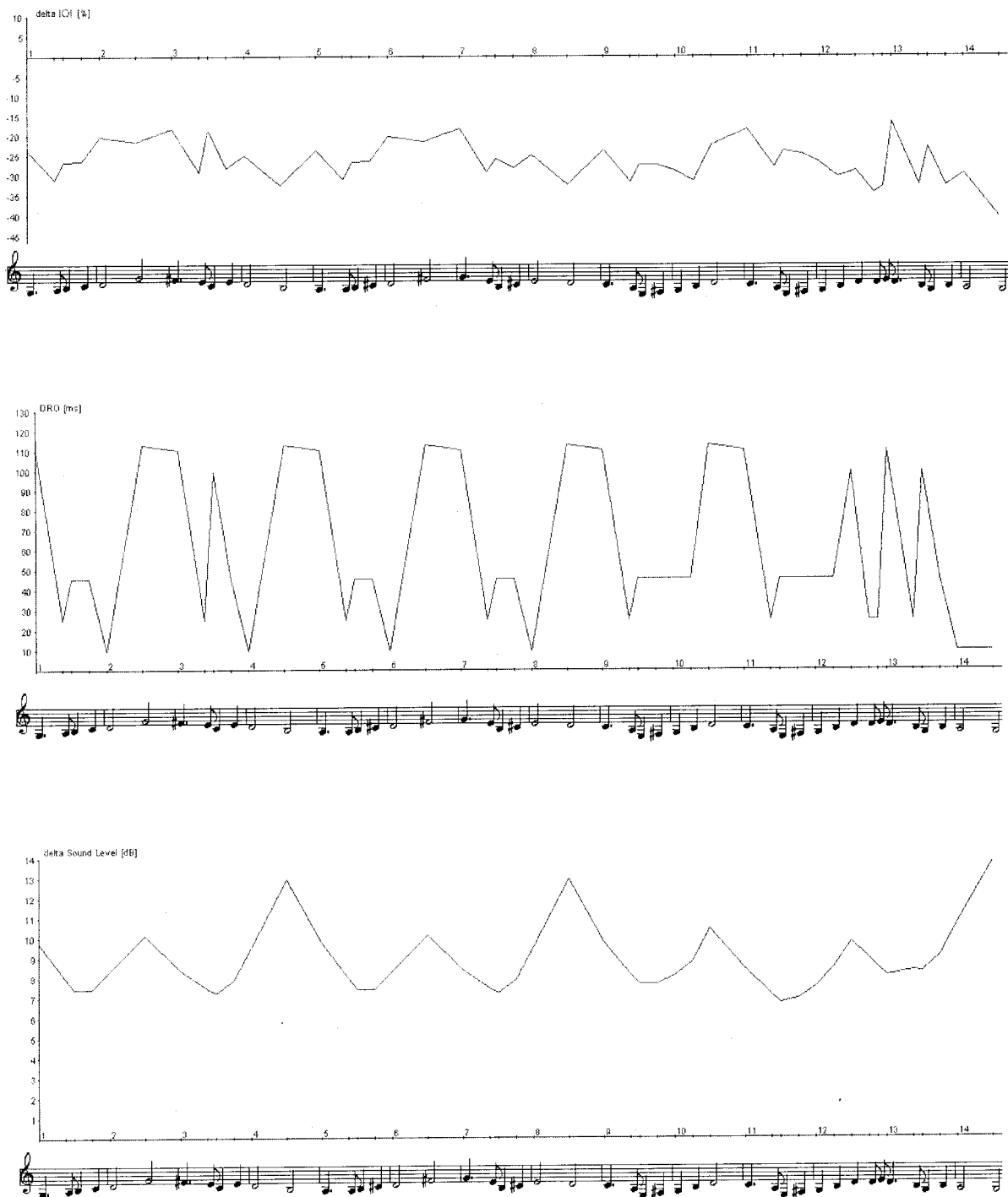


Figure 3. The melody Ekorn. The brackets above the staff correspond to different phrase levels; piece (4), phrase (5), and sub-phrase (6).



Figure 3



Figure 4

These versions represented our modeling of "fear," "anger," "happiness," "sadness," "solemnity," and "tenderness." In addition, a "dead-pan" version, henceforth called the "no-expression" version, was generated for the purpose of comparison. The tempo for the "no-expression" version was 187 quarter notes per minute for *Ekorn* and 96 quarter notes per minute for *Mazurka*. The same macro-rules were used for both *Ekorn* and *Mazurka*.

The polyphonic structure of *Mazurka*, represented as a multi-track MIDI file, required the use of a synchronization rule: the Melodic Sync rule. This rule generates a new track consisting of all tone onsets in all tracks. When there are several simultaneous onsets, the note with the maximum melodic charge is selected. All rules that modify the IOI are applied to this synchronization track, and the resulting IOIs are transferred back to the

Figure 4. The first 12 (of 23) measures of Mazurka. The brackets mark the phrase structure as in Figure 3. The entire compo-

sition can be found on the Internet at arts.ucsc.edu/faculty/cope/home.

original tracks (Sundberg, Friberg, and Frydén 1989). Rules that act on parameters other than IOI are applied to each track independently.

Figures A1 and A2 in the Appendix show the relative deviations of IOI and the variation of DRO and sound level for all six performances of each piece. Negative DRO values indicate a *legato* articulation of two subsequent notes, while positive values indicate a *staccato* articulation. Negative values of sound level deviation indicate a softer performance as compared with the "no-expression" performance.

Fourteen performances (7 emotions \times 2 examples)—originally stored in MIDI files—were recorded as 44.1 kHz, 16-bit, uncompressed sound files. A grand piano sound (taken from the Kurzweil sound samples of the Pinnacle Turtle Beach sound card) was used for the synthesis.

Figure 5. Ordered average percentage of "same" responses given by each subject for the repetitions of Ekorn and Mazurka.

Listeners

Twenty subjects, 24–50 years in age and consisting of five females and 15 males, volunteered as listeners. None were professional musicians or music students, although 18 of the subjects stated that they currently or formerly played an instrument on a non-regular basis. The subjects all worked at the Speech Music Hearing Department of the Royal Institute of Technology, Stockholm.

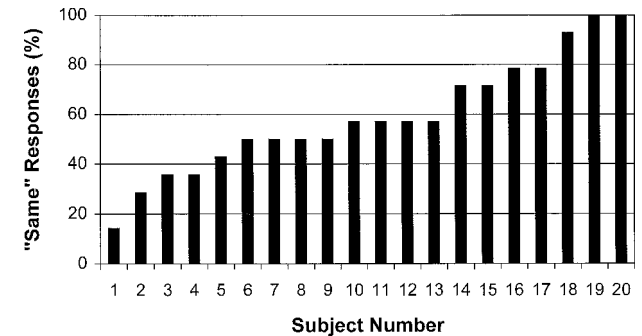
Procedure

The subjects listened to the examples individually over Sennheiser HD435 Manhattan headphones, adjusted to a comfortable level. Each subject was instructed to identify the emotional expression of each example as one out of seven alternatives: "fear," "anger," "happiness," "sadness," "solemnity," "tenderness," or "no-expression." The responses were automatically recorded with the Visor software system, specially designed for listening tests (Granqvist 1996). Visor presents the sound stimuli as anonymous boxes in random order and provides seven empty columns corresponding to the seven different emotions into which the boxes may be placed. Subjects were instructed to (1) double-click on each of the boxes and listen to its music, (2) identify the emotion conveyed by the performance, (3) drag the box to the corresponding column, and (4) continue until all the boxes were placed. Subjects could listen as many times as needed to each music sample, and they were given the opportunity to associate any number of boxes with the same emotion/column.

Each session contained two sub-tests repeated once, each presenting the seven performances of (1) *Ekorn*, (2) *Mazurka*, (3) repeat of *Ekorn*, and (4) repeat of *Mazurka*. The stimuli order within each sub-test was automatically randomized for each subject by the Visor program. The average duration of an experiment session was approximately eleven minutes.

Analysis

Consistency of the subjects' answers was analyzed from the repeated stimuli in the test. Four sub-



jects, who gave the same answers for repeated stimuli in fewer than 36% of the cases, were eliminated from subsequent analysis (see Figure 5).

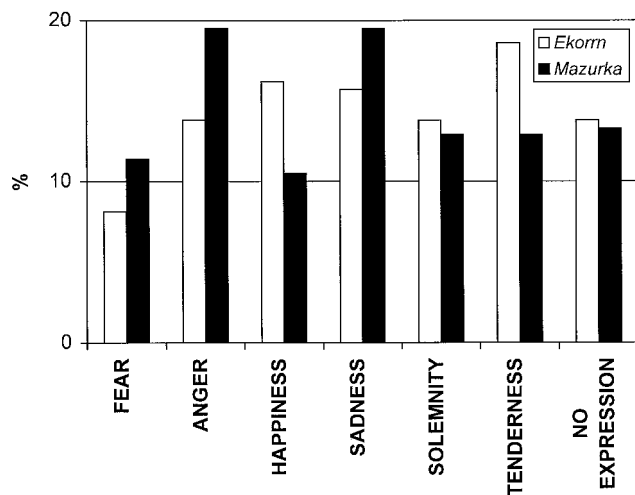
Statistical analysis of the results from the listening test was conducted. We conducted a three-way repeated-measures analysis of variance (ANOVA), with fixed factors Intended Emotion (7 levels), Perceived Emotion (7 levels), and Piece of Music (2 levels) and with listeners' choices as the dependent variable. For each intended emotion, listeners' choices were codified by marking the subjects' chosen emotions with the value one and the remaining six emotions with the value zero. In addition, a two-way ANOVA was conducted for each intended emotion separately with fixed factors Perceived Emotion (7 levels) and Piece of Music (2 levels).

The synthesis of each emotional performance was achieved by using a special macro-rule, each including a specific subset of DM rules, as already mentioned (see Table 1, Table 2, and Figure 1). A total of 17 parameters were involved in these macro-rules, some of which were interrelated. An attempt was made to reduce the number of dimensions of this space by means of a principal component analysis of these parameters.

Results and Discussion

The three-way ANOVA revealed that the main factor Perceived Emotion was significant ($F(6,3038) = 3.917, p < 0.0007$), implying that listeners overall preferred "angry" and "happy." However, the main factors Intended Emotion and

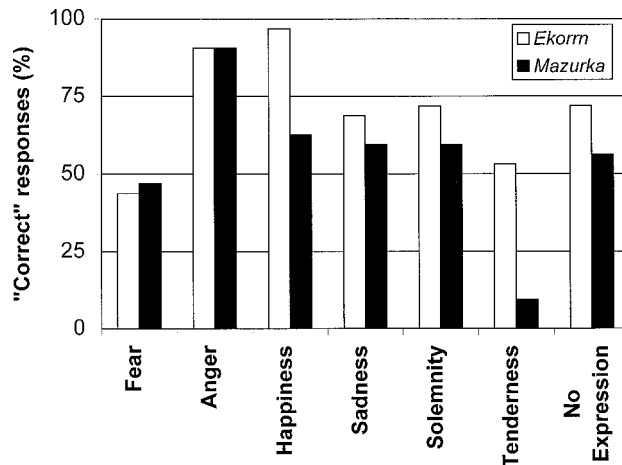
Figure 6. Percentage distribution of subjects' choices over the two pieces Ekorn and Mazurka.



Piece of Music were not significant. The interaction between Intended Emotion and Perceived Emotion was significant ($F(42, 3038) = 55.505, p < 0.0001$). The interaction between Intended Emotion, Perceived Emotion, and Piece of Music was also significant ($F(42, 3038) = 2.639, p < 0.0001$), showing an influence of the score on the perception of the intended emotion. An interesting result emerging from the three-way ANOVA was a significant interaction between Piece of Music and Perceived Emotion ($F(6, 3038) = 3.404, p < 0.0024$): subjects rated performances of *Mazurka* to be mainly "angry" and "sad," while performances of *Ekorn* were perceived as more "happy" and "tender." Both pieces were perceived with almost the same rate of "solemnity" and "no-expression," as shown in Figure 6. This result confirms the typical association of "happiness" with major mode and "sadness" with minor mode.

The significance of the main factor Perceived Emotion in the three-way ANOVA induced us to run a two-way ANOVA for each of the seven Intended Emotions in order to verify that each of them was significant. In each of the seven two-way ANOVAs, the effects of Perceived Emotion were, as expected, significant ($p < 0.0001$), and the effects of music were insignificant. Figure 7 shows the percentage of "correct" responses for the two pieces. In all cases but one, these percentages well exceeded the 14% chance level and were in all

Figure 7. Percent "correct" responses for the two test examples obtained for each of the indicated emotions.



cases but one the highest value. To facilitate a more detailed analysis, the subjects' responses for each version of *Ekorn* and *Mazurka* are presented as a confusion matrix in Tables 3a and 3b. An interesting outcome in both confusion matrixes is their high degree of symmetry along the main diagonal, except for a few terms, which implies consistency in the subjects' choices.

The "tenderness" version of *Mazurka* was the only case that was not correctly identified (9%). From the two-way ANOVA for the intended emotion "tenderness" emerged a significant interaction between Piece of Music and Perceived Emotion ($F(6,434) = 7.519, p < 0.0001$). To explain this significance, a *post hoc* comparison using the Scheffé test was applied to all possible pairs of the seven levels of the main factor Perceived Emotion for each Piece of Music separately. The Scheffé test revealed that *Ekorn* was perceived as "tender," while *Mazurka* was perceived as "sad." According to Juslin (1997b), the tenderness version is often performed with a higher sound level than the "sadness" version, while the same sound level for both versions was used in this experiment. This observation may help us achieve better macro-rules for "sadness" and "tenderness" performances in future work.

For happiness, the responses differed substantially between *Ekorn* (97%) and *Mazurka* (63%). This was confirmed by a significant interaction be-

Table 3.

Confusion matrix (%) for the classification test of seven synthesized performances of (a) *Ekorrn* and (b) *Mazurka*.

(a)							
Intended Emotion	<i>Fear</i>	<i>Anger</i>	<i>Happiness</i>	<i>Sadness</i>	<i>Solemnity</i>	<i>Tenderness</i>	<i>No Expression</i>
<i>Fear</i>	44	0	3	3	0	41	9
<i>Anger</i>	3	91	0	0	6	0	0
<i>Happiness</i>	0	3	97	0	0	0	0
<i>Sadness</i>	3	0	0	69	0	28	0
<i>Solemnity</i>	3	3	3	6	72	0	13
<i>Tenderness</i>	6.5	0	0	28	3	53	9.5
<i>No-Expression</i>	0	0	9	0	13	6	72
(b)							
Intended Emotion	<i>Fear</i>	<i>Anger</i>	<i>Happiness</i>	<i>Sadness</i>	<i>Solemnity</i>	<i>Tenderness</i>	<i>No Expression</i>
<i>Fear</i>	47	0	0	13	3	31	6
<i>Anger</i>	6	91	0	3	0	0	0
<i>Happiness</i>	6	28	63	0	3	0	0
<i>Sadness</i>	3	0	0	59	0	38	0
<i>Solemnity</i>	3	16	0	13	59	0	9
<i>Tenderness</i>	13	0	0	50	3	9	25
<i>No-Expression</i>	3	0	9.5	0	22	9.5	56

tween Piece of Music and Perceived Emotion ($F(6,434) = 11.453, p < 0.0001$). For the remaining five Intended Emotions, the interaction between Perceived Emotion and Piece of Music was not statistically significant.

The fear version of *Ekorrn* was classified both as "afraid" and as "tender": 44% of the subjects perceived it as "afraid," and 41% of them perceived it as "tender." The fear version of *Mazurka* was significantly perceived as more "afraid" and less "tender."

These differences in the perception of *Ekorrn* and *Mazurka* performances could be partly because *Mazurka* is written in a minor mode, which

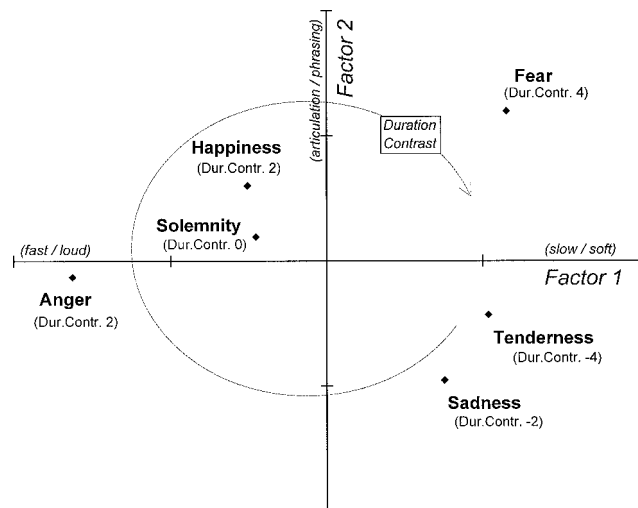
in the western music tradition is often associated with moods of sadness, fear, or anger.

From the principal component analysis, two principal factors emerged, explaining 61% (Factor 1) and 29% (Factor 2) of the total variance. Figure 8 presents the main results in terms of the distribution of the different setups in the two-dimensional space.

Factor 1 was closely related to deviations of sound level and tempo. Louder and quicker performances had negative values of Factor 1 and were delimited by the coordinates of "anger" and "solemnity." Softer and slower performances had

Figure 8. Two-dimensional space of the indicated emotions derived from a principal component analysis of all rule parameters in all rules used for the different emotional

rule setups. The dashed curve represents an approximation of how the k -value of the rule Duration Contrast varied between the setups.



positive values of Factor 1 and were delimited by the coordinates of "sadness" and "fear."

Factor 2 was closely related to the articulation and phrasing variables. This distribution resembles those presented in previous works and obtained using other methods (De Poli, Rodà and Vidolin 1998; Orio and Canazza 1998; Canazza et al. 1998). The quantity values (k -values) for the duration contrast rule in the macro-rules, also shown in Figure 8, increased clockwise from the fourth quadrant to the first. Figure 8 also shows an attempt to qualitatively interpret the variation of this rule in the space. It shows that in the "tender" and "sad" performances, the contrast between note durations was very slight (shorter notes were played longer), strong in the "angry" and "happy" performances, and strongest in the "fear" version.

The acoustical cues resulting from the six synthesized emotional performances are presented in Figure 9. The average IOI deviations are plotted against the average sound level together with their standard deviations for all six performances of each piece. Negative values of sound level deviation indicate a softer performance compared to the "no-expression" version. "Anger" and "happiness" performances were thus played quicker and louder, while "tenderness," "fear," and "sadness" performances were slower and softer relative to a "no-expression" rendering. Negative values of IOI

deviation imply a tempo faster than the original, and vice versa for positive values. The "fear" and "sadness" versions have larger standard deviations obtained mainly by exaggerating the duration contrast but also by applying the phrasing rules. These acoustic properties fit very well with the observations of Gabrielsson and Juslin and also with performance indications in the classical music tradition. For example, in his *Versuch über die wahre Art das Clavier zu spielen* (1753, reprinted 1957), Carl Philippe Emanuel Bach wrote ". . . activity is expressed in general with staccato in Allegro and tenderness with portato and legato in Adagio . . ."). Figure 9 could be interpreted as a mapping of Figure 8 into a two-dimensional acoustical space.

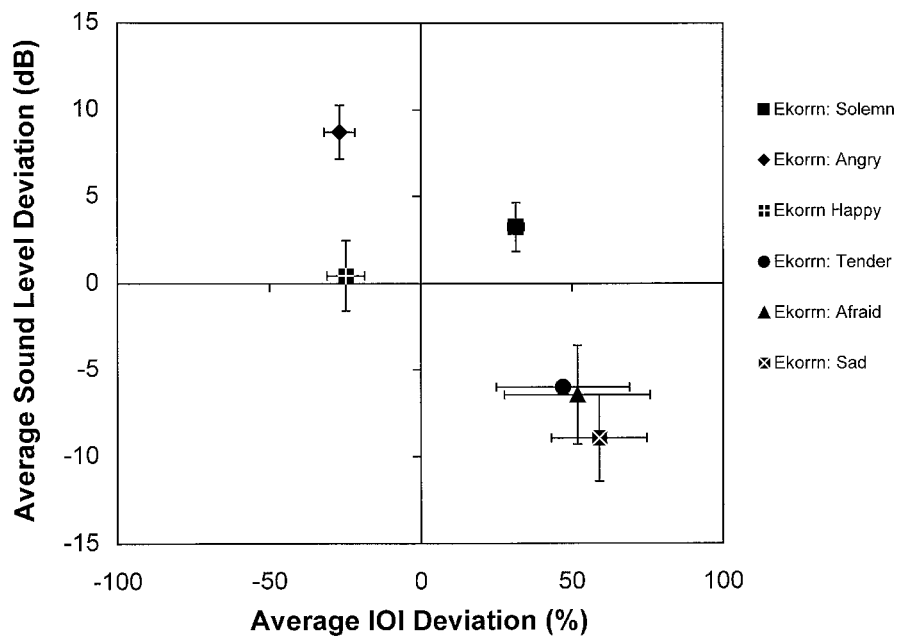
The Gabrielsson and Juslin experiment included the cues tone onset and decay, timbre, and vibrato. None of these can be implemented in piano performances and thus had to be excluded from the present experiment. Yet our identification results were similar to or even better than those found by Gabrielsson and Juslin. A possible explanation is that listeners' expectations regarding expressive deviations vary depending on the instrument played.

Finally, the SL and time deviations used here in the seven performances can be compared to those observed in other studies of expressiveness in singing and speech. Figure 10 presents such a comparison. For tempo, the values were normalized with respect to "happy," because "neutral" was not included in all studies compared. In many cases, striking similarities can be observed. For example, most investigations reported a slower tempo and a lower mean sound level in "sad" than in "happy." The similarities indicate that similar strategies are used to express emotions in instrumental music, singing, and speech.

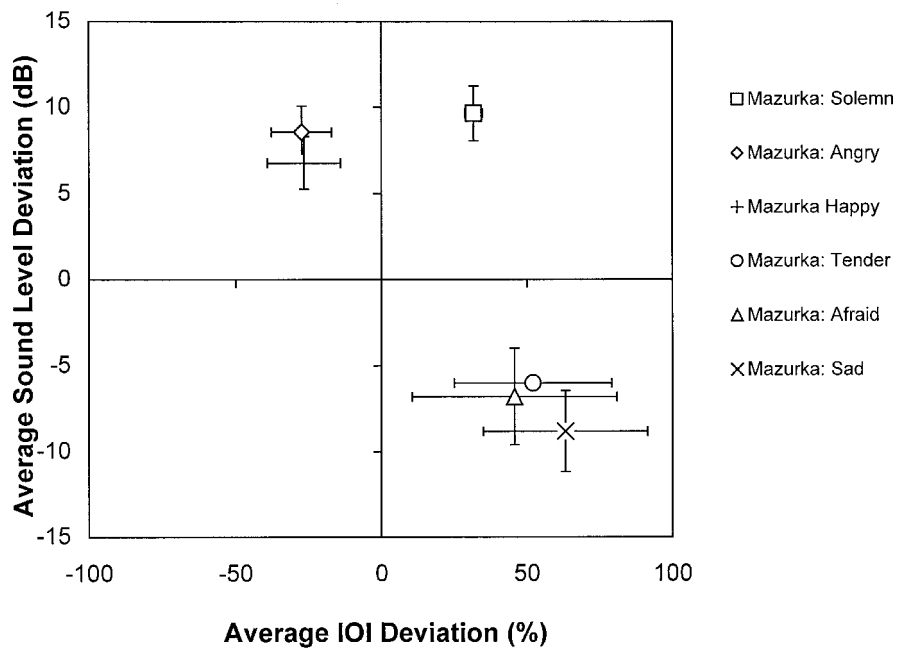
General Discussion

The main result of the listening experiment was that the emotions associated with the DM macro-rules were correctly perceived in most cases. This suggests that DM performance rules can be grouped into macro-rules to produce performances that can be associated with different emotional states.

Figure 9. Average IOI deviations versus average sound level deviations for all six performances of (a) Ekornn and (b) Mazurka. The bars represent the standard deviations.



(a)

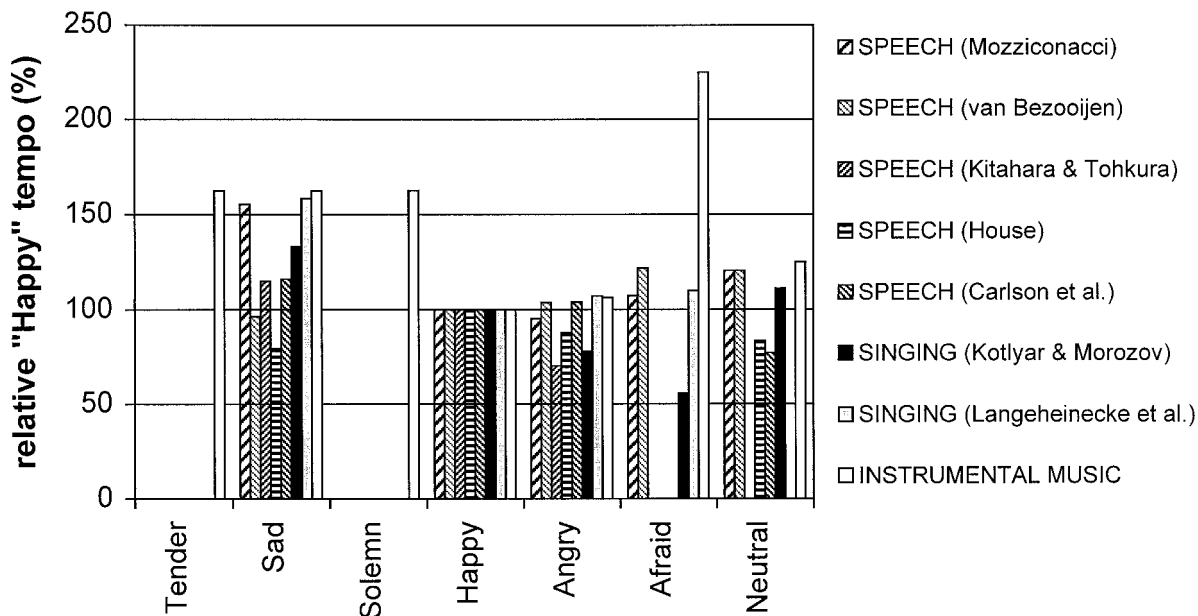


(b)

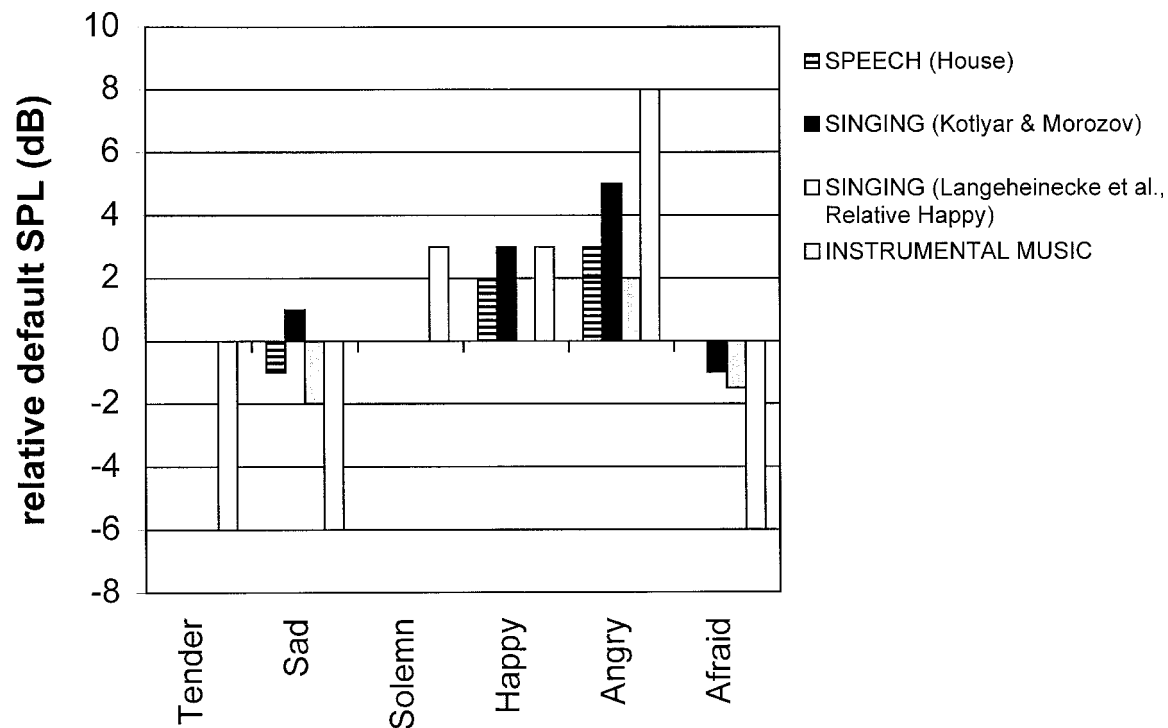
Figure 10. Relative deviations of (a) tempo and (b) sound pressure level (SPL) as reported in speech research (Mozziconacci 1998; van Bezooijen 1984;

Kitahara and Tohkura 1992; House 1990; Carlson, Granström, and Nord 1992), singing voice research (Kotlyar and Morozov 1976;

Langeheinecke et al. 1999), and in the present study (indicated as "instrumental music" in the figure).



(a)



(b)

An important observation is that the same macro-rules could be applied with reasonably similar success to both *Ekorrn* and *Mazurka*—two compositions with very different styles. This is not surprising, because the performance rules have been designed to be generally applicable. On the other hand, still better results may be possible if somewhat different versions of the macro-rules are used for scores of differing styles. For example, a style-dependent variation of the quantities for the *staccato/legato* and the phrase-marking rules may be introduced.

The magnitude of each effect, i.e., the k -values, plays an important role in the differentiation of emotional expression. Quantities greater than unity imply an amplification of the effect of the rule, values between zero and unity a reduced effect, and negative values the inverse of the effect. While the default value is $k = 1$, higher values have been found optimal if a rule is applied in isolation (Sundberg, Friberg, and Frydén 1991). In general, quantities deviating from unity have been found appropriate mainly when many rules are applied in combination and for the purpose of adapting the rules to the intrinsic expressive character of the composition. Values smaller than unity may be needed when the net effect of many rules produces exaggerated effects. In the present experiment, exaggerated values for some of the rules were found useful for producing emotionally expressive performances where emphasis or warping of the musical structure was needed. In studies comparing expressive and neutral renderings of the same musical examples, greater expressive deviations have been found in the expressive versions (Sundberg 2000).

The misclassification of *mazurka* as "sad" rather than the intended alternative mood of "tender" may be related to tonality. According to Meyer (1956), the minor mode is typically associated with intense feeling and "sadness, suffering, and anguish in particular." Meyer argues that this association "is a product of the deviant, unstable character of the mode and of the association of sadness and suffering with the slower tempi that tend to accompany the chromaticism prevalent in the minor mode." Associations between minor mode with sadness and major mode with happiness have been experimentally confirmed in stud-

ies of both children and adults (Gerardi and Gerken 1995; Kastner and Crowder 1990). According to this reasoning, the minor mode of *mazurka* may explain why this piece was much less successful than *Ekorrn* in its "happiness" versions.

Applications

Our findings indicate interesting new potential for the DM system. Our results clearly demonstrate that, in spite of its deterministic structure, the DM system is capable of generating performances that differ significantly in emotional quality. This invites some speculation regarding future applications.

One immediate application is the design of new plug-ins for commercial music sequencers and editors. Users would be able to produce emotionally colored performances of any MIDI file while still using their favorite software tool.

Another possibility would be to use the new macro-rules as tools for objective performance analysis. Macro-rules could be used in reverse order to analyze the emotional content of a performance. It has been demonstrated that the rule parameters can be automatically fitted to a specific performance (c.f. Friberg 1995b); this would allow the classification of deviations in various performance parameters in terms of plausible intended emotion. This presents a new perspective in the field of music performance didactics, where DM has already been successfully applied (Friberg and Battel forthcoming).

A third possibility is to design an "emotional toolbox" capable both of recognizing the emotion of players/singers and of translating it back in automatic performance. This could be applied to smart Karaoke applications included in the future MPEG-7 standard for improving human-machine interaction. The user might sing the melody of a song to retrieve it from large MIDI file databases on the Internet, and the Karaoke system could play back the file with the same emotions used by the singer (Ghias et al. 1995).

A fourth appealing application is that of consumer electronics involving automatic perfor-

mance of synthesized sounds, for instance in cellular phones. Rule-based techniques can be used to improve the quality of the ringing tone performance of cellular phones. This is probably the most popular music synthesizer, at least with the highest diffusion, and we listen to its mechanical and dull performances every day. In a current project, we are using our technique to design an emotional cellular phone. The caller will be able to send a code to another's phone and, by influencing the performance of the receiver's ringing tone, include an emotional content. In this way, it will be possible to have a musical interaction between the caller and the receiver. Furthermore, some cellular phone producers include the possibility of using MIDI standard as a format for ringing tone files.

Conclusions

The results demonstrate that, in spite of the deterministic nature of the DM program, different performances of the same score can be achieved by applying different combinations of rules. By interpreting the results of Gabrielsson and Juslin in terms of rules incorporated in the DM program, emotionally differing performances can be produced. Because all DM rules are triggered by the structure as defined by the score in terms of, for example, a combination of note values or intervals, the representation of musical structure seems to play a decisive role in creating emotional expressiveness. According to a principal component analysis applied to the rule parameters, the two most important dimensions for emotional expressiveness are mean SL and tempo (Factor 1), and phrasing and articulation (Factor 2). Therefore, the Duration Contrast rule plays a particularly productive role. Mean SL and mean tempo seem to produce similar contributions in performance of instrumental music as in speech and singing. The results from the listening test reflect the typical association of happiness with major mode and sadness with minor mode, and the association of tenderness with major mode and anger with minor mode. In future research, we will ex-

amine the formulation of new rules for the rendering of expressive articulation. In fact, in a recent study, Bresin and Battel (in press) showed how *legato*, *staccato*, and repeated tone articulation varies significantly in performances biased by different expressive adjectives.

Acknowledgments

This work was supported by the Bank of Sweden Tercentenary Foundation. The authors are grateful to Johan Sundberg, who assisted in editing the manuscript, to Peta White, who edited our English, to Diego Dall'Osto and Patrik Juslin for valuable discussions, to the reviewers, and to the twenty subjects who participated in the listening test.

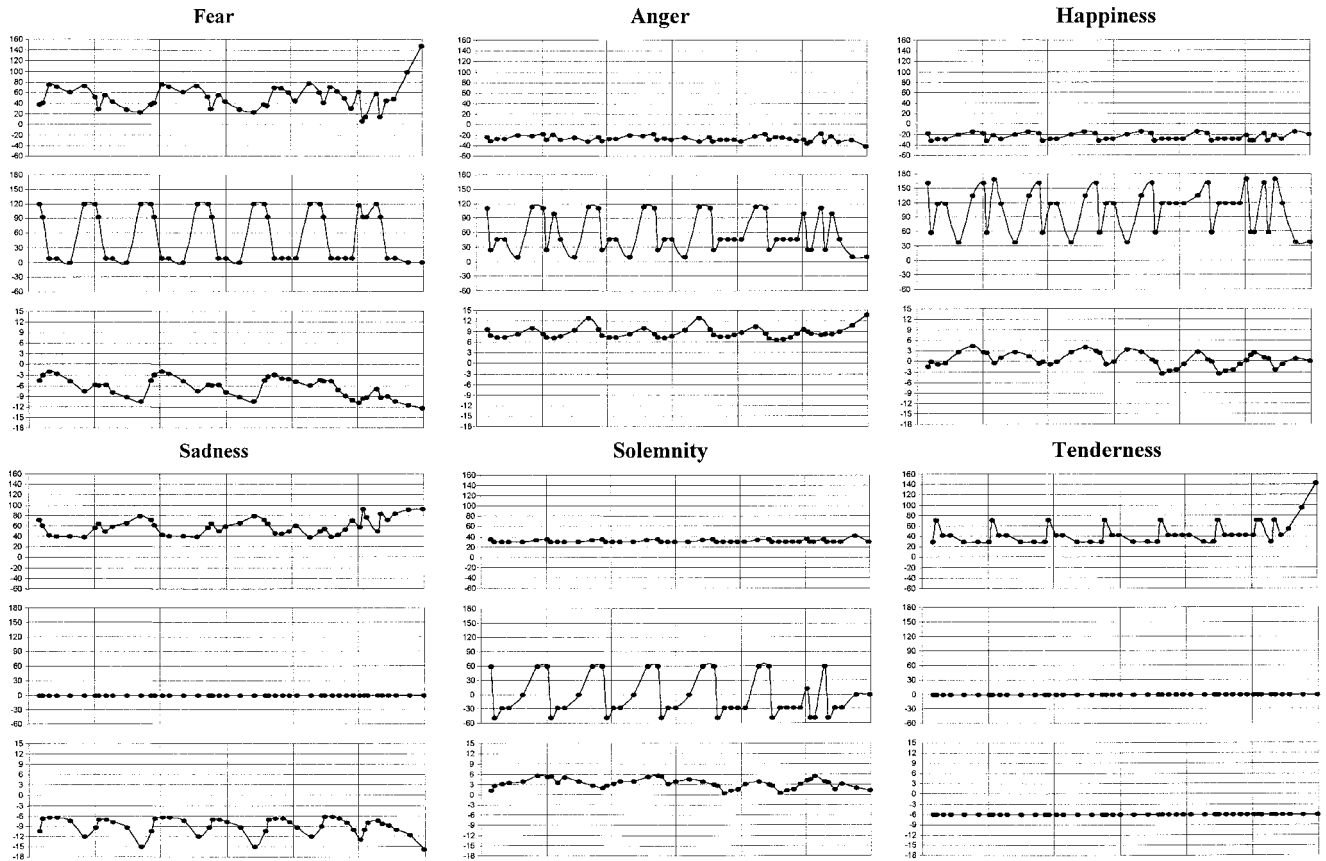
References

- Bach, C.P.E. 1957. Facsimile of first edition, 1753 (Part One) and 1762 (Part Two). Ed. L. Hoffmann-Erbrecht. Leipzig: Breitkopf & Härtel.
- Battel, G. U., and R. Fimbiani. 1998. "How Communicate Expressive Intentions in Piano Performance." In A. Argentini and C. Mirolo, eds. *Proceedings of the XII Colloquium on Musical Informatics*. Udine: AIMI, 67–70.
- Bresin, R., and G. U. Battel. In press. "Articulation Strategies in Expressive Piano Performance." *Journal of New Music Research*.
- Bresin, R. and A. Friberg. 1998. "Emotional Expression in Music Performance: Synthesis and Decoding." *TMH-QPSR, Speech Music and Hearing Quarterly Progress and Status Report* 4:85–94.
- Bresin, R., and A. Friberg. 2000. "Rule-Based Emotional Colouring of Music Performance." *Proceedings of the 2000 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 364–367.
- Canazza, S. et al. 1997. "Sonological Analysis of Clarinet Expressivity." In M. Leman, ed. *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology*. Berlin: Springer Verlag, 431–440.
- Canazza, S. et al. 1998. "Adding Expressiveness to Automatic Musical Performance." In A. Argentini and C. Mirolo, eds. *Proceedings of the XII Colloquium on Musical Informatics*. Udine: AIMI, 71–74.

- Carlson, R., B. Granström, and L. Nord. 1992. "Experiments with Emotive Speech: Acted Utterances and Synthesized Replicas." *Proceedings ICSLP 92*. Banff, Alberta, Canada: University of Alberta. 1:671–674.
- Cope, D. 1992. "Computer Modeling of Musical Intelligence in Experiments in Musical Intelligence." *Computer Music Journal* 16(2):69–83.
- De Poli, G., A. Rodà, and A. Vidolin. 1998. "A Model of Dynamic Profile Variation, Depending on Expressive Intention, in Piano Performance of Classical Music." In A. Argentini and C. Mirolo, eds. *Proceedings of the XII Colloquium on Musical Informatics*. Udine: AIMI, 79–82.
- Friberg, A. 1991. "Generative Rules for Music Performance." *Computer Music Journal* 15(2):56–71.
- Friberg, A. 1995a. "A Quantitative Rule System for Musical Expression." Doctoral dissertation. Stockholm: Royal Institute of Technology.
- Friberg, A. 1995b. "Matching the Rule Parameters of Phrase Arch to Performances of 'Träumerei': A Preliminary Study." In A. Friberg and J. Sundberg, eds. *Proceedings of the KTH Symposium on Grammars for Music Performance*. Stockholm: Royal Institute of Technology, Speech Music and Hearing Department, 37–44.
- Friberg, A. et al. 1998. "Musical Punctuation on the Microlevel: Automatic Identification and Performance of Small Melodic Units." *Journal of New Music Research* 27(3):271–292.
- Friberg, A. V. et al. 2000. "Generating Musical Performances with Director Musices." *Computer Music Journal* 24(3):23–29.
- Friberg, A., and G. U. Battel, Forthcoming. "Structural Communication: Timing and Dynamics." In R. Parncutt and G. McPherson, eds. *Science and Psychology of Music Performance*. Oxford: Oxford University Press.
- Friberg, A., and J. Sundberg. 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(3):1469–1484.
- Gabrielsson, A. 1985. "Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience." *Music Perception* 3(1):59–86.
- Gabrielsson, A. 1994. "Intention and Emotional Expression in Music Performance." In A. Friberg et al., eds. *Proceedings of the Stockholm Music Acoustics Conference 1993*. Stockholm: Royal Swedish Academy of Music, 108–111.
- Gabrielsson, A. 1995. "Expressive Intention and Performance." In R. Steinberg, ed. *Music and the Mind Machine: The Psychophysiology and the Psychopathology of the Sense of Music*. Berlin: Springer Verlag, 35–47.
- Gabrielsson, A., and P. Juslin. 1996. "Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience." *Psychology of Music* 24:68–91.
- Gerardi, G. M., and L. Gerken. 1995. "The Development of Affective Responses to Modality and Melodic Contour." *Music Perception* 12(3):279–290.
- Ghias, A. et al. 1995. "Query by Humming: Musical Information Retrieval in an Audio Database." *Proceedings of ACM Multimedia '95*. New York: Association for Computing Machinery, 231–236.
- Granqvist, S. 1996. "Enhancements to the Visual Analogue Scale, VAS, for Listening Tests." Royal Institute of Technology, Speech Music and Hearing Department *Quarterly Progress and Status Report* 4:61–65.
- House, D. 1990. "On the Perception of Mood in Speech: Implications for the Hearing Impaired." Lund University, Dept. of Linguistics. *Working Papers* 36:99–108.
- Juslin, P.N. 1997a. "Emotional Communication in Music Performance: A Functionalist Perspective and Some Data." *Music Perception* 14(4):383–418.
- Juslin, P. N. 1997b. "Can Results from Studies of Perceived Expression in Musical Performances Be Generalized Across Response Formats?" *Psychomusicology* 16:77–101.
- Juslin, P.N. 1997c. "Perceived Emotional Expression in Synthesized Performances of a Short Melody: Capturing the Listener's Judgment Policy." *Musicae Scientiae* 1(2):225–256.
- Kastner, M. P., and R. G. Crowder. 1990. "Perception of the Major/Minor Distinction: IV. Emotional Connotation in Young Children." *Music Perception* 8(2):189–202.
- Kitahara, Y., and Y. Tohkura. 1992. "Prosodic Control to Express Emotions for Man-Machine Interaction." *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 75:155–163.
- Kotlyar, G.M., and V.P. Morozov. 1976. "Acoustical Correlates of the Emotional Content of Vocalized Speech." *Acoustical Physics* 22(3):208–211.
- Langeheinecke, E. J. et al. 1999. "Emotions in the Singing Voice: Acoustic Cues for Joy, Fear, Anger, and Sadness." *Journal of the Acoustical Society of America* 105(2), part 2, 1331.
- Meyer, L.B. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

-
- Mozziconacci, S. 1998. *Speech Variability and Emotion: Production and Perception*. Eindhoven: Technische Universiteit Eindhoven.
- Orio, N., and S. Canazza. 1998. "How Are Expressive Deviations Related to Musical Instruments? Analysis of Tenor Sax and Piano Performances of 'How High the Moon' Theme." In A. Argentini and C. Mirolo, eds. *Proceedings of the XII Colloquium on Musical Informatics*. Udine: AIMI, 75–78.
- 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(5):2546–2568.
- Repp, B. H. 1998. "A Microcosm of Musical Expression: I. Quantitative Analysis of Pianists' Timing in the Initial Measures of Chopin's Etude in E Major." *Journal of the Acoustical Society of America* 104:1085–1100.
- Sundberg, J. 2000. "Emotive Transforms." *Phonetica*, Vol. 57, No. 2–4: 95–112.
- Sundberg, J. 1993. "How Can Music Be Expressive?" *Speech Communication* 13:239–253.
- Sundberg, J., A. Friberg, and L. Frydén. 1989. "Rules for Automated Performance of Ensemble Music." *Contemporary Music Review* 3:89–109.
- Sundberg, J., A. Friberg, and L. Frydén. 1991. "Threshold and Preference Quantities of Rules for Music Performance." *Music Perception* 9(1):71–92.
- van Bezooijen, R. A. M. G. 1984. *The Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht, The Netherlands: Foris Publications.

Figure A1. (1) Percent IOI deviation, (2) off-time deviation in msec, and (3) sound level deviation in dB for each performance of Ekorn.



The following is a list of World Wide Web sites relevant to this article.

Sound and MIDI files of the seven versions of Ekorn and Mazurka:

www.speech.kth.se/~roberto/emotion

KTH performance rules description:

www.speech.kth.se/music/performance/

www.speech.kth.se/music/publications/thesisaf/sammfa2nd.htm

www.speech.kth.se/music/publications/thesisrb/

Director Musices software (Windows and Mac OS):

www.speech.kth.se/music/performance/download

Figure A2. (1) Percent IOI deviation, (2) off-time deviation in msec, and (3) sound level deviation in dB for each performance of the first 45 notes of Mazurka.

