Rule-Based Emotional Coloring of Music Performance
Roberto Bresin and Anders Friberg

Department of Speech, Music and Hearing (TMH) - Royal Institute of Technology (KTH)
Drottning Kristinas väg 31, 10044 Stockholm, Sweden
{roberto, andersf}@speech.kth.se
http://www.speech.kth.se/music

ABSTRACT
Previously unexplored potentials offered by the KTH rule system highlight the possibility to obtain musical and emotionally different performances of the same score by hyper-articulating the performance of the score itself, i.e., by marking the structure of the composition. This new application of the KTH performance rule system is presented here. It is shown how particular combinations of performance rules, called rule palettes, can produce music performances with specific emotional coloring. In particular the principles and choices behind the design of a rule palette for angry colored performances are illustrated.

INTRODUCTION
This paper presents part of a new computational model of expression in music performance—The GERM model (Juslin et al., 1999). Drawing on previous research on performance, the GERM model is based on the hypothesis that expression in performance derives from four primary sources of variability, namely (1) Generative rules that function to convey the musical structure in an appropriate manner (e.g., Clarke, 1988; Friberg, 1991; Palmer, 1989); (2) Emotional expression governed by the performer’s expressive intention (e.g., Juslin, 1997a); (3) Random fluctuations reflecting internal timekeeper variance and motor variance (e.g., Shaffer, 1982; Wing & Kristofferson, 1973); and (4) Movement principles dictating that particular aspects of the performance should be shaped in accordance with our perception of physical motion (Bresin & Battel, forthcoming; Friberg et al., 2000; Friberg & Sundberg, 1999; Shove & Repp, 1995).

In the following we focus on the Emotional expression source of variability in music performance and illustrate how it can be simulated by means of special combinations of generative rules.

RULE BASED EMOTIONAL EXPRESSION
An important contribution in music performance is given by the emotional component, often marked in the score by the composer with Italian words, such as Con fuoco, Tenero, Vivace, Brillante. Juslin and Gabrielsson (1996) asked musicians (playing piano, violin, flute, and guitar) to communicate different emotions in performances of the same piece. They found that each emotion was coupled to a particular set of acoustic cues. In subsequent listening tests, listeners, both musically trained and untrained, used the same set of cues for decoding these emotions. The importance of different performance cues in the identification of emotional qualities was tested in synthesis experiments (Juslin, 1997b). Each expressive cue was set by hand to different values in synthetic performances on a commercial sequencer. Listeners could recognize and identify the intended emotions also for these stimuli. Continuing in the same direction we used Director Musices (DM), a software for automatic performance of music that has been developed in our group (Bresin and Friberg, 2000), to automatically control each cue in emotionally expressive performances. The main idea was that particular combinations of the DM performance rules, henceforth rule palettes, could give rise to performances that differ with respect to emotional expression. Six different rule palettes, each corresponding to an emotion (fear, anger, happiness, sadness, tenderness, and solemnity) were designed using the qualitative cue descriptions by Gabrielsson and Juslin. These rule palettes produce variations of the performance variables IOI (Inter-Onset Interval), OOI (Offset-Onset Interval) and L (Sound Level) for each note in the score. In a listening test with 16 listeners two pieces were performed without rules and with the six rule palettes for a total of 14 performances. The emotions associated with the DM rule palettes were correctly classified in most cases (Bresin and Friberg, forthcoming).

A more detailed description can be found in Bresin and Friberg (forthcoming). In this paper we want to illustrate the principles and choices that guided us in simulating a specific emotion, anger.

SIMULATING ANGER
It has been observed that the mean tempo and the mean sound pressure level are two of the most important expressive cues for communicating emotion in music performance are mean tempo and sound pressure level (SPL). These cues exhibit similar qualitative deviations in performance of instrumental music as in speech and singing, see Figure 1. For these reasons two new rules was introduced in the DM system. The Tone duration rule, for shortening or lengthening all notes in the score by a constant percentage, and the Sound level rule, for decreasing or increasing the SPL of all notes in the score by a constant value in dB. Gabrielsson and Juslin observed that players trying to communicate anger are using very rapid tempo, loud SPL, non-legato articulation, moderate time deviations, structural reorganization, and
increased contrast between long and short notes (Table 1). This qualitative description was translated to the generative rules using an analysis-by-synthesis method. The result is in the form of a rule palette in DM for each emotion defining the rule selection and its parameters.

For achieving very rapid tempo and loud SPL the Tone duration and Sound level rules presented above were used. Non-legato articulation was obtained with the Duration contrast articulation rule set to a detached articulation. Articulation seems to convey high degree of information in expressive playing (Bresin & Battel, forthcoming). Results from some informal listening tests conducted by the authors show that dead-pan performances with staccato articulation are classified as happy and brilliant, and dead-pan performances with legato articulation are classified as sad and dark. Moderate time deviations were obtained by using moderate values of the rule quantity parameters.

The structural reorganization described by Gabrielsson and Juslin has been realized with an unconventional use of the Phrase arch rule. In a natural performance phrasing is characterized by an accelerando and crescendo followed by a rallentando and decrescendo. Here, the rule is applied in the opposite way yielding a rallentando-decrescendo in the beginning of a phrase and a accelerando-crescendo in the end of the phrase.

The Duration contrast rule, applied with a positive rule quantity, was used for increasing the contrast between long and short notes. It has been shown that Duration Contrast rule plays an important role in communicating different emotions in music performance ranging from very high duration contrast for Afraid-performances and no contrast at all for Sad- and Tender-performances.

The First note in measure rule stresses the first note in each measure by playing it louder. Metrical patterns alter mainly the motional quality (Friberg & Battel, forthcoming) and in this case creates a dominantly downward motion similar to a forceful type of walking (c.f. Friberg et al., 2000).

The rule palette was applied to different scores. As an example, time and SPL deviations for each note in a Swedish nursery tune (Ekorrn satt i granen, henceforth Ekorrn, "The squirrel sat on the fir-tree", composed by Alice Tegnér) are displayed in Figure 2. The rule palette (without the First note in measure rule) was validated in a listening test. In total, 91% of the 16 subjects classified the performances as communicating anger (Bresin & Friberg, forthcoming).

**APPLICATIONS**

One immediate application is the introduction of new features in commercial music sequencers and editors by programming new plug-ins. In this way users will be able to obtain emotionally colored performances of any MIDI file and can continue to use their favorite software tool.

A second application is in performance analysis. By applying the performance rules in a reverse mode it is possible to extract some information about the emotional content of a performance. This opens new perspective in the field of music performance didactics, where the KTH rule system has already been successfully applied (Friberg and Battel, forthcoming).

A third appealing but maybe underestimated field of application is that of consumer electronics involving automatic performance of synthetic sounds like in mobile phones. Rule-based techniques can be used to improve the quality of ringing tone performances in mobile phones.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Expressive Cue</th>
<th>Gabrielsson and Juslin</th>
<th>Director Musices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tempo</td>
<td>Very rapid</td>
<td>Tone duration is shortened by 15%</td>
<td></td>
</tr>
<tr>
<td>SPL</td>
<td>Loud</td>
<td>Sound level is increased by 8 dB</td>
<td></td>
</tr>
<tr>
<td>Articulation</td>
<td>Mostly non-legato</td>
<td>Duration contrast articulation rule (k = 1)</td>
<td></td>
</tr>
</tbody>
</table>

- **Anger**
  - **Time deviations**
    - Moderate
    - Structural reorganizations
    - Punctuation rule (k = 2)
    - Phrase rule (k = -0.7, turn. position = 0.5, next = 1.3, amp = 4)
    - Sub-phrase rule (k = -0.7, turn. position = 0.3, amp = 4, last = 1)
    - Duration contrast rule (k = 2, amp = 0)
  - **SPL deviations**
    - Increased contrast between long and short notes
    - First note in measure (amp = 1)

Table 1. Playing a music score with Anger. Each performance rule involved has a particular setting in order to comply with the observations made by Gabrielsson and Juslin on each expressive cue.
This is probably the most popular music synthesizer, at least with the highest diffusion, and we are listening to its mechanical and dull performances every day. In a current project we are using our technique to design an emotional mobile phone. The caller will be able to send emoticons to the receiver in order to express different emotional situations, influencing the performance of ringing tones on the receiver’s phone. In this way it will be possible to have a musical interaction between the caller and the receiver. Furthermore some mobile phone producers are included the possibility of using the MIDI standard as format for ringing tone files.

A further application could be in the new Mpeg-7 standard. Mpeg-7 includes smart Karaoke applications, where the user will sing the melody of a song to retrieve it from a database. An emotional toolbox, capable of both recognizing the emotion of the singer and of translating it back in the Karaoke performance would probably improve the human-computer interaction. (Ghias et al., 1995).

CONCLUSIONS

The main result in this investigation is that emotional coloring of performances can be obtained by using rules for the communication of the musical structure. This was obtained by defining rule palettes in DM. It also opens the doors to the design of new rule palettes for other expressive situations usually expressed in Italian terms in classical music performance. The same rule palette can be applied to different compositions since the rules were designed to handle different musical scores. This was also validated in the listening test for different emotions. On the other end, the rules themselves are triggered by the musical structure. This suggests that the structure of the composition plays a primary role in expressive automatic playing.

As a final consideration, we think that a performance system, based on rules or other techniques, should be included in the next MPEG-7 standard in order to enhance the performance possibility in interactive systems involving music.

LINKS

KTH performance rules description:
http://www.speech.kth.se/music/performance/
Sound and MIDI examples of emotional performances:
http://www.speech.kth.se/~roberto/emotion/

REFERENCES


