ABSTRACT

The way a performer phrases music is important for the way the music is perceived by listeners. Phrasing influences the structural interpretation of the music by listeners as well as his or her emotional engagement with the music; at least, those are the hypotheses that this study likes to confirm. An experiment was run in which twelve listeners heard three performances of an Etude of Skriabin. They pressed a button to indicate the phrase-boundaries of the music and moved a slider to indicate their emotional engagement with it. Two main expressive parameters for a pianist were measured, tempo and key-velocity, and related to the responses of the listeners. It was found that tempo correlated with the indication of phrase boundaries, while key-velocity was especially related to the listeners’ emotional engagement. Further examination showed that both aspects can actually be related to phrasing, though tempo to local phrasing and dynamics to global phrasing. The pianist used dynamics to express the music’s overall form and to modulate the tension of the music throughout the entire piece. The listeners responded emotionally to this.

1. INTRODUCTION

Three findings from previous research are especially relevant for this study. The first is that expressive variations in, for example, tempo and dynamics are often related to the musician’s structural interpretation of the music [1], [2]. The second is that these variations are also often related to the expression of a certain emotion by the performer [3], [4]. And the third is that these aspects of a performance influence the structural or emotional interpretation of the music by listeners depending on which interpretation is asked for [2], [3], [4].

This study aims to continue and reconcile these findings and hypothesizes that it is the musician’s phrasing that influences simultaneously the listener’s structural and emotional interpretation. The way a musician phrases the music influences the attention of the listener and so modulates to what extent the listener is emotionally engaged. It is especially when a listener gets the idea of a larger phrasing, a transgression of the local fixation, that he is emotionally moved. As shown in previous literature [5], [6], the pianist especially uses two expressive variables, tempo and dynamics, to express the phrase-structure of the music. Accordingly, he also uses these to move the listener.

2. METHOD

2.1 Musical Performances

An explorative experiment was run to investigate the hypotheses. A professional pianist was asked to perform an emotionally engaging piece of his choice at a concert that was organised for the experiment’s purpose. He performed the piece first without public in a normal manner (to be referred to as p1) and an exaggerated manner (to be referred to as p2) and then performed the piece with public in a normal, concert manner (to be referred to as p3). He performed on a Yamaha Disklavier, which made it possible to register MIDI information of the performance. In addition, audio recordings were made and presented to the participants of the experiment.

The pianist chose to perform Etude Op. 8 no. 11 by Alexander Skriabin, which is a slow and lyrical piece (Andante cantabile) in a late Romantic style that has a considerable amount of modulations. According to the pianist, the piece can be played with a lot of freedom. Theoretically, the piece has a simple A B A with coda structure (A A’ B A’’ A’’’ C to be more precise), but the pianist interpreted the line of the music differently: The first main target of the music is a release of tension halfway the B section. Everything preceding this target point is a preparation for this tension release. The A section is anyway preparatory; it leads towards the start of the B section, which is the real beginning of the piece. After this release of tension, the music builds up towards the dramatic return of the theme of the A section. This prepares for the second possible point of tension release halfway the coda at a general pause. The release is however not continued and the piece ends most sad.

2.2 Participants

Twelve people participated in the experiment among them were four musicians. The participants varied greatly in musical experience. Some of them never had had music lessons and hardly listened to classical music, while others basically performed classical music already their entire live.
2.3 Procedure

The participants sat behind a desk with a slider and a joystick before them. They heard the three performances of the Skrjabin Etude twice in random order. The first time they heard the music, they indicated the phrase boundaries in the music by pressing the button of the joystick. The second time they heard the music, they indicated to what extent they were emotionally engaged with it by moving a MIDI-slider up and down. The whole procedure was explained to them by a written instruction and a practice trial.

3. RESULTS

3.1 Performance Data

The key-velocity and onset-times of notes were extracted from the MIDI files. From this, the average key-velocity for each quarter note, which roughly corresponds to the dynamics of the performance, was calculated as well as inter-onset-intervals (IOI’s) between successive quarter notes, which is a measure for local duration. The quarter note was taken as unit, because it gives both sufficiently detailed information about the performances and sufficient consistency between listener response data for which synchronisation is an issue (see below).

The resulting profiles of quarter note key-velocity and quarter note IOI are plotted in Figure 1, top panels. Separate graphs are plotted for p1, p2 and p3. Vertical dotted lines indicate section boundaries. Bar numbers are given at the bottom. The profiles were highly similar for the three performances: they all started in a slow tempo and with soft dynamics, had considerable crescendi and accelerandi in the A section, a diminuendo and crescendo in the B section accompanied by first a highly variable tempo and thereafter an accelerando, a fast and loud return of the A section with limited variation in tempo and dynamics, a soft and slower repeat of the theme, and a coda that fades away in dynamics and tempo (see Figure 1 top panels).

In addition to this global pattern, the IOI-profile shows the characteristic peaks of phrase-final lengthenings. It shows this at a fairly high density and large magnitude. There is no large-scale hierarchy in the phrase-final lengthenings that has larger lengthenings at major boundaries and short lengthenings at minor boundaries. Instead rubato is quite steep throughout the piece, except in the forte return of the A section (A’’'). The key-velocity profile shows drops in velocity at most phrase-boundaries, though these are balanced by strong crescendi in most sections.

This is the basic pattern of all three performances. Differences between them are that performance 2 is clearly an exaggerated version of the other performances: all variations in tempo and dynamics are larger. Performance 1 is the more modest version of the three, while performance 3 is in between. Performance 1 and 3 further contain a crescendo and decresendo in the coda that performance 2 lacks.

3.2 Listeners’ Data

The indication of phrase-boundaries was measured at a sampling rate of 10 Hz. The measure was 0 when the participant did not press the button and 1 if the he or she pressed the button to indicate a phrase boundary. The data was filtered to be 1 only at the onset-time of a phrase-boundary indication. For the rest of the time, the measure was put to 0.

For each quarter note in the performance, the number of people who indicated a phrase-boundary was calculated by summing the number of boundary indications per quarter note over participants. The resulting “segmentation measure” was expressed as a multiple of chance-level. Chance-level was defined as the number of boundary indications per quarter note if the total of boundary indications would have been equally distributed over all quarter notes of the piece. From this recalculation, the quarter-note level turned out to be the most statistically reliable unit for this measure (better than the 8th-note or half-bar level for example). It should be noted that some kind of data-reduction was necessary, since the participants would never react exactly at the same time and could have been indicating the same boundary, even if onset-times of the key-press differed 1 or 2 seconds.

The indication of emotional engagement was also measured at a sampling rate of 10 Hz using a MIDI-slider that had a range from 0 to 127. The average level of the MIDI-slider (“emotion measure”) per quarter note was calculated for each participant separately and averaged over participants.

The first result from these measures is that participants disagreed quite strongly, though the averages per performance are statistically reliable (see bottom panels of Figure 1). On average the indication of section boundaries was 1.6 times chance level, which is significantly above chance, but not very much. The indication of emotional engagement was diverse between subjects to such extent that no significant differences can be found between the three performances: the standard deviation of the difference between the emotion measure of two performances is larger than the average difference itself. The average profile within pieces is statistically reliable: the standard deviation per quarter note is relatively small with respect to the average per quarter note (on average 1:2.7, 1:2.2 and 1:2.4 for performance 1, 2 and 3 respectively).

The second and third results are the profiles themselves: The segmentation measure shows most boundary indications at the start and in the B section. It shows a considerable drop in the number of boundary indications in the return of the theme. Highest peaks are between repeats of the theme in the first and second A section and at two places that the pianist mentioned explicitly: a general pause halfway the coda and a point of release of tension halfway the B section (see downwards pointing arrows in Figure 1).
The emotion measure increases towards the B section. It decreases and increases within the B section, reaches a maximum at the return of the theme, and decreases at the repeat of the theme and in the coda. For performances 1 and 3 it shows a local revival in the coda.

3.3 Relation Performance and Listeners’ Data
The quarter note IOI and key-velocity measures were correlated with the emotion and segmentation measures per quarter note. This was done directly and with a time-delay of one, two and three quarter notes of the performance data with respect to the listeners’ data. The best correlations were obtained between the performance data and the segmentation data if the performance data was not delayed or delayed for only one quarter note. For the emotion data, however, the best correlations were achieved if the performance data was delayed by three quarter notes. The optimal correlations are shown in Table 1.

Generally, the emotion measure was highly correlated with key-velocity and much less negatively correlated with duration. The segmentation measure was on the other hand more highly correlated with duration than (negatively) with key-velocity. A direct correlation between the segmentation measure and the emotion measure at the quarter note level was not significant. The number of boundary indications per section did however highly correlate negatively with the average indication of emotional engagement per section.

Table 1: Correlations at the quarter note level (column 2 and 3) and at the level of sections (last column).

<table>
<thead>
<tr>
<th>Performance 1</th>
<th>IOI</th>
<th>Velocity</th>
<th>Segm. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>-.50</td>
<td>-.34</td>
<td></td>
</tr>
<tr>
<td>Segm. M</td>
<td>.50</td>
<td>-.34</td>
<td></td>
</tr>
<tr>
<td>Emotion M</td>
<td>-.40</td>
<td>.74</td>
<td>-.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance 2</th>
<th>IOI</th>
<th>Velocity</th>
<th>Segm. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>-.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segm. M</td>
<td>.55</td>
<td>-.34</td>
<td></td>
</tr>
<tr>
<td>Emotion M</td>
<td>-.49</td>
<td>.77</td>
<td>-.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance 3</th>
<th>IOI</th>
<th>Velocity</th>
<th>Segm. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>-.49</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>Segm. M</td>
<td>.47</td>
<td>-.34</td>
<td></td>
</tr>
<tr>
<td>Emotion M</td>
<td>-.43</td>
<td>.75</td>
<td>-.80</td>
</tr>
</tbody>
</table>

It should not be taken as a surprise that the correlations with the segmentation measure are relatively lower than with the emotion measure, since the segmentation measure still contains the between participant variability, while the emotion measure is a pure average. In addition, if we use a multiple regression analyses that takes key-velocity and IOI as independent variables and segmentation measure or emotion measure as independent variables, the contributions of velocity is reduced to non-significant for the segmentation measure, while the contribution of IOI also cancels out for the emotion measure. Duration becomes the main predictor of segmentation and key-velocity of emotion.

3.4 Relation Phrasing and Emotion
The correlations presented in Table 1 and the multiple regression analyses suggest that, for these performances, tempo was especially a cue for phrase-boundaries, while dynamics was especially a strong cue for the intensity of emotion. This may in turn suggest that phrasing did not play an important role for the emotional engagement of listeners. Still there seems a relation between phrasing and emotion, since the correlation between the number of phrase boundary indications per section and the average emotional engagement per section was very high. How should we interpret these results?

From the descriptions of the performance data, it became clear that phrase-final lengthenings occurred at a high density, which means that they indicated rather local phrase-boundaries, and that there was limited hierarchy within these lengthenings. It became also clear that the dynamics and the tempo profiles showed an overall pattern of increase and decrease that was especially apparent in the key-velocity profile, but masked by the large phrase-final lengthenings in the profile of the quarter note IOI. The interpretation is that the pianist used both dynamics and tempo to phrase the music, though tempo was especially used to locally phrase music, while dynamics was more saliently used to express the larger form of the music. The listeners responded emotionally to this expression of the larger form.

And the overall form is as described by the pianist: the first part is an introduction and builds up to the B section, which he considered as the real beginning of the piece. This beginning is again a preparation for the first target of the piece: the release of tension at the middle of the B section. Hereafter tension builds up towards the dramatic return of the theme, which leads via a repeat of the theme in contrasting dynamics to the second important target of the theme: the second possible release of tension at the general pause. After the general pause, the release is not given and all hope is lost. The piece ends most sad. The pianist most skilfully expresses this interpretation in the patterning of dynamics (see arrows in the key-velocity panel of Figure 1). The resulting phrasing is over the entire piece with subdivisions at measures 22 and 36. The return of the theme is the culminating point of the piece where after tension can release. According to the pianist, this tension cannot however be fully resolved.

CIM-3
4. SUMMARY

This study has investigated the expressive functioning of two acoustic cues, tempo and dynamics, in three performances of an étude of Skriabin and has shown that both tempo and dynamics were used to express the phrasing of the music, though tempo expressed local phrasing and dynamics the larger form. The local phrasing was especially reflected in the segmentation indications of the listeners. The global phrasing was reflected in the ratings of emotional engagement of the listeners. This followed the pattern outlined by the dynamics and reached a maximum at the turning point of the piece and local minima at section divisions.

5. DISCUSSION

This paper has investigated the relation between experience of emotional intensity and expressive features of a performance. It did not consider other factors that could have contributed to the experience such as melodic movement, harmony or rhythm. In a follow up study such a comparison is planned. For now it suffices to see the extent to which these two expressive relate to emotional experience. We see promising evidence in this study that the way of phrasing influences the structural interpretation of the music by listeners, but moreover also the emotional experience of it. A relation between tempo and phrasing and dynamics and emotion has also been shown in previous studies and this study provides a confirmation of these findings [5], [6], [7], [8]. However, this study has also given insight into the simultaneity of these relations and it has demonstrated how skilfully the pianist modulated these. This skillfulness is probably crucial for the emotional experience. An aspect that should be investigated further in future research. Just a crescendo or diminuendo without proper preparation would not have given the strongly emotional responses that this study got. Although the participants varied strongly in responses, they did mention to be highly moved and were made sure to give responses according to their feelings.

6. REFERENCES


Figure 1: Graphs of performance and listeners’ data