PERFORMANCE CUES FOR LISTENERS’ EMOTIONAL ENGAGEMENT

Renee Timmers (1, 2), Antonio Camurri (2), Gualtiero Volpe (2)

(1) Austrian Institute for Artificial Intelligence
(2) Infomus, DIST, University of Genova

ABSTRACT

A variety of factors may influence listeners’ emotional engagement with a musical performance. It may depend on the sight of the pianist, the performers’ interpretation of the music, or the musical experience of the listener. In an exploratory study, we investigated what aspects of a performance are most strongly related to listeners’ emotional engagement and considered cues present in sound as well as in the performer’s movement. Twelve participants were presented audio and video recordings of three performances of an Etude of Skriabin. Their task was to indicate their emotional engagement with the music by moving a slider up and down. The results can to a large extent be described as listeners’ synchronization with performance aspects: higher energy levels in sound and movement of the performance corresponded to higher levels of emotional engagement, while lower levels had lower engagement. Considerable overlap existed between the cues of the movement and the audio suggesting an unambiguous interpretation of the music by the performer. This interpretation was however most clearly revealed by the dynamics and tempo of the performances, where tempo communicated local phrase structure and dynamics the overall pattern of tension and release.

1. INTRODUCTION

With the perfection of audio playback systems and with the rise of computer-generated music, one could expect the role of the musical performer and the need for live concert attendance to decay. Fortunately this is not the case and people still wholeheartedly enjoy concert visits. One of the reasons for this might be, besides social and cultural reasons, the communication that goes on between performer and listener and the excitement of witnessing the moment-to-moment realization of the performance.

Communication from performer to listener has been demonstrated in several studies. Performers communicate a personal interpretation of the music, which concerns not only the structure of the music [1] [2] [3], but also the expressive or emotional content [4] [5] [6] [7]. When asked to choose between different possible interpretations, listeners are rather accurate at choosing the one expressed by the pianist.

What has been less investigated and what is probably harder to investigate is the communication of emotional engagement from performer to public, but also from public to performer. The performer may show emotional engagement in his movements, facial expressions and in his sound production, and also may explicitly express a pattern of tension – release or variations in emotional intensity that guide and influence listener’s emotional engagement. The responses of listeners may influence and inspire the performance.

We do not expect to be able to investigate this process quite accurately, but we do think it is possible to investigate at least one aspect of it: the communication of emotional engagement from performer to listener. To investigate this, we adopted a conceptual framework previously defined for the analyses of expressive communication in dance that consists of four layers [8]. (1) a performance data preprocessing layer, (2) a performance data feature extraction layer, (3) a performance data segmentation, mapping and interpretation layer, and (4) a relation to listener’s responses layer. In other words, we used a bottom-up approach that assumes that measurable features from video and audio recordings of a performance that is measurable aspects of the movement and the sound of a performance are able to account for the observed communication. Layers 1 and 2 are technical processing layers, but layers 3 and 4 are aimed to have psychological relevance as well. Layers 3 and 4 could be compared with the distinction that Juslin [7] makes between the expression of emotion by performers and the understanding of emotion by listeners: the performer may or may not be clear and distinct in his emotional expression and listeners may or may not be accurate and sensitive to this expression.

2. METHOD

2.1. Musical performance

A professional pianist performed Etude Op. 8 no. 11 by Alexander Skriabin at a concert that was organised for the experiment’s purpose. He first performed the piece 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 it 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.

Etude Op. 8 no. 11 is a slow and lyrical piece (Andante cantabile) in a late Romantic style. 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, 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 (left downward pointing arrow in Figure 1). 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 (right downward pointing arrow in Figure 1). The release is however not continued and the piece ends most sad.

2.2. Participants and procedure

24 people participated in the experiment among them were eight musicians. The participants were divided in two groups. One group saw the performances at a computer screen and heard the performances over speakers (referred to as the video condition). The other group only heard the performances over speakers (referred to as the audio condition). The participants sat behind a desk with a slider and a joystick before them. They saw and/or heard the performances two times. 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, participants indicated to what extent they were emotionally engaged with it by moving a MIDI-slider up and down while listening to the music. The whole procedure was explained in a written instruction and a practice trial.

3. RESULTS

3.1. Layers 1 and 2: preprocessing and feature extraction of performance data

Auditory data. The note-onset times and the key-velocity per note were taken from the MIDI recordings from the Yamaha Disklavier. The note-offset information was not used, because the pianist used too much pedal for articulation measures to be reliable.

Two features – local duration and local dynamics – were extracted from this information. Local duration was measured as the inter-onset-interval (IOI) between successive quarter note onsets of a middle voice that had a rhythm of repeated 8th notes. Local dynamics was measured as the average key-velocity of all note onsets within a time-span of a quarter note.

Visual data. For the analysis of the movement of the pianist, we concentrated on the movement of the head, which closely relates to the movement of the back of the pianist. Techniques for motion detection and tracking were applied to the recorded images for segmenting the silhouette of the head, individuating its bounding box, and tracking its movement.

Three features were extracted. First of all, the position of the head was measured in two directions: x-direction showing left-right position and y-direction showing backward-forward position. We used the Lucas and Kanade feature-tracking algorithm [9] that assigns and tracks a specified number (in our case 40) of randomly assigned moving points within a region.

Figure 1: Quarter note IOI and key-velocity per quarter note for p1, p2 and p3. Dotted vertical lines indicate section boundaries. Arrows are explained in the text.

Figure 2: Left-right movement and Backward-forward movement of the head of the pianist. High values correspond to a right and forward position, while low values correspond to left and backward position. Separate plots for the three performances show especially consistent left-right movement.

Thirdly, velocity was calculated for each trajectory using the symmetric backward technique for the numeric derivative. Average values of position and velocity among the forty trajectories were calculated for both the x and y component. In addition, the velocity values were integrated for the x and y direction to get a general measure of movement velocity.
Redundancy in the number of points (i.e., forty points instead, for example, of just the barycentre of the blob) allowed us to get more robust and reliable values for velocity. A low-pass filter was applied to smooth the obtained data. Measures were summarized per quarter note in order to be comparable to the other measures.

3.2. Layer 3: Mapping and first interpretation of features

Correlations between performance measures were calculated to check the coherence between them. Key-velocity and IOI are negatively correlated ($r = -0.51$ on average). Velocity of head movement is positively correlated with key-velocity ($r = 0.45$ on average) and negatively with IOI ($r = -0.25$ on average). The low correlation between values is partly due to the asynchrony between the periodicity of the measures. If peak-values (maximum for key- and movement velocity and minimum for IOI) per two bars are correlated, agreement between movement and sound measures becomes higher. Especially the two velocity measures turn out to be highly correlated ($r = 0.79$ on average for key and movement velocity, versus $r = -0.38$ on average for movement velocity and IOI).

Backward-forward position is negatively correlated with key-velocity ($r = -0.39$ on average) and positively with IOI ($r = 0.26$ on average), which means that the performer moves forward in soft and slow passages and backwards in louder and faster and therefore more intense passages. Left-right position is also negatively correlated with key-velocity and positively with IOI, but correlations are fairly low ($r = -0.16$ and $r = 0.15$ on average, respectively). Notice however that left-right movement was highly consistent over the three performances, which suggests that these movements were systematic and intended. The negative correlation with key-velocity and positive correlation with IOI suggests a tendency for a right position to coincide relatively often with longer durations and softer notes as at the end of phrases and a left position to coincide relatively often with shorter durations and louder notes as in the middle of phrases.

Notably, the contour of the key-velocity pattern closely relates to the pianist’s interpretation of the musical piece. In the interview, he described the structure of the piece as starting with a moderate introduction that 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. Key-velocity reflects this interpretation by showing higher values for tense passages and diminuendi at points of tension release (see arrows at the bottom of Figure 1). Only the soft ending does not correlate with the description of tension-release given by the pianist: this soft ending is not a release, but is loaded with emotion. In addition, this global pattern in key-velocity could be seen as communicating the overall form of the piece that subdivides the entire piece at measures 22 and 36. The return of the theme is the culminating point of the piece where after tension can release.

All performance measures show a periodic increase and decrease. To check the relation between these periodicities and the musical structure, the location of minimum in key-velocity, and maxima in IOI, left-right and backward-forward position were compared to the location of phrase-boundaries. Generally, the Skriabin etude has a local structure of two-bar phrases. The forward and the right position of the performer were taken as start/end point for periodicity. IOI was most systematically related to the two-bar phrasing of the Skriabin piece, followed by key-velocity. 55% of the phrase-boundaries were joined by a slowing down in tempo. The other phrase-boundaries were directly followed by a slowing down in tempo (a delay of 1 quarter note). For the key-velocity, 42% of the phrase-boundaries coincided with a minimum in key-velocity, 15% was anticipated by a minimum and 28% followed by a minimum. The period-boundaries of the movement of the pianist hardly synchronized with the score-phrasing, instead the location of these boundaries varied greatly with respect to the two-bar score-phrasing.

These results indicate a differentiation in function between the performance measures, where local duration especially communicates local phrasing and key-velocity and movement velocity communicate the pattern of tension-release of the music. Minima in key-velocity reinforce the segmentation indicated by local duration now and then. And local duration joins key-velocity to some extent in communicating the musical tension.

3.3. Layer 4: Conceptual layer and relation to responses of listeners

The participants of the experiment indicated the end of phrases by pressing a button and the degree of emotional engagement by moving a MIDI slider. To get a measure of subjective segmentation of the piece (SM), the number of people indicating a phrase boundary for each quarter note was calculated and expressed as a multiple of chance level. The average indicated emotional engagement per quarter note was taken as measure of emotional engagement (EM).

To check the relation between extracted performance features and listeners’ responses, a series of multiple regression analyses were done. In the first analysis, quarter note IOI, key-velocity, and movement velocity were used to predict SM. In the second analysis, the same variables were used to predict EM. In the third analysis, peak values per hyper-bar were used to predict EM per hyper-bar. All analyses were 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 $R^2$ obtained are reported in Tables 1 and 2. These were obtained with a delay of either zero or one for SM, and either two or three for EM.

SM was significantly predicted by the model. From this model, IOI was the only significant variable. In other words, duration was a predictor of the variation in number of
participants indicating a section-boundary. More participants indicated a phrase-boundary for longer durations.

EM was well predicted by the quarter note model, but even better by the second model that took the peak-values per hyper-bar to predict EM per hyper-bar. Key-velocity was in the majority of cases the only significant variable, but was sometimes joined by movement velocity. IOI reached significance only for p2, for which variations in location duration are relatively large. Notably, all explained variances were higher for the non-musicians than for the musicians. The non-musicians seem especially to have paid attention to the performance variables, while the musicians either paid less attention to them or provided a more abstract response.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Audio</th>
<th>Audio</th>
<th>Video</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mus</td>
<td>Non-mus</td>
<td>Mus</td>
<td>Non-mus</td>
</tr>
<tr>
<td>SM</td>
<td>.18</td>
<td>.23</td>
<td>.24</td>
<td>.26</td>
</tr>
<tr>
<td>EM</td>
<td>.32</td>
<td>.57</td>
<td>.29</td>
<td>.56</td>
</tr>
<tr>
<td>Quarter</td>
<td>.48</td>
<td>.71</td>
<td>.40</td>
<td>.71</td>
</tr>
</tbody>
</table>

Table 1: Average $R^2$ for each condition obtained for the regression analyses explaining the variance in SM and EM on the basis of quarter note or peak per two bars IOI, key-velocity, and, for the video conditions, movement velocity.

As a comparison, the analyses were repeated with x-position and y-position as independent movement variables instead of the more general movement velocity variable. The results did not improve or change from this alteration, instead x and y-position did not contribute significantly to any of the regressions.

These results confirm a differentiation between expressive means: tempo primarily communicates musical tension, which is interpreted especially by the non-musicians as variations in emotional intensity. Velocity of the movement is correlated with dynamics and may therefore also reflect musical tension, but the sounding parameters seem to be the main communicative factors.

4. SUMMARY AND CONCLUSION

In an explorative study, the communication of emotional engagement from performer to listener was investigated. A analysis framework was used where extracted cues from recordings of performances were analyzed for their communicational function.

Features were extracted from the video and MIDI data of three performances of a Skriabin etude that characterized the position and movement velocity of the head of the pianist and the tempo and dynamics of the performance. Correlation between the features and their relation to the score-phrasing suggested a differentiation in function between the features. Movement measures were most strongly related to the dynamics of the performance and less strongly to the tempo of the performance and especially peak-values correlated highly. The periodicity of rise and fall in tempo was systematically related to the score phrasing. This was less so for the dynamics of the performance and hardly the case for the movement of the head of the pianist. The close relation of dynamics to the description of the pianist of tension and release finally suggested that the dynamics of the performance communicated the pianist’s interpretation of musical tension, while tempo communicated local phrasing.

The differentiation in function between tempo on the one hand and dynamics and movement velocity on the other hand was substantiated by the responses of the listeners. Tempo was the only significant factor that could explain the subjective segmentation of the participants, while dynamics was the main factor that explained emotional engagement. These results can to a large extent be described as listeners' synchronization with performance aspects: higher energy levels in sound and movement of the performance corresponded to higher levels of emotional engagement, while lower levels had lower engagement. The non-musicians showed these tendencies most strongly, while the musicians often deviated from this pattern. An aspect that seems worthwhile to be investigated further in future research.

5. REFERENCES