Listeners' emotional engagement with performances of a Scriabin étude: an explorative case study

Psychology of Music

Psychology of Music
Copyright © 2006
Society for Education, Music
and Psychology Research
vol 34(4): 481–510 [0305-7356
(200610) 34:4; 481–510]
10.1177/0305735606067165
http://pom.sagepub.com

RENEE TIMMERS
UNIVERSITY OF NIJMEGEN, THE NETHERLANDS
MATIJA MAROLT
UNIVERSITY OF LJUBLJANA, SLOVENIA
ANTONIO CAMURRI
UNIVERSITY OF GENOVA, ITALY
GUALTIERO VOLPE
UNIVERSITY OF GENOVA, ITALY

ABSTRACT In an explorative study, the variation in listeners' judgments of the emotionality of three performances of a Scriabin étude was investigated. Three performances of the Scriabin étude were recorded. Video and/or audio recordings were presented to 24 listeners who segmented the music in short phrases and indicated their emotional engagement with the music using a slider. The relation between the performance data and the listeners' responses was analysed as well as the effects of musical training, medium and musical structure. The analyses were done using multiple regression analyses and decision trees. The results confirmed the hypothesized influence of the performer's interpretation, the listener's background and the global phrase structure, though not always in the expected way. In particular the dynamics of the performance correlated with listeners' judgments of emotionality, while tempo correlated more strongly with the indications of phrase structure. This was more clearly the case for non-musicians than for musicians. The movements of the pianist were related to the dynamics of the performance and seemed to aid the communication of emotional intensity.

KEYWORDS: *decision trees, motion, musical training, piano performance, serial correlation*

Introduction

The emotional response to music seems unpredictable and therefore hard to study and yet there have been an increasing number of studies that do so. The explorative study of Sloboda (1991) is exemplary by its investigation of the relation between the emotional responses and aspects of the musical structure: the responses were clearly emotional and responses were remembered to follow certain musical passages that were analysed for common structural characteristics. A similar exploration of the causes of remembered emotional responses to music was done by Scherer et al. (2001–02), who found that aspects of musical structure were more important for emotional responses when listening to classical music than when listening to non-classical music.

Recently, several studies have investigated the perception of emotion in music more experimentally. The responses to the music did not need to be truly emotional, but the music should be consciously perceived as communicating a certain emotion such as happiness, sadness or anger. Parameters of the musical structure were shown to be responsible for the communication such as its harmony, rhythm, or melodic profile (for an overview see Gabrielsson and Lindström, 2001), and also features of the musical performance were shown to influence the perception of emotions, such as tempo, dynamics and articulation (for an overview see Juslin, 2001). The distinction between the investigation of 'perceived' emotion and 'felt' emotion is important to note. To perceive music as being expressive of an emotion does not necessarily mean that a listener experiences this emotion. And different factors may play a role in experiencing an emotion than in attributing an emotion to music. Although most researchers assume that there is a positive relation between perceived emotion and experienced arousal or emotion, other relationships have been found as well (Gabrielsson, 2001–02).

Besides the type of emotion associated with musical fragments, emotional intensity of music has been investigated using continuous measurement of self-reported emotional responses. The use of continuous measurement provides the possibility of zooming in and investigating the relation between music and emotional response locally. For example, Krumhansl (1996) related continuous ratings of musical tension to music theoretical predictions of tension and release and to phrase structure. Sloboda and Lehmann (2001) investigated the intensity of perceived emotion with performances and found that the points of high emotion often coincided with points that the performers had explicitly mentioned in an interview. Schubert (2004) investigated the relationship between acoustic characteristics and perceived emotional valence and arousal and found an especially strong relationship between arousal and loudness.

But, as Scherer and Zentner (2001) pointed out, emotional responses (and the perception of emotion) are not only a function of the musical structure. Context factors are important such as the musical background of the listener, the familiarity of the listener with the music, and the presence of a performer with whom a listener can sympathize. Seeing the performer may help to understand the emotional expression. Indeed, visual cues are sometimes found to be stronger than auditory cues for the communication of emotions (Ohgushi and Hattori, 1996; Thompson and Russo, 2004).

From this complexity of interacting factors, we chose to study the relationship between a subset of them and the varying emotional engagement of listeners throughout the piece. Most experimental studies focus on 'perceived' emotion. Instead, we asked listeners to report 'felt' emotional engagement with the music. The interest is to see how much some of the findings for 'perceived' emotion generalize to 'felt' emotion. Although we cannot be sure that listeners actually reported 'felt' emotion in an experimental set up such as the one we used, we expected them to be more subjective and introspective than they would be if they were instructed to report the emotional intensity of the music. However, in interpreting the results, it should be born in mind that a self-reported measure of felt emotion is not entirely reliable, especially when giving responses continuously, and listeners could easily have adopted a strategy of reporting 'perceived' emotion rather than 'felt' emotion. In this study, we speak about reported emotional engagement, which is in line with the instruction to the participants, but we must be aware of the ambiguous meaning of this measure.

The music used in the study is the étude Op. 8, No. 11 by Alexander Scriabin. The music of Scriabin is little used in empirical studies of this kind and his music is not well known. Familiarity with this specific piece is unlikely to be a factor. The factors that we did take into account are: (1) the musical training of the participants; (2) the appearance and movements of the performer; (3) the expressiveness of the performance and the microstructure of the performance (e.g. variations in tempo and dynamics); and (4) the musical structure, especially phrase structure, and the performer's and listeners' interpretation of these.

In the study, MIDI, video and audio recordings were made of three performances of the Scriabin étude by a professional pianist that varied in degree of expressiveness. The decision to limit the study to one pianist was made to facilitate an in-depth study. Twelve participants heard the performances over loudspeakers and 12 participants saw and heard the performances on a computer screen and over speakers. With the first hearing, they indicated the phrase boundaries within the music by pressing a button and, in a repeated hearing, they indicated their emotional engagement with the music by moving a slider up and down. Among the participants were 8 musicians and 16 non-musicians. In the analyses, the tempo and dynamics of the performances were examined as well as the velocity of the movements of the pianist. The responses of the listeners were analysed per subject group, and their relation to the performance data was examined using multiple regression analyses (MRAs).

One characteristic of continuous responses is that the responses show autocorrelation, or serial-correlation. The presence of serial correlation influences regression analyses. This issue is addressed by focusing on the absolute values as well as the derivative of the responses (see Schubert, 2001, 2001–02 for a discussion). As an alternative solution, the use of decision trees is explored. In these analyses, the emotional engagement responses are categorized as flat or changing.

Our expectation is that emotional responses depend to a large extent on the communication between performer and listener. The performer is a mediator between listener and composer and influences the tension and excitement of a moment. In addition, the performer expresses emotional engagement in sound and movements to which the listeners respond. We expect the sensitivity to the performance to vary with musical training and musical background. The emotional engagement of listeners varies throughout the piece and is expected to reach a peak at important culmination points in the music. If a performer engages the listener's attention, he will guide the listener towards these moments. As a working hypothesis, we expect that variations in the intensity of the performance will correlate with the pianist's interpretation of emotional intensity and tension in the music and that this will correlate with listeners' judgments of emotional engagement.

Method

MUSICAL PERFORMANCE

A professional pianist performed étude Op. 8, No. 11 by Alexander Scriabin at a concert that was organized for the purpose of the experiment. He first performed the piece without an audience 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 an audience in a normal, concert manner (to be referred to as p3). The choice of the Scriabin piece was made in consultation with the pianist. Scriabin is one of the pianist's areas of expertise and the étude is part of his concert repertoire. We considered the music to be emotionally moving and stylistically interesting. Few Scriabin performances have been the subject of empirical studies and being relatively unknown facilitates comparison between the responses of musicians and non-musicians. The étude is short (a little over 4 minutes), but not too short for emotional engagement, and it leaves a lot of room for interpretation by the performer. The pianist used a specific interpretation for the instruction to perform with exaggerated expression: he aimed to perform in the style that was fashionable among Scriabin's contemporaries who used a lot of rubato and great freedom in performance.

The étude is a slow piece (Andante cantabile) in a minor mode (B flat minor) – see the score in the Appendix. Theoretically, it has a simple A B A with coda structure (A A' B A" A"' C to be more precise). The A sections have a two-bar phrase structure that group into four- (A"'), six- (A') or eight-bar phrases (A and A") and have the main key of B flat minor. The B section also consists of two bar phrases, though the boundary between groups is phase shifted (it does not align with the bar lines). Each two-bar phrase is in a different major key.

However, the pattern of tension-release of the music as explained by the pianist in an interview does not strictly follow the ABA structure. The first main target of the music is a release of tension halfway through the B section (bar 23). Everything preceding this target point is a preparation for this tension-release. The A section is in any case preparatory; it leads towards the start of the B section, which is the real beginning or main part of the piece according to the pianist. 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 through the coda at a general pause (bar 46). The release is, however, not continued and the end of the piece is very sad, again according to the pianist. The major key at the end of the piece is not a release or a happy ending.

The pianist performed on a Yamaha Disklavier, which made it possible to register MIDI information of the performance. In addition, audio and video recordings were made and presented to the participants of the experiment. Selections of the recorded material can be found on the internet at www.infomus.org

PARTICIPANTS

Twenty-four people participated in the experiment. Among them were eight musicians, who studied music or were music graduates. The non-musicians were humanity and engineering students.

PROCEDURE

The musicians and non-musicians were equally divided over 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 twice. The first time they heard the music, they indicated the phrase boundaries in the music by pressing the button of the joystick. This task was added to study the relation between emotional engagement and subjective segmentation or phrasing. The second time they heard the music the participants indicated to what extent they were emotionally engaged with it by moving a MIDI-slider up and down. A high position of the slider meant high emotional engagement, while a low position of the slider meant low emotional engagement. The order of the repeated performances was randomized over participants. The whole procedure was explained in a written instruction in Italian and a practice trial that used different musical material. To limit the duration of the experiment, the participants only heard each performance twice. And they did not hear any of the performances more often than other performances, because the practice trials were done using different materials. In other words, the task of the participants was carried out in a spontaneous and unpractised manner, which might sometimes have led to confusion or inconsistencies in their responses, especially as the music of Scriabin is not easy. This point may need improvement in a follow-up study.

Results

PERFORMANCE DATA

In the analyses of the performer's data, we assume that the main information conveyed by the pianist is expressed in variations in tempo, dynamics, and body movement. Of course, other sources may be important as well, such as timbre, articulation, and facial expressions, but we focus on these three aspects as they are among the important expressive parameters for a pianist. In addition, we focus on the analyses of increases and decreases in tempo, dynamics and movement velocity, and on the location of points of relative release. In doing so, we use the simple hypothesis that variations in the intensity of the performance correlate with the pianist's interpretation of emotional intensity and tension in the music and that this correlates with listeners' judgments of emotional engagement.

Auditory data

To get a measure of loudness and tempo, the key-velocity and onset times of notes were extracted from the MIDI files. From this, the average key-velocity over all note-onsets of the melody and accompaniment was calculated per quarter note as well as inter-onset-intervals (IOIs) between successive quarter notes of the top voice of the accompaniment. The quarter note was taken as unit, because it gave both sufficiently detailed information about the performances and sufficient consistency between listener response data for which synchronization was an issue (see the 'Subjective Segmentation' section later).

The resulting profiles of quarter note IOI and quarter note key-velocity are plotted in Figure 1. Separate graphs are plotted for p1, p2 and p3. Vertical dotted lines indicate section boundaries and 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 forte 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.

In addition to this global pattern, the IOI profile shows the characteristic peaks of final phrase lengthenings. It shows quite large peaks at least every two bars. Rubato is steep throughout the piece with a tempo change of 25 percent (small peaks in first A section) to above 80 percent (peaks in B section) from the mean. The forte return of the A section (A") has the smallest variations in tempo. The key-velocity profile shows drops in velocity at most phrase boundaries, but these are counter-balanced by strong crescendi in most sections.

This is the basic pattern of all three performances. Differences between them are that p2 is an enlarged version of the other performances: most variations in tempo and dynamics are larger.

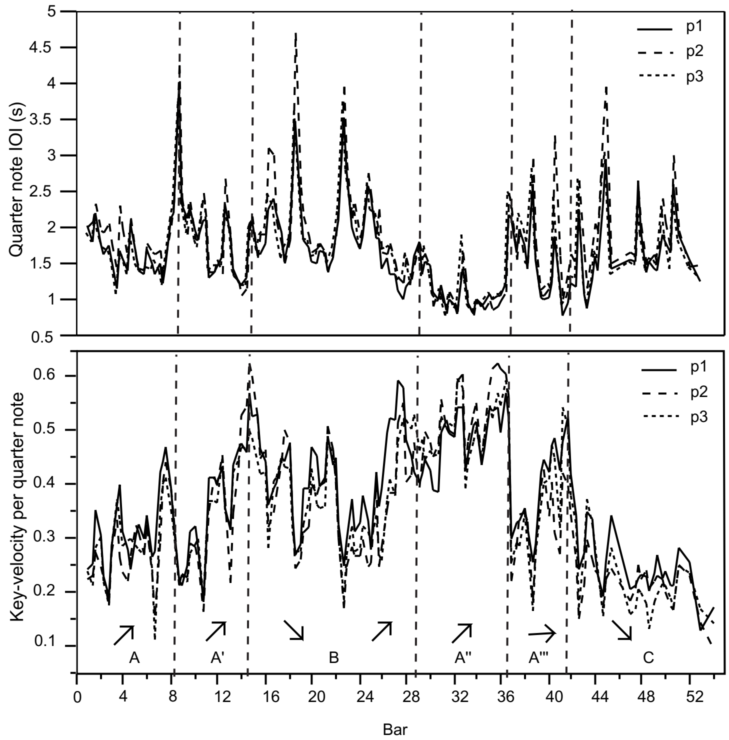


FIGURE 1 *IOI and key-velocity per quarter note for the three performances of the Scriabin étude.*

Visual data

For the analysis of the movement of the pianist, we concentrated on the movement of the head, which shows both backward–forward movement (y-direction) and left–right movement (x-direction). The head was chosen for practical reasons, because it was a relatively easy point to track from the video recordings. Although it is a limitation to focus only on the head movements, the head is also a sensible point to track, since it also reflects the movements of the upper body.

Video recordings were made from the top, the right and the left side of the pianist. The recordings from the right side of the pianist were shown to the participants in the experiment, but the recordings from the top were more appropriate for movement extraction. From these recordings from the top, the position of the head was measured using the Lucas and Kanade feature-tracking algorithm (Lucas and Kanade, 1981) that assigns and tracks a specified number (in our case 40) of randomly assigned moving points within a region. Velocity and acceleration has been calculated for each trajectory using the symmetric backward technique for the numeric derivative. Average values of position and velocity among the 40 trajectories were calculated for

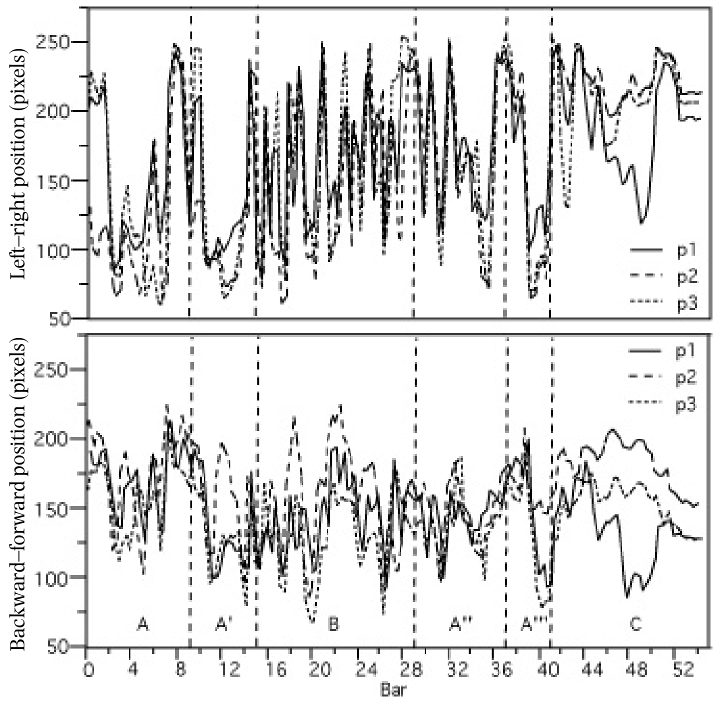


FIGURE 2 *Position of the head in left–right and backward–forward direction. The average position is plotted for each quarter note in the performance.*

both the x and y component. In addition, the velocity values were integrated for the x and y movement to get a general measure of amount of movement over time. Redundancy in the number of points (i.e. 40 points instead of, for example, just the centre of the head circle) 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.

The position of the head is plotted in Figure 2 for two dimensions: left–right (upper panel) and backward–forward (bottom panel) seen from the perspective of the pianist. So left is the left-hand side of the pianist and right the right-hand side. Backward is away from the piano, while forward is towards the piano. The movement of the head was especially pronounced and especially consistent over performances in the left–right direction (correlation between p1 and p2 and between p2 and p3 was 0.79; it was 0.83 between p1 and p3). The backward–forward movement becomes more pronounced for the later performances (p2 and p3). The periodic movement is relatively fast in the middle parts of the piece (B and A") and slower in the outer parts. This suggests intensification towards the middle followed by a relaxation towards the end.

INTERPRETATION OF PERFORMANCE DATA

The variations in tempo, dynamics and movement-velocity are related to various degrees. Key-velocity and IOI have the highest absolute correlation (r = –0.51 on average), followed by movement velocity and key-velocity (r = 0.45 on average). Variations in movement velocity and IOI showed the lowest correlation (r = –0.25 on average). These correlations are modest 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, the agreement between movement and sound measures becomes higher. Especially, the two velocity measures are highly correlated (r = 0.79 on average for key- and movement-velocity, versus r = –0.38 on average for movement velocity and IOI).

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 piece: the second possible release of tension at the general pause. Key-velocity reflects this interpretation (see the 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 velocity could be seen as communicating the overall form of the piece that subdivides the entire piece at bars 23 and 37. The return of the theme is the culminating point of the piece after which tension can release.

All performance measures show a periodic increase and decrease. The relationship between these periodicities and the musical structure was analyzed by comparing the location of minima in key-velocity, and maxima in IOI, x-position and y-position to the location of phrase boundaries. Generally, the Scriabin étude has a local structure of two-bar phrases. The forward and the left position of the performer's head were taken as start/end point for periodicity. IOI was most systematically related to the two-bar phrasing of the Scriabin piece, followed by key-velocity. Fifty-five percent 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 one quarter note). For key-velocity, 42 percent of the phrase boundaries coincided with a minimum in key-velocity, 15 percent was anticipated by a minimum and 28 percent 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 phrase structure. The strong relation between periodicities in tempo and phrase boundaries confirms previous research (Todd, 1985; Palmer, 1989). A more flexible periodicity of dynamics is in contrast with Todd's (1992) model, but was also observed in other studies (Clarke and Windsor, 2000).

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

LISTENERS' DATA

The participants listened to each of the three performances twice. The first time they indicated phrase boundaries by pressing a button and the second time they judged their emotional engagement with the music.

Subjective segmentation

The indication of phrase boundaries was measured at a sampling rate of 10 Hz. The measure was 0 when participants did not press the button and 1 if they 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. This was done separately for the different conditions: musicians and non-musicians hearing only or hearing and seeing p1, p2 or p3.

This sum per quarter note was expressed as a multiple of chance level. Chance level was defined as the number of boundary indications per quarter note if the total number of boundary indications had 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: it had a relatively high average value of multiple of chance level (higher than the 8th-note or half-bar level, for example). Note that some kind of time reduction was necessary, since the participants would never respond exactly at the same time and could have been indicating the same boundary, even if onset times of the key-press differed by 1 or 2 seconds. The resulting segmentation measure (SM) shows for each quarter note the agreement between participants expressed as a multiple of chance level.

To examine the effect of performance, medium and training on SM per quarter note, the correlations between SM profiles were compared for different groups. Table 1 shows a summary of these pair-wise correlations. It shows the average correlations between conditions that differ in performance (column and row 1), medium (column and row 2), and/or training (column and row 3). As can be seen, the correlations between conditions with

TABLE 1 *Average correlation between SM and profiles of conditions that differ in performance (column and row 1), medium (column and row 2) and/or training (column and row 3)*

	Performance	Medium	Training
Performance	0.59		
Medium	0.42	0.41	
Training	0.47	0.43	0.46

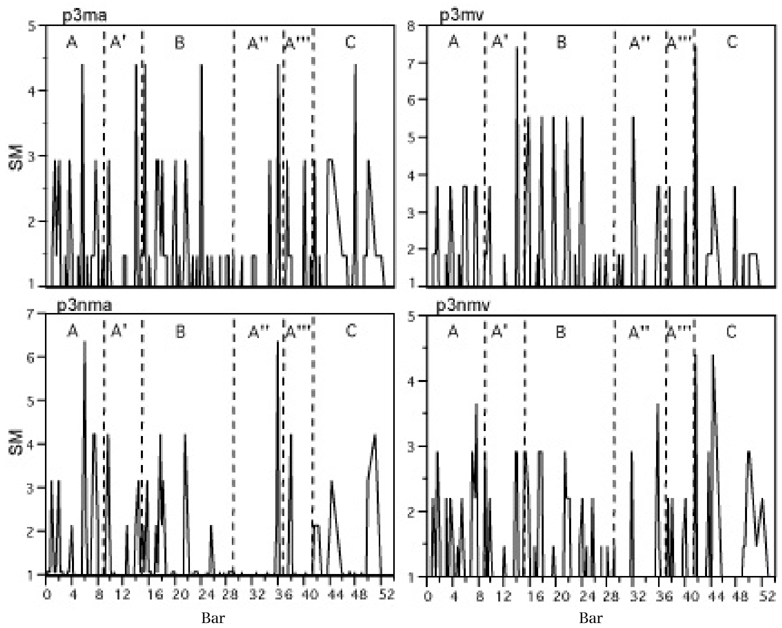


FIGURE 3 *SM per quarter note for musicians (top) and non-musicians (bottom) of the audio (left) and video (right) condition of p3.*

different performances are relatively high, while those between conditions with different mediums are relatively low. In other words, the effect of the performance seems to have been small compared to the effect of medium and training.

Figure 3 illustrates what the effect of musical structure, training and medium has been. It shows SM per quarter note for the musicians and non-musicians of the video and audio condition of p3 with a threshold of chance level. This chance level should be interpreted with care for the musicians, since the number of musicians per condition was small (only four). Some of the peaks in SM coincide with section boundaries, such as between A and A', between A' and B, and between A''' and C. Peaks occur at the points of relaxation indicated by the pianist halfway through the B section (bar 23) and halfway through the coda (bar 46). The video condition has a higher SM at the boundary between A''' and C than the audio condition, but a lower SM at the transition to A'''. The musicians marked the transition from A'' to B more clearly than the non-musicians and they more consistently marked the two-phrase structure of the piece in sections A, B, and A'''.

Some of this fluctuation in SM is related to the phrasing of the pianist as suggested by his use of tempo variation. The correlation between SM and IOI is on average 0.43, which is higher than the (absolute) correlation between SM and key-velocity, which is on average –0.27, and between SM and movement velocity, which is on average –0.08.

Emotional engagement

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 (EM)) per quarter note of each performance was calculated for each participant separately. The grid of 127 points is a small grid that was used to make the measure continuous. It does not mean that the difference of one point is meaningful. The interpretation of this measure is not entirely straightforward. On the one hand, there is the reference of the minimum and the maximum and absolute levels of emotional engagement in between. On the other hand, the scale might have been used primarily to indicate increases and decreases in emotional engagement with the difficulty that the physical size of the scale is limited. If we look at the grand average of EM over all participants and conditions in the left panel of Figure 4, we see that EM changes almost continuously throughout the piece without strong plateaus. On the other hand, the profile does clearly show regions of high EM and lower EM. If we look at the derivative of this grand average as shown in the right panel, we see that the most pronounced characteristics are the fast decreases in EM, while the maxima or highpoints in EM are not distinguishable.

This possible difference in characteristics becomes important when we consider the problem of serial correlation. Continuous measures such as EM consist of data points that correlate serially. The height of one point is partially dependent on the height of previous points. When such a continuous measure is correlated with another measure, correlations will always be artificially high (Schubert, 2001–02). One solution is to use the derivative of the measure, in other words to focus on the amount of increase and decrease, instead of the absolute levels of the measure (Schubert, 2001, 2001–02). However, as mentioned before, it might be that the participants controlled the absolute level of EM rather than the relative level of EM and the points of change could have been less meaningful than the points of stability. If so, this would make explanations of the derivative misleadingly hard. Therefore to get a balanced point of view, we will report the results of MRAs using both the absolute and the relative levels of EM.

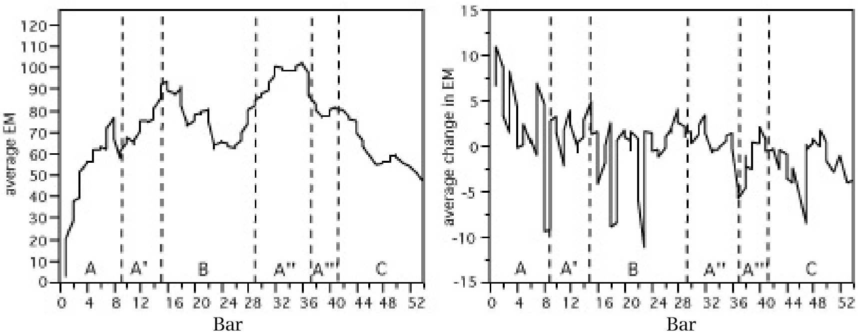


FIGURE 4 *Grand average of EM per quarter note (left) and grand average of change in EM per quarter note (right).*

Effect of training, medium and performance

To see if training, medium and performance had a systematic effect on the profile of EM, we compared the correlations between EM profiles of individual participants of the same group with the correlations between EM profiles of participants of different groups. EM profiles should correlate more with EM profiles from the same condition than with EM profiles from a different condition.

Table 2 shows the average correlation between profiles of conditions that differ in performance (row and column 1), medium (row and column 2), and/or training (row and column 3). The numbers between brackets show the average correlations between the derivative of EM profiles. These numbers should be compared with the correlations between EM profiles from single conditions, which were on average 0.37. This is not any higher than the correlations between profiles of different conditions, except for the correlations between profiles of musicians and non-musicians that are on average slightly lower ($r = 0.32$). This means that only training seems to have had an influence. Overall, however, the correlations between profiles are low.

Because of these large individual differences, we will treat participants separately in the following analyses of the relationship between EM and performer's data.

TABLE 2 *Average correlation between EM profiles, and in brackets between the derivative of EM profiles, of conditions that differ in performance (column and row 1), medium (column and row 2) and/or training (column and row 3)*

	Performance	Medium	Training
Performance	0.39 (0.17)		
Medium	0.37 (0.13)	0.37 (0.13)	
Training	0.32 (0.11)	0.34 (0.12)	0.32 (0.12)

RELATION BETWEEN EM AND PERFORMER'S DATA

Performance cues and emotional engagement

Tempo, dynamics and movement velocity contain cues that may influence listeners' emotional engagement. As said before, we use the simple hypothesis that variations in the intensity of the performance correlate with the pianist's interpretation of emotional intensity and tension of the music and that this correlates with listeners' judgments of emotional engagement. High intensity in tempo, dynamics and movement-velocity that is fast tempo, loud dynamics, and fast movements were expected to correspond with high emotional engagement, while low intensities were expected to correspond with points of emotional relaxation. We therefore expected negative correlations between EM and IOI and positive correlations of EM with key-velocity and movement-velocity.

MRAs were done with quarter note IOI, key-velocity, and movement-velocity as predictors of EM. The analysis was done for each participant and each condition separately. Because there might be a time delay between the performer's cue and the response of the participants, all analyses were repeated with a time delay of one, two and three quarter notes of the performance data with respect to the listeners' data. The results of the analyses with the highest explained variance for a participant will be reported irrespective of the time delay. These were generally obtained with a delay of either two or three quarter notes, which approximates 2–3 seconds, which is similar to lag times found by, for example, Krumhansl (1996) and Schubert (2004).

The analyses were done for the entire piece and for each section separately; see Table 3 for a summary. Section A''' is however too short to show sufficient variation in EM for the MRA to be reliable. It is therefore combined with the coda. Note that the explained variances are rather low. They are lower than found in related studies (Schubert, 2001–02). The main reason for this difference is that Table 3 reports the average explained variance for analyses fitted onto data of individual participants instead of averaged data (compare with the results reported in Table 4). Another possible reason is that participants in the current study reported 'felt' instead of 'perceived' emotions and the responses were therefore more subjective and less predictable. Nevertheless, the low explained variance does point to a limited applicability of the model to the data.

An analysis of variance (ANOVA) tested the effects of, and interactions between, performance, training and medium on the explained variance gained for the separate sections of individual participants. The main effects of training and medium were significant. The explained variances were on average higher for the video condition than for the audio condition (F (1, 351) = 6.0, p =.01) and higher for the non-musicians than for the musicians (F (1, 351) = 13, p < .001). There was no effect of performance and there were no significant interactions.

TABLE 3 *Average R^2 per condition of the regression analyses explaining the variance in EM per section of individual participants on the basis of quarter-note IOI, key-velocity and, for the video conditions, movement-velocity*

Section	Audio		Video	
	Musical	Non-musical	Musical	Non-musical
Whole	0.22	0.30	0.19	0.35
A	0.42	0.43	0.46	0.56
A'	0.46	0.53	0.49	0.52
B	0.25	0.36	0.22	0.36
A"	0.27	0.34	0.39	0.49
AC	0.29	0.38	0.37	0.41

For both audio and video conditions, key-velocity was the main predictor of emotional engagement in these analyses. For the whole piece, key-velocity was significant in 83 percent and IOI in 28 percent of the audio conditions. For the whole piece in the video condition, key-velocity was significant in 77 percent of the cases, IOI in 43 percent of the cases, and movement velocity in 37 percent of the cases. The directions of the effects were generally as expected: emotional engagement increased with velocity and decreased with duration. The individual sections showed a similar pattern, though overall the parameters were less often significant.

In Table 4, the results of four different models, or rather analyses, are compared. The first is the same model as reported in Table 3, but fitted onto the average EM per condition. The second is a fit of peak values of key- and movement-velocity and minimum values of quarter note IOI per two bars onto the average EM per two bars per condition. To take the peak values of velocity and IOI per two bars is a way of smoothing the performer's data and bringing it closer to the contour of EM. The third fits the quarter note key-velocity, IOI and movement-velocity onto the average change in EM per quarter note per condition. And the fourth fits the change per quarter note in key-velocity, IOI, and movement-velocity onto the average change in EM per quarter note per condition. The explained variances are highest if peak-values per two bars are used. They are very low for the fits on EM change.

The most important parameter for the fits of the first, second and fourth analyses is key-velocity. Key-velocity is significant for most of the analyses followed at a distance by IOI and movement-velocity. Only for the third analyses, where EM change is explained on the basis of absolute levels of performance variations is IOI the most important parameter. It is the only parameter that reaches significance. This suggests that change in EM and the absolute level of EM are determined by different processes.

TABLE 4 *R^2 per condition of the regression analyses that explain the average EM (rows 1 and 2) and the change in EM (rows 3 and 4) per participant group on the basis of quarter note (rows 1 and 3), peak level per two bars (row 2) or change per quarter note (row 4) of IOI, key-velocity and, for the video conditions, movement-velocity*

Model	Audio		Video	
	Musical Average	Non-musical Average	Musical Average	Non-musical Average
EM	0.32	0.57	0.29	0.56
EM (peak)	0.48	0.71	0.40	0.67
EM change	0.06	0.14	0.03	0.21
EM change (change)	0.07	0.09	0.04	0.20

RELATION BETWEEN PHRASING AND EMOTIONAL ENGAGEMENT

Krumhansl (1996; Krumhansl and Schenck, 1997) found a relation between tension, emotion and phrase-structure. Her observation was that emotion and tension tend to increase towards the end of the phrase and decrease with the presentation of new material, which was generally at the start of the new phrase. This is in line with the theory of Meyer (1956) that predicts an emotional response to a delay of resolution of expectations: musical tendencies within a phrase increase in complexity and expectation towards the middle and are resolved at the end of a phrase or section. Similarly, we expected to find a modulation of emotional engagement that would be confined by the phrase structure of the music. In other words, we expected specifically that larger changes in EM would coincide with phrase boundaries.

We ran several analyses to test this hypothesis. The first was an ANOVA that tested for each participant separately whether changes in EM were significantly larger around phrase boundaries than in the middle of phrase boundaries. The two-bar score phrasing was taken as reference. Around phrase boundaries meant directly preceding, following or at a two-bar phrase ending, while the quarter notes in the middle of the phrase did not directly adjoin a phrase boundary. This effect of phrase position was significant for only a minority of the participants. A second ANOVA tested whether larger absolute changes in EM (larger than average) occurred significantly more often at/around phrase boundaries than in the middle of phrases. This was significantly so for all participants in all conditions ($p < 0.05$). A third analysis was a direct correlation between SM per condition and EM per condition and between SM per condition and the absolute change per quarter note in EM per condition. The previous correlations were not high, but the latter correlations were significant for all conditions of the non-musicians and for half of the conditions of the musicians: the absolute change in

EM tended to increase with SM. In other words, stronger boundaries corresponded with larger changes in EM and no boundaries or weaker boundaries corresponded with smaller changes in EM.

This relates to previous observations that larger changes tend to occur around phrase boundaries and, specifically, section boundaries. Together with the relatively high explained variance for the peak per two bars model, this suggests that modulations of EM tend to follow the global contours of phrasing.

MODELLING EMOTIONAL RESPONSES WITH DECISION TREES

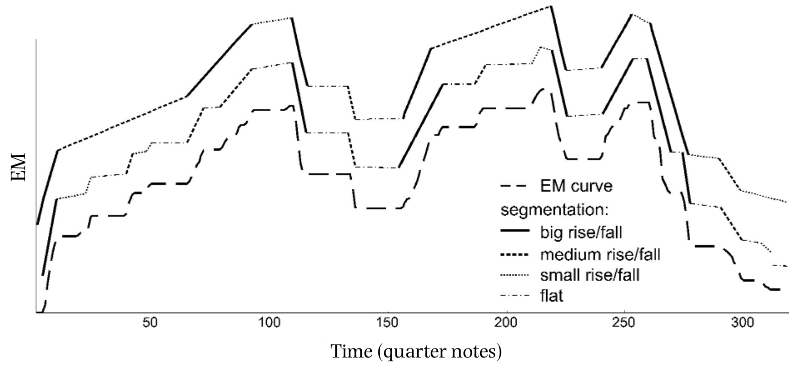
The above analyses suggest different processes for larger changes in EM, which occur around major phrase boundaries, and the absolute level of EM, which seems to especially correlate with loudness. Therefore to substitute EM by change in EM to filter out serial correlation within EM seems undesirable. To explore an alternative to MRAs for the investigation of the relation between stimuli and response, we close this article with a re-analysis of the data using decision trees. As further improvements to the analyses, the melodic contour of the music was added as a factor, the loudness measure was based on the recorded audio material, and the duration measure was transformed into a tempo measure (see Friberg and Sundberg, 1999 for a discussion of the choice between tempo or duration).

Decision tree learning is a common method used in data mining (see Mitchell, 1997; for an application of decision trees, see Camacho, 1998). A decision tree is a model that is both predictive and descriptive. It is called a decision tree because the model obtained is presented in a tree structure, where leaves represent classifications (predicted values) and branches represent conjunctions of features that lead to those classifications (descriptive values). The visual presentation makes the decision tree very easy to understand; as a result, decision trees have become a very popular data mining technique. They are most commonly used for classification (predicting what group a case belongs to), but can also be used for regression (predicting a specific value). Decision trees can handle both continuous and categorical variables, and can automatically infer interactions between variables, as well as identifying important variables.

We used classification decision trees to find relations between auditory cues and measures of emotional engagement. The question addressed was: what causes the changes in emotional engagement? So, the task of decision trees was to extract rules that would explain the changes in EMs on the basis of changes in auditory cues.

Construction of training sets for machine learning

Decision trees are built by supervised learning from a set of training examples. Our training data consisted of features extracted from three auditory cues: loudness, tempo and melodic contour. Loudness was extracted



Note: The three curves show the same EM curve segmented into smaller (bottom) to larger (top) linear segments; depending on the slope of the linear segment, the segment is categorized as flat, or small, medium or big rise/fall.

FIGURE 5 *Segmentation and labelling of EM curve.*

from audio recordings, the other two from MIDI data. These three cues were used to predict changes in listeners' emotion measures.

As a first step, the continuous EM data were split into segments that were categorized as either flat EM, small, medium or big rise in EM, or small, medium or big fall in EM. Categorization was performed by splitting an EM curve into a number of linear segments and then labelling each segment according to the categorization scheme. The linearization was performed with a least squares algorithm, fitting a number of linear segments to an EM curve, minimizing the error. The number of segments was chosen adaptively; depending on the nature of the curve, we chose a coarser or finer segmentation scheme. This procedure is illustrated in Figure 5. It shows an EM curve and a finer and a coarser segmentation of it. Segmentations are shown above the actual curve for clarity. In the analyses, we preferred the finer scheme to include local variations.

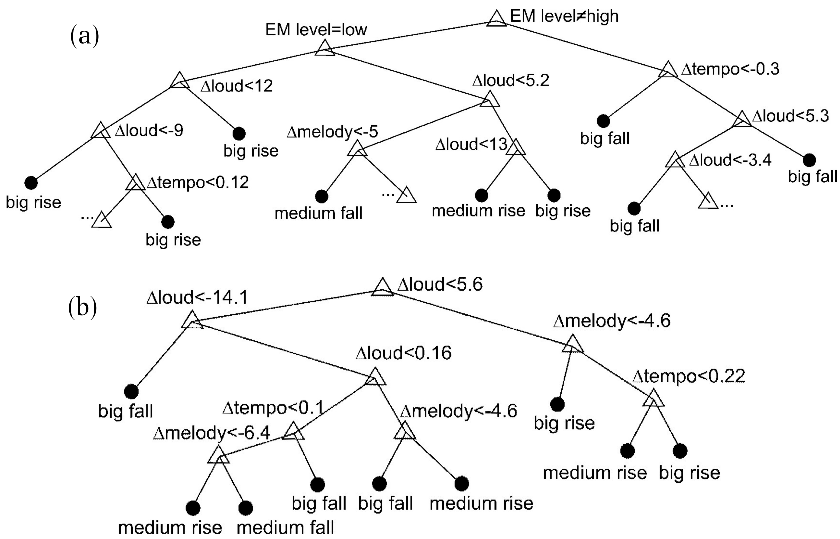
The training data consisted of the categorized EM curves and corresponding levels of performance cues. Performance variables included were loudness, tempo, and melodic contour. For each EM segment, the amount of change, the steepness and the variance of the performance variable was calculated and included as separate performance cue. Additionally, the current level of emotional engagement, labelled as high, medium or low, was added as a context cue.

The gathered data were used as training data to construct binary classification decision trees that predicted changes in emotional engagement by classifying each input vector (consisting of performance and context cues) into one of seven categories: no change in EM and big, medium, or small rise or fall. To prevent overfitting, the optimal tree and pruning level were determined by 10-fold cross-validation. The training was done and trees were constructed for each participant separately (but including the data of three performances) and for each group of participants taking the data of different participants from one group together.

Results decision trees

Due to inconsistencies of responses, satisfying trees were produced for only about half of the participants, while, for others, trees degenerated into very complex or trivial structures, which we ignored during our analysis of results. For an illustration of the results derived by the decision trees, we focus on the trees obtained for the non-musicians of the audio condition (Figure 6(a)) and the non-musicians of the video condition (Figure 6(b)).

Each node in a tree has two branches and splits the tree according to the rule linked to the node. For the left branch the rule is valid, and for the right branch the inverse of the rule is true. The leaves of the trees are linked to changes in emotional engagement. As can be seen in Figure 6, all rules found for the non-musicians relied on changes in auditory cues, either loudness (Δloud), tempo (Δtempo) or melody (Δmelody). In addition, the current level of emotional engagement (EM level) was part of the tree for the non-musicians of the audio condition.



Note: Each node in a tree (open triangles) has two branches and splits the tree according to the rule linked to the node. For the left branch the rule is valid, and for the right branch the inverse of the rule is true. The leaves of the trees are linked to changes in emotional engagement (filled circles).

FIGURE 6 *Decision tree for EM responses of non-musicians of the audio condition (a) and of the video condition (b).*

The percentage of correctly predicted changes in EM for the tree based on the audio condition of the non-musicians was 73 percent. The correct distinction between rise or fall in EM was 89 percent. For the video condition of the non-musicians this percentage correct was less good – only 52 percent. The distinction between rise and fall in EM was, however, almost as accurate, namely in 78 percent of the cases.

Our main findings are summarized as follows:

- *Loudness as the dominant cue*: most trees confirmed the dominant role of loudness and its overall positive relation to emotional engagement. Loudness was found to be the most common top rule, discriminating between branches that classify rises and falls of EMs. This can be observed in Figures 6(a) and 6(b). In the latter, loudness is the top predictor cue. Rising loudness (increase over 5.6 dB) leads to rises in EM, large falls in loudness (over 14.1 dB) lead to big falls in EM. A similar situation can be observed in the middle branch of the tree in Figure 6(a) (EM level is medium), where loudness separates between rises and falls.
- *Saturation of emotional engagement*: in the responses of a number of listeners we observed a phenomenon related to saturation of emotional engagement, which changes the overall nature of a cue–response relationship. When the level of emotional engagement of a listener was very high or low, any large change in a cue could trigger a significant change in emotional engagement. This is shown in the tree depicted in Figure 6(a). Though the overall relation between loudness and EM for audio non-musicians is positive (increase in loudness leads to increase in EM), this relationship changes when the level of EM is low or high. In both cases, large increases or drops in loudness contribute to changes in EM (rises when level is low or falls when level is high). This is reflected in the left- and right-most branches of the tree in Figure 6(a). Saturation was observed in the responses of several listeners, irrespective of the condition and the overall nature of their cue–response relationships. We mainly observed saturation as being related to loudness and less so to other cues.
- *Tempo and melody as supporting cues*: while loudness was the most prominent cue and was mostly responsible for rise/fall distinctions, tempo and melody became important when loudness was static, and discriminated between various levels of rises or falls when combined with changes in loudness. Analysis of trees showed that it is the absolute change in tempo that causes changes in engagement, while its direction is not so relevant. This can also be seen in the tree of Figure 6(b), where increases in tempo add to increases in EM when loudness rises, but also add to drops in EM when loudness falls. Such a relationship between tempo and EM explains why the direction of tempo has not been consistent in regression models; they are unable to capture the varying nature of cue–response relationships.

The melodic line of the chosen piece contains several large drops over consecutive bars, which also contributes to explanations of changes in emotional engagement. As with tempo, drops in melody seem to accentuate the effects of loudness and lead to more pronounced rises or falls. This can be observed in both trees (Figure 6(a) and 6(b)).

Summary and discussion

As explained in the introduction, the aim of this study was to investigate the emotional engagement of listeners with performances of a slow étude of Scriabin and to explore the effects of: (1) musical training of the listeners; (2) seeing the pianist and the movements of the pianist; (3) the expressiveness of the performance and the microstructure of the performance (e.g. variations in tempo and dynamics); and (4) the musical structure, especially phrase structure, and the performer's and listeners' interpretation of this. Below we summarize and discuss the results for each of these points.

The main difference between the responses of the musicians and the non-musicians was that the indications of emotional engagement by the non-musicians showed a stronger relation to the variations in dynamics, loudness and movement-velocity than those of the musicians. While the majority of the variance for the non-musicians could be explained on the basis of loudness, only a minor part was explained by the responses of the musicians in general. Also changes in EM occurred more systematically around phrase boundaries for non-musicians than for musicians. This suggests that non-musicians tended to pay closer attention to, or were more influenced by, the surface structure when indicating emotional engagement, while the musicians' responses were more detached from the surface structure.

An effect of the medium was found in the explained variances for the musicians and non-musicians together. On average, the explained variance was significantly higher for the video condition than for the audio condition. This higher explained variance was not due to the additional independent variable in the video condition of movement-velocity, since movement-velocity hardly reached significance in the MRAs. Instead loudness was the most important variable for the explanation of EM for both the audio and video condition. The higher explained variance might indicate a more successful communication between performer and listener with the addition of visual information, which would be in line with previous findings (e.g. Ohgushi and Hattori, 1996; Thompson and Russo, 2004).

The expressiveness of the performance was not found to have a consistent effect on emotional engagement; instead, the results often generalized over performance mode. This is in contrast with other studies that varied expressiveness (e.g. Davidson, 1993; Kendall and Carterette, 1990; Shimosako and Ohgushi, 1996). One reason for the difference could be that an inexpressive version was not present in the current study. Often the difference is largest between the inexpressive and normal version and only small between the normal and exaggerated version (both in Davidson, 1993 and Shimosako and Ohgushi, 1996).

The expressive variations within the performance did show a clear relationship with EM and changes within EM. Peak values in key-velocity (loudness) per two bars were most highly related to variations in EM in such a way that (absolute levels of) EM increased with loudness. Variations in duration (local tempo) were related to the size of the changes in EM; larger changes in EM occurred at positions with greater lengthenings of duration. This was confirmed in the analysis of EM with decision trees. The direction of EM change is only related to the direction of changes in loudness, but not to the directions of changes in tempo.

The difference in the explanation of absolute levels of EM and changes in EM is a reason to be careful with substituting regression analyses based on absolute levels with regression analyses based on derivatives, as Schubert (2001, 2001–02) proposes. The use of decision trees as an analytic tool may be of advantage here.

The importance of loudness for EM and tempo for change in EM seems to relate to the use of tempo and dynamics by this pianist. Dynamics reflected the pattern of tension-release as described by the pianist, while variations in quarter-note IOI related most strongly to the phrase structure of the music. The suggested relation between change in EM, quarter-note IOI and phrase structure was confirmed in the analysis of the relation between emotional engagement and phrasing, which was most apparent for the location of changes in emotional engagement; changes in emotional engagement occurred significantly more often around phrase boundaries than in the middle of phrases. Moreover, the size of the changes in emotional engagement correlated positively with the strength of phrase boundaries. In the analysis of EM with decision trees, melodic contour was added as an additional feature of musical structure. It seemed to have a similar effect to tempo: it increased the effect of loudness on EM at positions with descending melodic contour.

The relation between the pianist's interpretation, phrasing and emotional engagement seems an important observation, since we do not expect just any increase or decrease in dynamics to change the emotional engagement of listeners, but only carefully modulated dynamics that outline the intensity of the music, as was observed by Sloboda and Lehmann (2001).

A more general aim of this study was to see if findings from studies that investigated 'perceived' emotional intensity generalize to this study, which asked listeners to report 'felt' emotional engagement. We cannot know the extent of the difference in the task for the participants and can only assume a difference. We did find participants to be highly subjective, in the sense that agreement between participants was very low, which might point towards introspection rather than outward observation. To the extent that listeners agreed, they showed similar patterns to those in studies on 'perceived' emotional intensity and tension. For example, Krumhansl (1996) also showed a relation between higher-level phrasing and intensity of perceived emotion and tension and found, as well, that video and audio information of a ballet performance communicate similar patterns of tension-release (Krumhansl and Schenck, 1997). Moreover, the importance of dynamics for emotional intensity was suggested by several authors (see, e.g. Schubert, 2001).

In future studies, it seems of special interest to see how models on perceived categories of emotions such as developed by Juslin (2001) and Friberg et al. (2002) can be combined with observations about fluctuations in emotional intensity, emotional engagement or tension and release as in the current and related studies. An important step will be to measure valence besides arousal, and to see how different aspects of the music interact and lead to an emotional response. More effort should be put into really measuring 'felt' emotions and trying to model these. The addition of physiological measurements may be needed, such as skin conductance and temperature (e.g. Krumhansl, 1997) or frontal brain activation (e.g. Schmidt and Trainor, 2001). A qualitative investigation should accompany quantitative exploration to understand better what drives individual participants, and perhaps a larger scale study with a variety of music would be needed to characterize preferences and emotional responses for individual or groups of participants.

The effect of the performer and his influence on profiling the emotional engagement with the music was suggested by the current study. However, obviously a comparison between emotional responses to performances by different pianists of the same piece would be necessary to be sure about this effect.

Despite the limitations of this study, the study did take a step towards a more detailed understanding of the relation between movement cues, auditory performance cues, phrasing and the self-reported emotional engagement of listeners. It further explored the effects of musical training and the sight of the performer. It certainly suggested a need for further innovation in research methods and exemplified one of the possibilities (for a promising alternative see also Vines et al., 2004).

Conclusion

The conclusions on emotional engagement from this study are, in short, that especially non-musicians seemed to closely follow local variations in performance parameters when indicating their emotional engagement with music, while musicians had more varying emotional responses to the performances that were less predictable on the basis of variations in performance cues. Audio cues were in general more often significant in their explanation of emotional responses than video cues, though the movements of the pianist may have emphasized audio cues through their correlation with, for example, dynamics and therefore increased the predictability of responses. Changes in levels of emotional engagement consistently occurred more often around phrase boundaries than in the middle of phrases for both musicians and non-musicians and were greater at stronger phrase boundaries. Loudness was overall a good predictor of continuous variations in emotional engagement. Tempo and melodic contour increased the effect of loudness as secondary factors. The emotional engagement increased and decreased with the global lines of tension and release within a global phrasing that was outlined by the pianist and expressed in his use of dynamics. The correlation of emotional engagement with key-velocity meant therefore also a relation to global phrasing.

ACKNOWLEDGEMENTS

This study was realized with the financial support of the EU Mosart IHP Network and of the EU-IST Network of Excellence HUMAINE (Human–Machine Interaction Network on Emotion). We would like to thank Barbara Mazzarino, Andrea Pertino and Riccardo Casazza for their help and Massimiliano Damerini for his outstanding performances.

REFERENCES

Camacho, R. (1998) ‘Inducing Models of Human Control Skills’, in *Lecture Notes in Computer Science, Vol. 1398: Proceedings of the 19th European Conference on Machine Learning*, pp. 107–18. London: Springer-Verlag.

Clarke, E.F. and Windsor, L.W. (2000) ‘Real and Simulated Expression: A Listening Study’, *Music Perception* 17: 277–314.

Davidson, J.W. (1993) ‘Visual Perception of Performance Manner in the Movements of Solo Musicians’, *Psychology of Music* 21: 103–13.

Friberg, A. and Sundberg, J. (1999) ‘Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners’, *Journal of the Acoustical Society of America* 105: 1469–84.

Friberg, A., Schoonderwaldt, E., Juslin, P.N. and Bresin, R. (2002) ‘Automatic Real-Time Extraction of Musical Expression’, in M. Nordahl (ed.) *Proceedings of the International Computer Music Conference 2002*, pp. 365–7. San Francisco, CA: International Computer Music Association.

Gabrielsson, A. (2001–02) ‘Perceived Emotion or Felt Emotion: Same or Different?’, *Musicae Scientiae* special issue: 123–47.

Gabrielsson, A. and Lindström, E. (2001) ‘The Influence of Musical Structure on Emotional Expression’, in P. Juslin and J.A. Sloboda (eds) *Music and Emotion: Theory and Research*, pp. 223–48. Oxford: Oxford University Press.

Juslin, P.N. (2001) ‘Communicating Emotion in Music Performance: A Review and a Theoretical Framework’, in P. Juslin and J.A. Sloboda (eds) *Music and Emotion: Theory and Research*, pp. 309–37. Oxford: Oxford University Press.

Kendall, R.A. and Carterette, E.C. (1990) 'The Communication of Musical Expression', *Music Perception* 8(2): 129–64.

Krumhansl, C.L. (1996) 'A Perceptual Analysis of Mozart's Piano Sonata K. 282: Segmentation, Tension, and Musical Ideas', *Music Perception* 13(2): 401–32.

Krumhansl, C.L. (1997) An Exploratory Study of Musical Emotions and Psychophysiology', *Canadian Journal of Experimental Psychology* 51: 336–52.

Krumhansl, C.L. and Schenck, D.L. (1997) 'Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchines's Choreography of Mozart's Divertimento No. 15', *Musicae Scientiae* 1: 63–85.

Lucas, B. and Kanade, T. (1981) 'An Iterative Image Registration Technique with an Application to Stereo Vision', in P.J. Hayes (ed.) *Proceedings of the International Joint Conference on Artificial Intelligence, 1981*, pp. 674–9. Vancouver: William Kaufmann.

Meyer, L.B. (1956) *Emotion and Meaning in Music*. Chicago: The University of Chicago Press.

Mitchell, T.M. (1997) *Machine Learning*. New York, London: McGraw-Hill.

Ohgushi, K. and Hattori, M. (1996) 'Emotional Communication in Performance of Vocal Music – Interaction between Auditory and Visual Information', in B. Pennycook and E. Costa-Giomi (eds) *Proceedings of the Fourth International Conference for Music Perception and Cognition*, pp. 269–74. Montreal: Faculty of Music, McGill University.

Palmer, C. (1989) 'Mapping Musical Thought to Musical Performance', *Journal of Experimental Psychology* 15(12): 331–46.

Schmidt, L.A. and Trainor, L.J. (2001) 'Frontal Brain Electrical Activity (EEG) Distinguishes Valence and Intensity of Musical Emotions', *Cognition and emotion* 15(4): 487–500.

Scherer, K.R. and Zentner, M.R. (2001) 'Emotional Effects of Music: Production Rules', in P. Juslin and J.A. Sloboda (eds) *Music and Emotion: Theory and Research*, pp. 361–92. Oxford: Oxford University Press.

Scherer, K.R. , Zentner, M.R. and Schacht, A. (2001–02) 'Emotional States Generated by Music: An Exploratory Study of Music Experts', *Musicae Scientiae* special issue: 149–71.

Schubert, E. (2001) 'Continuous Measurement of Self-Report Emotional Response to Music', in P. Juslin and J.A. Sloboda (eds) *Music and Emotion: Theory and Research*, pp. 393–414. Oxford: Oxford University Press.

Schubert, E. (2001–02) 'Correlation Analysis of Continuous Emotional Response to Music: Correcting for the Effects of Serial Correlation', *Musicae Scientiae* special issue: 213–36.

Schubert, E. (2004) 'Modeling Perceived Emotion with Continuous Musical Features', *Music Perception* 21(4): 561–85.

Shimosako, H. and Ohgushi, K. (1996) 'Interaction between Auditory and Visual Processing in Impressional Evaluation of a Piano Performance', in *Proceedings of the Acoustical Society of America and Acoustical Society of Japan, Third Joint Meeting*, pp. 357–62.

Sloboda, J.A. (1991) 'Music Structure and Emotional Response: Some Empirical Findings', *Psychology of Music* 19: 110–20.

Sloboda, J.A. and Lehmann, A.C. (2001) 'Tracking Performance Correlates of Changes of Perceived Intensity of Emotion during Different Interpretations of a Chopin Piano Prelude', *Music Perception* 19(1): 87–120.

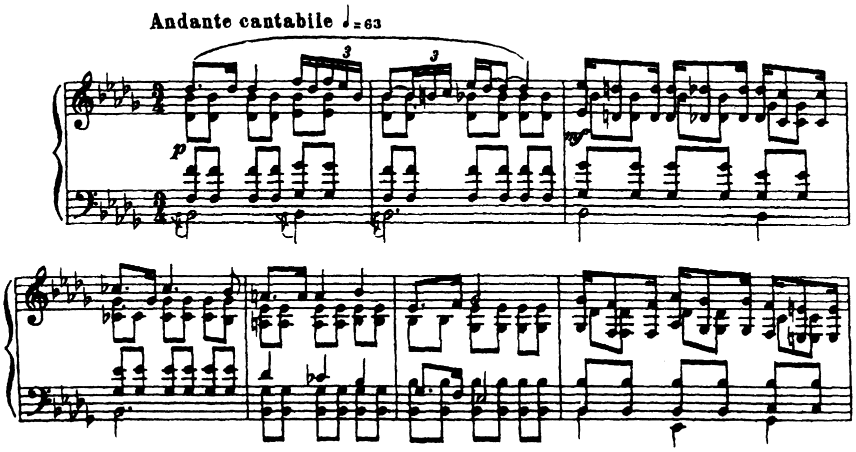
Thompson, W.F. and Russo, F.A. (2004) 'Visual Influences on the Perception of Emotion in Music', in S. Lipscomb, R. Ashley, R. Gjerdingen and P. Webster (eds) *Proceedings of the Eighth International Conference on Music Perception and Cognition*, pp. 198–99.

Todd, N.P. (1985) 'A Model of Expressive Timing in Tonal Music', *Music Perception* 3: 33–51.

Todd, N.P. (1992) 'The Dynamics of Dynamics: A Model of Musical Expression', *Journal of the Acoustical Society of America* 91(6): 3540–50.

Vines, B., Nizzo, R., Krumhansl, C.L. , Wanderley, M. and Levitin, D. (2004) 'Visual Music: The Perceptual Impact of Seeing a Clarinettist Perform', in S. Lipscomb, R. Ashley, R. Gjerdingen and P. Webster (eds) *Proceedings of the Eighth International Conference on Music Perception and Cognition*, pp. 200–3.

Appendix: Scriabin, Étude Op. 8, No. 11




2
mf
dim.
pp ppp
dim.
pp
ppp
cresc.


3
cresc.
f
dim.
pp
cresc.
cresc.
EVERYNOTE.com


4
rit.
a tempo
3
3
pp
ppp
1

RENEE TIMMERS received a MA in Musicology from the University of Amsterdam in 1997 and a PhD in Psychology from the University of Nijmegen in 2002. She works as a researcher at the Music, Mind, Machine group (University of Nijmegen), and has been affiliated to, among others, the Centre for the History and Analysis of Recorded Music (King's College London) and the Austrian Research Institute for Artificial Intelligence (Vienna). She teaches Psychology of Music at the Royal Conservatory of the Hague, the Netherlands. Her research interests include performance of music, perception of performance, and music cognition and emotion.
Address: Nijmegen Institute for Cognition and Information, University of Nijmegen, Nijmegen, The Netherlands. [email: reneetimmers@solcon.nl]

MATIJA MAROLT received BS and PhD degrees, both in computer science, from University of Ljubljana, Slovenia in 1995 and 2002 respectively. Since 1995, he has been working in the Laboratory of Computer Graphics and Multimedia, at the Faculty of Computer and Information Science, University of Ljubljana, where he is currently assistant in computer science. His research interests include music information retrieval, specifically extraction and segmentation of melody, audio transcription and music similarity, as well as audio-visual interaction.
Address: InfoMus, DIST, University of Genova, Viale Causa 13, 16145 Genova, Italy. [email: matija.marolt@fri.uni-lj.si]

ANTONIO CAMURRI was awarded a master's degree in Electric Engineering in 1984 and PhD in Computer Engineering in 1991. He is Associate Professor at DIST-University of Genova (Faculty of Engineering), where he teaches Software Engineering and Multimedia Systems. He is founder and Scientific Director of the DIST-InfoMus Lab at the University of Genova (infomus.dist.unige.it). Results from the InfoMus Lab have been used in several international artistic productions (e.g. Salzburg Festival 1999, Teatro La Scala, Milan, 1996), in museum and science centre interactive exhibits, entertainment multimedia, and multimedia systems for therapy and rehabilitation. His research interests include sound and music computing, multimodal intelligent interfaces, computational models of non-verbal expressive gesture, interactive multimodal-multimedia systems for museums, theatres, music, entertainment, therapy and rehabilitation.
Address: InfoMus, DIST, University of Genova, Viale Causa 13, 16145 Genova, Italy. [email: Antonio.Camurri@unige.it; website: www.infomus.org]

GUALTIERO VOLPE is a research scientist in the DIST-InfoMus Lab at the University of Genova. His research interests include intelligent and expressive human–machine interaction, modelling and real-time analysis and synthesis of expressive content in music and dance, Kansei information processing and expressive gesture communication. He has a PhD in computer engineering from the University of Genova. He was co-chair of the 5th International Gesture Workshop and guest editor of a special issue of the *Journal of New Music Research on Expressive Gesture in Performing Arts and New Media*.
Address: Infomus, DIST, University of Genova, Viale Causa 13, 16145 Genova, Italy. [email: Gualtiero.Volpe@unige.it]