# How Momentary Affect Impacts Retrospective Evaluations of Musical Experiences


## Abstract

Music is a temporal experience that can elicit fluctuating moment-to-moment intensities of affect, yet the relationship between moment-to-moment affect during a musical experience and subsequent retrospective evaluations (REs) of the experience is unclear. Three aspects of this relationship were investigated: over-weighting of specific moments (peak and end), segmentation of an experience (cohesive (individual pieces) vs. segmented (collection of pieces)), and trend of experience (increasing vs. decreasing trends of affect intensity). Across two studies, participants ($N$ = 123) listened to a recital (set) of six pieces and provided moment-to-moment evaluations of emotional intensity, as well as global REs of the pieces and the entire set. Trend was manipulated (between-subjects) by ordering pieces by increasing (*Low-High*) or decreasing (*High- Low*) emotional intensity. The peak-end did not contribute substantially to REs for individual pieces. REs of the recital relied on averages of global ratings of individual pieces rather than momentary affect, suggesting that segmented and cohesive experiences are evaluated differently. The *Low-High* group produced higher REs of emotional intensity than the *High-Low* group, demonstrating a trend effect. The average is proposed as the most appropriate predictor for REs in affective – including musical – experiences, with over-weighting of certain moments based on memorability (rather than the peak-end).


*Keywords*: peak-end, trend, segmentation, music, emotion

Music has the ability to convey and impact various emotions to various degrees of intensity (Ali & Peynircioglu, 2010; Baltazar & Saarikallio, 2016; Juslin & Sloboda, 2001). Although music is a temporal experience that prompts fluctuating moment-to-moment emotional intensity, listeners are also able to quickly make retrospective judgments of overall global emotional intensity for musical experiences spanning a few seconds (e.g., a piece excerpt) to hours (e.g., a live concert or recital). Listeners forming retrospective evaluations must rely on their memories of their experience (Kahneman et al., 1997). These memories, and in particular the *emotional experience* of these memories, form the basis for whether or not the individual chooses to repeat the experience. For example, in a study of dance festival visitors, positive emotion was one of the strongest predictors of whether individuals were likely to purchase festival tickets again and recommend the festival to others (Lee et al., 2008). There is thus not only theoretical but also practical impetus to better understand the way in which individuals' moment-to-moment experiences of music are integrated to form an overall impression that can later be recalled.

In considering how individuals integrate moment-to-moment experiences, it is no surprise that a simple average of a person's moment-to-moment experiences across a time period typically serves as a good predictor of global impressions (Strijbosch et al., 2019). However, an averaging model assumes that all moments are given equal weighting and overlooks factors such as order of intensity and segmentation. In contrast, other research on temporal experiences suggests that certain moments are given more weighting than others – notably, the maximum (peak) and final (end) intensities. For example, an individual's retrospective rating of pain intensity is primarily determined by averaging the *peak* pain intensity and the intensity of pain felt at the *end* of the experience – known as the *peak-end effect* (Kahneman et al., 1993).

In addition to the peak-end effect, the trend of an experience may also affect an individual's global evaluations of the experience. For instance, individuals perceive sound sequences of increasing loudness to be retrospectively more annoying than sequences that decrease in loudness (Ariely & Zauberman, 2000), suggesting a preference for declining sequences of negative affect. Similarly, when presented with different price profiles of monthly installments, car buyers preferred descending monthly installments (e.g., 60, 50, 40, 30, 20) to constant (e.g., 40, 40, 40, 40, 40) or ascending ones (e.g., 20, 30, 40, 50, 60) (Peine et al., 2012).

Another important factor to consider is whether experiences are perceived as segmented or cohesive, especially where the averaging model is not applied. Consider, for example, an individual who watches a theatrical play and provides ratings of emotional intensity at nine regular intervals during the play on a scale ranging from 1 (least emotional) to 10 (most emotional). The individual might generate the following profile of moment-to-moment intensity ratings across time: [3, 3, 5, 5, 9, 6, 6, 6, 8]. Applying the peak-end effect to the moment-to-moment ratings would yield a global rating of 8.5 (averaging 9 and 8). In contrast, if a play were segmented into three acts, the profile might appear as: [{3, 3, 5}, {5, 9, 6}, {6, 6, 8}]. Applying the peak-end effect to the individual acts *first* would yield the profile: [{4.5}, {7.5}, {7}]; applying the peak-end effect again to this profile would yield a global rating of 7.25 for the play, which would be somewhat lower than if the peak-end effect were applied directly to the moment-to-moment ratings. As such, the same "integration models" might produce different global evaluations depending on whether experiences are cohesive or segmented. Discussions about the unit of analysis extends beyond segmentation of musical experiences, and may also be relevant to the experience of visual arts. For example, aesthetic experiences of art could be examined in terms of experiences of individual paintings or the experience of an entire exhibition of paintings. To further illustrate, the

museum visit in totality may even be considered the unit of analysis, which prompts broader and more integrated considerations of how each part of the museum can impact the rest of the artistic experience (Smith, 2014).

The overweighting of peak-end, trend, and segmentation are but a few of the "order effects" that have been observed to affect the way an experience is remembered; others include its rate of change (Hsee et al., 1991), primacy-recency effects (Murdock, 1962), and calibration effects (Fasold et al., 2015). Although these may also influence listeners' evaluations of music, we believe that examining trend, segmentation, and peak-end effects are of highest priority and relevance for musical experiences, for several reasons. First, trend overlaps with many other order effects, such as primacy-recency and calibration, given that it accounts for changes in emotional intensity from the beginning to the end. Second, segmentation is particularly relevant, given that many musical experiences (such as a recital or concert) may be segmented into intervals. Understanding the role of trend and segmentation may prove highly beneficial to musicians and event curators; to our knowledge, no empirical study has yet investigated these factors in the context of a musical recital. Third, although the peak-end effect has been rigorously observed across various non-musical experiences, few studies have compared its effectiveness to other models in predicting global retrospective evaluations of musical experiences. Further, the peak-end effect appears to predominantly apply to aversive experiences (e.g., Kahneman et al., 1993; Stone et al., 2000; Varey & Kahneman, 1992), whereas studies investigating positive affect have produced mixed results (Do et al., 2008; Hui et al., 2014; Kemp et al., 2008; Rode et al., 2007).

Given that music is consistently ranked one of the top ten sources of pleasure (Dubé & Le Bel, 2003), it is important to establish whether the peak-end effect, a common heuristic, applies to segmented musical experiences. Further, understanding the relationship between moment-to-moment experiences and global evaluations of a musical performance may inform

musicians on the best use of the order of musical selections to maximize their emotional impact on an audience–an important goal, given the strong relationship between emotional arousal and ratings of pleasure (Salimpoor et al., 2009).

A review of previous studies investigating musical experiences reveals mixed results regarding the importance of the *peak-end* effect. While the average across moment-to-moment evaluations has been shown to consistently predict global ratings (Duke & Colprit, 2001; Rozin et al., 2004; Schäfer et al., 2014; Sloboda & Lehmann, 2001), whether the peak is a stronger predictor or significant contributor to global ratings is unclear. For example, Schäfer et al. (2014) found the average to be the strongest single predictor of global ratings, with peak-end contributing significantly, although the $R^2$ change with the addition of the peak-end effect is not shown in the regression. In contrast, Rozin et al. (2004) found global ratings to correlate most strongly with the peak. As such, it yet remains unclear the extent to which the peak-end influences global ratings. Additionally, the literature currently seems to lack empirical investigations on the potential interaction with the impact of trend and segmentation.

Further, the bulk of previous studies used musical clips or individual pieces that did not control for genre, duration, and instrumentation, rendering participant responses susceptible to individual differences across trials. Often, a musical experience will consist of an entire recital or collection of pieces; thus, presenting musical experiences as clips or individual pieces contributes to the issue of ecological validity, which is a moderator of emotional responses produced by music. For example, in a study aiming to demonstrate the role of activated art schemas in increasing the enjoyability of negative emotions such as anger, presenting participants with an entire live theatre performance was more effective at activating art schemas than presenting them with single pictures or film clips (Wagner et al., 2016).

Taking into account gaps in earlier research, the aim of this study is to understand the extent to which the peak-end, trend, and segmentation of musical experiences influence global retrospective evaluations of these experiences. In this work, we build upon paradigms used by prior research by presenting musical pieces as a recital, simulating a real-world experience, and additionally controlling for various individual factors that may influence decisions, such as degree of musicality and musical experience. Crucially, trend will be manipulated by ordering recital pieces by increasing (*Low-High*) or decreasing (*High-Low*) emotional intensity. The contribution of the *peak-end* will be compared to other parameters (*average*, *beginning*, *peak*, and *end*) in a regression analysis to identify a model that best predicts global ratings. The effect of segmentation (whether the recital set is perceived as one cohesive experience or segmented by each piece) will be observed by comparing moment-to-moment ratings across the set to global ratings for each piece as predictors of global ratings of the recital.

We propose three hypotheses:

*Hypothesis 1*: The average of moment-to-moment emotional intensity (EI) ratings across a piece will be the single best predictor of the overall EI rating for the piece; adding the peak-end rating will significantly improve the prediction.

*Hypothesis 2*: Individual pieces will be considered segments, such that the average of the global EI ratings for individual pieces, rather than the average of moment-to-moment ratings across a recital, will be the single best predictor of the global EI rating for the set.

*Hypothesis 3:* The trend of EI across the recital will influence global EI ratings for the set, such that intensity ratings for the *Low-High* condition will be higher than ratings for the *High-Low* condition.

## Experiment 1

### Method

***Participants***

Sixty-three participants (49 females and 14 males[1]) were recruited via a university participant pool. Participants ranged between 18 and 50 years ($M$ = 21.38, $SD$ = 4.36). Of the 33 participants who considered themselves to be musicians, 19 participants had 10 or more years of experience, and 12 participants identified at ABRSM/equivalent Grade 8 level or above.[2] Music was rated to be an important part of life for most participants (Min = 0, Max = 10, $M$ = 7.81, $SD$ = 2.30), and participants varied in self-reported musicality (Min = 0, Max = 10, $M$ = 5.67, $SD$ = 2.49).

***Procedure***

To control for individual differences in preferences for instrumentation and genre, we sought to use pieces that only consisted of one instrument and that varied in emotional intensity but were consistent in style. Stimuli consisted of six complete solo piano pieces in the Romantic style[3] performed by professional pianists. The pieces were selected to be similar in length (approximately 2-2.5 minutes) and for variation in tempo, articulation, and key. These pieces were expected to be unknown to most participants and to vary in emotional intensity (Table 1).

A pre-test was conducted to determine the global emotional intensity of each piece using a separate cohort of participants.[4] The mean emotional intensity ratings obtained from the pre-test were ranked, and two conditions of recital order were created by ordering pieces

[1] The sample was drawn from a pool consisting primarily of female psychology students.
[2] ABRSM (Associated Board of the Royal Schools of Music) is a musical examinations board recognized internationally and by most musicians in the UK. Grade 8 is the prerequisite for obtaining a teaching-level diploma in music, and is the highest number-grade level for most music certification boards in the UK.
[3] The Romantic style was chosen for emotional, expressive, and dynamic qualities. Romantic pieces were expected to elicit more varied responses in emotional intensity than, for example, pieces in the Baroque style.
[4] Pieces were presented to 14 participants. The methodology was identical to that used in the main study; pieces were presented in a random order. The global emotional intensity for each piece was averaged across participants, and a repeated-measures ANOVA revealed a significant main effect of piece; $F(5, 65) = 4.61$, $p < .001$, $h_p^2$ = 26.2%, suggesting that the pieces varied in emotional intensity in a consistent pattern (ranking) across participants.

from low to high intensity (*Low-High* condition) and high to low intensity (*High-Low* condition) (Table 1).

This study utilized a continuous evaluation slider, which measured moment-to-moment ratings of participants' subjective experiences of emotional intensity. Participants operated the slider as they listened to each piece by dragging it left and right according to their moment-to-moment subjective experiences of emotional intensity.[5] The slider ranged from *0 – not emotionally intense at all* to *100 – very emotionally intense*. Values were captured every 200ms[6] (5 data points/s) to reflect continuous moment-to-moment ratings throughout each piece.

Participants completed the study on a laptop or desktop computer (an error message was displayed if it was accessed via other devices, such as cell phones/tablets) and were instructed to make sure they were in a quiet space and wearing headphones. They were randomly assigned to listen to the *Low-High* recital or the *High-Low* recital. The slider (ranged from *0 – not emotionally intense at all* to *100 – very emotionally intense* and set to a default value of 50) captured participants' moment-to-moment ratings of emotional intensity. After each piece was played, the screen automatically progressed to a screen displaying three sliders. The first slider corresponded to overall emotional intensity, ranging from *0 – not emotionally intense at all* to *100 – very emotionally intense*. The second slider corresponded to enjoyment, ranging from *0 – did not enjoy at all* to *100 – enjoyed very much*. The third slider corresponded to familiarity, ranging from *0 – not familiar at all* to *100 – very familiar*.

[5] Participants were asked to move the sliders to reflect their felt emotions. No drastic measures were taken to distinguish between felt and perceived emotions, since participants typically conflate the two (Konečni, 2008), which are highly correlated (Hunter, Schellenberg, & Schimmack, 2010), and we were primarily interested in the *relationship* between the moment-to-moment evaluations rather than the type of emotions themselves.
[6] We found this sampling rate to be sufficient in producing dynamic emotional intensity ratings as well as revealing distinct peaks and troughs in the contour graphs.

Participants were asked to adjust the sliders to reflect their views for the piece to which they had just listened.

Once participants had listened to all six pieces, subjects progressed to a demographic questionnaire that asked for their gender, age, education level, nationality, and whether or not they played a musical instrument. Following the demographic survey, participants were asked to indicate how emotionally intense they perceived the overall recital to be and how much they enjoyed the overall recital by dragging a slider. They also indicated whether they wanted to listen to the recital again (clicking "yes" provided participants with a link to the recital at the end of the study; this link has since been deactivated). Lastly, participants completed a questionnaire on their self-identified musicianship and musical preferences.[7] We obtained IRB approval for these studies at University College London.

**Results**

Each participant generated an emotional intensity profile for each piece. These profiles were averaged across all participants to create a visual display of emotional intensity contours for each piece (*Figure 1*).

As illustrated in Figure 1, the contours of moment-to-moment emotional intensity varied for each piece, with some pieces having a more distinct peak (e.g., Chopin-Waltz) than others.

***Factors contributing to ratings of individual pieces***

To investigate the factors that predicted participants' overall ratings of emotional intensity for individual pieces, *average, beginning, end, peak,* and *peak-end* values were calculated for each piece for each participant by running a script written in Python. The

[7] Participants reported musicality (0 = not musical at all, 10 = very musical) and importance of music in one's life (0 = not important at all; 10 = very important), and whether or not they considered themselves to be musicians. For those who considered themselves to be musicians, participants answered further items indicating years of experience in their instrument and grade level (ABRSM or equivalent). Participants who did not identify themselves as musicians were considered to have 0 years of experience in the analysis.

*average* was calculated by averaging moment-to-moment intensity ratings across an individual piece. The *beginning* was defined as the interval between 5s and 15s for each piece, and the values between this interval were averaged to obtain a value. This 5-15s interval was modelled after previous studies (Nagel et al., 2007; Schäfer et al., 2014; Sloboda & Lehmann, 2001), which have suggested 5s to be sufficient for generating an initial response to perceived emotions in music. The *end* was defined as the last 10s of the piece, and the values in this interval were averaged to obtain a value. The *peak* was defined as the maximum value that was recorded from the participant for each piece. The *peak-end* value was obtained by averaging the *peak* and *end* values.

Pearson correlations between the parameters (*global rating*,[8] *average, beginning, end, peak* and *peak-end*) across all pieces and all participants ($n$ = 63; number of trials: 378) were calculated (Table 2). Although not central to our hypothesis, enjoyment and familiarity[9] were also included to illustrate the general relationship among variables.

All correlations among parameters (apart from familiarity, a control parameter) were significant at the $p < .001$ level. As expected, the *average* correlated most highly with the overall rating ($r$ = .829). The *peak-end* correlation followed closely ($r$ = .806) and was higher than the correlations of the individual *peak* and *end* ratings. The beginning was least correlated with the average and the other parameters.

Our first hypothesis ($H_1$) predicted that the *average* would be the single best predictor of the overall EI rating for an individual piece and that adding the *peak-end* rating would significantly improve the prediction. We tested this hypothesis using a hierarchical regression analysis, predicting that the average would produce the highest $R^2$ change and that the addition of peak-end in the following step would induce a significant $R^2$ change. To prevent

[8] Of emotional intensity
[9] Control variables and enjoyment are discussed in the last section of the results.

conflating separate trials with separate participants, we calculated regressions on a piece-by-piece basis, performing six hierarchical regression analyses in total. To control for individual differences, musicality, importance, musicianship,[10] years of musical experience, gender,[11] age, and familiarity were entered in the first step.

To first identify the single best predictor of overall emotional intensity, all of the variables (*average*, *peak*, *peak-end*, *beginning*, and *end*) were first entered individually in the second step to ensure that the *average* yielded the highest $R^2$ change.[12] Once this was confirmed, the *average* was entered in the second step, and the *peak-end* was entered in the third step. The beta coefficients and $R^2$ changes (as a result of adding the *average* and *peak-end*) illustrate the influence of these parameters in the overall prediction (Table 3).

Consistent with our first hypothesis, for each of the pieces, the *average* was the best single predictor of overall emotional intensity, accounting for between .411 (Chopin-Etude) and .683 (Schumann-Kinder) of the total $R^2$ value. However, the *peak-end* did not consistently significantly improve the prediction: the $R^2$ change ranged from .002 (Chopin-Waltz) to .06 (Chopin-Etude) and was only significant for four of the pieces (Liszt, Chopin-Etude, Rachmaninoff, and Schumann-Carnaval).[13]

Overall, results partially supported the hypothesis; the finding that the *peak-end* did not reliably improve the prediction suggests that certain moments (such as the peak and end) sometimes do appear to be overweighted in forming global evaluations, but they cannot be reliably encapsulated by the peak-end, for reasons which will be explored in the discussion.

[10] Non-musicians were coded as 0; musicians were coded as 1.
[11] Females were coded as 0; males were coded as 1.
[12] In the case where the first step yielded a significant $R^2$ value, all control variables (musicality, importance, musicianship, years of musical experience, gender, age, and familiarity) were also entered individually to determine whether one of these variables could be the best predictor. For all the pieces in Experiment 1, the *average* was the single highest predictor.
[13] Because the peak-end did not consistently significantly improve the prediction in the third step, we substituted the other parameters (*beginning*, *peak*, and *end*) to see if there was a better predictor. The *end* proved to be a better predictor than the *peak-end* for two of the pieces: Chopin-Etude and Schumann-Carnaval (see below Table 3).

***Segmentation of the recital***

The second aim of our study was to determine whether a recital is perceived as a segmented or cohesive experience. To investigate this, we compared the effectiveness of two parameters in predicting global emotional intensity ratings of a recital: *average-global* and *average-momentary*. *Average-global* was calculated by averaging the six global ratings of the individual pieces. *Average-momentary* was calculated by averaging all of the moment-to-moment ratings across the recital without distinguishing across pieces.[14] To compare the effectiveness of these parameters, we ran two hierarchical regressions: in both, controls (musicality, importance, musicianship, years of musical experience, gender, and age) were entered in the first step. In the first regression, *average-global* was entered in the second step, and *average-momentary* in the third step. In the second regression, *average-momentary* was entered in the first step and *average-global* in the third step (Table 4).

Results indicate that *average-global* accounted for a larger proportion of the variance (.440 versus .345). In the first regression, the addition of *average-momentary* did not significantly improve the overall prediction, suggesting that *average-momentary* does not explain additional variance unaccounted for by *average-global*. Furthermore, in the second regression, the addition of *average-global* in the third step significantly improved the prediction, suggesting that the prediction of global intensity ratings is significantly improved by the inclusion of *average-global* even when *average-momentary* has already been factored in.

Given that global recital ratings were more strongly predicted by the average of global ratings of individual pieces, our results suggest that, as hypothesized, the recitals were perceived as a segmented (rather than cohesive) experience.

[14] Hence, the values making up the *average-momentary* parameter in Table 4 are equivalent to the values making up the *average* parameter in Table 3 combined across pieces. Similarly, in Experiment 2, the values making up *average-momentary* in Table 9 are equivalent to the combined *average* values in Table 8.

***Comparison of the Low-High and High-Low groups***

Our third hypothesis predicted that the *Low-High* group would provide greater global intensity ratings for the set than the *High-Low* group. As predicted, the *Low-High* group rated the overall recital as more emotionally intense ($M$ = 66.77, $SD$ = 10.42) than the *High-Low* group ($M$ = 57.63, $SD$ = 16.46). To test the statistical significance of this result, we entered data into a box-plot and removed two outliers. A one-way independent-samples ANOVA revealed a significant difference between the two groups, $F(1,59) = 6.761$, $p = .012$, $\eta^2_p$ = 10.3%,[15] supporting the prediction that increasing patterns of emotional intensity result in higher overall ratings of emotional intensity than decreasing trends.

***Enjoyment, "listen again" responses, and control variables***

Our study included a number of control variables, measuring participants' self-reported musicality, the importance of music in their lives, self-reported musicianship, years, of musical experience, gender, and age. Familiarity was also measured for individual pieces. While certain demographic traits[16] were related to each other and to emotional intensity ratings (Table 5), these control variables did not completely account for differences in emotional intensity, when combined accounting for between 5.3% - 24.5% for individual pieces and 12.4% for the overall recital. However, these variables were related to enjoyment, with familiarity significantly related to enjoyment for individual pieces ($r = .132$, $p = .010$) and importance, musicianship, and years of experience significantly related to enjoyment ratings ($r = .283$, $p = .025$; $r = .300$, $p = .017$; and $r = .335$, $p = .007$ respectively) for the overall recital (Tables 2 and 5).

[15] A Levene test revealed that the data may have lacked equality variance; however, a non-parametric Mann-Whitney test yielded similar results, so the initial analysis was reported for ease of interpretation.

[16] For instance, musicianship was positively correlated with emotional intensity ($r = .200$), but this was not significant. Musicality was negatively correlated with age ($r = -.300$, $p = .017$).

Interestingly, perceived emotional intensity had the strongest correlation with enjoyment ($r$ = .409, $p$ = .001), and enjoyment exhibited the strongest correlation with participants choosing to listen to the recital again ($r$ = .507, $p$ < .001).

## Discussion

Our first experiment yielded results both consistent with previous work and novel to our understanding of how moment-to-moment experiences are integrated in musical experiences. The results replicated those of earlier work showing that the *average* was the best single predictor for individual pieces, but our piece-by-piece analysis revealed that *peak-end* did not consistently significantly contribute to the overall prediction, and in some cases the *end* served as a better parameter, suggesting no clear *peak-end* effect. *Average-global* was a better predictor of global recital ratings than *average-momentary*, suggesting that participants integrated ratings from segments (global ratings for pieces) rather than from moment-to-moment ratings throughout the recital. Finally, trend influenced global set ratings such that intensity ratings were higher when pieces were ordered in increasing emotional intensity (*Low-High*) as compared to when pieces were ordered in decreasing intensity (*High-Low*).

# Experiment 2

The results from the first study demonstrated that trend could be used to influence listener global evaluations of musical experiences and that segmentation should be taken into account when considering how moment-to-moment experiences are integrated to form overall impressions. However, while the *average* was consistently the best single predictor, the *peak-end* did not consistently improve the prediction, only improving the impression for four out of the six pieces. Furthermore, the *end* actually served as a better predictor than the *peak-end* for two of the pieces. If it were the case that the parameter *depended* on the pieces

themselves, then one would expect that in a replication of the study, the same parameters would improve the prediction of the same pieces. Additionally, we recognize that the population sample in Experiment 1 consisted primarily of university psychology students, who may have been more motivated to pay close attention to the music throughout the study due to speculating about the purpose of the study and possible demand characteristics. To account for these issues raised by Experiment 1, a second cohort was recruited.

## Method

### *Participants*

Participants were recruited using Amazon Mechanical Turk, an online subject pool, and were paid $2.40 for their participation. Sixty adults (39 male, 21 female) took part, ranging from 20 to 58 years ($M$ = 32.85, $SD$ = 8.64). Eight played an instrument and considered themselves to be musicians. Music was rated to be an important part of life for most participants (Min = 0, Max = 10, $M$ = 7.68, $SD$ = 1.77), and they varied in their self-reported musicality (Min = 0, Max = 10, $M$ = 6.53, $SD$ = 2.44).

### *Procedure*

The previous study found that most of the stimuli produced rather high emotional-intensity ratings, with few pieces being rated as less emotional. To increase the variety of emotional-intensity ratings, two of the pieces (which were previously in the middle range) were changed to pieces that were expected to be perceived as less emotional in order to increase the range of the emotional intensity across the pieces.

As in Experiment 1, a pre-test was conducted with a separate cohort of participants.[17] The mean emotional intensity ratings obtained from the pre-test were ranked, and two

[17] Pieces were presented to 15 participants recruited through Amazon Mechanical Turk and were paid $2.40 for their participation. The methodology was identical to the pre-test conducted in Experiment 1. The global emotional intensity for each piece was averaged across participants, and a repeated-measures ANOVA again revealed a significant main effect of piece; $F(5, 70) = 4.63$, $p < .001$, $h_p^2$ =24.9%

conditions were created by ordering pieces from low to high intensity (*Low-High* condition) and high to low intensity (*High-Low* condition) (Table 6).

The procedure replicated the one used in Experiment 1.

**Results**

The parameters were defined and the data was analyzed consistent with those in Experiment 1. The contours of moment-to-moment emotional intensity for the individual pieces are displayed in Figure 2. For pieces also used in Experiment 1, the contours for both cohorts are displayed in the same graph.

As shown in Figure 2, similar contour patterns are seen between Experiments 1 and 2, suggesting that perceived changes in emotional intensity over time are consistent across different cohorts of participants.

***Factors contributing to ratings of individual pieces***

Correlations between all parameters were significant at the $p < .001$ level, as seen in Table 7, with the exception of the correlation between *familiarity* and *peak*, which was significant at the $p < .05$ level. The *average* correlated most highly with the overall rating ($r = .753$) as in Experiment 1, but the *peak-end* correlation ($r = .725$) was not higher than the *end* correlation ($r = .727$). In contrast to Experiment 1, *familiarity* was significantly correlated with all other parameters, although it was the least correlated with the other parameters.

As in Experiment 1, parameters were first entered individually to determine the single best predictor.[18] Contrary to results in Experiment 1, the average was not always the single best predictor. As seen in Table 8, the *average* was the single best predictor for three of the pieces: Chopin-Waltz ($R^2$ change = .295, $p = .016$), Beethoven-Fast ($R^2$ change = .160, $p$

[18] In the case of four of the pieces (Chopin-Waltz, Beethoven-Fast, Beethoven-Slow, and Schumann-Kinder), the control variables entered together in the first step resulted in a significant $R^2$ value (Table 8); these variables were then entered individually to determine whether one of the control variables could be the single best predictor.

= .031), and Schumann-Kinder ($R^2$ change = .128, $p$ = .028); however, the addition of the *peak-end* did not improve the prediction for any of these pieces. Furthermore, the *peak*, and *peak-end* were the single best predictors ($R^2$ change = .264, $p$ = .002 and $R^2$ change = .156, $p$ = .001) for Chopin-Etude and Liszt, respectively. Even more surprisingly, familiarity was the overwhelmingly strongest predictor for Beethoven-Slow, alone accounting for 62.7% of the variance ($p < .001$); no other parameter improved the prediction.

While the *average* was consistently the best single predictor in Experiment 1, it does not appear to be the case in Experiment 2, for reasons which will be considered in the discussion. Similar to Experiment 1, the *peak-end* did not consistently improve predictions, suggesting that the *peak-end* is not consistently overweighted when forming evaluations of musical pieces.

***Segmentation of the recital***

Results relating to this hypothesis were replicated in Experiment 2 (Table 9), with *average-global* accounting for a larger proportion of the variance than *average-momentary* ($R^2$ change = .134 versus .035). The control variables accounted for a larger proportion of the variance in Experiment 2 than in Experiment 1 ($R^2$ = .203 versus .124 respectively), for reasons which will be explored in the discussion.

***Comparison of the Low-High and High-Low groups***

Consistent with Experiment 1, the *Low-High* group rated the overall recital as more emotionally intense ($M$ = 73.00, $SD$ = 16.79) than the *High-Low* group ($M$ = 63.83, $SD$ = 23.02). A one-way independent samples ANOVA revealed a significant difference between the two groups: $F(1, 58) = 3.10$, $p$ = .042 (one-tailed), $h_p^2$ = 5.1%. The results support the prediction that positive slopes of emotional intensity result in higher overall ratings of emotional intensity than negative slopes do.

***Enjoyment, "listen again" responses, and control variables***

Similar to Experiment 1, certain demographic traits related to each other,[19] but only self-reported musicality was correlated with emotional intensity. As in Experiment 1, emotional intensity exhibited the strongest correlation with enjoyment ($r$ = .544, $p$ < .001). As mentioned previously, individual differences had a larger impact on overall recital ratings than they did in Experiment 1. For individual pieces, these control variables combined accounted for between 12.3% - 71.6% of the variance for individual pieces (Table 8), with familiarity in particular having a large effect, being significantly correlated with the *average*, *beginning*, *end*, *peak*, and *peak-end* (Table 7).

As in Experiment 1, perceived emotional intensity was strongly correlated with enjoyment ($r$ = .544); however, all of the variables (Table 10) apart from age were significantly correlated with decisions to listen to the recital again, unlike Experiment 1.

**Discussion**

The results of Experiment 2 replicate the results confirming our second and third hypotheses in Experiment 1, providing further support for the effects of trend and segmentation. In neither Experiment 1 nor Experiment 2 did the peak-end consistently improve predictions, further suggesting that the peak-end is not necessarily overweighted when forming evaluations of musical experiences. Furthermore, the *average* was not consistently the best predictor in Experiment 2; other strongest predictors included *peak*, *peak-end*, and interestingly, familiarity, suggesting that there may be factors not captured by moment-to-moment emotional intensity that influence overall ratings of pieces.

## General Discussion

This study aimed to clarify the relationship between moment-to-moment experiences and global retrospective evaluations of emotional intensity for musical experiences. Across

[19] Musicality was positively correlated with musicianship ($r$ = .734 $p$ < .001) and experience ($r$ = .432, $p$ < .001). Age was negatively correlated with musicianship ($r$ = -.446, $p$ < .05)

two experiments, the *average* of momentary ratings of emotional intensity was often the most consistent single predictor of the subsequent evaluation of global emotional intensity for individual pieces, although not without exception. The *peak-end* rating did not reliably improve the prediction across all pieces, suggesting that individuals do not necessarily give specific moments more weight than others when forming retrospective evaluations. The average of *global* ratings of pieces, in addition to being the strongest predictor of global ratings for the recital, was better than the average of the *moment-to-moment* ratings across all pieces, supporting the prediction that participants perceived the recital as a segmented rather than cohesive experience. Additionally, when pieces were ordered by increasing emotional intensity, ratings of global emotional intensity for a recital were higher than when pieces were ordered by decreasing emotional intensity, suggesting that global evaluations of emotional intensity for musical experiences are influenced by trend.

**Implications of the peak-end effect and integration models for musical experiences**

Whereas previous studies have suggested that the peak (Rozin et al., 2004) or peak-end (Schäfer et al., 2014) is given more weight than other moments in evaluations of musical experiences, this study demonstrated that, while the peak and peak-end *can* improve predictions for certain musical pieces, they are not consistently over-weighted across a set of pieces, even when controlled for genre, duration, and instrumentation.

The findings of this work provide support for the notion that the peak-end effect may be more applicable to aversive rather than pleasurable experiences. From a theoretical perspective, the peak and end moments are hypothesized to be important because they give individuals an indication of the capacity requirements needed to survive a threatening experience (Frederickson, 2000). In contrast, when recalling musical experiences, it is not just the peak and end moments that provide "value." Fundamental to music is its "directional" nature (Meyer, 1956, p. 161) and "unfolding temporal form" (Albert & Bell,

2002, p. 576), whereby "events are understood as consequences of earlier events" (Kramer, 1988, p. 453). As such, music cannot be understood by just considering "snapshot" moments. This may explain why the *average*, which integrates all moments of a piece, is a stronger predictor of global evaluations.

While controlling for several individual factors including musicality, musicianship, importance of music in one's life, gender, and age improved the overall prediction, it is still unclear why there was such a disparity in the predictive power across parameters. The observation that predictions for certain pieces were improved with the addition of one of the parameters (e.g., *peak-end* or *end*) perhaps suggests that certain moments *do* appear to be over-weighted, albeit not necessarily defined by parameters such as *peak*, *end*, and *peak-end*.

If certain moments are indeed over-weighted when forming retrospective evaluations, how are these moments determined? An explanation may lie in the *memorability* of these moments: emotional intensity experienced at the most memorable moments may contribute to the participant's retrospective global evaluation because they are the most likely to be recalled (Ochsner, 2000). This is supported by the finding that in Experiment 2, familiarity was strongly correlated with emotional intensity, and was even the single strongest predictor for one of the pieces (Beethoven-Slow).

Further support is found in existing literature; for example, in a study investigating the factors predicting overall happiness of a vacation, researchers found that after the average, the happiness at the vacation's most memorable and distinctive moments correlated the most highly with remembered happiness of the overall experience (Kemp et al., 2008). Previous studies of pleasurable experiences have also reported a possible "end" effect (Do et al., 2008; Hui et al., 2014). Considering that recency effects typically occur in tests of memory (Watkins, 1972), these "end effects" would also be consistent with this memorability model.

Future exploration of the factors influencing the memorability of moments, and whether these in turn influence retrospective evaluations, may prove promising.

**Global retrospective evaluations of the recital: Segmentation and trend effects**

In addition to illuminating issues regarding the over-weighting of specific moments, findings from this study demonstrated that recitals are perceived as segmented (rather than cohesive) experiences, thus audiences rely on the average of global ratings for individual pieces rather than moment-to-moment ratings accumulated throughout the entire recital. This is consistent with Ariely and Zauberman's (2000) theory that, as a result of segmentation, global evaluations of segments, rather than the relationship between all moments of an experience across segments, determine global evaluations of the entire experience. To our knowledge, this is the first time segmentation has been demonstrated empirically with a set of musical pieces.

A limitation of this finding is that it is unclear if the segmentation of the experience came about organically; it could be that segmentation was induced by requiring participants to form global ratings for individual pieces. For example, Ariely and Zauberman found that simply *asking* participants to provide momentary ratings reduced the impact of the trend on global ratings. An explanation for this is that segments become more apparent when asked to evaluate them. A future study might therefore compare the effect of presenting all the pieces consecutively without asking participants to provide a rating between pieces.

A final aim of this study was to demonstrate trend effects for musical experiences. The results supported the prediction that ordering pieces by increasing emotional intensity resulted in higher overall intensity ratings than ordering pieces by decreasing emotional intensity. Because trend effects had not previously been empirically demonstrated for musical experiences in this way, we compared only two trend patterns: *High-Low* and *Low-High*. As such, we cannot rule out the possibility that differences between the two groups were due to

recency effects (i.e., sets having endings of different emotional intensities) rather than trend effects.

One way to rule out alternative explanations for the trend effect is to introduce additional trend patterns such as *High-Low-High* (whereby the intensity of a sequence could decrease halfway and then increase again) and *Low-High-Low*. For example, a previous study demonstrated that sounds sequences were rated more annoying when the trend was constant (e.g., Low-High) than when the trend varied (e.g., High-Low-High) (Ariely & Zauberman, 2000).

A step further in investigating the effect of trend would be to manipulate the slope of change (Hsee et al., 1991). Previously, Rozin et al., (2004) found a strong linear fit between rankings of slope and correlations with remembered intensity ($r^2 = .92$), suggesting that moments with steeper slopes resulted in larger changes in emotional intensity. Observing an increase in the difference between *High-Low* and *Low-High* groups as a result of an increase in the slope of the trend would lend support to the hypothesis that retrospective evaluations are affected by trend.

**Practical implications, limitations, future research, and conclusions**

The results of this study may be of practical significance to musicians and listeners alike. Given the role of music as a source of pleasure (Dubé & Le Bel, 2003), the finding that emotional intensity was strongly related to enjoyment provides impetus for musicians to maximize the emotional impact of their music. Given the trend effects observed, professional musicians planning a recital may choose to maximize the audience's perceived emotional intensity by ensuring a positive slope of affect intensity across their recital. Furthermore, participants who enjoyed the recital were likely to be willing to listen to the recital again; understanding the relationship between emotional intensity and enjoyment may inform how

musicians might select repertoire to perform in front of audiences, as well as how event curators and arts organizations manage artistic programs.

Given that one of the primary reasons people listen to music is to manage or regulate their moods (Lonsdale & North, 2011), the finding that global evaluations can be affected by trend may be of interest to music listeners. The suggestion that memorability may determine the specific moments that influence subsequent retrospective ratings may be useful in developing methods to aid in mood regulation; a limitation of this study was that familiarity was not a controlled variable. Future research might minimize the influence of familiarity by using new compositions and in other genres.

An interesting observation was that self-reported musicians appeared to provide slightly higher ratings of emotional intensity. Although this finding was not significant, it may suggest that musicians are slightly more sensitive to the emotional intensity of music. This is consistent with findings that musicians are more sensitive to emotions in voices (Strait et al., 2009) and changes in musical features that may convey emotional content (Liu et al., 2018), and have higher levels of granularity of emotional responses to music (Kantor-Martynuska & Horabik, 2015). The results may have been insignificant because "musicianship" was defined by self-reported measures rather than by objective measures of musical ability. Future research could investigate this further by controlling for these measures and other variables, such as familiarity or preference (Pereira et al., 2011).

A further limitation was that control variables (musicality, musicianship, years of experience, importance of music in one's life, familiarity, age, and gender) accounted for a greater proportion of the variance in Experiment 2 than in Experiment 1. This may be because Experiment 2 reflected a more diverse population, but it may also be because participants recruited through Amazon Mechanical Turk had less of an investment or interest in the study (given that most participants in Experiment 1 were more motivated psychology

students) which could affect generalizability. Participants in Experiment 2 may have relied less on moment-to-moment emotional intensity and instead more on their schemas surrounding music and their familiarity with the pieces. This is supported by the fact that familiarity was a stronger predictor in Experiment 2 than in Experiment 1, and in one case was even stronger than the *average*. The degree to which audiences may rely on pre-existing ideas of music (rather than moment-to-moment experiences) depending on audience investment in the experience may offer promising avenues for marketers of music to consider, such as when promoting music on the radio or in public spaces.

To conclude, this study has demonstrated several aspects novel to our understanding of retrospective evaluations of musical experiences. Individuals segment musical experiences presented as a collection of musical pieces and are affected by positive and negative trends of affect intensity. Crucially, even when controlling for duration, instrumentation, and genre, the peak-end does not reliably or consistently influence global evaluations of the emotional intensity of musical experiences. This finding offers an opportunity to examine the possibility that memorability plays a role. Future research into these areas may prove promising in the endeavor to understand the relationship between momentary affect and retrospective evaluations of affective experiences.

## References

Albert, S., & Bell, G. G. (2002). Timing and Music. *The Academy of Management Review*, *27*(4), 574-593. https://doi.org/10.5465/AMR.2002.7566078

Ali, S. O., & Peynircioglu, Z. (2010). Intensity of emotions conveyed and elicited by familiar and unfamiliar music. *Music Perception*, *27*(3), 177–182. https://doi.org/10.1525/mp.2010.27.3.177

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making*, *13*(2), 219–232. https://doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<219::AID-BDM331>3.0.CO;2-P

Baltazar, M., & Saarikallio, S. (2016). Toward a better understanding and conceptualization of affect self-regulation through music: A critical, integrative literature review. *Psychology of Music*, *44*(6), 1500–1521. https://doi.org/10.1177/0305735616663313

Do, A., Rupert, A., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*(1), 96–98. https://doi.org/10.3758/PBR.15.1.96

Dubé, L., & Le Bel, J. L. (2003). The content and structure of laypeople's concept of pleasure. *Cognition and Emotion*, *17*(2), 263-295. https://doi.org/10.1080/02699930302295

Duke, R. A., & Colprit, E. J. (2001). Summarizing listener perceptions over time. *Journal of Research in Music Education*, *49*(4), 330–342. https://doi.org/10.2307/3345616

Fasold, F., Memmert, D., & Unkelbach, C. (2015). A theory-based intervention to prevent calibration effects in serial sport performance evaluations. *Psychology of Sport and Exercise*, *18*, 47–52. https://doi.org/10.1016/j.psychsport.2015.01.001

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808

Hsee, C. K., Abelson, R. P., & Salovey, P. (1991). The relative weighting of position and velocity in satisfaction. *Psychological Science*, *2*(4), 263–267. https://doi.org/10.1111/j.1467-9280.1991.tb00146.x

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47–56. https://doi.org/10.1037/a0016873

Hui, S. K., Meyvis, T., & Assael, H. (2014). Analyzing moment-to-moment data using a Bayesian functional linear model: Application to TV show pilot testing. *Marketing Science*, *33*(2), 222–240. https://doi.org/10.1287/mksc.2013.0835

Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. Oxford University Press.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375-405. https://doi.org/10.1162/003355397555235

Kantor-Martynuska, J., & Horabik, J. (2015). Granularity of emotional responses to music: The effect of musical expertise. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(3), 235–247. https://doi.org/10.1037/a0039107

Kemp, S., Burt, C. D. B., & Furneaux, L. (2008). A test of the peak-end rule with extended autobiographical events. *Memory & Cognition*, *36*(1), 132–138. https://doi.org/10.3758/mc.36.1.132

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 115–129. https://doi.org/10.1037/1931-3896.2.2.115.

Kramer, J. D. (1988). *The time of music: New meanings, new temporalities, new listening strategies*. Schirmer Books.

Lee, Y. K., Lee, C. K., Lee, S. K., & Babin, B. J. (2008). Festivalscapes and patrons' emotions, satisfaction, and loyalty. *Journal of business research*, *61*(1), 56-64. https://doi.org/10.1016/j.jbusres.2006.05.009

Liu, Y., Liu, G., Wei, D., Li, Q., Yuan, G., Wu, S., Wang, G, & Zhao, X. (2018). Effects of musical tempo on musicians' and non-musicians' emotional experience when listening to music. *Frontiers in Psychology*, *9*, 2118. https://doi.org/10.3389/fpsyg.2018.02118

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, *102*(1), 108–134. https://doi.org/10.1348/000712610X506831

Meyer, L. B. (1956). *Emotion and meaning in music.* University of Chicago Press.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. https://doi.org/10.1037/h0045106

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, *39*(2), 283–90. https://doi.org/10.3758/BF03193159

Ochsner, K.N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129*(2), 242–261. https://doi.org/10.1037/0096-3445.129.2.242

Peine, K., Wentzel, D., & Herrmann, A. (2012). Getting better or getting worse? Consumer responses to decreasing, constant, and ascending multi-dimensional price profiles. *Review of Managerial Science*, *6*(1), 81–101. https://doi.org/10.1007/s11846-010-0050-5

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PloS One*, *6*(11), e27241. https://doi.org/10.1371/journal.pone.0027241

Rode, E., Rozin, P., & Durlach, P. (2007). Experienced and remembered pleasure for meals: Duration neglect but minimal peak, end (recency) or primacy effects. *Appetite*, *49*(1), 18–29. https://doi.org/10.1016/j.appet.2006.09.006

Rozin, A., Rozin, P., & Goldberg, E. (2004). The feeling of music past: How listeners remember musical affect. *Music Perception: An Interdisciplinary Journal*, *22*(1), 15–39. https://doi.org/10.1525/mp.2004.22.1.15

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE*, *4*(10), e7487. https://doi.org/10.1371/journal.pone.0007487

Schäfer, T., Zimmermann, D., & Sedlmeier, P. (2014). How we remember the emotional intensity of past musical experiences. *Frontiers in Psychology*, *5*(AUG), 911. https://doi.org/10.3389/fpsyg.2014.00911

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano Prelude. *Music Perception*, *19*(1), 87–120. https://doi.org/10.1525/mp.2001.19.1.87

Smith, J. K. (2014). *The museum effect: How museums, libraries, and cultural institutions educate and civilize society.* Rowman & Littlefield.

Stone, A. A., Broderick, J. E., Kaell, A. T., DelesPaul, P. A. E. G., & Porter, L. E. (2000). Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? *The Journal of Pain*, *1*(3), 212–217. https://doi.org/10.1054/jpai.2000.7568

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *The Neurosciences and Music III: Disorders and Plasticity: Annals of the New York Academy of Sciences*, *1169*, 209–213. https://doi.org/10.1111/j.1749-6632.2009.04864.x

Strijbosch, W., Mitas, O., van Gisbergen, M., Doicaru, M., Gelissen, J., & Bastiaansen, M. (2019). From experience to memory: On the robustness of the peak-and-end-rule for complex, heterogeneous experiences. *Frontiers in Psychology*, *10*, 1705. https://doi.org/10.3389/fpsyg.2019.01705

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*(3), 169–185. https://doi.org/10.1002/bdm.3960050303

Wagner, V., Klein, J., Hanich, J., Shah, M., Menninghaus, W., & Jacobsen, T. (2016). Anger framed: A field study on emotion, pleasure, and art. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 134–146. https://doi.org/10.1037/aca0000029

Watkins, M. J. (1972). Locus of the modality effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 644–648. https://doi.org/10.1016/S0022-5371(72)80048-3

**Table 1**

Pieces used in Experiment 1, along with abbreviated piece names, composer, performer, and length, and order played in the *High-Low* and *Low-High* conditions.

| Piece | Abbreviated piece name | Composer/ Performer | Length (min: sec) | Piece order in *High-Low* condition* | Piece order in *Low-High* condition |
|---|---|---|---|---|---|
| 12 Études d'exécution transcendante, S. 139: No. 2, Molto vivace | Liszt | Franz Liszt/ Daniil Trifonov | 2:06 | 1 | 6 |
| 12 Études, Op. 25: No. 12 in C minor | Chopin-Etude | Frederic Chopin/ Maurizio Pollini | 2:32 | 2 | 5 |
| Prelude in G-Sharp minor, Op. 32, No. 12 | Rachmaninoff | Sergei Rachmaninoff/ Vladimir Ashkenazy | 2:26 | 3 | 4 |
| Carnaval, Op. 9: I. Preambule | Schumann-Carnaval | Robert Schumann/ Varvara Tarasova | 2:29 | 4 | 3 |
| Kinderszenen, Op. 15, No. 13 | Schumann-Kinder | Robert Schumann/ Martha Argerich | 2:12 | 5 | 2 |
| Waltz for piano in E-Flat Major, KK IVa/14 (B. 46) | Chopin-Waltz | Frederic Chopin/ Alice Sara Ott | 2:33 | 6 | 1 |

*The position of the pieces themselves may influence how moment-to-moment ratings are integrated to form retrospective ratings for individual pieces. Having both *Low-High* and *High-Low* conditions, where the order of the pieces is reversed, should counterbalance any such effects.

**Table 2**

Pearson correlations between parameters and the global ratings of emotional intensity for individual pieces ($n$ = 378)

| Parameter | *Average* | *Beginning* | *End* | *Peak* | *Peak-end* | Enjoyment | Familiarity |
|---|---|---|---|---|---|---|---|
| Global rating | .829** | .567** | .776** | .684** | .806** | .523** | .090 |
| ***Average*** | | .694** | .834** | .775** | .883** | .429** | .062 |
| ***Beginning*** | | | .466** | .609** | .569** | .251** | -.015 |
| ***End*** | | | | .669** | .948** | .420** | .053 |
| ***Peak*** | | | | | .871** | .360** | .015 |
| ***Peak-end*** | | | | | | .432** | .041 |
| ***Enjoyment*** | | | | | | | .132* |

*Correlation is significant at the $p < .05$ level (2-tailed).

**Correlation is significant at the $p < .001$ level (2-tailed).

**Table 3**

Hierarchical regression coefficients of control parameters, the *average*, and the *peak-end* for the six pieces ($n$ = 63) for each step with corresponding $R^2$ changes

| Piece | Step | Parameter | $\beta_1$† | $p_1$ | $\beta_2$ | $p_2$ | $\beta_3$ | $p_3$ | $R^2$ change | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Liszt | 1 | Musicality | .110 | .415 | -.002 | .981 | .019 | .780 | | |
| | | Importance | -.035 | .787 | .097 | .181 | .107 | .116 | | |
| | | Musicianship | .182 | .404 | .113 | .342 | .027 | .811 | | |
| | | Experience | .156 | .490 | .077 | .528 | .157 | .182 | | |
| | | Gender | .044 | .757 | -.067 | .385 | -.037 | .612 | | |
| | | Age | -.013 | .932 | .018 | .824 | .016 | .834 | | |
| | | Familiarity | .131 | .328 | .165 | .025 | .133 | .055 | .141 | .141 |
| | 2 | *Average* | | | .815 | <.001 | .455 | .001 | .611*** | .753*** |
| | 3 | *Peak-end* | | | | | .412 | .004 | .037** | .789*** |
| Chopin-Etude†† | 1 | Musicality | -.096 | .483 | -.067 | .503 | -.041 | .662 | | |
| | | Importance | .075 | .565 | .184 | .059 | .106 | .264 | | |
| | | Musicianship | .249 | .252 | .083 | .603 | .081 | .585 | | |
| | | Experience | -.037 | .868 | .002 | .988 | .037 | .806 | | |
| | | Gender | .183 | .200 | .140 | .177 | .141 | .149 | | |
| | | Age | -.129 | .383 | -.109 | .312 | -.085 | .400 | | |
| | | Familiarity | .175 | .196 | .089 | .369 | .051 | .586 | .145 | .145 |
| | 2 | *Average* | | | .674 | <.001 | .228 | .209 | .411** | .556** |
| | 3 | *Peak-end** | | | | | .510 | .006 | .060* | .616** |
| Rachmaninoff | 1 | Musicality | -.225 | .081 | -.118 | .106 | -.094 | .191 | | |
| | | Importance | .164 | .184 | .192 | .007 | .220 | .002 | | |
| | | Musicianship | .649 | .002 | .374 | .002 | .375 | .002 | | |
| | | Experience | -.499 | .018 | -.223 | .064 | -.219 | .061 | | |
| | | Gender | -.117 | .381 | -.014 | .849 | -.024 | .741 | | |
| | | Age | .073 | .591 | -.073 | .347 | -.070 | .353 | | |
| | | Familiarity | .134 | .271 | .114 | .098 | .128 | .058 | .243* | .243* |
| | 2 | *Average* | | | .105 | <.001 | .492 | .001 | .521*** | .764*** |
| | 3 | *Peak-end* | | | | | .310 | .037 | .019* | .783*** |
| Schumann-Carnaval††† | 1 | Musicality | -.016 | .910 | -.094 | .284 | -.084 | .308 | | |
| | | Importance | -.040 | .768 | .104 | .227 | .132 | .107 | | |
| | | Musicianship | .397 | .081 | .130 | .362 | .174 | .199 | | |
| | | Experience | -.216 | .345 | -.057 | .691 | -.125 | .358 | | |
| | | Gender | -.097 | .506 | .027 | .767 | .022 | .795 | | |
| | | Age | .034 | .819 | -.071 | .444 | -.038 | .667 | | |
| | | Familiarity | .029 | .823 | .062 | .454 | .093 | .233 | .078 | .078 |
| | 2 | *Average* | | | .802 | <.001 | .446 | .004 | .572*** | .650*** |
| | 3 | *Peak-end** | | | | | .425 | .005 | .048** | .698*** |
| Schumann-Kinder | 1 | Musicality | -.080 | .562 | -.136 | .056 | -.159 | .025 | | |
| | | Importance | .067 | .617 | .036 | .593 | .032 | .634 | | |
| | | Musicianship | .163 | .464 | .019 | .868 | .060 | .596 | | |
| | | Experience | -.238 | .297 | .007 | .950 | -.022 | .851 | | |
| | | Gender | -.064 | .658 | .053 | .473 | .060 | .409 | .088 | .088 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Age | .229 | .127 | -.032 | .682 | -.024 | .750 | | |
| | | Familiarity | -.070 | .599 | -.063 | .351 | -.065 | .326 | | |
| | 2 | *Average* | | | .866 | <.001 | .591 | <.001 | .683*** | .772*** |
| | 3 | *Peak-end* | | | | | .301 | .058 | .015 | .787*** |
| Chopin-Waltz | 1 | Musicality | -.027 | .847 | .053 | .574 | .052 | .577 | | |
| | | Importance | -.018 | .896 | .047 | .604 | .035 | .708 | | |
| | | Musicianship | .360 | .132 | .073 | .647 | .056 | .735 | | |
| | | Experience | -.377 | .127 | -.121 | .460 | -.106 | .530 | | |
| | | Gender | -.115 | .443 | -.129 | .191 | -.116 | .263 | | |
| | | Age | .021 | .891 | .013 | .895 | .009 | .925 | | |
| | | Familiarity | .112 | .429 | -.022 | .815 | -.017 | .856 | .053 | .053 |
| | 2 | *Average* | | | .766 | <.001 | .705 | <.001 | .545*** | .599*** |
| | 3 | *Peak-end* | | | | | .075 | .652 | .002 | .600*** |

*Significant at the $p < .05$ level.

**Significant at the $p < .01$ level.

***Significant at the $p < .001$ level.

†Subscripts represent the *beta*-coefficient and *p*-value corresponding steps.

††For Chopin-Etude, the addition of the *Peak-end* was not significant, but the *End* was significant ($\beta = .485$, $p < .001$, $R^2$ change = .092, total $R^2 = .648$).

†††For Schumann-Carnival, the addition of the *Peak-end* was significant but the addition of the *End* resulted in a greater $R^2$ change ($\beta = .401$, $p = 0.003$, $R^2$ change = .056, total $R^2 = .706$).

**Table 4**
Hierarchical regression coefficients of control parameters, *average-global*, and *average-momentary* for the overall recital ($n$ = 63)

| Model | Step | Parameter | $\beta_1$[†] | $p_1$ | $\beta_2$ | $p_2$ | $\beta_3$ | $p_3$ | $R^2$ change | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 (assuming segmentation) | 1 | Musicality | .091 | .506 | .170 | .087 | .157 | .128 | | |
| | | Importance | .087 | .517 | -.008 | .934 | .010 | .921 | | |
| | | Musicianship | .412 | .063 | -.053 | .753 | -.039 | .819 | | |
| | | Experience | -.239 | .284 | -.001 | .996 | -.005 | .977 | | |
| | | Gender | -.038 | .789 | -.025 | .803 | -.024 | .812 | | |
| | | Age | .098 | .499 | .030 | .772 | .018 | .869 | | |
| | | Condition | -.246 | .071 | -.025 | .802 | -.015 | .885 | .124 | .124 |
| | 2 | *Average-Global* | | | .767 | .000 | .682 | .001 | .440* | .565* |
| | 3 | *Average-Momentary* | | | | | .098 | .614 | .002 | .567* |
| 2 (assuming cohesion) | 1 | Musicality | .091 | .506 | .062 | .563 | .157 | .128 | | |
| | | Importance | .087 | .517 | .139 | .193 | .010 | .921 | | |
| | | Musicianship | .412 | .063 | .156 | .381 | -.039 | .819 | | |
| | | Experience | -.239 | .284 | -.086 | .628 | -.005 | .977 | | |
| | | Gender | -.038 | .789 | -.021 | .850 | -.024 | .812 | | |
| | | Age | .098 | .499 | -.037 | .749 | .018 | .869 | | |
| | | Condition | -.246 | .071 | -.011 | .925 | -.015 | .885 | .124 | .124 |
| | 2 | *Average-Global* | | | .662 | .000 | .098 | .614 | .345* | .469* |
| | 3 | *Average-Momentary* | | | | | .682 | .001 | .097* | .567* |

*Significant at the $p < .001$ level

**Table 5**

Pearson correlations between control factors, the global rating of emotional intensity and enjoyment for the recital, and "listen-again" responses in Experiment 1 ($n$ = 63).

| Measure | Musicality | Importance | Musician-ship | Experience | Age | Gender | Enjoy-ment | Listen again |
|---|---|---|---|---|---|---|---|---|
| Emotional intensity | .046 | .021 | .200 | .070 | .019 | -.013 | .409** | .194 |
| Musicality | | -.056 | -.039 | -.109 | -.300* | .057 | -.098 | -.078 |
| Importance | | | .004 | .018 | .083 | .245 | .283* | .212 |
| Musician-ship | | | | N/A | .033 | -.025 | .300* | .312* |
| Experience | | | | | .103 | -.087 | .335** | .455** |
| Age | | | | | | .332** | .085 | -.076 |
| Gender | | | | | | | -.072 | -.077 |
| Enjoyment | | | | | | | | .507** |

*Correlation is significant at the $p < .05$ level (2-tailed).

**Correlation is significant at the $p < .01$ level (2-tailed).

**Table 6**

Pieces used in the study, along with abbreviated piece names for simplified referral, composer, performer, length, and order played in the *High-Low* and *Low-High* conditions

| Piece | Abbreviated piece name | Composer/ Performer | Length (min: sec) | Piece order in *High-Low* condition | Piece order in *Low-High* condition |
|---|---|---|---|---|---|
| ***12 Études, Op. 25: No. 12 in C minor*** | Chopin-Etude | Frederic Chopin/ Maurizio Pollini | 2:32 | 1 | 6 |
| ***12 Études d'exécution transcendante, S. 139: No. 2, Molto vivace*** | Liszt | Franz Liszt/ Daniil Trifonov | 2:06 | 2 | 5 |
| ***Waltz for piano in E-Flat Major, KK IVa/14 (B. 46)*** | Chopin-Waltz | Frederic Chopin/ Alice Sara Ott | 2:33 | 3 | 4 |
| ***33 Variations in C Major on a Waltz by Anton Diabelli, Op. 120: Variation XXIV: Fughetta (Andante)*** | Beethoven-Fast | Ludwig van Beethoven/ Maurizio Pollini | 2:50 | 4 | 3 |
| ***33 Variations in C Major on a Waltz by Anton Diabelli, Op. 120: Variation XX (Andante)*** | Beethoven-Slow | Ludwig van Beethoven/ Maurizio Pollini | 2:06 | 5 | 2 |
| ***Kinderszenen, Op. 15, No. 13*** | Schumann-Kinder | Robert Schumann/ Martha Argerich | 2:12 | 6 | 1 |

**Table** 7

Pearson correlations between parameters and the global rating of individual pieces ($n$ = 360)

| Parameter | Average | Beginning | End | Peak | Peak-end | Enjoyment | Familiarity |
|---|---|---|---|---|---|---|---|
| Global rating | .753** | .585** | .727** | .535** | .725** | .617** | .509** |
| Average | | .837** | .893** | .663** | .892** | .503** | .440** |
| Beginning | | | .669** | .497** | .669** | .400** | .446** |
| End | | | | .605** | .940** | .473** | .424** |
| Peak | | | | | .840** | .347** | .111* |
| Peak-end | | | | | | .470** | .337** |
| Enjoyment | | | | | | | .509** |

*Correlation is significant at the $p < .05$ level.

**Correlations are significant at the $p < .001$ level.

**Table 8**

Hierarchical regression coefficients of control parameters, the *average*, and the *peak-end* for the six pieces ($n$ = 60) for each step with corresponding $R^2$ changes

| Piece | Step | Parameter | $\beta_1$† | $p_1$ | $\beta_2$ | $p_2$ | $\beta_3$ | $p_3$ | $R^2$ change | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Chopin-Etude†† | 1 | Musicality | -.371 | .149 | -.107 | .670 | -.073 | .765 | | |
| | | Importance | .265 | .237 | .224 | .281 | .176 | .387 | | |
| | | Musicianship | -.098 | .705 | -.007 | .977 | -.093 | .700 | | |
| | | Experience | .322 | .109 | .104 | .596 | .104 | .588 | | |
| | | Gender | .174 | .234 | .063 | .648 | .056 | .679 | | |
| | | Age | .151 | .346 | .168 | .256 | .119 | .419 | | |
| | | Familiarity | .256 | .202 | .127 | .501 | .175 | .349 | .123 | .123 |
| | 2 | *Average* | | | .459 | .003 | .247 | .184 | .145** | .268* |
| | 3 | *Peak-end* | | | | | .308 | .075 | .045 | .313* |
| Liszt††† | 1 | Musicality | .188 | .463 | .224 | .344 | .271 | .253 | | |
| | | Importance | -.166 | .445 | -.094 | .643 | -.062 | .758 | | |
| | | Musicianship | -.016 | .949 | .093 | .698 | .076 | .749 | | |
| | | Experience | -.015 | .937 | -.089 | .624 | -.134 | .464 | | |
| | | Gender | .221 | .124 | .166 | .214 | .129 | .334 | | |
| | | Age | .398 | .013 | .237 | .126 | .187 | .230 | | |
| | | Familiarity | .163 | .423 | .090 | .635 | .163 | .402 | .153 | .153 |
| | 2 | *Average* | | | .426 | .003 | .194 | .344 | .137** | .290* |
| | 3 | *Peak-end** | | | | | .346 | .138 | .031 | .321* |
| Chopin-Waltz | 1 | Musicality | .189 | .397 | .270 | .121 | .274 | .118 | | |
| | | Importance | -.108 | .571 | .022 | .882 | .018 | .902 | | |
| | | Musicianship | -.008 | .971 | -.148 | .409 | -.148 | .412 | | |
| | | Experience | -.041 | .819 | -.026 | .849 | -.034 | .806 | | |
| | | Gender | .044 | .733 | -.023 | .819 | -.029 | .778 | | |
| | | Age | .117 | .416 | .093 | .398 | .069 | .573 | | |
| | | Familiarity | .504 | .006 | .200 | .176 | .205 | .169 | .294** | .294** |
| | 2 | *Average* | | | .671 | <.001 | .572 | .016 | .295*** | .589*** |
| | 3 | *Peak-end* | | | | | .105 | .625 | .002 | .591*** |
| Beethoven-Fast | 1 | Musicality | .053 | .804 | -.008 | .967 | -.004 | .985 | | |
| | | Importance | -.082 | .650 | -.049 | .758 | -.056 | .726 | | |
| | | Musicianship | .281 | .224 | .192 | .343 | .191 | .349 | | |
| | | Experience | -.180 | .300 | -.109 | .475 | -.098 | .529 | | |
| | | Gender | .048 | .693 | -.045 | .679 | -.042 | .700 | | |
| | | Age | .029 | .827 | .109 | .362 | .126 | .316 | | |
| | | Familiarity | .428 | .023 | .248 | .138 | .226 | .195 | .374*** | .374*** |
| | 2 | *Average* | | | .525 | <.001 | .654 | .031 | .160*** | .534*** |
| | 3 | *Peak-end** | | | | | -.124 | .631 | .002 | .536*** |
| Beethoven-Slow†††† | 1 | Musicality | -.100 | .526 | .017 | .909 | .027 | .851 | | |
| | | Importance | .026 | .848 | -.072 | .559 | -.072 | .561 | | |
| | | Musicianship | .299 | .097 | .193 | .237 | .205 | .213 | | |
| | | Experience | -.002 | .989 | .020 | .864 | .022 | .850 | .671*** | .671*** |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Gender | .010 | .909 | -.022 | .781 | -.027 | .736 | | |
| | | Age | -.037 | .706 | .051 | .571 | .039 | .675 | | |
| | | Familiarity | .609 | <.001 | .366 | .014 | .329 | .035 | | |
| | 2 | *Average* | | | .430 | .001 | .270 | .242 | .070*** | .740*** |
| | 3 | *Peak-end* | | | | | .181 | .419 | .003 | .744*** |
| Schumann-Kinder | 1 | Musicality | .076 | .600 | .078 | .472 | .101 | .345 | | |
| | | Importance | .089 | .468 | .046 | .615 | .019 | .831 | | |
| | | Musicianship | .146 | .339 | -.087 | .471 | -.120 | .314 | | |
| | | Experience | -.050 | .659 | .042 | .626 | .061 | .474 | | |
| | | Gender | .039 | .635 | .020 | .742 | -.006 | .928 | | |
| | | Age | -.105 | .250 | -.061 | .371 | -.094 | .177 | | |
| | | Familiarity | .628 | <.001 | .335 | .002 | .284 | .008 | .716*** | .716*** |
| | 2 | *Average* | | | .601 | <.001 | .364 | .028 | .128*** | .844*** |
| | 3 | *Peak-end* | | | | | .301 | .080 | .009 | .853*** |

*Significant at the $p < .05$ level.

**Significant at the $p < .01$ level.

***Significant at the $p < .001$ level.

†Subscripts represent the *beta*-coefficient and *p*-value corresponding steps.

††For Chopin-Etude, the *peak* was the single best predictor: when entered in Step 2, $\beta = .576$, $p = .002$, $R^2$ change = .264, total $R^2 = .387$. No other parameter significantly improved the prediction when added in Step 3.

†††For Liszt, the *peak-end* was the single best predictor: when entered in Step 2, $\beta = .511$, $p = .001$, $R^2$ change = .156, total $R^2 = .309$. No other parameter significantly improved the prediction when added in Step 3.

††††For Beethoven-Slow, familiarity was the single best predictor, accounting for 62.7% of the variance ($\beta = .792$, $p > .001$).

**Table 9**

Hierarchical regression coefficients of control parameters, *average-global*, and *average-momentary* for the overall recital ($n$ = 60)

| Model | Step | Parameter | $\beta_1$ † | $p_1$ | $\beta_2$ | $p_2$ | $\beta_3$ | $p_3$ | $R^2$ change | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 (assuming segmentation) | 1 | Musicality | .411 | .064 | .220 | .140 | .236 | .116 | | |
| | | Importance | -.261 | .172 | -.142 | .268 | -.135 | .291 | | |
| | | Musicianship | .406 | .044 | -.198 | .196 | -.224 | .149 | | |
| | | Experience | -.365 | .029 | -.118 | .301 | -.112 | .323 | | |
| | | Gender | .092 | .487 | .000 | .999 | -.013 | .885 | | |
| | | Age | .076 | .607 | -.021 | .834 | -.009 | .929 | | |
| | | Condition | -.168 | .182 | .033 | .707 | .032 | .714 | .268* | .268* |
| | 2 | *Average-Global* | | | .867 | .000 | .783 | <.001 | .412** | .680** |
| | 3 | *Average-Momentary* | | | | | .131 | .299 | .007 | .687** |
| 2 (assuming cohesion) | 1 | Musicality | .411 | .064 | .399 | .037 | .236 | .116 | | |
| | | Importance | -.261 | .172 | -.182 | .268 | -.135 | .291 | | |
| | | Musicianship | .406 | .044 | .039 | .837 | -.224 | .149 | | |
| | | Experience | -.365 | .029 | -.238 | .100 | -.112 | .323 | | |
| | | Gender | .092 | .487 | -.003 | .979 | -.013 | .885 | | |
| | | Age | .076 | .607 | .087 | .492 | -.009 | .929 | | |
| | | Condition | -.168 | .182 | -.088 | .422 | .032 | .714 | .268* | .268* |
| | 2 | *Average-Global* | | | .571 | .000 | .131 | .299 | .203** | .471** |
| | 3 | *Average-Momentary* | | | | | .783 | <.001 | .216** | .687** |

*Significant at the $p < .05$ level

** Significant at the $p < .001$ level

**Table 10**

Pearson correlations between control factors, the global rating of emotional intensity and enjoyment for the recital, and "listen-again" responses in Experiment 2 ($n$ = 60)

| Measure | Musicality | Importance | Musician-ship | Experience | Age | Gender | Enjoy-ment | Listen again |
|---|---|---|---|---|---|---|---|---|
| Emotional intensity | .310* | .318 | .183 | -.046 | -.057 | .910 | .544** | .381** |
| Musicality | | -.056 | .734** | .432** | -.217 | .017 | . 299* | .339** |
| Importance | | | .288* | .236 | .006 | -.172 | .055 | .306* |
| Musician-ship | | | | N/A | -.446* | .224 | . 310* | .493* |
| Experience | | | | | -.139 | .103 | .013 | .317* |
| Age | | | | | | -.376** | -.206 | -.193 |
| Gender | | | | | | | .175 | .274* |
| Enjoyment | | | | | | | | .324* |

*Correlation is significant at the $p < .05$ level (2-tailed).

**Correlation is significant at the $p < .01$ level (2-tailed).

**Figure 1**

*Contours of Moment-To-Moment Emotional Intensity Ratings for the Six Pieces*

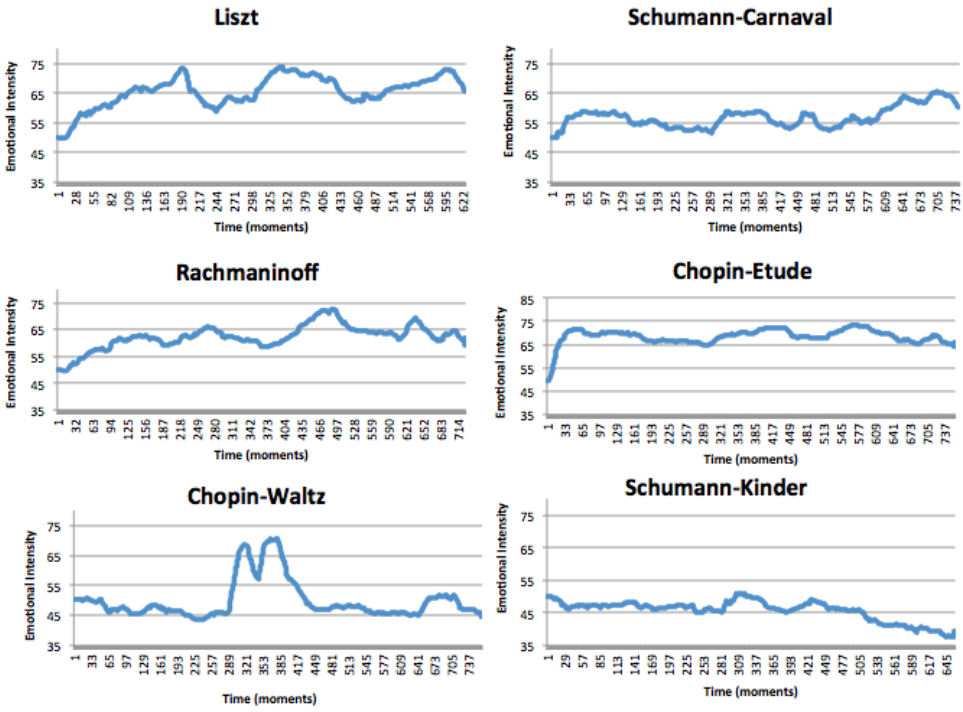
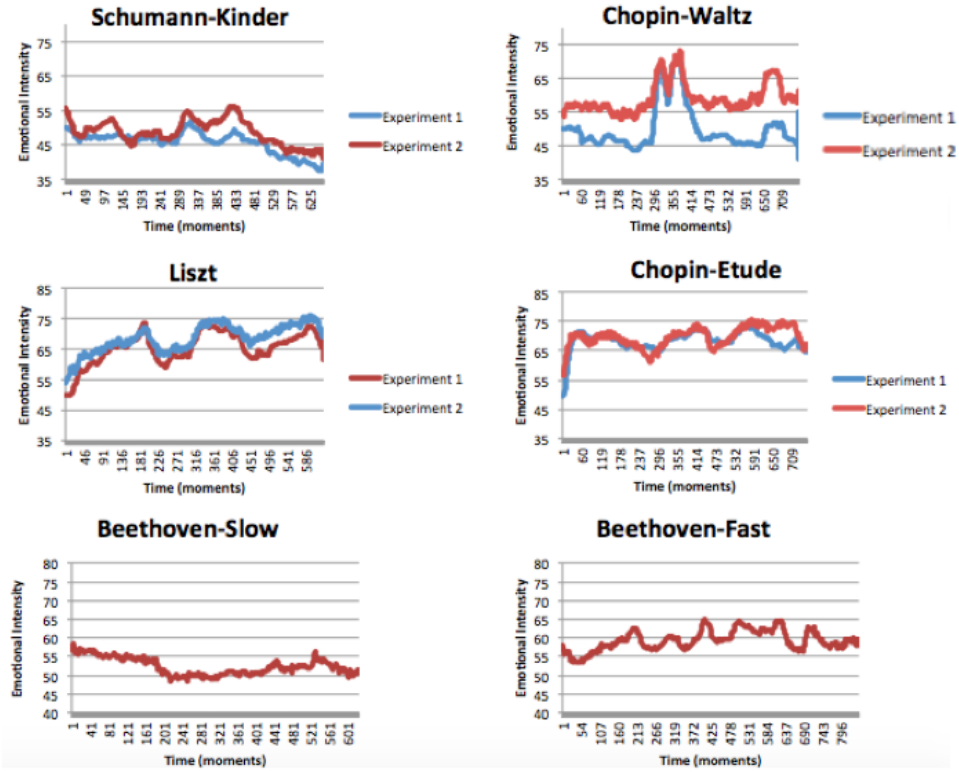


**Figure 2**

*Contours of Moment-To-Moment Emotional Intensity for the Individual Pieces*




Note. For pieces that were repeated from Experiment 1, contours from Experiment 1 are displayed in the same graphs.