Expressive Meaning and the Empirical Analysis of Musical Gesture: The Progressive Exposure Method and the Second Movement of Beethoven's *Pathétique* Sonata [*]

Albrecht, Joshua D.

ABSTRACT: Many investigations of the expressive meaning of musical works rely only on the musical interpretations and intuitions of the author. While invaluable, theorists' analyses are often biased or contradict one another. This paper presents a novel empirical approach to analyzing musical expression, in which the interpretations of individual theorists are balanced with listener reception in a broader audience, in this case a group of 110 music students from two universities. This new paradigm, which I have termed "the progressive exposure method," presents a larger excerpt in shorter discrete segments. An exploratory case study illustrates the progressive exposure method through an analysis of the expressive meaning of the second movement of Beethoven's *Pathétique* Sonata. When the results are amalgamated, a diachronic portrait emerges of cognitively complex emotions blended together as they unfold throughout the movement. This article provides readers with a hands-on, interactive tool for examining all of the results of the study. By presenting short musical gestures to listeners, a bottom-up, data-driven analysis of the expressive meaning of musical gestures and topics in the movement is possible. The consequent analytical results intersect in unique ways with more traditional theoretical and analytical practices, illustrating original applications of empirical methods to existing theories of musical expression as a means of providing converging evidence for those theories. Specifically, the results of this intersubjective analysis are discussed in light of theories of musical meaning by Hatten, Meyer, Narmour, Huron, and Margulis, and the results provide a new opportunity to directly and empirically testing a number of these authors' hypotheses.

A Sample Expressive Analysis and the Problem of Intersubjectivity

[1.1] In his review of recent developments in topic theory, Nicholas McKay (2007, 159) states that,

"formal analysis is, by its very nature, disinclined towards—if not incapable of—reading expressive intentions encoded in music's gestures or voices." While the point may be overstated, it is certainly the case that in the era of music theory as an academic discipline most formal analysis has not focused on decoding the expressive meanings of the music it analyzes. Why should this be? One of music's most immediate characteristics is its ability to express emotion, an attribute of music recognized by the entire spectrum of listeners, whether musically trained or not. Juslin and Laukka (2004), in a survey of 141 listeners, report that a remarkable 100% of respondents believed that music can express emotion, with 76% of listeners reporting that they believed music expressed emotions often. Similarly, one of the most commonly stated goals professional musicians have in pursuing a career in music is to provide "a means to generate positive emotional experiences mostly for one's own satisfaction" (Persson 2001, 277). Not only are analysts who ignore the expressive meaning of music in danger of misrepresenting the experience of listening to music and so doing a disservice to their readers, but they also are in danger of telling incomplete stories about the music they analyze.

[1.2] Emotional expression is not a central goal for all styles of music. For some music, however, to ignore its expressive meaning is to miss a big part of what makes it work. Consider, for example, Beethoven's *Pathétique* Sonata, a work that stands out from Beethoven's other early piano sonatas, both for being a uniquely powerful example of musical *pathos* and for formal and harmonic innovation.[1] The expressive power of the sonata was immediately recognized by its contemporaries and it has remained a perennial favorite. **Example 1** shows mm. 29–44 from the second movement, including the rondo's second refrain, the beginning of the second episode, and a modulation to the key of the episode's submediant. What might an analysis of the expressive meaning of this excerpt look like?

[1.3] First, I hear the refrain as a lovely, bittersweet melody supported by a gently rocking accompaniment. Set in an unchanging *piano* dynamic and centered around middle C, the melody is serene and contemplative on the surface. I find that one of the most noteworthy elements of the melody is the persistent use of cross-barline leaps. Beyond simply suggesting compound melody, the leaps seem to me to carry poignant expressive overtones, regularly opening up sudden space, both literally and expressively. As one example, the tritone descending leap of m. 35 rends the otherwise calm final descent of the refrain with a discordant note of longing, capped off by a delayed resolution on the downbeat of m. 37. The hint of unrest signaled in the refrain surfaces instantly in the second episode. The calm A♭ major is immediately displaced by the marked parallel minor, while the rhythm also immediately intensifies by replacing the gentle rocking accompaniment with agitated repeated triplet-sixteenth notes. Moreover, the rich texture thins out and is coupled with a higher tessitura, marked with a drop to an even quieter *pianissimo*, the combination of which signifies to my ear that the now explicit grief is being forcefully bottled up.

[1.4] Although the anguished melodic tritone reappears in mm. 38, this time it is answered by a distinctly new voice with a different character counterpointed agaiperfect nst it in the bass—one that almost bounces in a staccato-like way with a playful chromatic-lower neighbor before counteracting the soprano's upward tritone leap with an upward fourth. The right hand's plaintive melody sounds again, but this time without the upward tritone leap, followed by a modified answer in the left hand. The transformation of the melody is coupled with a thickening of the texture, punctuated with dramatic *szforzandi*, in what I hear as a victorious transformation. While the agitated triplets continue, the harmonies outline a string of applied dominant seventh chords, resulting in a strong modulation to the submediant (notated as "E major" in the score). Cumulatively, I hear this episode as expressing a quiet internalized grief being discovered and shared by a second voice and then transformed through that experience into a glorious resolution. The exultation of the cadential arrival is, however, colored by the agitation of the muddy texture and insistent triplet rhythm.

[1.5] What is the value of this kind of expressive analysis of the music's structures? Although often written with feints toward objective language, it is important to remember that expressive analyses typically reflect one person's interpretation of the music's expressive meaning. There are always many ways one might analyze a particular musical piece, even with more "objective" elements of musical structure, and it is not hard to find a spectrum of perspectives and strong disagreements

about the same music. Even if there is nearly universal agreement that this music is emotionally expressive, one might imagine the problem of analytical consensus would be even greater when discussing what exactly the expressive meaning of this music is. After all, what right do I have to make claims about how others might hear the expressive meaning of this piece?

[1.6] Ultimately, any analysis should serve the music, both capturing aspects of how the reader already hears the music and opening up viable new paths for understanding it. One useful metric for evaluating the contribution of an analysis is the degree to which it is consistent with a communally defined sense of musical meaning, either retrospectively or prospectively (or both). Appealing to a communally defined sense of meaning has been the approach of most recent analytical approaches to expressive meaning, evident in those writers loosely comprising topic theory and narrative theory. The typical approach is to argue that the interpretations given are essentially intersubjective in nature, meaning that *historically informed* audiences would generally perceive the meanings outlined in the analysis.

[1.7] For example, Hatten (2004, 6) defends his analyses of the expressive meaning of musical gestures and topics by saying that, "in the face of postmodernism, I maintain that the 'aesthetic' is no illusion. . . . I maintain that we still have access to *relatively objective* (by which I mean *intersubjectively defensible*) historical meanings." Allanbrook (1986, 2–3) also argues that analysts can identify specific expressive meanings of musical structures shared by typical listeners: "[Composers were] in possession of something we can call an expressive vocabulary . . . [that] provides a tool for analysis which can mediate between the [works] and our individual responses to them. . . . By recognizing a characteristic style, [the analyst] can identify a configuration of notes and rhythms as having a particular stance. . . . In short he [*sic*] can articulate within certain limits the shared response a particular passage will evoke." This tactic represents the most common strategy, either explicit or implicit, in defending assessments of the emotional meanings of the identified gestures are intersubjectively defensible, or broadly agreed upon by listeners competent in the style.

[1.8] Appeals to intersubjective meaning notwithstanding, expressive analyses are typically still the product of one person generalizing their impressions of the music's expressive meanings to some vague notion of "intersubjectivity." However convincing an analysis is, it is inherently problematic for a single writer to claim that their views reflect a truly intersubjective understanding of a work. By definition, any "intersubjective" analysis is going to involve a group of subjects and a common stimulus. When contemporary music theorists make intersubjective claims, they tend to be arguing that their analysis approximates the experience—or potential experience—of some group of contemporary musically sophisticated listeners.

[1.9] In this article, I will treat the question of intersubjective evaluations of expressive meaning of music directly. Rather than trying to argue abstractly that my expressive reading of the second movement of Beethoven's *Pathétique* Sonata must reflect some notion of intersubjective perception, I will create an analysis that is literally inter-subject generated. My approach will approximate the community of "musically sophisticated listeners" by recruiting over one-hundred undergraduate music majors. By collecting the pertinent data about how listeners perceive the emotional meaning of this work, I can sidestep the philosophically difficult task of logically defending my perceptions as reflective of a broad community and instead let the data generate a bottom-up picture of intersubjective expressive meaning in the movement.

[1.10] My method will be centered around a novel approach I call the *progressive exposure method*, in which listeners progressively hear and rate the emotional expression of short excerpts from the movement. Although my interest is in examining the expressive meaning of musical gestures, my methodology attempts to remain neutral by minimizing assumptions about what constitutes relevant gestures before analysis. Automatically slicing the recording up into short segments allows both for the relative isolation of musical gestures, and for the data themselves to suggest what the relevant gestures are in a bottom-up manner. I have chosen to present the excerpts in random order in this study to eliminate maturation and fatigue effects. Data collected are then amalgamated into a diachronic portrait, in which individual observations are stitched together to build one narrative analysis tracing unfolding emotions over the course of the movement. The literal multi-subject analysis is then compared to the above analysis, along with other analyses

produced by several contemporary theorists.

[1.11] Intersubjectively analyzing the movement using an empirical methodology also provides new opportunities for testing longstanding theories of musical expression.[2] Many of these theoretical claims take the form of hypotheses that could, in principle, be tested empirically. Relatively recent developments in statistics, methodology, and computing now make it possible to test claims that have historically remained untested. At the same time, it is important to remember that empirical tests do not essentialize theoretical claims, as if some empirical test of data in an isolated study could somehow be said to "verify" or "refute" a theory. Collecting a database of expressive evaluations of the movement, however, provides the ability to put theoretical ideas in dialogue with listener perceptions in an empirical domain to look for converging evidence. I will use the data to directly test a number of hypotheses from the theories of Meyer (1956), Hatten (1994, 2004), Narmour (1990, 1992), Huron (2006), and Margulis (2013). A number of hypotheses from these theories will be directly tested.

Methodological considerations

[2.1] Before exploring the method and results of the study, I would like to examine some of the methodological issues underlying any empirical investigation of expressive meaning in music.[3] One question revolves around the distinction between emotions *induced,* or personally felt in the listener in response to the music, and emotions *perceived* by the listener, or recognized as expressed by the music. In many circumstances, we might imagine these emotions are fairly similar. However, this may not always be the case. For example, a listener might perceive a particular song as expressing happiness, but if the song is associated with negative autobiographical memories (perhaps connected to a failed relationship), it may evoke sadness in the listener. It is even possible that the distance between perceived and felt emotion may influence the degree to which music is heard as ironic or sarcastic.

[2.2] Although both induced and perceived emotion are important elements of musical expression, there are significant differences between them.[4] For the purposes of this study, therefore, I will focus on perceived emotion.[5] This is not to deny that felt emotion is an important component of the experience and likely influences perceived emotion. Rather, this is a pragmatic consideration; my goal is to try as much as possible to isolate the question under study and avoid skewing the results with other confounding factors. Also, because the study was conducted in a laboratory environment, it is likely that listeners' felt emotions in response to music were significantly reduced.

[2.3] For this study, I developed a new measurement paradigm called the *progressive exposure method*. In this paradigm, a long excerpt is divided into discrete segments. The segments used in the progressive exposure method can be of any duration deemed appropriate by the researcher; segment lengths should be long enough for participants to be able to make a judgment about the emotion of the segment, but short enough to not host several dramatic changes of emotion. For this study, I was interested in the expressive implications of short musical gestures, and how various different musical parameters are combined in different segments to create emergent meaning.[6] For the second movement of Beethoven's *Pathétique*, five-second segments seemed to be about the right length to capture most gestures, including a little bit of musical context, without incorporating too many different gestures in the same segment.

[2.4] Another relevant concern for empirical studies that involve listening to music is the selection of a recording. There are a large number of commercially available recordings of the *Pathétique* Sonata, raising the question of which recording to select. While the interpretive decisions performers make in encoding emotional expression is a worthwhile topic of study, recall that the focus of this study is on listeners' perception of expressive meaning.[7] One way to normalize the effect of performance would be to use a computer to generate a MIDI performance. However, the result would be a mechanical realization of the work with greatly reduced expressive potential. A better approach would be to select an interpretation that is not "extreme" in its realization; a recording that reflects the interpretive decisions of a large proportion of the performers of a work might represent a sort of "mean" of the "population" of recordings.[8] To choose a recording that

was not too extreme in interpretation, two trained researchers were given instructions to independently listen to each of the recordings of the movement on compact disc in music library of the Ohio State University, a total of 23 recordings, and select some that seemed average in its expressive decisions. Once these selections were made, the researchers came to a mutual conclusion about which was most average.

[2.5] The selected performance was by John O'Conor and appeared on the album *Breathe®: relaxing piano for lovers* (O'Conor 2008). As an aside, it is noteworthy that the most middle-of-the-road performance would be from a CD compilation of the type that proliferated during the first decade of the twenty-first century, in which the most marketed feature of the music is not the composer or performer, but on how useful the music is for creating some sort of non-musical social effect. The length of the recording was 4'47". As mentioned above, this recording was divided into five-second segments for use with the progressive exposure method. To avoid abrupt onsets and offsets, a 500 ms fade-in and fade-out were applied to each segment, and to mitigate the effect of arbitrary boundaries, two collections of segments were generated. The first used a 0" offset and the second used a 2.5" offset. Even though the segments were 5" in length, this approach allows for a dovetailed 2.5" resolution of data, providing a more detailed examination of how emotion unfolds throughout the movement, and allowing every 2.5" of the movement to be heard in two different excerpts.

[2.6] Finally, any empirical study of musical emotion must settle on a choice of emotion categories to examine. Empirical music emotion studies tend to use one of only a few approaches to selecting emotion terms.[9] Rather than using any of the existing approaches wholesale, or relying on my own intuitions of musical emotion categories and risk jeopardizing the intersubjective nature of my investigation, I opted to use an empirical approach to developing a list of discrete emotion terms. I recruited four doctoral students in music theory and music composition and one applied music professor from Ohio State University's music department to listen to a selection of five-second excerpts from the movement and freely describe what emotions they thought the passage expressed. To encourage the participants to speak at length, they spoke freely while I recorded their comments on a laptop. The free responses of the five participants were divided into discrete comments, resulting in 592 discrete comments. These comments were subjected to an informal content analysis by two independent researchers, who grouped comments together that were similar and provided each group with a label descriptive of the contents of the group. Categories with similar labels between the two researchers were combined, and categories with similar content but different labels were amalgamated. A number of additional categories were discarded, because they either had small representation or the comments referred more to musical structures than to emotion. The resulting fifteen emotion categories are shown in **Table 1**, along with how many discrete comments were considered to fit into each category. Several comments could not easily be grouped, but the table reflects 453 out of 592 comments (77%). While many terms common to music and emotion studies were included, such as happy/joyful, sad/depressed/tragic, or calm/serene, there were also a number of terms that are less commonly tested. Dark, weighty, carefree, and striving/yearning are all infrequently used terms but were deemed to be reflective of the expressive content of the movement. These fifteen terms were used in the analysis of expressive meaning in the movement.

Empirical Study of Expressive Meaning

Some Preliminary Detail

[3.1] Using these fifteen emotion terms (Table 1), an intersubjective empirical study was conducted using the progressive exposure method to capture a diachronic portrait of perceived emotional expression in the second movement of the *Pathétique* Sonata. I recruited 110 participants, 51 undergraduate music majors from the participant pool at Ohio State University, and 59 undergraduate music majors from Westminster Choir College. The mean age for participants was 22.1 years (standard deviation [sd] = 6.4), and the mean number of years of musical training was 14.3 (sd = 6.2). The recordings used were 112 five-second segments in two offset groups of 56 each (0" and 2.5") spanning the entire duration of the 4'47" John O'Conor recording. Participants were

assigned randomly to one of the two offset groups.

[3.2] Participants listened to all 56 five-second segments from their offset group in random order and rated the extent to which each excerpt "conveyed or expressed" each of the fifteen emotion categories.[10] Additionally, participants heard five randomly selected segments a second time as a test for within-participant reliability, resulting in 61 total excerpts per participant. Unfortunately, the task was too long for one participant, so the fifteen emotion categories were divided into five groups of three emotions each. Participants were randomly assigned to one of the five emotion groups. An example of the interface is given in **Example 2**.

Intersubjective reliability

[4.1] In order to evaluate the expressive ratings of the various participants as a group of listeners, some way of assessing all of the responses as roughly equivalent is needed. This is typically accomplished through averaging responses. One criticism of this kind of empirical analysis of expressive meaning might be that listeners experience music in highly personalized ways, and so it is overly reductionistic or misleading to average results together, or that personal judgments are themselves unreliable. Before averaging together the intersubjective data gathered from the affective analysis, it is helpful to ask the more fundamental question: To what extent does a community of listeners actually agree about the expressive meaning of this movement? To test the degree to which participants were personally consistent (intra-participant reliability) and the degree to which all participants provided similar responses (inter-participant reliability), a fairly in-depth analysis of reliability metrics was conducted on the responses. To summarize the findings in the Appendix, although there was naturally some variation in the responses, participants were widely consistent in evaluating the expressive meaning of the segments, lending confidence to considering the responses collected as part of a shared response from a community of listeners. A full description of the results from the reliability tests can be found in the Appendix.

[4.2] However, not all expressive categories were equally reliable. As a result of this assessment, four emotion categories were discarded from further analysis. In the first case, cheeky/sassy was deemed not appropriate for this movement.[11] Secondly, sincerity/truthful, important/serious, and emotional/moody all provided low measures of within- and between-participant reliability. Post-experiment interviews suggested that these three compound affective categories were either not well defined and confusing to participants, or combined two different emotion categories that resulted in different strategies from different participants. Moreover, many of these categories were expressively neutral, and so might have been treated differently by different participants.[12] Again, further details can be found in the Appendix.

Sample expressive analysis

[5.1] In order to examine how the community of participants understood the expressive meaning of the movement, an intersubjective analysis of the movement was conducted. Individual responses were amalgamated and averaged to form a diachronic portrait depicting ways in which each emotion unfolds over the course of the movement. For the purposes of analysis, the data were normalized across subject-scale, by measuring responses in terms of standard deviations away from the mean for that emotional category for that participant. This decision was made to compensate for differences in the ways that different participants might use the scale, and for differences in the ways the scale is used for each category.[13] High ratings for an emotion are not absolutely high, but high in relation to other segments for that emotion.

[5.2] The entire set of participant evaluation data in the study is accessible through the provided interactive tool. A static image corresponding to the following discussion, for those who do not wish to use the tool, is provided in **Example 3**, which combines elements from the single emotion and multiple emotions explorer.[14] To investigate individual emotion ratings, select the "single emotion explorer" and choose the emotion category you are interested in. In this display, scores are represented as normalized, or as standard deviations from the mean, so 0 is the average rating for that emotion and +1 is one standard deviation higher than the average for that emotion. Each box represents ratings for the segment heard directly under the box, which can be heard by clicking on

the boxplot. The median of the responses for each segment is displayed as the line in the middle of the box. The box itself encompasses 50% of the responses, and the lines that extend beyond the box show the entire range of scores, with any statistical outliers shown as points outside of the lines. Connecting orange lines show the mean rating for each segment. To explore several emotion categories, click on the "multiple emotions explorer." Here, you can select whichever emotion categories you are interested in comparing. Connecting lines show the mean rating for each segment. Upon opening the interactive tool, the excerpt spanning mm. 34–44 is highlighted, which mirrors most of the excerpt reproduced in Example 1, although the range of the excerpt displayed can be controlled through changing the values of the slider. You can listen to any of the 5" segments by clicking on its data point on the graph, or the entire excerpt displayed by clicking on "Play Excerpt." To start, only six of the eleven emotion categories are displayed.

[5.3] One benefit of the progressive exposure method is the opportunity to track how the interleaved data connect to trace the development of any emotion category over the course of the excerpt. However, because the data are still discrete (rather than continuous), it is possible to examine exactly which surface features or musical gestures are correlated with these expressive changes. For example, notice that the ratings for unsettled/anxious show a local peak corresponding to the suspension on the downbeat of measure 36. When the suspension is resolved, unsettled/anxious ratings decrease, consistent with traditional notions of the emotional effect of suspension figures. The same localized peak occurs simultaneously for weighty, sad/depressed /tragic, and suspense/anticipation, not shown in this example.

[5.4] At measure 37, when the music moves directly to the parallel minor, the texture thins considerably, the tessitura moves higher, and the rhythm gets noticeably faster with driving and repetitive triplet-sixteenth notes, all of the positively valenced expressive categories decrease while ratings for dark and unsettled/anxious increase.[15] Notice, however, that lonely ratings do not significantly change at measure 37, indicating that these changes to the musical surface do not strongly influence that dimension, perhaps because other musical dimensions may counterbalance the effect of mode on lonely ratings. The musical gesture that *is* correlated with a change to lonely ratings, however, involves a second voice in a different range with different character (more staccato) entering the texture at measure 38 and again at measure 40.[16]

[5.5] Finally, consider the end of this passage. At this point, the texture thickens substantially as four-note chords low in the left hand produce a muddier texture. At the same time, the register opens up, the agitated rhythms continue, and the harmony shifts to the major mode through a string of applied harmonies. This move is highlighted by a series of punctuated *sforzandi*, accenting the contrast. Notice, consistent with traditional associations of mode, that happy/joyful and contentment ratings increase significantly with the modulation while lonely and dark ratings decrease. Tellingly, calm/serene ratings, which had been highly correlated with happy/joyful ratings to this point, continue to remain low while unsettled/anxious ratings remain relatively high. This can likely be explained by the persistence of the triplet-sixteenth notes and the loud *sforzandi*. Despite the modulation to the major mode, carefree ratings (not pictured in Example 3) also remain below average in this excerpt, consistent with a general trend throughout the movement for significantly lower carefree ratings for passages with triplet-sixteenth rhythms rather than sixteenth notes.[17]

[5.6] The intersubjective expressive analysis for this excerpt interacts with my own intuitions about the movement (paragraphs [1.3-1.5]) in interesting ways. At a basic level, the data map squarely onto traditional expressive connotations of mode. The refrain and the modulation to the submediant elicits higher positive emotions and lower negative emotions than the beginning of the episode, and vice versa. The refrain exhibits high levels of contentment and calm/serene ratings, but also shows high ratings for lonely and dark, consistent with my perception of an uneasy, bittersweet mixture of emotions—a surface calm with darker undertones. However, my perception that the otherwise calm descent was interrupted by the descending tritone leap in m. 35 does not receive any support from the data. The explicit unrest of the episode, on the other hand, signaled by the increase of rhythmic motion and persistent repetitions is clearly reflected in the precipitous rise in unsettled/anxious ratings and drop in calm/serene ratings. Likewise, my perception of the new 'voice' entering the texture in m. 38 and 40 is mirrored by a corresponding drop in lonely

ratings. Another benefit of using this sort of approach is the way in which expressive categories can blend together to create subtle mixtures of emotional expression.[18] In addition to the complex bittersweet nature of the refrain, another potent example of complex emotional blending comes at the end of the excerpt (mm. 42–44). This passage was analyzed as expressing high levels of happy/joyful and calm/serene at the same time as it elicits high lonely ratings, perhaps suggesting a complex expressive mix that is happy but still unsettled—perhaps a sort of edgy, manic happiness.

Parallel passages

[6.1] This movement of the *Pathétique* provides a unique opportunity to investigate the effect of parallel musical passages with a fairly high degree of ecological validity.[19] While the progressive exposure method is not like a real listening environment in its presentation of unordered short segments, it nevertheless presents a sort of imperfect middle ground between the two goals of experimental control and ecological validity. In this rondo movement, the theme returns four times after the initial presentation. However, each time the theme returns, there are slight variations. While almost all of the musical parameters remain the same (melodic contour, harmonic progression, rhythmic and melodic accent patterns, etc.), slight compositional differences permit the analysis of the effect of small changes to one or two parameters in real musical situations. For example, the second statement of the theme is the same as the first with only minor exceptions: the melody is transposed up one octave and there is a slightly thicker texture with the addition of an extra inner voice. The fourth statement of the theme returns to the original melodic tessitura, but uses the faster sixteenth-triplet accompaniment. In the final statement of the theme, the higher register, thicker texture, and faster accompaniment are combined. By collapsing all ratings for any given emotional dimension across each statement, direct comparisons can be made between the refrains and hypotheses can be tested.

[6.2] The third statement of the theme even offers a sort of "control group" of segments by repeating the first statement exactly. In a real listening environment, repetition is always musically significant, even if repeated material is exactly the same as earlier material.[20] In this study, however, each five-second segment is played in random order. The result of random ordering is that participants have little ability to tell from where in the movement a given segment is derived. Therefore, the exact repetition of the original theme serves as a sort of control group, and the expectation is that there will be no statistically significant differences between ratings for the first and third statements of the theme.

[6.3] In her book on repetition in music, Margulis (2013, 173) wonders: "Are repetitions of the theme in a rondo, for example, expressively cumulative, such that later iterations assume a memory not only for the theme, but also for its performance the first time around?" Because this study's methodology explicitly scrambles segment order, there is an interesting way of approaching this question. Ordering can refer to three different things: 1) The way that Beethoven encoded ordering effects into his composition; e.g., refrain 5 is clearly composed differently than refrain 1. 2) The way that O'Conor encoded ordering effects into his performance; refrain 1 and refrain 3 are technically literal repeats compositionally, but because refrain 3 was performed after refrain 1, O'Conor may have performed with different types of microtiming and microdymanic subtleties. 3) The way that listeners hear the music sequentially; even were the movement to be performed precisely by a computer, we would expect refrain 3 to be heard slightly differently from refrain 1, in consequence of its position after episode 1. In most empirical studies, it is completely impossible to disentangle the effects of the above three ordering effects, because participants hear the music in real time. Consequently, it is difficult to know the extent to which participants' hearing of the music is influenced by temporality; it may be that listeners always perceive music as more intense emotionally because they expect a dramatic progression, or it may be that listeners always level off their emotional intensity due to the limitations of attention spans. Because excerpts are heard in random order, this study effectively eliminates listening order and tests only compositional and performative ordering. In a sense, we might say that the random ordering of the progressive exposure method allows for a more controlled examination of compositional and performative decisions of ordering.

[6.4] Other theorists' analyses of the emotional qualities of this movement can also be tested. For example, Hatten (1994, 207) claims this movement is associated with "assurance or reassurance" and Sisman (1994, 85) describes the movement as "consoling or healing." Both of these expressive characteristics are consistent with the development of expressive meaning over time—a "reassuring" or "healing" emotional quality in the movement might parallel a progression from higher ratings of negatively valenced emotion and lower ratings of positive-valenced emotions in the beginning of the movement to the opposite at the end of the movement. Because of the methodology's use of random ordering, then, we can test a hypothesis in which cumulative effects of compositional decisions on perceived emotion produce decreases in negatively valenced affects and increases in positively valenced affects across the movement.

[6.5] For the purposes of discussion, I'd like to highlight just three expressive categories (**Example 4**), although means for all emotions by theme statement are provided in Table 3. First, there is a significant decrease in sad/depressed/tragic ratings over the movement, both between the first and fourth theme and the first and fifth theme ($p < .0001$ in both cases). Likewise, carefree ratings show a significant increase over the movement, between the first and second theme ($p = .01$) and the first and fifth theme ($p < .0001$). Notice that the effect of higher register on carefree ratings is not significantly different ($p = .89$) than the effect of faster rhythms. The effect of higher register and faster rhythms combined, however, evidenced in Theme 5, results in significantly higher carefree ratings. Weighty ratings for faster rhythms in Theme 4, however, are not significantly different ($p = .11$) than the original statement of the theme. On the other hand, higher register significantly lowers weighty ratings ($p < .0001$).

[6.6] Additionally, as can be seen in **Table 2**, there is a significant increase in happy/joyful ratings between Theme 1 and all three of the other themes ($p < .0001$ for all three), and there are significant decreases in dark ratings from the first to fourth and fifth themes ($p < .0001$ in both cases) and lonely ratings from the first to fourth and fifth themes $p < .0001$ for both). As expected, none of the eleven expressive categories display a significant difference between the first and third themes. These results are not only consistent with Margulis's implied hypothesis that there are cumulative expressive effects of theme restatements throughout a rondo movement, but they are also consistent with Sisman's and Hatten's interpretation of a process of reassuring, consolation, or healing over the course of the movement, evidenced by decreasing negatively valenced emotions and increasing positively valenced emotions over the course of the movement. On the other hand, calm/serene and contentment ratings are significantly lower and unsettled/anxious ratings are significantly higher at the end of the movement. These ratings are likely tied to increased rhythmic activity, but these changes do not seem to support Sisman's and Hatten's interpretations of the expressive character in the movement.

"Extreme" emotion ratings

[7.1] Since Meyer (1956, 1967) hypothesized that stronger emotions would be evoked from passages that are less predictable, musical expectation has played a central role in theories of how music communicates emotion.[21] By measuring participant responses as standard deviations from the average response, we can examine what we might call "extreme ratings," or ratings that are more than one standard deviation away from that emotion's mean. If the theory that deviations from strong musical expectations lead to increased emotional expression (Meyer 1956 and 1967; Narmour 1990 and 1992; Huron 2006), one would expect to see more extreme ratings in sections of the *Pathétique* in which the latent probabilities of continuation in the music are less clear, or where the actual continuation of the music is more surprising.

[7.2] In rondo movements, it is often the case that episodes are the most musically unpredictable and harmonically adventurous, whereas refrains tend to be more predictable.[22] Under this assumption, an application of the theories put forward by Meyer, Narmour, and Huron would hypothesize an uneven distribution of extreme ratings, such that episode and coda passages would elicit more extreme ratings (ratings more than one standard deviation from the mean) than refrain passages. This hypothesis is directly testable with the data collected here.

[7.3] For each segment, the number of extreme ratings, whether positive or negative, was tallied for

each expressive dimension and the segment was classified as either refrain or non-refrain (i.e., episode, coda, or retransition). In this movement, there are three refrain passages, two episodes each with a retransition, and one coda. The results from this tally are shown in **Table 3**. Counting all of the extreme ratings collapsed across categories, the three refrain passages resulted in 1,398 extreme ratings in 152.5 seconds of excerpts, or an average of 9.17 extreme ratings per second. In the two episodes and coda passage, there are a total of 2,297 extreme ratings in 140 seconds, a significantly higher average of 16.41 extreme ratings per second.[23] These results are consistent with the hypothesis that less predictable formal sections— in this case episode and coda sections in Rondo form— evoke stronger emotion ratings.

[7.4] Indeed, the passages with the most extreme ratings are the two retransitions, passages that serve to delay the expected resolution back to tonic and the main theme. This finding is also consistent with Meyer's theory (1956, 1967) that delaying expected resolution gives rise to increased emotional response. The passage with the highest density of extreme ratings in the movement happens during the first retransition, measures 20-29, viewable through the interactive data exploration tool. The measures in question, along with ratings for happy/joyful, carefree, contentment, unsettled/anxious, weighty, and dark are displayed immediately in **Example 5**, although other emotion categories and longer excerpts also be explored in the tool. The shown passage elicited 517 extreme ratings, averaging 14.77 extreme ratings per second, but the single segment that elicited the most extreme ratings in the entire movement (with 98), is the excerpt corresponding to measures 23–24. Although there is technically harmonic support in this excerpt, the dominant harmony is struck before the excerpt begins, and so without hearing the harmonic context attacked, listeners only hear a single melodic line. In this rendition, the performer adds a noticeable amount of rubato. The melodic line chromatically surrounds the dominant pitch in a low register at a low dynamic level, a moment of high tension before the expected resolution. This moment, more than perhaps any other in this rendition of the movement, exemplifies Meyer's (1967) first category of deviation from implication that enhances emotional arousal, in which an outcome that is clearly implied (tonic resolution from the dominant chord aligning with a return of the theme) is maximally delayed (through chromatic play around the dominant pitch emphasized with rubato).[24]

Modeling the expressive meaning of musical gestures

[8.1] Finally, the perceptual data collected in this study provide an opportunity to build a theory of musical gesture in this movement from the bottom up by correlating particular musical structures with perceived expressive meaning. Again, this is not a new concept, but although several theories have been proposed for how different musical gestures combine in an emergent way to take on complex expressive meaning, these ideas have received very little empirical testing. For example, Agawu (1991, 15) draws an analogy between language and music to illustrate that it is not specific structures of music that have fixed meanings, but rather the relationships between them.[25] Likewise, according Hatten (2004, 220), the meaning that adheres to music is not simply a matter of identifying topics and moving on, but rather it emergently arises out of the interplay and combination of topics.[26] For Hatten, various musical gestures do not have absolute meaning, but they have meanings that are contingent on how they are combined with other musical gestures. For example, falling melodic contours might be thought of as abnegation, relaxation, or several other meanings depending on other mitigating musical factors, like mode, dynamic, etc.[27] Again, it is the combination of particular musical gestures in the set that create emergent meaning rather than essentialized meanings: after all, quiet dynamics are not always calm and minor mode music is not always sad.[28]

[8.2] In order to engage with the above theories of emergent meaning arising from a set of musical gestures, the empirical data were analyzed in a bottom-up way to take into account not only many surface-level musical elements, but also the interactions between these elements. The statistical method of multiple regression provides an appropriate analytical tool for this purpose. Multiple regression works by looking for correlations between several predictor variables (musical features of segments) and the output (emotion ratings for the segments). In order to conduct the regression analysis, sixteen musical elements were analyzed for each five-second segment, which served as predictors of the participants' ratings. These sixteen predictor variables, common parameters

associated with musical expression, are shown in **Table 4**, along with the way these variables were encoded.[(29)]

[8.3] By using regression to analyze how the movement's musical gestures are related to inter-participant evaluations of perceived emotion, a model can be built for how each emotion is expressed through interactions between musical gestures in the movement. This model can be used to test previous theories of how musical gestures are used to convey emotion. For example, in various places Hatten identifies upward melodic direction with "yearning" and downward motion with "resignation" (1994, 57), harmonic dissonance (especially the diminished seventh) with anguish or grief (2004, 15), increased dissonance and turns to the minor mode with increasing agitation (2004, 16), and the pastoral topic's simplicity with slow harmonic rhythms, simple melodic contours, compound meter, and the major mode (2004, 56).

[8.4] The results of the regression model for each emotion dimension are presented in **Table 5**, showing the musical gestures that combined as a set to be significantly predictive of listener evaluations of emotional expression.[(30)] First, observe that many of the musical parameters are aligned with intuitive notions of how emotion is expressed in Romantic music. For example, sad/depressed/tragic is correlated with the minor mode, slow surface rhythms, thin textures, and lower pitches. Happy/joyful is correlated with the major mode, higher pitches, and faster rhythms. Less commonly tested emotions also correspond with intuitions. Weighty is correlated with lower pitches, more tendency tones, louder dynamics, and dense textures. Lonely is correlated with thin textures, slow rhythms, lower pitches, and the minor mode. Unsettled/anxious is correlated with staccato articulations, fast rhythms, the minor mode, dense harmonies, and crescendos. Relating these models to Hatten's gestural theories above is less straightforward, although several similarities tentatively seem to hold.[(31)]

[8.5] Second, in looking at the correlations in Table 5, it is important to remember that the musical parameters listed in the center column reflect emergent meanings from a complex interaction of musical parameters that are correlated with emotional expression for the movement, rather than absolute meanings. For example, although calm/serene is correlated with the major mode, it is correlated with the major mode in situations in which the music *also* includes legato articulations, faster surface rhythms, faster harmonic tempo, lower pitches, less dense textures, diminuendo, common melodic successions, and a lack of tendency tones. The correlation between calm/serene and the major mode breaks down when the major-mode excerpts use staccato articulations, thick textures, and chords with tendency tones, as in the "E major" cadence in the second episode (see Example 3). In this passage (mm. 43–44), although the cadence is unambiguously in a major key, calm/serene ratings are significantly lower than average, even *extremely* low (see "extreme" ratings section above). Therefore, it is important to remember that in large measure these models truly point to musical *gestures* that are composed of combinations of the musical parameters provided in the table.

Summary of results

[9.1] So, what is the value of *this kind* of expressive analysis of the music's structures, one that is empirical and intersubjective in nature? The first observation, which may perhaps come as something of a surprise, is that *in general* we can trust undergraduate students' understanding of expressive meaning in music, at least when taken as a large enough group. Though there are certainly individual differences between students, and though certainly some student responses proved unreliable (see Appendix), as a whole participant groups of undergraduate music majors converge on an intersubjective understanding of musical gestures that can be generalized to a broader community of listeners. These responses were robust enough to generate a narrative analysis of expressive meaning consistent with prior theorists' analyses, to detect expressive differences between distinct formal sections, to identify moments of heightened expression related to musical expectation, to significantly differentiate between parallel passages with subtle musical differences, and to build complex models of the emergent expressive meaning of musical gestures consisting of combinations of musical parameters. The robustness of the results suggests that this kind of analysis may be a profitable way to investigate expressive meaning in larger repertoires, and that the analyses of experts can be brought into dialogue with a more crowd-sourced

understanding of expressive meaning.

[9.2] A related point is that this empirical analytical approach allows for the opportunity to test important theories that have not been thoroughly tested. Of course, not all theories should be tested with intersubjective data of perceptions of expressive meaning. However, many of the theories examined in this paper have made claims about how listeners *in general* will hear music and think about the expressive implications of that music. If these theories capture meaningful aspects of music listening, then we should expect that collected data will be consistent with the outcomes hypothesized. In this article, the outcomes of this study were consistent with important theories proposed by Meyer (1956), Narmour (1990, 1992), and Huron (2006), who hypothesized that less predictable music should elicit stronger emotional reactions.

[9.3] Even theories that do not form explicit empirically testable hypotheses make claims about intersubjective perceptions of expressive meaning of musical gestures. By actually collecting the pertinent data, these claims can be tested to determine if communities of actual listeners hear the expressive implications of musical gestures suggested by these theories. For example, the progressive exposure method with random ordering provides a unique opportunity to isolate the differences between compositional and performative decisions between repeated passages. The results were consistent with an expressive progression from the beginning to the end of the movement, consistent with Sisman's and Hatten's interpretations of the movement. Also, the empirical results are largely consistent with topic theory or a theory of musical gesture, although caution should be advised in reading complex topical or tropological interpretations from broad statistical correlation. Finally, through a multiple regression analysis, a model can be built that explicitly defines correlations between specific musical gestures and perceived emotional expression in the analyzed music.

[9.4] I do not find it surprising that this paradigm provided results that were consistent with most of the prior theoretical claims tested. It is encouraging that empirical approaches provide converging evidence with theoretical approaches in building knowledge about music. But beyond simply bolstering prior claims, I find the new insights provided by this type of empirical investigation exciting. For example, beyond simply finding correlations between surface musical elements and perceived emotion consistent with prior ideas, the models in section 8 suggest new combinations of effects for further investigation. It is even possible that new avenues of expression might discovered. By examining emotions one at a time, rather than asking for participants to identify the most prominent emotion of a passage, this approach also allows for investigating subtle mixtures of emotion. This method also appears to be sensitive to detecting moments of uncertainty and complexity in music by pointing out those areas of intense emotionality. Future research could be used to model more subtle formal boundaries (like ambiguous phrase endings) using perceived emotion data, or conversely to model perceptions of emotion using complexity measures like cross entropy.

[9.5] However, while these initial analyses provide a useful starting point for dialoguing about the relationship between the theory of gesture and empirical evidence in support of gesture, these results should not be accepted wholesale. Every research paradigm has strengths and drawbacks, and empirical approaches are particularly susceptible to particular types of misunderstandings. First, care should be taken in trying to overly generalize the results of the gestural analysis; discussing gesture in general in the classical style is very different from discussing how gesture is used in one specific movement.[32] We should not assume that there is only one language or one set of devices for expressing particular emotions in music in a given style, and one should not overly generalize about an entire style from the analysis of one movement.

[9.6] Moreover, the results of this study reflect the context of the respondents. It cannot be emphasized enough that while the progressive exposure method is an effective way to tie perceptual responses directly to short musical gestures, one casualty of the approach is any kind of longer musical context. Much of a passage's expressive impact must be understood within the context of the preceding music. There would undoubtedly be differences in some of the ratings if the passages were heard within their proper context. Even so, Tillman and Bigand (1996) found that small chunks of music played in backwards order did not affect the perceived expressiveness of the music for selections for Bach and Mozart, and Bigand et al. (2005) found that listeners can

accurately recognize emotions in classical music in as little as one second, suggesting that enough of the expressive meaning of the music is sufficiently encoded at a low level of structure to not invalidate this study. At the same time, as a well-known work, many listeners reported in post-experiment interviews that they recognized the work, and this prior knowledge undoubtedly brings with it some kind of context surrounding the isolated excerpts and complex cultural connotations that certainly affect the results to some degree. It is important to remember that there is a cultural and social component to these findings that reflects the reality of twenty-first century music major undergraduates in the United States. Beethoven's contemporaries may not have heard the music in quite the same way—but this is exactly the point! Expression is a cultural construct in response to particular musical gestures.

[9.7] Nevertheless, I believe this kind of intersubjective empirical analysis of music offers genuine value to the analytical enterprise. One of the most important goals of music analysis is to construct convincing narratives about music that reveal new insights about the music or direct our attention to important elements of the music that might otherwise have gone unnoticed. Truly compelling stories weave together disparate strands from different perspectives. When widely divergent methodologies provide converging evidence for the same theoretical position, the story told is more convincing.

[9.8] Empirical methods of analysis offer a different set of advantages and suffer from a different set of drawbacks than more traditional methods of analysis. One of the greatest benefits offered by empirical methods is the opportunity to counterbalance an analyst's individual biases with the perspectives of a large group of listeners. This is certainly easier to do in more substantial ways than is typically possible when the accountability of data is not present. When the results of empirical studies are consistent with longstanding theories of musical expression, the converging evidence lends greater credence to the theories.

Appendix: Reliability Metrics

[A.1] To test the degree to which participants were personally consistent (intra-participant reliability) and the degree to which all participants provided similar responses (inter-participant reliability), a number of metrics were examined.

Intra-participant reliability

[A.2] There are two ways in which responses that a particular participant provides might be unreliable. In the first instance, participants may simply not provide reliable responses in general, regardless of emotion category. This may reveal a problem with a particular participant's responses—perhaps they were haphazard in their answers or distracted. In the second instance, participants might be reliable in general, but may unreliably evaluate a particular emotion category. This may reveal a problem with a category, perhaps because that category is confusing or because the participant did not understand the label.

[A.3] Because a participant might be reliable in general but unreliable in relation to a particular emotion category, each participant-scale was analyzed separately for intra-participant reliability. With 110 participants and 3 emotion categories per participant, there were 330 total participant-scales. To check for reliability, participants provided two responses for 5 randomly selected segments for each emotion category. Correlations were calculated between the set of initial and second responses for each participant-scale. A histogram showing all 330 correlations is provided in **Example A1**. An *a priori* elimination criterion was set for a +.4 correlation, meaning that participant-scales that were less consistent than this standard would be eliminated. However, after examining the data, this standard seemed too strict, as it would result in the elimination of 100 participant-scales. Therefore, the decision was made *a posteriori* to move the elimination criterion to +.25, resulting in the elimination of 76 participant-scales.

[A.4] Note that correlations between -.5 and +.5 reveal a low level of consistency between responses. Interestingly, the emotion categories most heavily represented in this range are not evenly distributed. Of the 71 participant-scales with correlations between -.5 and +.5, dark, emotional/moody, and sincerity/truthful were most highly represented with 22 participant-scales.

To further investigate the difference in reliability between emotion categories, individual participant-scale correlations were averaged across category. The results are given in **Table A1**. The categories with the lowest average correlation were emotional/moody (mean = +.448; sd = .61) and sincerity/truthful (mean = +.362, sd = .50).

[A.5] Correlation averages were unable to be calculated for cheeky/sassy due to a mathematical property of correlations in which they are unable to be directly averaged without Fisher's z-transformation (Silver and Dunlap 1987). This transformation maps -1 onto -∞ and +1 onto +∞. Because correlations for three of the participant-scales for cheeky/sassy were exactly +1, averages were unable to be calculated. Post-experiment interviews revealed that many participants thought that the cheeky/sassy scale was inappropriate for the movement, and therefore rated that scale with the lowest possible rating for nearly every excerpt. This resulted in consistent 0 ratings, resulting in a +1 correlation. For these reasons, cheeky/sassy was discarded from further consideration for the remainder of the analysis.

<div align="center">Inter-participant reliability</div>

[A.6] In addition to testing intra-participant reliability, inter-participant reliability metrics also provide insights into the success of the various emotion categories. Low inter-participant reliability is more acceptable than low intra-participant reliability, however, as individual differences in listener ratings should be respected. Nevertheless, low inter-participant reliability may signal that participants misunderstood directions, were inattentive, or were operating under differing definitions for different emotion categories.

[A.7] After 76 participant-scales with low intra-participant reliability were eliminated, inter-participant correlations were calculated for the remaining scales. This was done by correlating the 56 ratings for each participant-scale with the 56 ratings for all other participant-scales for the same emotion and offset condition, and the correlations were averaged together using Fischer's z-transformation. As before, all participant-scales averaging a correlation below +.25 with the other participant-scales were examined. The results for each participant-scale that met this criterion are shown in Table A2. The participant-scales are identified with their emotion category, their offset group, and the mean inter-participant correlation.

[A.8] As is evident from **Table A2**, the categories that were most inconsistently used between participants were sincerity/truthful, emotional/moody, and important/serious.[33] These scales also demonstrated low intra-participant reliability and many of these scales had already been eliminated by that exclusion criterion. For example, only 15 participant-scales remained for important/serious, 12 for sincerity/truthful, and 10 for emotional/moody. Demonstrating low inter-participant reliability in addition to low intra-participatng reliability, these three scales were discarded from further consideration.

Albrecht, Joshua D.
The University of Mary Hardin-Baylor
Music Department
900 College St.
Belton, TX 76513
jalbrecht@umhb.edu

<div align="center">Works Cited</div>

Agawu, Kofi. 1991. *Playing with Signs.* Princeton University Press.

Allanbrook, Wye Jamison. 1986. *Rhythmic Gesture in Mozart: Le Nozze di Figaro and Don Giovanni.* University of Chicago Press.

Bigand, Emmanuel, Sandrine Vieillard, François Madurell, Jeremy Marozeau, and A. Daquet. 2005. "Multidimensional Scaling of Emotional Responses to Music: The effect of Musical Expertise and of the Duration of the Excerpts." *Cognition and Emotion* 19 (8): 1113–39.

Blood, Anne, and Robert Zatorre. 2001. "Intensely Pleasurable Responses to Music Correlate With Activity in Brain Regions Implicated in Reward and Emotion." *Proceedings of the National Academy of Sciences* 98: 118–23.

Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven.* Oxford University Press.

Cole, Malcolm S. 2001. "Rondo." In *The New Grove Dictionary of Music and Musicians.* 2nd edition, ed. Stanley Sadie and John Tyrell, 21: 649–56. Macmillan.

Crowder, Robert G. 1985. "Perception of the Major/Minor Distinction II: Experimental Investigations." *Psychomusicology* 5: 3–24.

Davies, Stephen. 1994. *Musical Meaning and Expression*. Cornell University Press.

de la Motte-Haber, Helga. 1968. *Ein Beitrag zur Klassifikation musikalischer Rhythmen*. Arno Volk Verlag..

Eerola, Tuomas. 2016. "Expectancy-Violation and Information-Theoretic Models of Melodic Complexity." *Empirical Musicology Review,* 11 (1): 2–17.

Ekman, Paul. 1992. "An Argument for Basic Emotions." *Cognition & Emotion* 6: 169–200.

Gabriel, Clive. 1978. "An Experimental Study of Deryck Cooke's Theory of Music and Meaning." *Psychology of Music* 6: 13–20.

Gabrielsson, Alf, and Erik Lindström. 2010. "The Role of Structure in the Musical Expression of Emotions." In *The Handbook of Music and Emotion: Theory, Research, Applications*, ed. Patrik N. Juslin, and John A. Sloboda, 367–400. Oxford University Press.

Gabrielsson, Alf. 2002. "Emotion Perceived and Emotion Felt: Same or Different?" *Musicae Scientiae,* 5 (1 supplement): 123–47.

Hatten, Robert S. 1994. *Musical Meaning in Beethoven: Markedness, Correlation, Interpretation*. Indiana University Press.

—————. 2004. *Interpreting Musical Gestures, Topics, and Tropes: Mozart, Beethoven, Schubert.* Indiana Univerisity Press.

Heinlein, Christian Paul. 1928. "The Affective Characters of the Major and Minor Modes in Music." *Journal of Comparative Psychology* 8: 101–42.

Hepokoski, James, and Warren Darcy. 2006. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late Eighteenth-Century Sonata.* Oxford University Press.

Hevner, Kate. 1936. "Experimental Studies of the Elements of Expression in Music." *American Journal of Psychology* 48: 246–68.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press.

Juslin, Patrik N., and Petri Laukka. 2004. "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening." *Journal of New Music Research* 33: 217–38.

Juslin, Patrik N., and Erik Lindström. 2011. "Musical Expression of Emotions: Modeling Listeners' Judgments of Composed and Performed Features." Special Issue: Music and Emotion. *Music Analysis* 29 (1–3): 334–64.

Juslin, Patrik N., and John A. Sloboda, ed. 2010. *The Handbook of Music and Emotion: Theory, Research, Applications.* Oxford University Press.

Kaminska, Zofia, and Jennifer Woolf. 2000. "Melodic Line and Emotion: Cooke's Theory Revisited." *Psychology of Music* 28: 133–53.

Kendall, Roger A., and Edward C. Carterette. 1990. "The Communication of Musical Expression." *Music Perception: An Interdisciplinary Journal* 8 (2): 129–63.

Kivy, Peter. 1980. *The Corded Shell: Reflections on Musical Expressions.* Princeton University Press.

Lartillot, Olivier, and Petri Toiviainen. 2007. "A MATLAB toolbox for musical feature extraction from audio." *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-01-08).* Bordeaux, France.

Margulis, Elizabeth H. 2013. *On Repeat: How Music Plays the Mind.* Oxford University Press.

McKay, Nicolas. 2007. "On Topics Today." *Zeitschrift der Gesellschaft für Musiktheorie* 4 (1–2): 159–83.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music.* University of Chicago Press.

—————. 1967. *Music, the Arts, and Ideas.* University of Chicago Press.

Narmour, Eugene. 1990. *The Analysis and Cognition of Basic Musical Structures: The Implication-Realization Model.* University of Chicago Press.

—————. 1992. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model.* University of Chicago Press.

O'Conor, John. 2008. *Breathe®: relaxing piano for lovers.* Telarc B0010DJ1YM.

Persson, Roland S. 2001. "The Subjective World of the Performer." In *Music and Emotion: Theory and Research*, ed. Patrik N. Juslin, and John A. Sloboda, 275–89. Oxford University Press.

Repp, Bruno. 1997. "The Aesthetic Quality of a Quantitatively Average Music Performance: Two Preliminary Experiments." *Music Perception: An Interdisciplinary Journal* 14 (4): 419–44.

Robinson, Jenefer. 2005. *Deeper than Reason: Emotion and Its Role in Literature, Music, and Art.* Oxford University Press.

Robinson, Jenefer and Robert S. Hatten. 2012. "Emotions in Music." *Music Theory Spectrum* 34 (2): 71–106.

Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 59: 899–915.

Sapp, Craig. 2007. "Comparative Analysis of Multiple Musical Performances." In *Proceedings of the Eighth International Conference on Music Information Retrieval.* September 23rd–27th, 2007, Vienna, Austria.

—————. 2008. "Hybrid Numeric/Rank Similarity Metrics for Music Performance Studies." In *Proceedings of the Ninth International Conference on Music Information Retrieval.* September 14th–18th, 2008, Drexel University, Philadelphia, PA.

Scherer, Klaus R., and James S. Oshinsky. 1977. "Cue Utilization in Emotion Attribution from Auditory Stimuli." *Motivation and Emotion* 1: 331–46.

Silver, N. Clayton, and William P. Dunlap. 1987. "Averaging correlation coefficients: Should Fischer's z-transformation be used?" *Journal of Applied Psychology* 72: 146–48.

Sisman, Elaine. 1994. "Pathos and the *Pathétique*: Rhetorical Stance in Beethoven's C-Minor Sonata, Op. 13." *Beethoven Forum* 3: 81–105.

Sloboda, John A. 1991. "Music Structure and Emotional Response: Some Empirical Findings." *Psychology of Music* 19 (2): 110–20.

Tagg, Philip. 2006. "Music, Moving Images, Semiotics, and the Democratic Right to Know." In *Music and Manipulation. On the Social Uses and Social Control of Music*, ed. S. Brown, and U. Volgsten, 163–86. Berghahn Books.

Tillman, Barbara, and Emmanuel Bigand. 1996. "Does Formal Musical Structure Affect Perception of Musical Expressiveness?" *Psychology of Music* 24 (1): 3–17.

Watson, K. Brantley. 1942. "The Nature and Measurement of Musical Meanings." *Psychological Monographs* 54: 1–43.

Wiggins, Geraint A., Daniel Müllensiefen, and Marcus T. Pearce. 2010. "On the non-existence of music: Why music theory is a figment of the imagination." *Musicae Scientiae*, Discussion Forum 5: 231–55.

Zentner, Marcel R., Didier Grandjean, and Klaus R. Scherer. 2008. "Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement." *Emotion* 8: 494–521.

<center>Footnotes</center>

* I would like to thank Sharon Morrow for her generous help in recruiting and running students at Westminster Choir College of Rider University.
Return to text

1. Sisman (1994) provides an excellent discussion of the role of *pathos* in the eighteenth century and how this relates directly to the form of musical expression taken in the *Pathétique* Sonata.
Return to text

2. For a more detailed discussion of the gap between music theories and scientific theories, focusing on the problem of falsifiability and rigorous scientific testing, see Wiggins, Müllensiefen, and Pearce 2010.
Return to text

3. Readers more generally suspicious about expressive analysis may take issue with my assumption that 'the music' expresses anything at all emotionally, which has famously been the subject of much philosophical discussion. For example, Davies (1994) and Kivy (1980) reject the notion that music expresses emotion directly as a reflection of some internal emotion or intention of the composer. Rather, they argue that musical features reveal the appearance of characteristics that typically co-occur with various emotions, but are not direct expressions of emotions themselves, in the same way that basset hounds look sad while not necessarily feeling sad. Under this paradigm, listeners who think the music is expressing emotion are simply assigning an emotion to acoustical patterns they *perceive* as emotional through association.

Instead of wrestling with the ontological question of musical agency, I will rather take a practical approach by focusing on the emotional expression *perceived by* the listener. If the focus of the study is then entirely on what the listener perceives, in what sense does an analysis regard the music at all? Would the emotional responses evoked in the listener by the music best be regarded as arbitrary? Hatten's (2004) answer to this question again lies in the notion of intersubjective agreement between listeners. Sidestepping the problem of agency, there can still be communally recognized expression attached to particular musical gestures. Even if one avoids the agential implication that musical gestures actually "communicate" these emotions, one can still speak about patterns of emotional expression that are correlated with these gestures reliably over time and between listeners. If particular musical structures reliably evoke expressive connotations, then one may say that those gestures are expressive of those emotional states.
Return to text

4. For example, listeners tend to be much more interpersonally and intrapersonally reliable when judging perceived emotion than induced emotion (Juslin and Laukka 2004). The emotion aroused by music is notoriously impacted by much more than just the music itself; time of day, personal circumstances, one's current mood and arousal levels, the environment in which the music is heard, and the degree to which a listener is attending to the music all can impact the emotion felt in response to music much more strongly than the emotion perceived in the music. There are also stronger differences between listeners regarding felt than perceived emotion (Gabrielsson 2002). For example, the research on *frisson*, the experience of music-induced goose bumps or chills,

indicates that while the experience is reliable for listeners in specific moments in music (Sloboda 1991), there are low levels of correlation between listeners for the same excerpt (Blood and Zatorre 2001) — music that induces *frisson* for one listener leaves another unmoved and vice versa.
Return to text

5. Although I will be focusing on perceived emotion, the instructions that I give to the participants do not always reflect this. Framing my prompts with these kinds of subtle philosophical distinctions would likely unnecessarily complicate the issue for the participants of the study. Although in my instructions I routinely refer to concepts like "the emotions you think the music is trying to express or convey," it is important to remember that my inquiry is limited to the perceptions of emotional expression that are correlated with the musical gestures heard, rather than any essentialized claim of the ontological reality of agency in the music.
Return to text

6. Hatten's (2004, 94) definition of musical gesture also suggests a minimum segment length. Hatten considers gestures to be "perceptually *synthetic* gestalts with *emergent* meaning" that consist, among other things, of "specific timbres, articulations, dynamics, tempi, pacing, and their coordination with various syntactic levels (e.g., voice-leading, metric placement, phrase structure)." Moreover, gestures are units "in the perceptual present (typically within two seconds)." According to this definition, segments should be at least two seconds in length to capture musical gestures.
Return to text

7. As Kendall and Carterette (1990) argue, one way of conceptualizing the emotional meaning a listener perceives a work to be expressing is as the intersection of at least the compositional intent of the composer and the interpretation of the performer. Similarly, Juslin and Lindström (2011) have shown that the way that listeners perceive emotional expression in music involves a complicated combination of structural and performative elements.
Return to text

8. This approach was systematized in a fascinating study by Repp (1997). In this study, Repp averaged together the mictrotiming decisions of ten different performances of Schumann's *Traumerei* and compared this "average" performance to the original ten recordings. The average was rated second highest in quality, but second lowest in individuality. He did the same with thirty performances of Chopin's Etude in E major and found similarly high ratings of quality and low ratings of individuality. Similar studies have been conducted by Sapp (2007 and 2008) on large collections of Chopin's mazurkas. See also: The Mazurka Project (http://www.mazurka.org.uk). Unfortunately, at the time of the study no "average" recording of the second movement of Beethoven's *Pathétique* Sonata was available.
Return to text

9. Interested readers are referred to Juslin and Sloboda (2010) for a detailed survey of different approaches with their strengths and weaknesses. By way of quick summary, common approaches include free response descriptions, looking at a small number of "basic emotions" (e.g. Ekman 1992), a dimensional model in which all emotions are plotted in a continuous space consisting of a small number of dimensions such as arousal and valence (e.g. Hevner 1936, Russell 1980), or compiling an eclectic list of terms deemed appropriate for the music under study (e.g. Zentner, Grandjean, and Scherer 2008). This paper uses a modified version of the strategy suggested by Zentner et al. (2008).
Return to text

10. See footnote 4.
Return to text

11. Intra-participant reliability could not be calculated for cheeky/sassy due to a mathematical property of correlations in which they are unable to be directly averaged when they are exactly +1. Because cheeky/sassy ratings were exactly 0 for all segments tested, the correlations were unable to be calculated. Moreover, post-experiment interviews revealed that many participants found this category was not relevant to the movement, and so it was discarded from further consideration.
Return to text

12. Robinson 2005 argues that categories like "being moved" might simply reflect *that* a listener is emotionally affected without specifying *in what way* they are affected.
Return to text

13. To illustrate this point, consider the most two extreme participants. For the sake of comparison, we will assign 1 to the extreme left of the scale and 7 to the extreme right of the scale, with a midpoint of 4. The most positive participant has an average rating of 5.2, and the most negative participant has an average rating of 2.6. The lowest rating for the positive participant is 2.3; although significantly low for that participant, it would only be slightly below average for the negative participant. Likewise, the highest rating for the negative participant is 6.2; although significantly high for that participant, nearly 20% of the positive participant's ratings are above that mark. For these reasons, using standard deviations provides a useful means of controlling participant scale-usage style and identifying which segments participants find exceptional against the context of the movement. By focusing on exceptional moments within the context of a movement, this empirically derived expressive analysis more closely mirrors traditional analyses.
Return to text

14. The tool works best on Google Chrome, and takes a few seconds to load the first time. For more help in using the tool, consult the "Detailed Instructions" tab of the tool.
Return to text

15. These results are consistent with traditional associations of the minor mode with negatively valenced emotion and faster rhythms with higher levels of arousal.
Return to text

16. This observation is consistent with Robinson and Hatten's (2012) notion of musical voices as agents or personae in a musical drama. In this case, it is likely that the two contrasting gestures may actually imply different agents or personae, as discussed in Hatten 2004: "In terms of agency, the two contrasting gestural types. . .suggest the roles of protagonist and antagonist in conflict dramas, or more neutrally, actant and negactant" (225).
Return to text

17. Formally, this observation is statistically significant, $t(1152.116) = -6.6$, $p < .0001$.
Return to text

18. Blending cognitively complex emotions is an important component of an effective emotion paradigm according to Robinson and Hatten (2012, 71): "We claim that sometimes music can appropriately be heard as containing a 'persona'. . . and that this persona can be experienced as expressing more *complex* emotions, such as hopefulness or resignation, as well as *blends* of emotion, and emotions that *develop and change* over time." Robinson and Hatten (2012, 71). Zentner et al. (2008) likewise argues that music is particularly suited to express cognitively complex emotions; their research found, for example, that tender-longing is a particularly common emotion in musical expression, although it is often neglected because it is more complex than the set of 'basic' emotions most often studied.
Return to text

19. Experimental paradigms usually force a tradeoff between ecological validity and experimental control, with more rigorous experimental designs manipulating specific musical features, but without any other musical context (for an overview, see Gabrielsson and Lindström 2010). Experimenters have manipulated mode (Heinlein 1928, Crowder 1985), rhythm and tempo (de la Motte-Haber 1968), melodic properties (Gabriel 1978, Kaminska and Woolf 2000), and synthesized tone sequences (Scherer and Oshinsky 1977) to attempt to assess the influence that these musical factors exert on perceived emotion. While permitting a high level of experimental control, the artificially constructed stimuli used in these experiments raise questions about the level to which this actually mirrors real listening situations.

By contrast, a different approach utilizes excerpts from real music and asks listeners to rate perceived emotion in these excerpts. Some approaches (e.g. Tagg 2006) ask listeners to write down free responses to real musical excerpts while others (Watson 1942) use a forced-choice paradigm.

While this approach more closely reflects real listening situations, the stimuli used are very complicated real pieces of music, composed of so many musical parameters that it can be difficult to generalize about any specific parameters.
Return to text

20. Even literally repeated notation does not result in identical perceptions. Of course, the performance may be different, but even acoustically identical literal repetition, such as the difference between an initiating or continuing function, will be perceived differently as a matter of form or function: "At a minimum, a repeated element will sound different from its initial presentation by virtue of coming later and having been heard before. More subtly, it will sound different as a function of its position within the unfolding series of metric projection" (Margulis 2013, 35).
Return to text

21. Narmour (1990 and 1992) theorized that reversals of implication or uncertain realizations would be accompanied by stronger emotional response. Similarly, Huron (2006) has outlined the role that unrealized expectation or uncertain continuation plays in heightening the emotional experience of a passage. For a more extensive overview, see Gabrielsson and Lindström 2010.
Return to text

22. Hepokoski and Darcy (2006, 398–399) describe the particular character of a rondo refrain as light, playful, and tuneful, also citing Cole (2001), who describes rondo themes are simple and tuneful. Episodes, by contrast, are much less predictable and permit a wide variety of formulae. In describing specifically the second episode, or couplet, of a 5-part rondo, Caplin (1998, 234) points out the unpredictability of this formal section by saying, "[s]uch a wide variety of formal procedures can be found at this point in the form that generalizations are difficult to make. Most such cases have a certain development-like quality about them. Indeed, a few are organized along the lines of a true development section."
Return to text

23. This difference was statistically significant, $t(16596.26) = -27.77$, $p < .0001$.
Return to text

24. "Three varieties of deviation may be distinguished. (1) The normal, or probable, consequent event may be delayed. Such a delay may be purely temporal or it may also involve reaching the consequent through a less direct route, provided that the deviation is understandable as a means to the end in view. (2) The antecedent situation may be ambiguous. That is, several equally probable consequents may be envisaged. When this takes place, our automatic habit responses are inadequate, for they are attuned only to a clear decision about probabilities. And (3) there may be neither delay nor ambiguity, but the consequent event may be unexpected – improbable in the particular context" (Meyer 1967, 10–11).
Return to text

25. "the relationships between units of language are more important than any intrinsic properties of those units."
Return to text

26. "the fusion of topics. . .is *emergent*, or beyond the mere sum of the correlations of each topic."
Return to text

27. "pitch contour alone does not provide uniform results, given the many variables affecting our interpretation of such contours: metric placement and rhythmic duration, harmonic setting, articulation, dynamics, timing (both tempo and pacing), orchestration, and the like" (Hatten 2004, 150).
Return to text

28. "a gestural accounting for all of these variables at least as a 'fuzzy set'. . .can help us evaluate their contribution to an emergent affect" (Hatten 2004, 150).
Return to text

29. See Lartillot and Toiviainen 2007 and Eerola 2016. There are fifteen entries in the table because the pitch height variable is actually two separately encoded values—the highest and lowest pitch in semitones.
Return to text

30. The "variance" values in a regression model (technically adjusted R2) indicate how much of the variance in listener response can be explained through the analysis of the given musical features alone. All models were able to explain 19-38% of the variance in participant response.
Return to text

31. An imperfect analogy can be made between the pastoral topic and the calm/serene category. According to Hatten, this topic would be correlated with the major mode and slower harmonic tempi. In this movement calm/serene is indeed correlated with the major mode, although calm/serene is actually associated with faster harmonic tempi in this movement. For Hatten, turns to the minor mode and marked dissonance lead to increased agitation. As predicted, unsettled/anxious is correlated with the minor mode. However, harmonic dissonance was not significantly correlated with unsettled/anxious ratings in this movement, although the results are not incompatible with Hatten's theory. Hatten associates yearning with upward melodic motion. While, contrary to expectation, pitch direction was not significantly related to striving/yearning, the emotion was associated with a lower lowest pitch. As one last example, Hatten associates greater harmonic dissonance (specifically the diminished seventh chord) with increased anguish or grief. While the analysis is not fine-grained enough to distinguish types of dissonance, in this movement sad/depressed/tragic was not significantly correlated with the level of dissonance in the harmony.
Return to text

32. This point extends to any applications of this research to topic theory more broadly. The correspondence between the expressive categories tested and topic theory discussed is not exact, which may account for some of the discrepancies. Additionally, the analysis used in the regression equations in many cases is not fine-grained enough to neatly map onto predictions about gestures in music broadly construed.
Return to text

33. Where True = truthful/sincere, Moody = emotional/moody, Joy = happy/joyful, Import = important/serious, Yearn = striving/yearning, Weight = weighty, and Sad = sad/depressed/tragic.
Return to text

Copyright Statement

Prepared by Michael McClimon, Senior Editorial Assistant