

## ACOUSTICALLY EXPRESSING AFFECT

---

AIMEE BATTCOCK & MICHAEL SCHUTZ  
*McMaster University, Hamilton, Canada*

**COMPOSERS CONVEY EMOTION THROUGH MUSIC** by co-varying structural cues. Although the complex interplay provides a rich listening experience, this creates challenges for understanding the contributions of individual cues. Here we investigate how three specific cues (attack rate, mode, and pitch height) work together to convey emotion in Bach's *Well Tempered-Clavier (WTC)*. In three experiments, we explore responses to (1) eight-measure excerpts and (2) musically "resolved" excerpts, and (3) investigate the role of different standard dimensional scales of emotion. In each experiment, thirty nonmusician participants rated perceived emotion along scales of valence and intensity (Experiments 1 & 2) or valence and arousal (Experiment 3) for 48 pieces in the *WTC*. Responses indicate listeners used attack rate, Mode, and pitch height to make judgements of valence, but only attack rate for intensity/arousal. Commonality analyses revealed mode predicted the most variance for valence ratings, followed by attack rate, with pitch height contributing minimally. In Experiment 2 mode increased in predictive power compared to Experiment 1. For Experiment 3, using "arousal" instead of "intensity" showed similar results to Experiment 1. We discuss how these results complement and extend previous findings of studies with tightly controlled stimuli, providing additional perspective on complex issues of interpersonal communication.

*Received: August 10, 2018, accepted June 28, 2019.*

**Key words:** emotion, perception, applied music cognition, valence, arousal

---

"Music can serve as a way of capturing feelings, knowledge of feelings, or knowledge about the forms of feeling, communicating them from the performer or the creator to the listener."

(Gardner, 1993, p. 124)

**MUSIC'S RELATIONSHIP WITH EMOTION** IS one of the central reasons for our engagement with it (Juslin & Laukka, 2004) and continues to fascinate composers, listeners, psychologists, and neuroscientists alike. Similar to their use in vocal

expression, listeners attend to and decode specific cues in lawful ways, with certain cues unique to music. Emotional communication is complex, governed by a multitude of factors both within the acoustic signal itself as well as from learned associations and experiences (i.e., national anthems, cultural conventions, etc.). The complexity and importance of this issue has generated sustained research interest (Hevner, 1936; Koelsch et al., 2004; Wiggins, 1998), finding consistent agreement in many aspects of its communicative abilities. Although some aspects are difficult to quantify precisely, a growing body of research on the relationship between psychophysical cues and their emotional associations has proven informative.

### Timing as a Cue for Emotional Expression

Timing is a powerful cue for emotional communication; however, understanding its effect is complex as timing encompasses several distinct musical properties such as tempo and rhythm (Balkwill & Thompson, 1999; Juslin & Madison, 1999; Schellenberg, Krysciak, & Campbell, 2000). Tempo, which describes the number of beats per minute, is of great importance for conveyed emotion (Balkwill & Thompson, 1999; Gagnon & Peretz, 2003; Scherer & Oshinsky, 1977). The role of musical tempo holds some parallels with articulation rate in speech, with fast and slow tempos associated with happiness and sadness respectively (Hevner, 1937, Juslin, 1997; Rigg, 1940). Sensitivity to tempo emerges at an early age, with children as young as four making affective judgments using tempo rather than familiarity (Mote, 2011). This develops earlier than their sensitivity to mode (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). Mote (2011) argues that the dependency on tempo suggests children may generalize associations between speed and emotion in human behavior—particularly speech—to music. This early sensitivity to timing may help explain why cues like tempo are found to have stronger effects than mode (Hevner, 1935, 1937).

Rhythm also plays a complex yet powerful role in emotional communication. The effect of rhythm is found to vary as a function of melody and intended emotion. In an experiment consisting of four measure melodies from unknown folksongs or experimentally composed melodies selected to express "happy," "sad," or "scary,"

listeners rated melodies higher in the appropriately expressed emotion when excerpts contained rhythmic variation. In addition, the effect of rhythmic variation interacted with pitch (Schellenberg et al., 2000). The authors suggest their selection of emotional exemplars resulted in melodies that differed on a number of structural dimensions (number of contour changes, mean pitch level, as well as meter), which can explain why the effect of rhythm appeared context specific. Furthermore, the effect of rhythm can be so powerful it extends cross-culturally, correlating with emotions like joy, sadness, and peace within Hindustani ragas presented to Western listeners (Balkwill & Thompson, 1999). There participants rated pieces expressing joy to contain more simple rhythms in contrast to sad pieces, which participants judged to have more complex rhythms. In addition, these naïve, Western listeners could accurately identify the intended emotions conveyed within the ragas. These findings demonstrated that despite unfamiliarity with the musical stimuli, the cue of rhythm remained a salient indicator of the conveyed emotion.

#### Mode as a Cue for Emotional Expression

Unlike timing and pitch the musical cue of mode is specific to music, referring to the structure of pitch information. Hevner's (1935) landmark work on mood associations with common Western modes (major and minor) illustrates that minor modes are associated with negatively valenced emotions such as "sad" and "melancholy," whereas major melodies are described as "cheerful" and "gay." In fact, mode is often a significant predictor of valence, with the major mode commonly associated with positively valenced emotions (Costa, Fine, Enrico, & Bitti, 2004; Crowder, 1985).

The connection between emotion and musical mode is well established (Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005; Quinto, Thompson, & Keating, 2013; Webster & Weir, 2005), showing major-minor distinctions are useful predictors of emotions such as happiness and sadness (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990). The impact of mode is so strong it can shape emotional responses more so than pitch or timing (Hevner, 1935, 1937). However the relative contributions of mode and tempo are complex (Gagnon & Peretz, 2003; Juslin & Lindström, 2010).

Although powerful, mode is a culture-specific cue that requires learning (Corrigall & Trainor, 2014). Meyer's (1956) proposed theory of deviations highlights the idea that relationships between major and minor keys stem from expectations of regular and normative

melodic progressions. In this regard, the associations and regularities must be internalized in order to form implicit and explicit musical expectations. Mode's power requires exposure: before the age of five, children are unable to identify this relationship between short melodies and emotional faces (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990).

#### Pitch as a Cue for Emotional Expression

Emotion can also be conveyed through the perceptual property known as pitch—the subjective "highness" or "lowness" of a tone. Despite its clear role in speech (Bachorowski & Owren, 1995; Breitenstein, Lancker, & Daum, 2001; Scherer, 1995), it's musical role is less straightforward. Pieces in higher octaves are generally found to be associated with more positive emotional adjectives such as happy, glad, and dreamy when assessing pairs of pitches (Eitan & Timmers, 2010), scales (Collier & Hubbard, 2001), commercially recorded works (Gundlach, 1935; Watson, 1942; Wedin, 1972), and transposed compositions (Hevner, 1937). Conversely, lower octaves are associated with negative emotions such as sad, agitated, and somber (Gundlach, 1935; Hevner, 1937; Scherer & Oshinsky, 1977; Watson, 1942; Wedin, 1972).

However, research on discrete emotions provides a different perspective. For example, high pitches are in some cases associated with negative emotions, as well as low pitches with positive emotions (Ilie & Thompson, 2006; Scherer & Oshinsky, 1977). Second, pitch information (specifically pitch range) does not emerge as a strong predictor of listeners' ratings of target emotions across different musical cultures—although other cues do seem to translate (Balkwill & Thompson, 1999). Those authors suggest this may have occurred given that pitch range plays an important role in expectancy, which can be generalized to emotional arousal, rather than specific emotions. Thus, the pitch information contained in Hindustani ragas did not provide useful information for listeners to interpret a specific, discrete emotion.

Research using the dimensional perspective of emotion also raises questions about pitch height's role. High-pitched music has been associated with both high and low-arousal emotional terms; listeners are found to associate high pitch with anger and fear (Scherer & Oshinsky, 1977; Wedin, 1972a), in addition to affective adjectives representing low arousal states such as graceful and serene (Hevner, 1937). Musical stimuli lower in pitch have been associated with sadness and boredom (Hevner, 1937; Scherer & Oshinsky, 1977), but also with

affective adjectives such as excitement and agitation (Hevner, 1937; Rigg, 1940).

Although a body of research suggest pitch height plays a role in musical emotion, its relationship appears less clear than cues such as tempo (Gabrielsson & Lindström, 2010) and mode. The varying effects of pitch may emerge, in part, from the range of stimuli used within experiments. Differences may occur not only as a result of the increased complexity of polyphony (Ilie & Thompson, 2006) over monophony (Balkwill & Thompson, 1999), but also with respect to performed versus synthesized and manipulated (Scherer & Oshinsky, 1977) musical stimuli. Monophonic and experimentally “controlled” stimuli are often used for studies exploring the cue-response relationships, therefore more work on the natural use of cues will shed light on the complex relationships between cues and listener perceptions of musical emotion.

### Measuring Emotional Communication

Assessments of musical emotion involve both discrete and dimensional models. Discrete models function as forced-choice paradigms based on the framework of Ekman’s (1992) theory of basic emotions. These models assume a limited number of fundamental emotions such as anger, joy, sadness, fear, etc., derivative of biologically determined emotional responses (Borod, 2000). Experimental procedures utilizing discrete emotional models often require participants to select which discrete emotion is represented (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013). Although discrete models facilitate paradigms involving recognition, they restrict the range of more complex but recognizable emotions (Eerola & Vuoskoski, 2013).

In contrast, the dimensional model of emotion can offer more reliable measurement with emotionally ambiguous stimuli (Eerola & Vuoskoski, 2010). For example, Russell’s (1980) popular circumplex model organizes emotional responses into two dimensions: valence and arousal. In this framework, valence represents the intrinsic positive or negative component of emotion and arousal represents the intensity or energy of the emotion. A number of studies have harnessed this view’s utility in music (Wedin, 1969, 1972a, 1972b). In these studies, factor analyses on the semantic contents of adjectives or words listeners associated to musical excerpts indicated arousal and emotional valence emerge as the two main dimensions.

Two dimensional models can account for a large proportion of variance (Schubert, 1999); however, the standard dimensions of valence and arousal alone fail to

fully explain responses (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005) leading to interest in alternatives. For example, Schimmack and Grob (2000) argue that the ambiguous definition of arousal introduces confusion, which can be interpreted as either an awake-tired or tense-relaxed state. As such, many studies explore variations from the standard dimensional model, using labels such as tension (Ilie & Thompson, 2006), activity (Leman, Vermeulen, De Voogdt, Moelants, & Lesafre, 2005), and strength (Luck et al., 2008). Consequently, here we assess emotion using different dimensional labels in order to contribute to ongoing discussions on this contested topic.

### Reflections on Stimuli Used to Explore Emotion

Several studies use polyphonic musical examples, such as one drawing upon stimuli chosen to represent specific quadrants of the circumplex model (Dibben, 2004). Others focus on film soundtracks designed to stir up emotion (Vuoskoski & Eerola, 2011), offering insight into the processing of highly emotional musical experiences. However the popularity and familiarity of this music introduce interesting challenges to interpreting results. Participants may be familiar with certain pieces of film music, having formed pre-existing associations with moments in the film, influencing their responses to the music. Furthermore, pieces from film soundtracks can contain sounds from multiple instruments in an orchestra or band, which introduce another layer of complexity (incorporating different timbres, pitch information, etc.).

The growing field of Music Information Retrieval (MIR) also extends the literature of perceived emotions in music by extracting features from stimuli to determine which predict emotion ratings. This approach has led to useful insight on a wide range of stimuli, such as polyphonic ringtones (Friberg, Schoonderwaldt, Hedblad, Fabiani, & Elowsson, 2014), film soundtracks (Eerola, 2011), and pop music (Yang & Chen, 2012). For example, Korhonen, Clausi, and Jernigan (2006) used five excerpts of a Western art music style, collecting the continual emotional appraisals for dimensions of valence and arousal. The authors used the overall median emotional appraisal across the response time-series to represent each piece in their analyses. They then created models of the emotional content for each piece as a function of time and the musical features extracted from excerpts. As music is a time and potentially emotion varying stimuli, these time-series approaches can prove powerful tools for exploration. At the same time, requiring participants to provide continuous responses affects participants’ cognitive load,

potentially affecting their emotional responses. Additionally, although stimuli used in MIR research on this topic is rooted in naturalistic music listening, a large proportion focus on either pop music or soundtrack music containing multiple instruments. The sheer complexity of these naturalistic examples complicates efforts to draw strong conclusions about specific musical cues. Finally, the degree to which automated analyses accurately reflect the structural cues recognized as significant by music theorists is in itself an open question (Byrd & Crawford, 2002). Consequently, additional work is needed to explore conveyed emotion in other musically polyphonic styles, and assessment of the effectiveness of feature extraction compared to score based cue quantification is crucial.

In order to provide a more focused perspective on the specific cues communicating emotional information, researchers often turn to monophonic (single-line) melodies affording rigorous quantification (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013). Others have turned to stimuli designed or composed to depict discrete emotions (Balkwill & Thompson, 1999; Hailstone et al., 2009). These approaches avoid the problems inherent with more naturalistic approaches such as studies of film music and/or MIR based analyses of large corpora of popular music, by offering precise control of multiple parameters. However, they are far removed from the types of music that so powerfully evoke strong emotions—such as the sounds heard in concert halls, home stereo systems, and personal listening devices. In addition, experimental designs independently manipulating cues such as pitch and timing to avoid confounds overlook the powerful cumulative effects of the ways in which great composers chose to co-vary certain cues (Schutz, 2017).

Previous work has offered useful insight into musical emotion utilizing naturalistic stimuli with considerable variation on many dimensions, or tightly controlled stimuli with controlled manipulations. Here we aim to fill a gap between these approaches by exploring perceptual consequences of specific cues in unaltered renditions of widely performed and studied music. In order to identify the independent contributions of “natural” cues lacking independence, we drew on our team’s previous extensive analysis and encoding of cues such as pitch timing and mode, as well as the technique of commonality analysis, or variance partitioning, on regression modeling. This provides novel insight into the unique and shared contributions of co-varying cues as manipulated by a renowned composer, offering useful new insight into how they work together to convey emotion.

## The Present Study

Here we assess the relationship between musical structure and emotion perception of unaltered music written by a historically distinguished composer, J. S. Bach (1685–1705). Building upon previous approaches manipulating cues such as pitch and timing, we explored the degree to which Bach’s choices of mode, pitch, and timing affect listeners’ emotional responses to complex polyphonic music routinely performed and enjoyed in a wide variety of musical settings. Specifically, we used J. S. Bach’s well known *Well-Tempered Clavier (WTC) Book 1* as performed by Friedrich Gulda (Bach, 1722/1973). Our approach complements and extends previous targeted explorations of manipulations to individual cues by exploring the perceptual consequences of the ways in which Bach naturally co-varied their use in a set of pieces still widely performed and studied. This preserves the musical complexity often experienced by listeners, offering an opportunity to assess generalizability of previous research on monophonic or experimentally designed acoustic stimuli, as well as previous studies of emotional excerpts that likely came with extra-musical associations (i.e., film scores, popular music excerpts, etc.)

Given the significance of mode (Dalla Bella et al., 2001; Heinlein, 1928; Juslin & Lindström, 2010; Quinto et al., 2013), we wanted to base this exploration on a “balanced” set of major and minor key pieces. This proved surprisingly difficult, as Western music is overwhelmingly written in major keys. Classical composers such as Haydn and Mozart display a bias towards the major mode (Tan, Pfordresher, & Harré, 2010), which can also be found in both jazz (Broze & Shanahan, 2013) and rock (Temperley & de Clercq, 2013). As such, Bach’s *WTC* is ideally suited for this exploration and offers a naturally balanced set of pieces with one Prelude and one Fugue in each major and minor key.

Emotional responses to the stimuli were encoded using a dimensional model in order to account for the complexity and richness of emotional affect within this set of pieces. We adapted Russell’s (1980) circumplex model of emotion to represent the emotional space with scales of valence and arousal. For comparison, we tested two versions—one incorporating dimensions of valence and intensity (Experiments 1 and 2), and another with dimensions of valence and arousal (Experiment 3). In order to generalize our results broadly, we chose to use participants with minimal music training. Although previous research indicates nonmusicians and musicians may perceive emotional connotations in music similarly (Bigand et al., 2005; Juslin & Laukka, 2003),

those of untrained participants allowed us to establish a consistent baseline that could be expanded upon in future research.

This study had two primary aims. First, to determine the relationship between timing, mode, and pitch on the perception of emotion, as they naturally vary in an ecologically valid polyphonic stimulus. Second, to determine the validity of an alternative affective dimension—intensity—in lieu of Russell’s (1980) dimension of arousal. Our hypotheses included predictions that: 1) timing, mode, and pitch cues will predict listener ratings of emotion; 2) musical mode will increase in its importance within musically “resolved” excerpts (excerpts cut to end in the piece’s starting nominal key); and 3) cues will vary to the extent they are important across valence and intensity/arousal. For our second aim, we predicted that listener responses of emotional intensity (Experiment 1) for perceived emotions would not be significantly different than ratings of perceived emotional arousal (Experiment 2).

### Experiment 1: Intensity

#### METHOD

*Participants.* We recruited thirty nonmusicians (< 1 year of music training) undergraduates (12 males,  $M = 19.7$  years,  $SD = 2.9$ ; 18 females,  $M = 19.1$  years,  $SD = 3.0$ ) from the McMaster University Psychology participant pool who reported normal hearing and normal or corrected-to-normal vision. The experiment met ethics standards according to the McMaster University Research Ethics Board. Participants received course credit in return for participation.

*Musical stimuli.* Experimental stimuli consisted of audio recordings of J. S. Bach’s *WTC (Book 1)* as performed by Friedrich Gulda ( $n = 48$ ). Excerpts contained the first eight measures of each piece, with a two-second fade out starting at the ninth musical measure. Although faster and slower pieces varied in duration, this approach provided consistency in terms of musical units (measure length). Stimuli lasted 7–64 s in duration ( $M = 30.2$  s,  $SD = 13.6$ ). We prepared all excerpts using Amadeus Pro.

*Cue quantification.* Beyond encoding the modality indicated in each piece’s key signature, our analysis required quantifying two additional cues: pitch height and timing. We calculated pitch height using methods based upon Huron, Yim, and Chordia (2010)—and later extended by Poon and Schutz (2015)—to weight notes according to their duration (similar to other music, these pieces included both long and short notes). In this approach, pitch height is calculated by summing

duration-weighted pitch values within each measure, then dividing by the sum of note durations within that measure. Previously, Poon and Schutz (2015) also used this method to calculate theoretical averages for the first eight measures of each of the 48 pieces using tempi noted in a score. Here we used that approach as a point of departure, adjusting the tempi used in the calculations to reflect those in the stimuli played for participants. We also re-calculated information as needed for Experiment 2, which involved excerpts of variable lengths rather than the eight-measure excerpts used in Experiment 1 and 3. This ensured that all attack rate information used for comparison in each experiment corresponded to the stimuli heard by participants. Additional technical details on pitch and timing quantification methods are available in Poon and Schutz (2015), including a figure annotating the exact pitch and timing values assigned to each note in the first measure of the C Major Prelude. We used this approach to calculate musical attack rate in part to allow for parallel comparisons of timing in speech, specifically with articulation rate (Johnstone & Scherer, 2000; Scherer, 2003).

Pitch height values varied from 33.13–53.00 ( $M = 43.90$ ,  $SD = 4.03$ ) corresponding to  $\sim F3$  to  $\sim C\#5$ ; attack rate information for eight-measure excerpts ranged from 1.30–10.13 attacks per s ( $M = 4.91$ ,  $SD = 2.18$ ). We operationalized mode as the tonal center of the piece, as indicated by the denoted key signature of each score, coded dichotomously (0 = minor, 1 = major). Admittedly, nominally minor excerpts in our experiment contained some major chords and vice versa, making for a less controlled treatment of mode than monophonic excerpts created to be either unambiguously major or minor. Nonetheless, this is entirely in keeping with normative practice in musical composition, where harmonic progressions typically include both major and minor chords. As each of these pieces starts in the nominal key, we believe it is a reasonable way to explore mode as it is experienced in concert halls and on recordings—rather than the more controlled (but uncommon in natural practice) approaches found in psychological experiments.

*Design and procedure.* Participants first completed a consent form and musical experience survey (see Appendix A), then entered a sound-attenuating booth where the research assistant verbally instructed participants on the rating task. After each excerpt, participants rated two aspects of perceived emotion, using scales for valence and intensity. Instructions emphasized the full use of each scale displayed. Research assistants told participants they would be asked to provide ratings of emotion based on what the music conveys on two scales

after listening to each piano excerpt. They described valence as how positive or negative the emotion sounds, ranging on a scale from 1 (*negative*) to 7 (*positive*). Intensity referred to the “energy” of the emotion, where high intensity pieces may sound excited or agitated, and low arousal pieces may sound dull or calm. The scale of intensity ranged from 1 (*low intensity*) to 100 (*high intensity*). We asked participants to make ratings based on emotion conveyed, rather than inquiring about emotions evoked (Gabrielsson, 2002). After hearing these instructions, participants completed four practice trials with alternate recordings not used in testing trials performed by Rosalyn Tureck (Bach, 1722/1953), where they could ask the research assistant for procedural clarifications. We conducted the experiment using PsychoPy (Peirce et al., 2019), a Python-based psychology program, and presented the experiment on a DELL monitor. Participants listened to the stimuli at a consistent and comfortable listening level through two Gateway 2000 speakers placed on either side of the computer monitor in a sound-attenuating booth (IAC Acoustics, Winchester, US). Each participant heard an individually randomized order of the 48 excerpts and provided responses via an Apple mouse connected to a 13-inch MacBook Pro located outside the booth.

## RESULTS

Visualizing participant data on Russell’s (1980) two-dimensional circumplex model provides a useful first step to understanding emotional responses in these stimuli. Figure 1a shows ratings for the first experiment, illustrating minor key pieces received lower valence ratings than major for both preludes (left column) and fugues (right column). In fact, of the 24 preludes, only one major piece (B major) fell in the lower half of valence ratings. Of the 24 fugues, only one (C major) clearly fell in the lower half of valence ratings (B major and D minor fugues tied for the 12th lowest valence rating). This is consistent with previous research indicating mode’s strong effect on emotion, and suggests our treatment of mode as a binary variable based on the nominal key of each piece provides a useful framework for understanding the emotional messages conveyed. However, as shown by Poon and Schutz (2015), composers co-vary cues in normative musical practice, making it difficult to understand the ultimate reason for this putative effect of mode. To explore this issue further, we turned to three separate statistical analyses. These both provide different perspectives on interpreting the data, as well as useful points of comparison with a rich literature on emotional communication in both speech and music.

In order to clarify cue contributions, we assessed participant ratings from three perspectives. First, we examined Pearson product-moment and Pearson point biserial correlations between the three acoustic cues (pitch, timing, mode) and the two dimensions of response (valence, intensity). Second, we assessed the relationship between acoustic cues (attack rate, mode, pitch) and listener responses, as captured by a two-dimensional model of valence and intensity, using a least squares standard multiple linear regression. Third, we further assessed relative cue contributions with commonality analyses to determine partitioned variance within the regression models. We determined Cronbach’s alpha for listener ratings across 48 excerpts to be  $\alpha = 0.98$  for both valence and intensity ratings, suggesting participant ratings are highly consistent. Valence ratings ranged from 1.90–6.13 ( $M = 4.12$ ,  $SD = 1.20$ ) and intensity ratings ranged 21.00–82.20 ( $M = 51.52$ ,  $SD = 19.69$ ).

*Correlations.* Within the three acoustic cues, we found a significant correlation only between attack rate and mode,  $r(46) = .43$ ,  $p < .01$ . Pitch height correlated significantly with neither attack rate,  $r(46) = -.14$ ,  $p = .35$ , nor mode,  $r(46) = .14$ ,  $p = .33$ . Independent-samples *t*-tests revealed significant differences in attack rate attack rate,  $t(46) = -3.24$ ,  $p < .05$ , but not pitch height,  $t(45) = -.98$ ,  $p = .33$ , between major and minor key pieces. This is consistent with finding a significant correlation between mode and attack rate, but a lack of significant correlation between mode and pitch height. Within participant ratings, we found a positive correlation between ratings of valence and intensity,  $r(46) = .80$ ,  $p < .001$ , indicating the dimensions of the standard two dimensional model did not function independently in this context.

Exploring the relationship between acoustic cues and participant ratings, both attack rate,  $r(46) = .71$ ,  $p < .001$ , and mode,  $r(46) = .76$ ,  $p < .001$ , correlated with valence ratings. This is consistent with the visualization in Figure 1a suggesting that mode plays a strong role in explaining valence ratings. Similarly, both attack rate,  $r(46) = .71$ ,  $p < .001$ , and mode,  $r(46) = .44$ ,  $p < .002$ , correlated with intensity ratings. In contrast, pitch height did not play a meaningful role, as it correlated with neither valence,  $r(46) = .17$ ,  $p = .24$ , nor intensity,  $r(46) = -.08$ ,  $p = .61$ , ratings. This analysis suggests that emotional responses are affected by only by timing and mode, with minimal role of pitch height. This outcome is helpful in drawing contrasts between previous work on the perceptual consequences of pitch and timing on emotional speech (Breitenstein, Van Lancker, & Daum, 2001). However, correlations amongst the cues

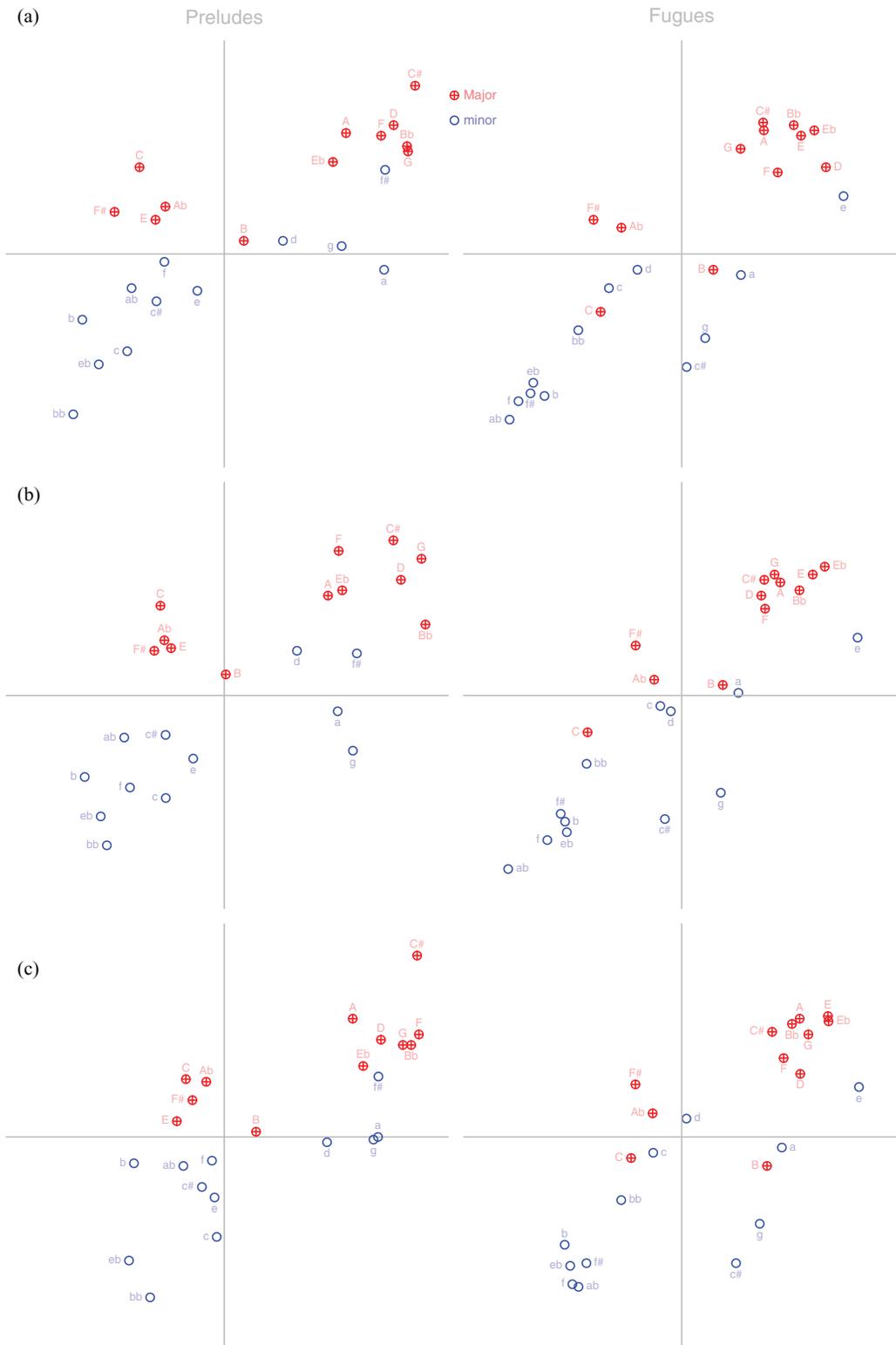


FIGURE 1. Mean ratings for all 48 pieces in the WTC (separated by preludes and fugues) plotted across the 2D circumplex space for (a) Experiment 1, (b) Experiment 2, and (c) Experiment 3. Major key pieces are shown with a cross through the circle, minor key pieces are shown with an open circle.

TABLE 1. Regression Model for Normalized Attack Rate, Mode, Pitch Height and Valence and Intensity Ratings (Experiment 1)

Predictor Coefficients	Valence				Intensity			
	B	SE	t	p	B	SE	t	p
Attack Rate	0.5031	0.0435	6.264	$p < .001$	0.6329	1.0479	5.356	$p < .001$
Mode	-0.5212	0.1911	-6.485	$p < .001$	-0.1708	4.6064	-1.445	$p = .156$
Pitch Height	0.0215	0.0215	2.277	$p < .01$	-0.0134	0.5171	-0.124	$p = .902$

Note: Beta values indicate strength and direction of relationship between each predictor variable and valence and intensity ratings. Default state for mode is Major.

themselves (e.g., timing correlates with mode)—which are likely common in music written for artistic purposes—complicates interpretation of simple correlations between cues and ratings. Consequently, we turned to additional analyses to better understand what cues predict listener ratings of emotions, as well as the specific contributions of individual cues to participant responses.

*Linear regression analysis.* We ran a standard linear multiple regression analyses on normalized predictor values using the R Statistical Package to assess predictors of mean ratings of valence and intensity. We chose the major mode as the reference level for mode, where the remaining level of the categorical variable (minor mode) is contrasted against it in analysis. The regression analysis revealed that all three acoustic cues—attack rate, mode, and pitch height—significantly predicted ratings of valence (Table 1). In contrast, only attack rate predicted ratings of intensity (Table 1). This approach illustrates two important insights beyond those available from the correlations alone. First, when examined with this more nuanced assessment, mode does not predict intensity ratings. Although it correlated with intensity ratings in our first analysis, the linear regression suggests its contribution stems from its correlation with attack rate. Conversely, although we did not find a simple correlation between pitch height and valence ratings in our first analysis, it did serve as a significant predictor here.

Overall, the 3-cue predictor model accounted for 77% of the variance in ratings of valence (adjusted  $R^2 = .765$ ),  $F(3, 44) = 52.13$ ,  $p < .001$ , and 49% of the variance in ratings of intensity (adjusted  $R^2 = .492$ ),  $F(3, 44) = 16.20$ ,  $p < .001$ . Tolerance and variance inflation factor (VIF) values indicated no issue of multicollinearity despite moderate correlation,  $r = .43$ ,  $p = .002$ , between attack rate and mode (attack rate, tolerance = .773, VIF = 1.293; mode, tolerance = .772, VIF = 1.295). The inclusion of interaction effects increased overall model predictability by a small amount for valence (adjusted  $R^2 = .771$ ),  $F(7, 40) = 23.55$ ,

$p < .001$ , and intensity (adjusted  $R^2 = .494$ ),  $F(7, 40) = 7.55$ ,  $p < .001$  (see Appendix B).

*Commonality analysis.* Finally, in order to more fully understand the overall contributions of each cue, we used commonality analysis to decompose the  $R^2$  of each model. This technique affords examination of contributions of both unique and shared variance for each of our predictors (Table 2 & 3). Commonality analysis allows for reporting on the multivariate relationships between predictors beyond beta values, however does not address potential interaction effects within the model. Here, “shared” variance between predictors (overlapping regions in Figure 2) represent the variance those variables have in common with the dependent variable (Ray-Mukherjee et al., 2014). The presence of negative commonalities occurs when correlations among predictor variables have opposite signs (Pedhazur, 1997), or in the case that a variable confounds the explained variance of another variable in the model (Capraro & Capraro, 2001), such as a suppressor variable. Suppressor variables remove error variance in other predictors. As a result, the variable “suppresses” irrelevant variance and increases the predictive ability of the other

TABLE 2. Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 1)

		$R^2_{y,123} = .7655$	% Explained Variance
Unique to $X_1$	<u>Attack Rate</u>	.1958	25.09%
Unique to $X_2$	<u>Modality</u>	.2099	26.89%
Unique to $X_3$	<u>Pitch Height</u>	.0259	3.31%
Common to $X_1$ and $X_2$	C (AR, Mo)	.3453	44.25%
Common to $X_1$ and $X_3$	C (AR, PH)	-.0218	-2.79%
Common to $X_2$ and $X_3$	C (Mo, PH)	.0477	6.12%
Common to $X_1$ , $X_2$ and $X_3$	C (AR, Mo, PH)	-.0224	-2.87%
Totals		.7655	100%

**TABLE 3. Commonality Analysis for Variance in Listener Ratings of Intensity (Experiment 1)**

		$R^2_{y,123} = .4939$	% Explained Variance
Unique to $X_1$	<u>Attack Rate</u>	.3098	59.03%
Unique to $X_2$	<u>Modality</u>	.0225	4.30%
Unique to $X_3$	<u>Pitch Height</u>	.0002	0.02%
Common to $X_1$ and $X_2$	C (AR, Mo)	.1867	35.57%
Common to $X_1$ and $X_3$	C (AR, PH)	.0196	3.74%
Common to $X_2$ and $X_3$	C (Mo, PH)	.0003	0.06%
Common to $X_1, X_2$ and $X_3$	C (AR, Mo, PH)	-.0143	-2.72%
Totals		.4939	100%

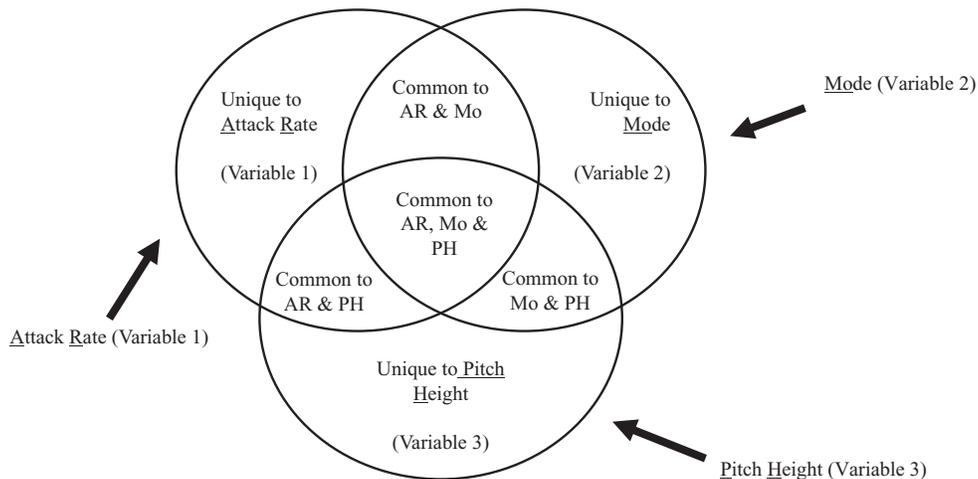
predictor and regression model overall (Cohen & Cohen, 1983; Capraro & Capraro, 2001).

*Cue contributions.* To further explore the relative strengths of each cue, we examined their unique and shared contributions to predictions of participant response (Figures 3 & 4) using commonality analysis. Uniquely, attack rate accounted for the largest amount of variance within both valence (25%) and intensity (59%) ratings. Mode uniquely accounted for 27% of variance within valence ratings, but only 4% in intensity ratings. Pitch height uniquely accounted for 3% of variance in valence ratings but did not meaningfully contribute (< 1%) to the intensity model.

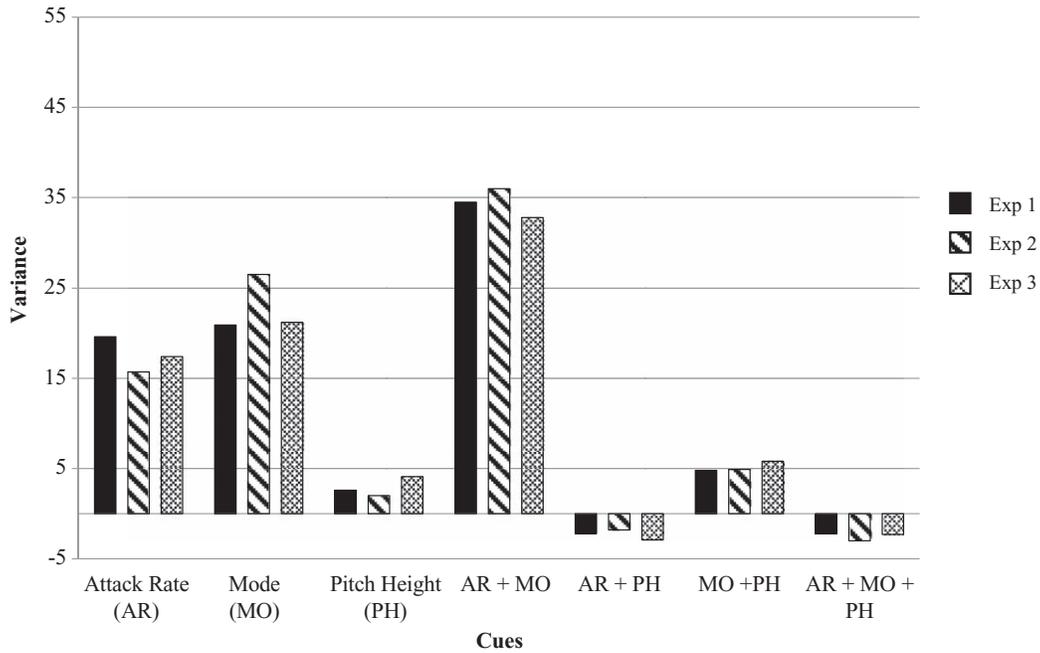
Shared variance accounted for a total of 45% of valence and 37% of intensity ratings, with the largest contribution from the relationship between attack rate and mode (44% contributed to valence model, 36% to intensity model). Mode and pitch height contributed 6% of shared variance to the ratings of valence (Table 2) but did not to contribute to ratings of intensity (Table 3). Attack rate and pitch height accounted for -3% of shared valence variance in contrast to 4% of shared intensity variance. Variance common between all three cues explained -3% and -3% of variance in the valence and intensity models respectively. Some researchers interpret negative commonalities as indicating confounding suppression effects (Beaton, 1973), whereas others postulate this suggests the predictor of interest has no influence (Frederick, 1999). Capraro and Capraro (2001) caution the interpretation of negative values for variance common to all predictors: they argue a negative commonality value for all cues combined suggests an inverse relationship to the dependent variable, in contrast to the direct relationships found for individual predictors. As this represents the first application of commonality analysis to the study of music, for our purposes we believe it best to follow the latter approach and focus on cues with positive values.

**DISCUSSION**

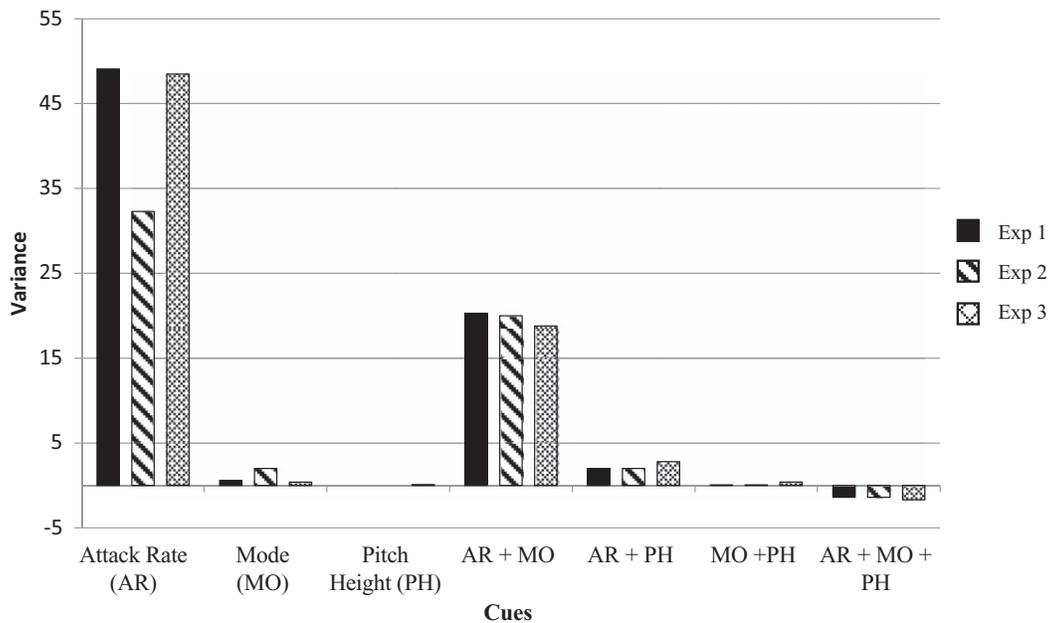
Our results are consistent with previous findings in both music and speech that faster attack rates lead to higher judgments of valence and intensity, suggesting faster delivery of acoustic information may convey more positive emotions (Breitenstein et al., 2001; Juslin, 1997). In



**FIGURE 2.** Visual representation of predictor relationships using Commonality Analysis as used here, adapted from original by Capraro and Capraro (2001).



**FIGURE 3.** Unique and shared variance of valence ratings by musical cue. The unique and shared contributions of attack rate, and mode cues explained the vast majority of variance across three experiments. The three bars for each cue depict ratings made of both (1) 8 measure excerpts (Experiment 1) and (2) variable length musically resolved excerpts using valence and intensity ratings (Experiment 2), as well as (3) 8 measure excerpts using valence and arousal ratings (Experiment 3).



**FIGURE 4.** Unique and shared variance of intensity/arousal ratings by musical cue. Attack rate's unique and shared contribution with mode explained the majority of variance across perceived ratings of intensity/arousal. The three bars for each cue depict ratings made of both (1) 8 measure excerpts (Experiment 1) and (2) variable length musically resolved excerpts using valence and intensity ratings (Experiment 2), as well as (3) 8 measure excerpts using valence and arousal ratings (Experiment 3).

contrast to work from Ilie and Thompson (2006) and Scherer and Oshinsky (1977), we found pitch height did not correlate with valence or intensity ratings, but appeared as a significant predictor within the three-cue regression model of valence ratings. Our analysis of the structural properties identified a correlation between mode and timing, consistent with previous findings that major key pieces tend to be faster than minor—both in these specific excerpts (Poon & Schutz, 2015), as well as more generally across a range of musical literature (Post & Huron, 2009). However, our results build on those outcomes by exploring perceptual evaluations of pieces varied in mode and timing. Additionally, they provide a useful converging measure to research using more constrained or systematically manipulated stimuli.

Attack rate significantly predicts listener ratings of both valence and intensity, indicating timing cues play an important role in both aspects of emotion. Both our linear regression and commonality analyses demonstrate timing as the most consistent predictor of emotional ratings. According to commonality analysis, attack rate uniquely predicted 25% of the total variance of valence ratings, and 59% of total intensity variance. Additionally, its shared contributions with mode predicted 44% of valence 35% of intensity variance. In contrast, pitch height contributed minimally (3% for valence, < 1% for intensity). While attack rate remained the most valuable cue for ratings of intensity, mode uniquely predicted more variance of valence ratings than attack rate. This holds important implications for performer's interpretation of the musical score, for unlike pitch and mode, performers' decisions regarding tempo directly affect timing cues such as attack rate, and a review of well-known recordings of this music demonstrates considerable disagreement in tempo interpretation. For example, Palmer (1994)'s review of tempi used in this set of pieces illustrates that Glenn Gould (Bach, 1722/1965) performed the E minor fugue (BWV 855) at twice the rate of Tureck (Bach, 1722/1953). Similarly, Newman (Bach, 1722/1973) performed the B minor prelude (BWV 869) at three times the rate of Gulda (Bach, 1722/1973). Finding that the cue most under control of performer interpretation plays a considerable role in emotion raises intriguing questions regarding the complex relationship between compositional structure and performer interpretation in shaping listeners' responses to musical passages.

Mode is typically regarded as an important cue for the perception of emotional valence (Hunter et al., 2008; Pallesen et al., 2005). Our findings are to some degree consistent with this view, as depicted by plotting the

mean rating of each piece across the circumplex space (Figure 1). Further, our statistical analyses illustrate that mode correlates with valence, with major key excerpts rated higher in valence and more intense. This is consistent with a large body of previous work, where major keys are commonly associated with positive valence in contrast to minor keys (Hevner, 1935). Regression analyses converge with the correlational results by finding this cue significantly predicted valence ratings. However, they also illustrate that it played little role in predicting intensity ratings despite a significant correlational trend (likely a reflection of Bach's use of faster attack rates for major key pieces). According to our assessment of relative cue strength, mode functioned as the strongest cue for valence ratings, and second for intensity ratings. Uniquely, it predicted more variance associated with valence ratings (26.89%) than intensity (4.30%). This demonstrates that while mode is important for distinctions of valence, it may not be informative in the perception of emotional intensity. Furthermore, our results suggest mode's contribution to listener ratings, specifically for emotional intensity, may be a function of its relationship to cues more crucial to this emotional dimension, such as attack rate.

The selective effect of mode is particularly intriguing given disagreement over mode's significance in emotional evaluation. Although many studies of musical emotion have found it plays a powerful role (Dalla Bella et al., 2001; Hunter et al., 2008; Webster & Weir, 2005), prominent music theorists suggest its role is minimal and may be the result of its correlation with other cues of musical structure (Hatten, 2004). Our results help to clarify some confusion over this important musical parameter by demonstrating mode's important role in listener perception of valence, but not intensity.

Research on music and speech suggest higher pitches correlate with positive valence (Breitenstein et al., 2001; Hevner, 1937). In contrast, here pitch height correlated with neither valence nor intensity. Furthermore, it had minimal predictive power in the commonality analysis. We suspect this difference may reflect the more complex role of pitch height in music with multiple voices and harmonic structure. Research on speech tokens often use a single voice for obvious reasons, and musical research exploring parallels often uses monophonic or single-voiced stimuli (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013). Although such simplified monophonic melodies provide a compelling parallel to speech, they share a tenuous connection to music that typically contains a great deal of pitch information beyond that of a single voice (i.e., polyphony, accompaniment, harmonic context, etc.).

Pitch height predicted valence ratings in the linear regression analysis (albeit to a lesser degree than other cues and cue combinations), but did not significantly predict intensity ratings. Although this might suggest some role for pitch height, our commonality analyses found it contributed minimally. Unique contributions of pitch height accounted for < 1% (intensity) and 3% (valence) of listener ratings. Therefore, we conclude pitch height holds limited predictive value within this corpus of complex, polyphonic music created by a renowned composer for musical—rather than research—purposes.

In summary, our regression findings are somewhat consistent with previous work indicating the role of timing, mode, and pitch in perceived emotion; however, we found minimal contribution of pitch for both valence and arousal ratings. Our findings also suggest cue importance varies as a function of emotional dimension. All three cues predicted valence ratings, yet only attack rate predicted intensity ratings. Mode and pitch height served as better predictors of valence rather than intensity. These results inform previous debates on the importance of timing and mode (Gagnon & Peretz, 2003; Juslin & Lindström, 2010), suggesting timing cues (quantified as attack rate) contribute more to expressed emotion than mode. Finally, the commonality analyses suggest attack rate is the strongest contributor to both dimensions,

Previous research suggests mode and timing cues are of high importance for the perception of emotion (Eerola, Friberg, & Bresin, 2013; Gagnon & Peretz, 2003), where mode strongly predicts emotional valence. Therefore, the dominance of timing contributions in both dimensions of Experiment 1 raises an important issue: Would better control over musical key changes improve the weight of mode in listener judgements of valence? To assess this issue, we conducted an additional experiment with musically ‘resolved’ excerpts.

### Experiment 2: Musically Resolved Excerpts

Our first experiment used eight-measure excerpts for all 48 pieces. Although this approach has the benefit of consistency, some pieces modulated to different keys by the end of the excerpt (i.e., the eighth measure of the C minor prelude outlines a C major chord). In order to explore whether this affected mode’s strength in Experiment 1, we ran a second experiment using excerpts ending in the piece’s nominal key (e.g., C major). This required variability across stimuli length (in measures) but offers a useful complementary perspective to the strict eight-measure durations of the first experiment, allowing for better insight into the relative strength of mode within this corpus of music. We then compared

these responses to revised pitch and timing information corresponding to the segment evaluated. For excerpts longer than eight measures we calculated the pitch height and timing of the additional measures; for excerpts shorter than eight measures we removed the measure in question from the pitch and timing calculations used to predict responses. For example, as this experiment used an 11-measure segment of the D minor prelude, we calculated pitch and timing information for three additional measures beyond the eight calculated previously.

### METHOD

We followed the same procedure as in the first experiment but used variable length (rather than eight measure) excerpts ending in the piece’s nominal key. Participants included 30 nonmusicians (< 1 year music training) undergraduate students (10 males,  $M = 18.3$  years,  $SD = 0.7$ ; 20 females,  $M = 18.8$  years,  $SD = 1.0$ ). They reported normal hearing and normal or corrected-to-normal vision. Musical stimuli ranged from 7–52 s ( $M = 25.4$  s,  $SD = 11.0$ ). Participants received course credit in return for participation.

*Cue quantification.* Pitch and timing information corresponded to the quantification of each cue within the specific number of measures required to reach a “resolution” back to the home key for each excerpt. In these excerpts, pitch height values varied from 33.13–53.13, corresponding to  $\sim F3$  to  $\sim C\#5$  ( $M = 43.87$ ,  $SD = 4.15$ ), attack rate information ranged 1.30–10.13 attacks/s ( $M = 4.87$ ,  $SD = 2.22$ ). We quantified mode the same way as in experiment 1 (0 = minor, 1 = major).

### RESULTS

Visualizations of ratings on Russell’s circumplex model appear in Figure 1b for ease of comparison with previous results. Similar to the first experiment, only one major prelude (B major) appeared in the lowest half of valence ratings, and only one major key fugue (C major) appeared in the lowest half of valence ratings. Similar to Experiment 2, Cronbach’s alpha for listener ratings appeared as  $\alpha = .97$  for both valence and intensity ratings, indicating high agreement across participants’ ratings. Participants’ valence rating ranged from 1.80–5.97 ( $M = 4.12$ ,  $SD = 1.20$ ) and intensity ratings ranged from 20.743–83.93 ( $M = 52.56$ ,  $SD = 18.20$ ).

*Correlations.* As we recalculated pitch and timing information for these variable length excerpts, we reran our original analysis of the acoustic cues. Despite these changes, we found similar to the first experiment a significant correlation between the cue of attack rate and mode,  $r(46) = .44$ ,  $p < .001$ . Pitch height

TABLE 4. Regression Model for Normalized Attack Rate, Mode, Pitch Height and Valence and Intensity Ratings (Experiment 2).

Predictor Coefficients	Valence				Intensity			
	B	SE	t	p	B	SE	t	p
Attack Rate	0.5031	0.0418	5.920	$p < .001$	0.4540	0.9521	5.513	$p < .001$
Mode	-0.5212	0.1831	-7.688	$p < .001$	-0.5863	4.1721	-1.546	$p = .129$
Pitch Height	0.1667	0.0203	2.121	$p < .05$	0.1457	0.4633	-0.044	$p = .965$

Note: Beta values indicate strength and direction of relationship between each predictor variable and valence and intensity ratings.

significantly correlated with neither attack rate,  $r(46) = -.17$ ,  $p = .26$ , nor mode,  $r(46) = .13$ ,  $p = .39$ . Similar to the first experiment,  $t$ -tests revealed a significant difference in attack rates,  $t(46) = -3.27$ ,  $p < .05$ , between the major and minor key pieces, but no significant difference in pitch height,  $t(45) = -0.86$ ,  $p = .39$ . Ratings of valence and intensity correlated significantly,  $r(46) = .78$ ,  $p < .001$ , which suggests these dimensions functioned in a dependent manner.

Attack rate,  $r(46) = .69$ ,  $p < .001$ , and mode,  $r(46) = .80$ ,  $p < .001$ , significantly correlated with valence ratings. Attack rate,  $r(46) = .72$ ,  $p < .001$ , and mode,  $r(46) = .46$ ,  $p < .001$ , also correlated significantly with intensity ratings. In contrast, pitch height significantly correlated with neither ratings of valence,  $r(46) = .15$ ,  $p = .32$ , nor intensity,  $r(46) = -.09$ ,  $p = .55$ .

*Regression analysis.* All three cues significantly predicted participants' valence ratings. However, only attack rate predicted intensity ratings (Table 4). In contrast to the correlational results, mode did not predict intensity ratings. We found no significant simple correlation between pitch height and ratings of valence; however, it significantly predicted listener judgements of valence in our regression model.

The three-cue predictor model accounted for 79% of the variance in valence ratings (adjusted  $R^2 = .789$ ),  $F(3, 44) = 59.73$ ,  $p < .001$ , and 51% of the variance in intensity ratings (adjusted  $R^2 = .514$ ),  $F(3, 44) = 17.58$ ,  $p < .001$ . Regression models investigating interaction effects show similar predictability in variance prediction for valence (adjusted  $R^2 = .790$ ),  $F(7, 40) = 26.22$ ,  $p < .001$ , and intensity (adjusted  $R^2 = .513$ ),  $F(7, 40) = 8.06$ ,  $p < .001$  (see Appendix B).

*Cue contributions.* As shown in Figures 3 and 4 (stripped bars), attack rate and mode accounted for the largest amount of unique variance within valence ratings (20% and 33% respectively). Attack rate remained the only important contributor of intensity ratings (58%), and mode uniquely accounted for 5% of the model's variance. Pitch height uniquely accounted for 3% of variance for valence ratings, and none for

intensity ratings. Shared variance explained 45% of total valence rating and 38% of total intensity rating variance. Attack rate and mode contributed the largest proportion of shared variance to both models (45% contributed to valence and 36% to intensity ratings). Variance shared between mode and pitch height contributed 6% to the valence model, but contributed less than 1% to intensity ratings. In contrast, calculations for the relationship between attack rate and pitch height produced a -2% contribution to valence and 4% to intensity ratings. The shared variance common between all three cues accounted for approximately -3% of the variance in valence and intensity models.

#### DISCUSSION

Similar to Experiment 1, Experiment 2 highlights the relationship between attack rate (timing) information and mode within listener ratings of emotion. Correlation and regression results followed the same trends as reported in Experiment 1. Attack rate and mode significantly correlated with valence and intensity, whereas pitch height significantly correlated with neither. Regression analyses indicated all three cues significantly

TABLE 5. Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 2)

		$R^2_{y,123} = .8029$	% Explained Variance
Unique to $X_1$	<u>A</u> ttack <u>R</u> ate	.1570	19.56%
Unique to $X_2$	<u>M</u> odality	.2648	32.99%
Unique to $X_3$	<u>P</u> itch <u>H</u> eight	.0202	2.51%
Common to $X_1$ and $X_2$	C (AR, Mo)	.3595	44.78%
Common to $X_1$ and $X_3$	C (AR, PH)	-.0181	-2.25%
Common to $X_2$ and $X_3$	C (Mo, PH)	.0493	6.14%
Common to $X_1$ , $X_2$ and $X_3$	C (AR, Mo, PH)	-.0299	-3.72%
Totals		.8029	100%

TABLE 6. Commonality Analysis for Variance in Listener Ratings of Intensity (Experiment 2)

		$R^2_{y,123} = .5452$	% Explained Variance
Unique to $X_1$	Attack Rate	.3141	57.61%
Unique to $X_2$	Modality	.0247	4.53%
Unique to $X_3$	Pitch Height	.0000	0.00%
Common to $X_1$ and $X_2$	C (AR, Mo)	.1987	36.44%
Common to $X_1$ and $X_3$	C (AR, PH)	.0217	3.97%
Common to $X_2$ and $X_3$	C (Mo, PH)	.0010	0.18%
Common to $X_1$ , $X_2$ and $X_3$	C (AR, Mo, PH)	-.0149	-2.73%
Totals		.5452	100%

predicted listener ratings of valence; however, only attack rate predicted arousal ratings. As in Experiment 1, this finding suggests the salience of cues as emotional indicators differ for the two dimensions.

Experiment 2 explored whether the influence of mode would increase when using excerpts starting and ending in the same nominal key. Although our findings here broadly mirrored those of the first experiment for valence ratings, the salience of musical mode increases when excerpts “resolve” (i.e., end in the same nominal key in which they began), with mode increasing in its predictive power and on attack rate decreasing in predictability (see Table 3). This manipulation did not affect all cues, as pitch height’s contribution remained small. As such, the results of Experiment 2 suggest mode’s predictive power is stronger when excerpts start and end in a consistent manner. This helps clarify mode’s power in complex passages containing chords outside the target mode (i.e., major chords in nominally minor keys and vice versa)—an approach that is common in actual musical practice, although complicated to rigorously assess under controlled laboratory conditions.

### Experiment 3: Arousal

The first two experiments quantified emotion employing an adaptation of Russell’s (1980) 2D circumplex model of affect, using dimensions of valence and intensity. The literature contains some disagreement over the best label for the non-valence dimension. For example, Trainor and Schmidt (2001) use “intensity,” whereas “arousal” is more common in other models (Russell, 1980; Schubert, 2004). As “intensity” is also used to describe the power, or physical characteristic of sound,

it is possible participants might have confused emotional intensity with sound intensity in our first two experiments. Therefore for the sake of thoroughness we ran a third experiment following the procedure and stimuli used in Experiment 1, but labeling the ratings scales valence and arousal rather than valence and intensity. This afforded exploration of the consequences of different approaches to labeling the dimension representative of emotional “energy,” and to ensure listeners’ understanding of the “energy” dimension in the first experiment had not been conflated with sound intensity.

### METHOD

We used an experimental procedure and cue quantification methods identical to the first experiment (matching excerpt length at eight measures); however, here participants rated perceived emotion on a scale of valence and a scale of arousal (rather than valence and intensity). In addition, cue quantification values remained identical to values calculated for Experiment 1. Although we used the label of “arousal” for the second dimension, the scale explanation given to participants remained identical to that of the “intensity” scale in Experiment 1. Participants included 30 undergraduate nonmusician (< 1 year music training) students (9 males,  $M = 20.8$  years,  $SD = 3.0$ ; 21 females,  $M = 21.2$  years,  $SD = 4.0$ ) with reported normal hearing and normal or corrected-to-normal vision. Musical stimuli ranged from 7–64 s ( $M = 30.2$  s,  $SD = 13.6$ ). Participants received course credit in return for participation.

### RESULTS

Visualizations of ratings on Russell’s (1980) circumplex model appear in Figure 1c for ease of comparison with previous results. Similar to the first experiment, only one major prelude (B major) appeared in the lowest half of valence ratings, however two (C major, B major) of the major key fugues appeared in the lowest half of valence ratings. Cronbach’s alpha for listener ratings demonstrated high internal consistency as in Experiment 1 and 2 ( $\alpha = .97$ ) for both valence and arousal ratings. Valence ratings from participants spanned from 1.97–6.30 ( $M = 4.09$ ,  $SD = 1.12$ ) and arousal ratings ranged 30.27–82.83 ( $M = 56.99$ ,  $SD = 16.95$ ).

*Correlations.* As the musical excerpts used in this experiment are identical to those of the first experiment, cue quantification analyses (intercue correlations and  $t$ -tests) are identical to those reported in Experiment 1. In terms of their relationship to perceptual ratings, similar to previous experiments, attack rate,  $r(46) = .67$ ,  $p < .001$ , and mode,  $r(46) = .76$ ,  $p < .001$ , correlated with valence ratings, with listeners giving higher rating to

TABLE 7. Regression Model for Normalized Attack Rate, Mode, Pitch Height on Valence and Arousal Ratings, as well as Mode (Experiment 3)

Predictor Coefficients	Valence				Arousal			
	B	SE	t	p	B	SE	t	p
Attack Rate	0.475	0.042	5.676	$p < .001$	0.630	0.915	5.253	$p < .001$
Mode	-0.524	0.186	-6.258	$p < .001$	-0.585	4.092	-1.276	$p = .272$
Pitch Height	0.210	0.021	2.759	$p < .01$	0.156	0.452	-0.426	$p = .672$

Note: Beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings.

faster, major-key pieces. Attack rate,  $r(46) = .70$ ,  $p < .001$ , and mode,  $r(46) = .42$ ,  $p < .01$ , also correlated with intensity ratings, suggesting listeners' judgement of intensity to be higher when pieces had higher attack rates and major key structures. As in Experiment 1 and 2, pitch height contrasted with the other cues, correlating with neither valence,  $r(46) = .22$ ,  $p = .13$ , nor intensity,  $r(46) = -.11$ ,  $p = .45$ , ratings.

*Regression analysis.* Standard linear multiple regression analysis revealed attack rate, mode, and pitch height contributed significantly towards valence ratings (Table 7). Analysis of arousal indicated attack rate as the only significant predictor (Table 7). Despite the correlation between mode and intensity, mode did not significantly predict intensity ratings in this regression analysis. However, although pitch height did not correlate significantly with valence, it significantly predicted valence ratings.

The 3-cue predictor model accounted for 48% of the variance in ratings of arousal (adjusted  $R^2 = .478$ ),  $F(3, 44) = 15.33$ ,  $p < .001$ , and 75% of the variance in ratings of valence (adjusted  $R^2 = .746$ ),  $F(3, 44) = 46.96$ ,  $p < .001$ . Despite a moderate correlation between attack rate and mode,  $r(46) = .445$ ,  $p < .01$ , tolerance and VIF values do not suggest multicollinearity (attack rate, tolerance = .773, VIF = 1.293; mode, tolerance = .772, VIF = 1.295). Regression models investigating interaction effects show a small increase in variance prediction for valence (adjusted  $R^2 = .771$ ),  $F(7, 40) = 23.55$ ,  $p < .001$ , and arousal (adjusted  $R^2 = .485$ ),  $F(7, 40) = 7.32$ ,  $p < .001$ , models; however, no interactions reached significance (see Appendix B).

*Cue contributions.* Attack rate accounted for the largest amount of unique variance within both valence (23%) and arousal (60%) ratings (Figures 3 and 4). Mode also contributed, accounting for 28% of unique valence variance, but only 4% of unique arousal variance. Uniquely, pitch height accounted for 5% of valence and less than 1% of arousal variance. The total shared variance across all cues accounted for 44% of valence and 36% of arousal ratings, primarily from the variance shared between attack rate and mode (43%

for valence ratings, 34% for intensity ratings). Mode and pitch height accounted for 8% of shared variance within the valence model and less than -1% within the arousal model. The shared variance contributed from attack rate and pitch height was -4% for valence ratings and 5% for ratings of arousal. Contribution from all three cues in variance contribution remained negative across both models, with -3% contributed to valence and -3% to arousal.

#### DISCUSSION

Our third experiment investigated the consequences of using different labels for the "energy" dimension of the circumplex model of emotion. Regression (Table 7) and commonality analyses (Tables 8 and 9) indicate minimal change from participant data collected in Experiment 1, where participants rated the valence and intensity of perceived emotion. The intensity regression model in Experiment 1 accounted for 49% (Table 3), whereas the regression result for arousal ratings in Experiment 3 accounted for approximately 48% (Table 9) of listener variance. This suggests both models

TABLE 8. Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 3)

		$R^2_{y.123} = .7620$	% Explained Variance
Unique to $X_1$	<u>A</u> ttack <u>R</u> ate	.1742	22.87%
Unique to $X_2$	<u>M</u> odality	.2119	27.80%
Unique to $X_3$	<u>P</u> itch <u>H</u> eight	.0412	5.40%
Common to $X_1$ and $X_2$	C (AR, Mo)	.3277	43.01%
Common to $X_1$ and $X_3$	C (AR, PH)	-.0285	-3.74%
Common to $X_2$ and $X_3$	C (Mo, PH)	.0581	7.62%
Common to $X_1$ , $X_2$ and $X_3$	C (AR, Mo, PH)	-.0226	-2.96%
Totals		.7620	100%

**TABLE 9.** Commonality Analysis for Variance in Listener Ratings of Arousal (Experiment 3)

		$R^2_{y,123} = .5111$	% Explained Variance
Unique to $X_1$	Attack Rate	.3066	59.98%
Unique to $X_2$	Modality	.0181	3.54%
Unique to $X_3$	Pitch Height	.0279	0.39%
Common to $X_1$ and $X_2$	C (AR, Mo)	.1740	34.04%
Common to $X_1$ and $X_3$	C (AR, PH)	.0279	5.45%
Common to $X_2$ and $X_3$	C (Mo, PH)	-.0018	-0.35%
Common to $X_1$ , $X_2$ and $X_3$	C (AR, Mo, PH)	-.0156	-3.06%
Totals		.5111	100%

similarly captured listener responses of perceived emotion within these musical excerpts.

Although this label change had little consequence, we felt it important to report for the sake of comparison with a wide range of existing research as both intensity (Trainor & Schmidt, 2001) and arousal (Russell, 1980; Schubert, 2004) appear in the literature. We believe this is helpful in contextualizing our results, for although some studies question the effectiveness of alternative 2D models to quantify emotion (Eerola & Vuoskoski, 2013; Schimmack & Grob, 2000), models based on valence and arousal are considered standard despite disagreement over the specifics of dimensional labels. Therefore, we simply conclude that our approach captures similar aspects of the perceived emotional “energy” in Experiments 1 and 3 regardless of the label used for the non-valence dimension.

### General Discussion

In this series of experiments, we explored the relationship between musical features and conveyed emotion using Bach’s *Well Tempered Clavier (WTC)*, a prominent composition by a well-respected composer. Here we build upon previous corpus analysis of Bach’s timing and pitch cues (Poon & Schutz, 2015) by empirically assessing their perceptual consequences. Complementing past work on highly emotive compositions such as film scores (Vuoskoski & Eerola, 2011) and familiar popular music (Yang & Chen, 2012), as well as tightly controlled manipulations to tone sequences (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013), our results shed new light on the ways in which listeners

respond to emotional cues when co-varying in a natural musical context. Cues such as attack rate and pitch height elicited affective consequences on listener judgments within musical stimuli in a manner complementing (though not always paralleling) those used in vocal expression. These findings are consistent with the view that music’s power to communicate emotion may derive from our capacity to process parallel features in speech.

According to our model, attack rate, mode, and pitch height significantly predict ratings of valence, consistent with work documenting the effects of mode and articulation on valence (Fabian & Schubert, 2003; Gabrielson & Lindström, 2001). Listeners also relied on attack rate (timing) cues to decode emotional intensity/arousal, common to results on speech and music (Ilie & Thompson, 2006; Schubert, 2004) in a manner aligned with past findings—higher pitch heights (Bachorowski, 1999; Hevner, 1937) and faster timings (Breitenstein, Van Lancker, & Daum, 2001; Juslin, 1997) are linked with positively valenced emotions in both speech and music. Our finding that attack rates predict both intensity and arousal is also consistent with previous work on music (Vieillard et al., 2008). This demonstrates that the relationships between cues and responses within unaltered passages of ecologically valid music is in some ways consistent with research using composed monophonic and polyphonic music (Schubert 2004; Vieillard et al., 2008). But here we document how Bach wove acoustic cues such as attack rate (timing) and mode together to shape emotional messages within complex polyphonic music.

A linear model built using only three cues derived from score-based analysis accounted for approximately 49–79% of the variance in participants’ ratings. Models incorporating more features such as loudness, tempo, melodic contour, texture and spectral centroid previously predicted 33–73% of perceived emotion within Romantic era music (Schubert, 2004). Our experiments employ music from a different era of musical style (Baroque), where relationships between cues such as mode and tempo differ from those in the Romantic era (Horn & Huron, 2015; Poon & Schutz, 2015). Despite differences in cue use across compositional styles, it is evident common cues such as attack rate (timing), and mode are pivotal in predicting participants’ perception of emotion within music.

Our models of emotional valence predicted more variance across experiments (approximately 75–79%) than of intensity/arousal (48–51%). This contrasts with work done on modelling listener’s perceived emotion, which predicts arousal better than valence (Eerola, 2011;

Eerola, Lartillot, & Toiviainen, 2009; Korhonen et al., 2006; Vuoskoski & Eerola, 2011). Cross validation analyses used to compare models across various empirically tested datasets including classical, film, pop, as well as mixed genre stimuli, also show higher prediction rates for the perceived arousal than perceived valence both across (16% valence, 43% arousal) and within (43% valence, 62% arousal) genres (Eerola, 2011). There, systematic feature selection and principal components analysis selected nine orthogonal features covering dynamic, rhythmic, timbral, and tonal aspects of the stimuli. Lower predictability for the intensity/arousal model may emerge in our results due to the lack of cues or features deemed “expressive.” We chose to quantify only three specific cues, two of which represent structural features within the music. Previous literature has shown a number of cues to be associated with emotional arousal, such as tempo (Husain, Thompson, & Schellenberg, 2002), articulation, and loudness (Schubert, 2004) or sound intensity (Dean, Bailes, & Schubert, 2011). Perhaps with the inclusion of these additional cues, our model of intensity/arousal might be more predictive.

Although Eerola’s (2011) analysis included more features, our models surprisingly demonstrated higher predictability—from essentially two cues. As mentioned above, the largest contribution occurred from the cue of attack rate, expressed as note attacks per second. Unlike that study, we extracted cues through score-based analysis, rather than via the MIRtoolbox program. Thus, even for theoretically similar features such as event density—determined from the detection of onsets from the peaks evident in the amplitude envelope with respect to attack time and slope—attack rate may capture something different. Additionally, within the datasets used in Eerola (2011), the “Classical” stimuli encompassed a large mix of orchestral/ensemble recordings as well as range of Western musical styles including Baroque, Romantic, etc. Our findings reflect perceived emotion from a set of polyphonic musical examples performed on one instrument, derived from one style and one composer. Furthermore, it is important to point out model comparisons across datasets using polyphonic music indicated a genre specificity for how well features predict valence, although less so for arousal. This highlights the difference between how valence and arousal can be expressed in music, but also the importance of exploring how cues function across styles of music, as core features of expressed emotion appear more effective for particular musical stimuli.

Our results also indicate the presence of interactions between musical features such as pitch height and attack

rate made only small contributions (approximately 1% to models in Experiments 1 and 3). Therefore, consistent with previous research investigating cues in monophonic stimuli (Eerola & Vuoskoski, 2013; Juslin & Lindström, 2010), the main driving effect of emotion perception appears driven by linear relationship between individual cues. However, it is possible that inclusion of other features would improve predictive power for intensity/arousal.

#### STRENGTH OF MUSICAL CUES

To assess relative cue strength, we used commonality analysis to calculate the unique and shared variance explained by three cues included in our model. Commonality analysis offers a powerful tool for picking apart contributions from the kinds of inter-related cues found in complex, composer created multi-voice stimuli. Variance partitioning of attack rate (timing), mode, and pitch height ultimately allow us to statistically compare how much each cue contributes and gives a sense of their musical importance in this experimental context.

*Timing.* Attack rate remained the strongest predictor of explained variance across valence, intensity, and arousal. This is consistent with research suggesting timing to be the most salient cue for emotion in music (Gagnon & Peretz, 2003), particularly for arousal (Schubert, 2004; Vieillard et al., 2008). The relationship observed between attack rate and arousal may stem from its general use in conveying information about energy. Attack rate describes the temporal rates of events, similar to rates of other behaviors such as speech, gait, etc. Thus, as faster speech and walking pace suggests more energy and energy expenditure from an individual (Gomez & Danuser, 2007), the rate at which the musical structure unfolds can reflect the energy expenditure of a performer giving the performance, or the association between event rate with the other biologically important rate cues may provide listeners with information about communicated energy.

Unlike pitch and mode, attack rate represents a cue reflecting contributions from both composer and performer. It accounts for the structural decisions of the composer in the form of number of note attacks per measure, as well as the performer’s choice of tempo for playing these rhythms. This suggests interesting future directions aimed at exploring the effect of different interpretation on musical communication. To some extent, the strength of timing here might reflect our use of musically untrained participants, who may have been less sensitive to mode, which requires specific musical knowledge or exposure to this type of music (Dalla Bella

et al., 2001). Thus, it remains an open question whether cue weights would differ substantially amongst musically trained individuals

*Pitch height.* In contrast to timing, pitch height played a smaller role (Figures 3 and 4), contributing minimally (0%–4.1% uniquely). It is possible that when hearing complex stimuli participants rely more on timing cues such as attack rate more than pitch height. Our use of “natural” stimuli admittedly poses challenges given the music complexity of polyphonic music created for artistic rather than scientific purposes. However, this approach arguably assesses the role of mode in a more realistic manner, as audiences frequently encounter music with more complex uses of mode mixing chord qualities than is found in tone sequences artificially constructed to focus on one type of mode.

Music with high pitch has previously been linked to affective terms of both high and low arousal (Scherer & Oshinsky, 1977; Wedin, 1972a). Models of listener responses to Bach’s *WTC* indicate pitch significantly predicted ratings of valence but not arousal. This is consistent with cross-cultural work using monophonic Hindustani ragas, showing pitch information in the form of pitch range did not help listeners outside of the musical culture interpret specific, discrete emotions (Balkwill & Thompson, 1999), as well as work comparing studies using multi-genre polyphonic music, revealing pitch did not fall in the top ten features predicting either dimension of affective space (Eerola, 2011).

Our results provide an interesting counterpart to a previous study by Schellenberg et al. (2000) showing pitch manipulations to be more influential than rhythmic manipulations on affect perception. Their results indicate pitch as more influential than timing when using novel, monophonic melodies performed by computers without harmonic context. Our contrary outcomes may reflect in part a different approach to timing; as our measure of attack rate considers the number of note onsets within the stimuli with respect to note durations, whereas they focus on manipulations of rhythmic structure. In addition, stimuli complexity may be a factor, as we employed the use of multi-voiced, polyphonic musical stimuli. Pitch height’s importance is likely greater in the context of single-lined melodies, when there are fewer voices and musical features. Therefore, our experiment provides insight as to how pitch height functions within a natural musical context. It is possible that attack density is an important aspect within conveyed musical emotion and as such, this study produces insight into how listeners disentangle communicated emotion within non-manipulated, passages containing the natural co-variation of cues.

*Mode.* Across all three experiments, the use of musically resolved excerpts (Experiment 2) led to the most accurate three cue model—perhaps in part due to increased predictive power of mode when ensuring excerpts started and ended in the same nominal key. Dalla Bella et al. (2001) reported that adults weighted mode more strongly than tempo, in contrast to children whose ratings seemed more reflective of tempo. Our results are inconsistent with those findings, as here ratings by adults showed timing had a similar influence as mode in valence, and a much more powerful influence on arousal/intensity. However, our task differs from theirs in many ways, including the structure of stimuli. Our musical excerpts contained the kinds of complex harmonic progressions characteristic of classical music, in contrast to short melodies designed to clearly signal major vs. minor. Further research clarifying mode’s role in harmonically complex passages similar to those written by great composers will help to clarify whether past findings on mode’s effect may not generalize to passages of natural music with complex harmonic structure.

Most pieces in this set mixed both major and minor chords, and some begin modulate (i.e., change their home key) within a few measures. Admittedly this makes analyses of mode more difficult than in excerpts constructed to clearly articulate only major or minor keys. These distinctions matter—our resolved excerpts in Experiment 2 attempted to control for these key changes, which resulted in an increase in mode’s role. There are of course, limitations in our method as it is not a perfect example of a singular mode/key; modulations and/or shifts may have occurred throughout the excerpt. Additionally, our musically resolved excerpts varied in duration from excerpts used in Experiment 1 and 2 (see Appendix C). As such, it is possible that stimuli duration played a role in the differences, and our conclusions should be interpreted in that light. Nonetheless, this trade-off is inevitable in evaluating music created for artistic rather than scientific purposes. These “problematically complex” passages are more representative of the kinds of progressions moving listeners in concert halls and home stereos on a regular basis. Consequently, we see our work balancing realism and control as a helpful complement to previous research on highly controlled tone sequences. For although our findings are to some degree consistent with previous demonstrations of manipulations to single line melodies lacking harmonic context, they raise interesting questions surrounding mode’s role in conveying affect within complex polyphonic compositions.

## MEASURING EMOTION

Two-dimensional models are frequently used to quantify emotion in research (Rodà, Canazza, & De Poli, 2014; Gomez & Danuser, 2004; Schubert, 1999, 2004). Using this method, emotions are broken down into components of varying degrees along the two dimensions, in contrast to discretely distinguishable categories. Russell's 2D circumplex model of affect (1980) is dominant in the field of emotion cognition, and considered a standard in emotional quantification. Despite general agreement on utilizing valence to evaluate musical affect (Carroll, Yik, Russell, & Barrett, 1999), researchers disagree over the best practice for additional dimension(s) (Schimmack & Grob, 2000; Vieillard et al., 2008). Previous studies use labels such as tension (Ilie & Thompson, 2006), activity (Leman et al., 2005) and/or strength (Luck et al., 2008).

To provide the most connection with the vast literature on musical emotion, our third experiment investigated the difference between various dimensions (arousal and intensity) as adapted from Russell's (1980) 2D model. Our results show each model accounted for similar amounts of variance across both dimensions. Small variations occurred between models where the valence/arousal model (Table 8 and 9) explained less of the variance within mean arousal and valence ratings than the valence/intensity model (Table 2 and 3). Overall these small differences suggest our experimental definition and use of "intensity" captures the second or "energetic" dimension of the 2D circumplex space, similar to "arousal." Given debate over the best measure of assessing emotion we believe this direct assessment of dimensional labeling is useful to note.

### Concluding Thoughts and Broader Implications

Together our experiments demonstrate the relative importance of attack rate (timing), mode, and pitch on emotional perception within a complex, composed musical corpus. This finding contributes to a growing literature on the relationships between cues and affective perception (Dalla Bella et al., 2001; Eerola, Friberg, & Bresin, 2013), by assessing cue contributions in a corpus of renowned musical pieces performed and heard frequently in concert halls around the world. The *WTC* was developed as a teaching tool for classical musicians, and is pedagogically in frequent use helping to refine a performer's expressive skills involving aspects like articulation, tempo and phrasing (Paggioli de Carvalho, 2016). Thus, a selection like the *WTC* affords further opportunity to explore cues expressed in its performance.

As our study focused on a corpus of music by one particular composer, future work using a broader corpus will further explore generalizations of these findings. However, this focused exploration of such a prominent set of pieces offers a unique opportunity to explore the effects of three structural cues for emotion as encountered in a natural musical context. Applying empirical scientific methods to assess emotional encoding and decoding of acoustic cues in complex, naturalistic music contributes to understanding listener perception in an everyday context. Previous literature reinforces careful consideration of mode's role in listener perception. It's significance in conveying emotion is frequently reported (Gagnon & Peretz, 2003; Hevner, 1935; Hunter, Schellenberg, & Schimmack, 2010); however, our results indicate mode's role is perhaps less straightforward in excerpts of natural music compared to musical cues such as attack rate. Intriguingly, this finding may help explain concerns voiced by music theorists that the view of major-as-happy is overly simplified, and is essentially an abstraction ignoring actual compositional practice (Hatten, 2004).

The results of the present study complement the relationships between perceived emotion and musical cues in the context of naturalistic musical stimuli. While the current study focused on Bach's *Well-Tempered Clavier*, future work could address music of other genres and time periods in order to determine whether these relationships change over centuries and continents. Insight into these changes can inform a deeper understanding between musical teaching practices and cognitive outcomes on a listener. In addition, the cues used within our models consist of predominantly composer-controlled features. Therefore, future studies should also consider performer-controlled cues and performer interpretation to further disentangle the connection between encoding and decoding within musical performances.

### Author Note

This research is supported in part by grants from Social Sciences and Humanities Research Council of Canada (SSHRC and the Canadian Foundation for Innovation (CFI)).

*Correspondence concerning this article should be addressed to Aimee Battcock, Department of Psychology, Neuroscience and Behaviour, Psychology Building (PC), Room 102, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada. E-mail: battcoae@mcmaster.ca*

## References

- BACH, J. S. (1953). *Bach: The Well-Tempered Clavier Book I - Prelude #24 in B minor, BWV 855* [CD; Recorded by R. Tureck]. RFA-Neef, Wittingen, Germany: Deutsche Grammonphon. (Original work published 1722)
- BACH, J. S. (1965). *Bach: The Well-Tempered Clavier, Book I - Prelude #10 in E minor, BWV 855* [CD; Recorded by G. Gould]. New York: Sony Classical. (Original work published 1722)
- BACH, J. S. (1973). *Bach: The Well-Tempered Clavier, Book I -* [CD; Recorded by F. Gulda]. MPS-Tonstudio, Villingen, Germany: Decca. (Original work published 1722)
- BACH, J. S. (1971). *Bach: The Well-Tempered Clavier, Book I - Prelude #10 in E minor, BWV 869* [CD; Recorded by A. Newman]. New York: Columbia Records. (Original work published 1722)
- BACHOROWSKI, J. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8, 53–57.
- BACHOROWSKI, J., & OWREN, M. (1995). Vocal expression of emotion: Intensity and context. *Psychological Science*, 6, 219–224.
- BALKWILL, L.-L., & THOMPSON, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, 17, 43–64.
- BEATON, A. E. (1973, March). *Commonality*. (ERIC Document Reproduction Service No. ED111829).
- BIGAND, E., VIEILLARD, S., MADURELL, F., MAROZEAU, J., & DACQUET, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and the duration of the excerpts. *Cognition and Emotion*, 19, 1113–1139.
- BOROD, J. C. (2000). *The neuropsychology of emotion* (1st ed.). Oxford, UK: Oxford University Press.
- BREITENSTEIN, C., VAN LANCKER, D., & DAUM, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, 15, 57–79.
- BROZE, Y., & SHANAHAN, D. (2013). Diachronic changes in jazz harmony: A cognitive perspective. *Music Perception*, 3, 32–45.
- CAPRARO, R. M., & CAPRARO, M. M. (2001) Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. *Multiple Linear Regression Viewpoints*, 27, 16–23.
- CARROLL, J. M., YIK, M. S., RUSSELL, J. A. & BARRETT, L.F. (1999). On the psychometric principles of affect. *Review of General Psychology*, 3, 14–22.
- COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- COLLIER, W. G., & HUBBARD, T. L. (2001). Musical scales and evaluations of happiness and awkwardness: Effects of pitch, direction, and scale mode. *American Journal of Psychology*, 114, 355–375.
- CORRIGALL, K. A., & TRAINOR, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, 17, 142–158.
- COSTA, M., FINE, P., ENRICO, P., & BITTI, P. E. R. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception*, 22, 1–14.
- CROWDER, R. G. (1985). Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society*, 23, 314–316.
- DALLA BELLA, S., PERETZ, I., ROUSSEAU, L., & GOSSELIN, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, 80(3), B1–B10.
- DEAN, R. T., BAILES, F., & SCHUBERT, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation. *PLoS ONE*, 6(4), e18591.
- DIBBEN, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception*, 22, 79–115.
- EEROLA, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, 40, 349–366.
- EEROLA, T., FRIBERG, A., & BRESIN, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, 4, 1–12.
- EEROLA, T., LARTILLOT, O., & TOIVIAINEN, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In K. Hirata, G. Tzanetakis, & K. Yoshii (Eds.), *Proceedings of the 10th International Conference on Music Information Retrieval* (pp. 621–626). Kobe, Japan: International Society for Music Information Retrieval.
- EEROLA, T., & VUOSKOSKI, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39, 18–49.
- EEROLA, T., & VUOSKOSKI, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, 30, 307–340.
- EITAN, Z., & TIMMERS, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114, 405–422.

- EKMAN, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 162–200.
- FABIAN, D., & SCHUBERT, E. (2003). Expressive devices and perceived musical character in 34 performances of Variation-7 from Bach's "Goldberg Variations." *Musicae Scientiae*, 7, 49–71.
- FREDERICK, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305–318). Stamford, CT: JAI Press.
- FRIBERG, A., SCHOONDERWALDT, E., HEDBLAD, A., FABIANI, M., & ELOWSSON, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *Journal of the Acoustical Society of America*, 136, 1951–1963.
- GABRIELSSON, A. (2002). Perceived emotion and felt emotion: same or different? *Musicae Scientiae*, 6, 123–148.
- GABRIELSSON, A., & LINDSTRÖM, E. (2010). The role of structure in the musical expression of emotions. In P. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). New York: Oxford University Press.
- GAGNON, L., & PERETZ, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition and Emotion*, 17, 25–40.
- GARDNER, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.
- GERARDI, G. M., & GERKEN, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception*, 12, 279–290.
- GOMEZ, P., & DANUSER, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, 53(2), 91–103.
- GOMEZ, P., & DANUSER, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7, 377–387.
- GUNDLACH, R. H. (1935). Factors determining the characterization of musical phrases. *American Journal of Psychology*, 47, 624–643.
- HAILSTONE, J. C., OMAR, R., HENLEY, S. M. D., FROST, C., KENWARD, M. G., & WARREN, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, 62, 2141–2155.
- HATTEN, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Bloomington, IN: Indiana University Press.
- HEINLEIN, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of Comparative Psychology*, 8, 101–142.
- HEVNER, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology*, 47, 103–118.
- HEVNER, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, 48, 246–268.
- HEVNER, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, 49, 621–630.
- HORN, K., & HURON, D. (2015). On the changing use of the major and minor modes 1750–1900. *Music Theory Online*, 21, 1–11.
- HUNTER, P. G., SCHELLENBERG, E. G., & SCHIMMACK, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, 22, 327–352.
- 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, 47–56.
- HURON, D., YIM, G., & CHORDIA, P. (2010). The effect of pitch exposure on sadness judgments: An association between sadness and lower than normal pitch. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition* (pp. 63–66). Seattle, WA: Causal Productions.
- HUSAIN, G., THOMPSON, W. F., & SCHELLENBERG, E. G. (2002). Effects of musical tempo and Mode on arousal, mood, and spatial abilities. *Music Perception*, 20, 151–171.
- ILIE, G., & THOMPSON, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23, 319–330.
- JOHNSTONE, T., & SCHERER, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotion* (Vol. 2, pp. 220–235). New York: Guilford.
- JUSLIN, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, 14, 383–418.
- JUSLIN, P. N., & LAUKKA, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814.
- JUSLIN, P. N., & LAUKKA, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33, 217–238.
- JUSLIN, P. N., & LINDSTRÖM, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, 29, 334–364.
- JUSLIN, P. N., & MADISON, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, 17, 197–221.
- KASTNER, M. P., & CROWDER, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception*, 8, 189–202.

- KOELSCH, S., KASPER, E., SAMMLER, D., SCHULZE, K., GUNTER, T., & FRIEDERICI, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7, 302–307.
- KORHONEN, M. D., CLAUSI, D. A., & JERNIGAN, M.E. (2006). Modeling emotional content of music using system identification. *IEEE Transactions on Systems, Man, and Cybernetics*, 36, 588–599.
- LAUKKA, P., EEROLA, T., THINGUJAM, N. S., YAMASAKI, T., & BELLER, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, 13, 434–449.
- LEMAN, M., VERMEULEN, V., DE VOOGDT, L., MOELANTS, D., & LESAFRE, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, 34, 39–67.
- LINDSTRÖM, E. (2006). Impact of melodic organization of melodic structure and emotional expression. *Musicae Scientiae*, 10, 85–117.
- LUCK, G., TOIVIAINEN, P., ERRKKILA, J., LARTILLOT, O., RIIKKILA, K., MAKELA, A., ET AL (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, 36, 25–45.
- MEYER, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.
- MOTE, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, 11, 618–622.
- PAGGIOLI DE CARVALHO, L. (2016). Bach's Well-Tempered Clavier: Pedagogical approaches and the different styles of preludes. *Per Musi*, 33, 97–115.
- PALLESEN, K. J., BRATTICO, E., BAILEY, C., KORVENOJA, A., KOIVISTO, J., GJEDDE, A., & CARLSON, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences*, 1060, 450–453.
- PALMER, W. A. (1994). *J. S. Bach: The Well-Tempered Clavier* (Third). Los Angeles, CA: Alfred Music Publishing.
- PEDHAZUR, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace.
- PEIRCE, J., GRAY, J. R., SIMPSON, S., MACASKILL, M., HÖCHENBERGER, R., SOGO, H., ET AL. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203.
- POON, M., & SCHUTZ, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, 6, 1–13.
- POST, O., & HURON, D. (2009). Western Classical music in the minor mode is slower (except in the Romantic period). *Empirical Musicology Review*, 4, 2–10.
- QUINTO, L., THOMPSON, W. F., & KEATING, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, 4, 1–8.
- RAY-MUKHERJEE, J., NIMON, K., MUKHERJEE, S., MORRIS, D. W., SLOTOV, R., & HAMER, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, 5, 320–328.
- RIGG, M. G. (1940). Speed as a determiner of musical mood. *Journal of Experimental Psychology*, 27, 566–571.
- RODÀ, A., CANAZZA, S., DE POLI, G., & ROD, A. (2014). Clustering affective qualities of classical music: Beyond the valence-arousal plane. *IEEE Transactions on Affective Computing*, 5, 364–376.
- RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- SCELLENBERG, E. G., KRYSIAK, A. M., & CAMPBELL, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception*, 18, 155–171.
- SCHERER, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, 9, 235–248.
- SCHERER, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227–256.
- SCHERER, K. R., & OSHINSKY, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1, 331–346.
- SCHIMMACK, U., & GROB, A. (2000). Dimensional model of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14, 325–345.
- SCHUBERT, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, 51, 154–165.
- SCHUBERT, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, 21, 561–585.
- SCHUTZ, M. (2017). Acoustic constraints and musical consequences: Exploring composers' use of cues for musical emotion. *Frontiers in Psychology*, 8.
- TAN, S. L., PFORDRESHER, P. Q., & HARRÉ, R. (2010). *Psychology of music: From sound to significance* (1st ed.). New York: Psychology Press.
- TEMPERLEY, D., & DE CLERCQ, T. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, 42, 187–204.
- TRAINOR, L. J., & SCHMIDT, L. A. (2001). Processing emotions induced by music. *Cognition and Emotion*, 15, 487–500.

- VIEILLARD, S., PERETZ, I., GOSSELIN, N., KHALFA, S., GAGNON, L., & BOUCHARD, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion*, 22, 720–752.
- VUOSKOSKI, J. K., & EEROLA, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, 15, 159–173.
- WATSON, K. B. (1942). The nature and measurement of musical meanings. *Psychological Monographs*, 54(2), i–43.
- WEBSTER, G. D., & WEIR, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, 29, 19–39.
- WEDIN, L. (1969). Dimension analysis of emotional expression in music. *Swedish Journal of Musicology*, 51, 119–140.
- WEDIN, L. (1972a). A multidimensional study of perceptual - emotional qualities in music. *Scandinavian Journal of Psychology*, 13, 241–257.
- WEDIN, L. (1972b). Evaluation of a three-dimensional model of emotional expression in music. *The Psychological Laboratories*, 54(349), 1–17.
- WIGGINS, G. A. (1998). Music, syntax, and the meaning of “meaning.” In *Proceedings of the First Symposium on Music and Computers* (pp. 18–23). Corfu, Greece: Ionian University.
- YANG, Y. H., & CHEN, H. H. (2012). Machine recognition of music emotion. *ACM Transactions on Intelligent Systems and Technology*, 3(3), 1–30.

Appendix A

Music Training Survey

Date: \_\_\_\_\_ Experiment Num: \_\_\_\_\_ Participant Num: \_\_\_\_\_  
(dd/mm/yy)

- 1) What is your age in years at the time of this study: \_\_\_\_\_
- 2) Please list your year of study at McMaster (i.e. first year undergraduate student, second year graduate student, etc). \_\_\_\_\_
- 3) Gender - I am (circle one): male      female      transgendered
- 4) Do you consider yourself proficient on a musical instrument, and if so, which one?  
    \_\_\_ Yes – if so, please list instrument(s): \_\_\_\_\_  
    \_\_\_ No
- 5) Have you taken private musical lessons  
    \_\_\_ Yes – if so, for how many years: \_\_\_\_\_  
    \_\_\_ No
- 6) Approximately how many hours a week do you spend practicing/performing/jamming (as opposed to listening) to music?  
    \_\_\_ (0-1)      \_\_\_ (1-5)      \_\_\_ (5-10)      \_\_\_ (10-15)      \_\_\_ (15-20)      \_\_\_ (> 20)
- 7) Approximately how many hours a week do you spend listening to (as opposed to practicing or performing) music?  
    \_\_\_ (0-1)      \_\_\_ (1-5)      \_\_\_ (5-10)      \_\_\_ (10-15)      \_\_\_ (15-20)      \_\_\_ (> 20)
- 8) Do you own an iPod or personal music listening device?  
    \_\_\_ Yes      \_\_\_ No
- 8a) If you answered yes to question 8, what kind of device do you own:  
    \_\_\_\_\_
- 8b) If you answered "yes" to question 8, approximately how many songs do you have on your device?  
    \_\_\_\_\_

## Appendix B

## Experiment 1 Regression Tables with Interaction Terms

TABLE B1.

Predictor Coefficients	Valence				Intensity			
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<u>Attack Rate</u>	0.4799	0.0459	10.45	$p < .001$	0.6678	0.0747	8.94	$p < .001$
<u>Mode</u>	-0.3283	0.0438	-7.50	$p < .001$	-0.0363	0.0711	-0.5	$p = .613$
<u>Pitch Height</u>	0.1387	0.0553	2.51	$p < .05$	0.0201	0.0899	0.22	$p = .824$
AR x Mo	0.0918	0.0460	2.00	$p = .053$	0.0751	0.0747	1.01	$p = .321$
AR x PH	0.0679	0.0443	1.53	$p = .133$	0.1856	0.0720	2.58	$p < .05$
Mo x PH	-0.0160	0.0553	-0.29	$p = .774$	0.1040	0.0899	1.16	$p = .254$
AR x PH x Mo	0.0096	0.0443	0.22	$p = .830$	0.0420	0.072	0.58	$p = .564$
Adjusted R <sup>2</sup>	0.879				0.693			

## Experiment 2 Regression Table with Interaction Terms

TABLE B2.

Predictor Coefficients	Valence				Arousal			
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<u>Attack Rate</u>	0.4071	0.0742	5.13	$p < .001$	0.6341	0.1181	5.37	$p < .001$
<u>Mode</u>	0.7044	0.0988	9.40	$p < .001$	0.3029	0.1116	2.71	$p < .01$
<u>Pitch Height</u>	0.0702	0.0988	0.71	$p = .48$	0.0272	0.1471	0.185	$p = .854$
AR x Mo	0.0226	0.0802	0.28	$p = .779$	0.0998	0.1194	0.836	$p = .408$
AR x PH	-0.0457	0.1015	-0.45	$p = .655$	-0.0266	0.1152	-0.176	$p = .861$
Mo x PH	0.0901	0.0998	0.90	$p = .372$	0.0588	0.1486	0.396	$p = .694$
AR x PH x Mo	0.0575	0.1023	0.56	$p = .578$	-0.1161	0.1528	-0.760	$p = .412$
Adjusted R <sup>2</sup>	0.7715				0.4934			

## Experiment 3 Regression Table with Interaction Terms

TABLE B3.

Predictor Coefficients	Valence				Arousal			
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<u>Attack Rate</u>	0.6440	0.0689	9.36	$p < .001$	0.8954	0.0999	8.96	$p < .001$
<u>Mode</u>	0.4399	0.0656	6.70	$p < .001$	0.0265	0.0952	0.28	$p = .782$
<u>Pitch Height</u>	0.2536	0.0829	3.06	$p < .01$	0.0226	0.1203	0.19	$p = .852$
AR x Mo	-0.1050	0.0695	-1.52	$p = .139$	-0.1014	0.1009	-1.00	$p = .321$
AR x PH	0.0938	0.0670	1.40	$p = .169$	0.1982	0.0973	2.04	$p < .05$
Mo x PH	0.0272	0.0838	0.33	$p = .747$	-0.0389	0.0122	-0.32	$p = .750$
AR x PH x Mo	-0.0382	0.0677	-0.56	$p = .576$	-0.0673	0.0983	0.684	$p = .498$
Adjusted R <sup>2</sup>	0.8558				0.6962			

## Appendix C

## Experiment Stimuli Durations (in seconds)

TABLE C1.

Piece	Experiment 1 & 3 Durations	Experiment 2 Durations
Fugue 1	00:54	00:35
Fugue 2	00:34	00:36
Fugue 3	00:21	00:17
Fugue 4	00:19	00:19
Fugue 5	00:36	00:20
Fugue 6	00:25	00:18
Fugue 7	00:22	00:21
Fugue 8	00:34	00:31
Fugue 9	00:18	00:17
Fugue 10	00:14	00:32
Fugue 11	00:10	00:14
Fugue 12	00:46	00:51
Fugue 13	00:35	00:28
Fugue 14	00:52	00:52
Fugue 15	00:16	00:28
Fugue 16	00:39	00:29
Fugue 17	00:41	00:43
Fugue 18	01:03	00:34
Fugue 19	00:17	00:17
Fugue 20	00:25	00:23
Fugue 21	00:15	00:20
Fugue 22	00:23	00:26
Fugue 23	00:49	00:49
Fugue 24	00:54	00:43
Prelude 1	00:30	00:14
Prelude 2	00:33	00:15
Prelude 3	00:08	00:08
Prelude 4	00:25	00:30
Prelude 5	00:17	00:24
Prelude 6	00:26	00:33
Prelude 7	00:25	00:30
Prelude 8	00:52	00:24
Prelude 9	00:30	00:27
Prelude 10	00:38	00:17
Prelude 11	00:27	00:07
Prelude 12	00:37	00:10
Prelude 13	00:29	00:29
Prelude 14	00:19	00:26
Prelude 15	00:24	00:30
Prelude 16	00:40	00:14
Prelude 17	00:20	00:21
Prelude 18	00:29	00:12
Prelude 19	00:19	00:18
Prelude 20	00:23	00:12
Prelude 21	00:24	00:28
Prelude 22	01:05	00:50
Prelude 23	00:27	00:20
Prelude 24	00:44	00:22

## Correlations between Stimuli Durations and Listener Ratings

TABLE C2.

Experiment	Valence Ratings		Intensity/Arousal Ratings	
	<i>r</i> Value	<i>p</i> Value	<i>r</i> Value	<i>p</i> Value
1	-.64	< .001	-.61	< .001
2	-.42	< .01	-.38	< .01
3	-.62	< .001	-.63	< .001