

PSYCHOPHYSIOLOGICAL INDICES OF MUSIC-EVOKED EMOTIONS IN MUSICIANS

MATTSON OGG & DAVID R. W. SEARS
McGill University, Montreal, Canada

MANUELA M. MARIN
University of Innsbruck, Innsbruck, Austria

STEPHEN McADAMS
McGill University, Montreal, Canada

A NUMBER OF PSYCHOPHYSIOLOGICAL MEASURES indexing autonomic and somatovisceral activation to music have been proposed in line with the wider emotion literature. However, attempts to replicate experimental findings and provide converging evidence for music-evoked emotions through physiological changes, overt expression, and subjective measures have had mixed success. This may be due to issues in stimulus and participant selection. Therefore, the aim of Experiment 1 was to select musical stimuli that were controlled for instrumentation, musical form, style, and familiarity. We collected a wide range of subjective responses from 30 highly trained musicians to music varying along the affective dimensions of arousal and valence. Experiment 2 examined a set of psychophysiological correlates of emotion in 20 different musicians by measuring heart rate, skin conductance, and facial electromyography during listening without requiring behavioral reports. Excerpts rated higher in arousal in Experiment 1 elicited larger cardiovascular and electrodermal responses. Excerpts rated positively in valence produced higher zygomaticus major activity, whereas excerpts rated negatively in valence produced higher corrugator supercilii activity. These findings provide converging evidence of emotion induction during music listening in musicians via subjective self-reports and psychophysiological measures, and further, that such responses are similar to emotions observed outside the musical domain.

Received: July 5, 2016, accepted January 30, 2017.

Key words: psychophysiology, emotion, measurement, arousal, valence

ANY DISCUSSION OF MUSIC'S UBIQUITY, utility, or privileged status among human societies must inevitably make mention of its apparent capacity to induce emotions in listeners. Emotional responses to music have been a popular area of research in many disciplines ranging from marketing and behavioral therapy to experimental psychology (Juslin & Sloboda, 2010a), where music's effectiveness in emotion regulation is frequently cited as its primary function (Egermann, Pearce, Wiggins, & McAdams, 2013; Juslin & Sloboda, 2010a; Zatorre & Salimpoor 2013). However, emotional reactions are often highly individualized experiences that remain difficult to quantify (Hodges, 2010, 2016; Juslin & Sloboda, 2010b; Juslin & Västfjäll, 2008). One prominent view espoused by Scherer (2004, 2005) emphasizes the need to link emotional experiences with bodily indices that can be objectively measured, such as facial expressions and psychophysiological reactions to music. Here, our aim is to study the nature of musical emotions through the use of rigorous control in selecting our stimuli and participants along with validation using behavioral data (Experiment 1). We also independently examine arousal and valence to identify a large set of psychophysiological measurements that can be used to index emotional experiences during music listening (Experiment 2).

Defining and Conceptualizing Musical Emotions

The development of reliable measurement techniques for music-evoked emotions has been marred by foundational issues facing emotion research in general, and particularly by the lack of a widely accepted definition for emotion (Juslin & Sloboda, 2010b; Mulligan & Scherer, 2012). This ambiguity is problematic for music-evoked emotions, as certain working definitions can be restrictive or exclusive, which perhaps stem from critiques that music by itself cannot evoke emotions at all (Konečni, 2008), or that music-evoked emotions are distinct from typical, everyday emotions due to their lack of a biological imperative (Scherer, 2004).

Other impediments to the development of reliable measures for music-evoked emotional responses follow

from unresolved issues regarding the conceptual framework underlying emotion. Discrete models of emotion focus on a small number of basic responses, such as joy, sadness, fear, and anger (Ekman, 1992; Izard, 1992). Conversely, dimensional models (Wundt, 1912) have demonstrated using factor analysis that emotions can be characterized by their position among orthogonal and bipolar axes representing valence (pleasure vs. displeasure) and arousal (high vs. low alertness) (Russell, 1980). While there are shortcomings with both the discrete and dimensional approaches (Scherer, 2004, 2005), studies of music-evoked emotions tend to support a two-dimensional framework (Hevner, 1936; Krumhansl, 1997), even when the two models are directly compared (Eerola & Vuoskoski, 2011). Critics of dimensional approaches also acknowledge their applicability to physiological investigations (Scherer, 2004). Thus, our study provides new knowledge by adopting a two-dimensional framework in concert with diverse psychophysiological measures (cf. Gingras, Marin, Puig-Waldmüller, & Fitch, 2015) to better examine multiple components of music-evoked emotional experiences.

Although there is evidence supporting a two-dimensional arousal-valence model of emotion in the musical domain, there is ambiguity regarding the semantic labels assigned to these dimensions. Some critics have suggested that the use of pleasantness as a label for the valence dimension (Russell, 1980) has connotations that may not be in line with the underlying structure of this dimension (Colombetti, 2005; Mulligan & Scherer, 2012). The term pleasantness emphasizes hedonic value or tone, which may not necessarily be endemic to a positively valenced emotional experience (Colombetti, 2005). Some investigations of emotions evoked by music have queried responses along a pleasantness-unpleasantness axis to index valence in behavioral paradigms (Gingras, Marin, & Fitch, 2014; Ilie & Thompson 2006; Khalfa, Roy, Rainville, Dalla Bella, & Peretz, 2008; Krumhansl, 1997; Marin, Gingras, & Bhattacharya 2012; Marin & Leder, 2013), whereas others simply use a negative-positive valence axis directly (Eerola & Vuoskoski 2011; Grewe, Nagel, Kopiez, & Altenmüller, 2007; Nagel, Kopiez, Grewe, & Altenmüller, 2007). Valence rating scales are also more common than pleasantness scales in psychophysiological studies of visual art (Gerger, Leder, & Kremer, 2014). Clearly, the issue of what semantic label to ascribe to this axis of affect as a means of navigating the underlying valence-arousal space is unresolved. Thus, the best course of action for further study is to adopt an exploratory, empirically guided approach by considering both pleasantness and valence in the research design.

Measuring Music-evoked Emotions: Limits of Subjective Self-reports

Despite a lack of consensus within the research community as to how we might define and conceptualize emotional reactions, many authors agree that an emotional episode consists of the presence and convergence of a series of measurable components, sometimes termed the emotion response triad: physiological changes, overt expression, and subjective feeling (Bradley & Lang, 2000; Izard, 1992; Scherer, 2004). During subjective tests of emotion in music listening, however, participants have difficulty distinguishing emotions *perceived* (or recognized) in the music from a genuine emotional response *felt* during listening (Kivy, 1990; Konečni, 2008). The prevalence of emotional responses in subjective rating tasks may also be over-reported due to the demand characteristics inherent in a forced-choice response (Konečni, 2008).

Measures that can achieve a more implicit index of emotion provide an indispensable complement to subjective self-reports of emotion (Hodges, 2010). Demonstrating that emotions elicited solely by music can provoke synchronized changes in the emotion response triad (Cochrane, 2010; Konečni, 2008) would constitute strong evidence that the nature of musical emotions is not fundamentally different from everyday emotions (Scherer, 2004), and further, that musical emotions are truly felt by the listener rather than solely perceived (Kivy, 1990). To date, these concerns have not been adequately addressed and warrant further investigation.

Measuring Music-evoked Emotions: Psychophysiological Approaches

To circumvent limitations in subjective, self-report responses and to address other components of the emotional experience, investigators have examined autonomic and somatic activity during affective reactions to musical stimuli. Most studies employ electrodermal, cardiovascular, and facial electromyographic measures (Bradley & Lang, 2007), which correspond to the physiological arousal (electrodermal and cardiovascular) and overt expression (facial electromyography) components of the emotion response triad (Scherer, 2004, 2005). Historically, these measures have provided quite reliable indicators of emotional reactions to pictures in the visual domain (Brown & Schwartz, 1980; Larsen, Norris, & Cacioppo, 2003).

Measures of electrodermal activity (also referred to as skin conductance) are common in the general psychophysiological literature on emotion. They enjoy a long

history as a reliable indicator of physiological changes in sympathetic nervous system activation during emotional arousal (Bouscein et al., 2012; Dawson, Schell, & Fillion, 2007). Cardiovascular measures have also been widely used as indices of autonomic activation and emotional responding, although this system is innervated by both the sympathetic and parasympathetic branches of the autonomic nervous system (Berntson et al., 1997; Berntson, Quigley, & Lozano, 2007).

Measures of electrodermal and cardiovascular activity capture arousal components of emotional responses, but generally fail to distinguish the valence axis of dimensional models or external expressions of emotion (Scherer, 2004). To measure these dimensions, investigators have employed electromyography recordings over the zygomaticus major and corrugator supercilii muscles (Bradley & Lang, 2007; Larsen et al., 2003; Van Boxtel, 2001). When recording electromyography as an index of emotional responses to music, nascent electrical activity of a smile (zygomaticus major) or furrowed brow (corrugator supercilii) often accompanies subjective positive or negative emotional reactions, respectively (Hodges, 2010; Tassinari, & Cacioppo, 1992). Use of these measures assumes that when psychologically engaged in a task or experience, minute variations in muscle activity occur outside of conscious awareness (Tassinari, & Cacioppo, 1992), in preparation for an overt response (Tassinari, Cacioppo, & Vanman, 2007).

The measures outlined above, although widely used to measure emotional responses in general, have not been as effective in characterizing music-evoked emotional responses (Hodges, 2010). Electrodermal measures are sensitive to a variety of musical characteristics, such as emotional expressiveness (Vieillard, Roy, & Peretz, 2012), tempo, genre (Dillman Carpentier & Potter, 2007), and unexpectedness (Egermann et al., 2013; Steinbeis, Koelsch, & Sloboda, 2006), but they have most commonly been shown to be associated with highly emotionally arousing musical stimuli (Gomez & Danuser, 2004; Khalfa et al., 2008; Lundqvist, Carlsson, Hilmerston, & Juslin 2009; Nater, Abbruzzese, Krebs, & Ehlert, 2006; Rickard, 2004). However, a number of studies have failed to find any influence of music on electrodermal activity (Blood & Zatorre, 2001), have reported an inconsistent pattern of response between activity and emotional ratings (White & Rickard, 2015), or have attributed electrodermal activity to orienting responses from novelty or audible change in the stimuli rather than to emotional arousal from the music per se (Chuen, Sears, & McAdams, 2016; Grewe et al., 2007). A similar picture exists concerning cardiovascular measures: some studies report effects of arousal

increasing cardiovascular activity (Blood & Zatorre, 2001; Rickard, 2004; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009; Witvliet & Vrana, 2007), whereas others report no effect (Guhn, Hamm, & Zentner, 2007; Lundqvist et al., 2009), or even cite decreased cardiovascular activity during reports of highly arousing emotions (White & Rickard, 2015). Conflicting and null results also permeate electromyographic findings: some studies find a complementary effect of valence between zygomaticus and corrugator activity (Witvliet & Vrana, 2007), whereas others find only an effect for corrugator (Viellard et al., 2012) or zygomaticus responses (Khalfa et al., 2008; Lundqvist et al., 2009), or find no significant response at all (Egermann et al., 2013; Grewe et al., 2007).

There is a clear discrepancy between the equivocality of findings in the musical emotion literature and the reliability of these psychophysiological measurements in the wider emotion literature. This may call into question whether music-evoked emotional experiences can be characterized by these psychophysiological measures at all and lends support to suggestions that music-evoked emotional responses may be of a different sort than so-called “everyday” emotions. However, we, along with others (Hodges, 2016), suggest that these issues may instead stem from a lack of standardization in the collection of psychophysiological data or the selection of stimuli, which may vary in genre, stimulus rendition style (recorded or synthesized), and duration. We address these issues in the present investigations.

Considerations Regarding Stimulus Selection

One potentially pervasive set of issues in the literature may stem from the stimuli used in previous studies. For example, familiarity and liking are frequently found to be associated (Grewe et al., 2007; Parncutt & Marin, 2006) and can mediate physiological responses (Grewe, Kopiez, & Altenmüller, 2009; Panksepp, 1995; Van den Bosch, Salimpoor, & Zatorre, 2013). However, studies often do not control for these effects (Grewe et al., 2007; Panksepp, 1995; Rickard, 2004). Additionally, peak emotional responses to music—particularly intense pleasurable reactions (Zatorre & Salimpoor, 2013)—are often highly individualized and associated with autobiographical events (Blood & Zatorre, 2001; Panksepp, 1995; Rickard, 2004; Salimpoor et al., 2009). It is therefore imperative that researchers control for familiarity, liking, and especially intense emotional responses in order to yield results based solely on musical factors and not on extramusical associations.

Studies in this area also frequently employ diverse stimulus sets ranging from loud rock music (Dillman

Carpentier & Potter, 2007; Gomez & Danuser, 2004) to dance (Grewe et al., 2007) to classical music of diverse orchestrations (Witvliet & Vrana, 2007). Previous studies on emotional effects of timbre and orchestral arrangement have shown these parameters to be a major factor in some emotional reactions to music (Hailstone et al., 2009; Nagel, Kopiez, Grewe, & Altenmüller, 2008). In an effort to control for these effects, some studies have generated stimuli specifically to meet the needs of their investigation (Lundqvist et al., 2009; Vieillard et al., 2012), but these stimuli run the risk of lacking ecological validity (Dowling, 1989).

It is also important that the full ranges (both positive and negative aspects) of the arousal and valence dimensions are represented among stimuli and that these dimensions can be analyzed independently (White & Rickard, 2015). Some studies have only examined “happy” and “sad” musical excerpts, which truncates the range of emotions that music and physiological measures can represent, and could contribute to the inconsistencies in the literature (Hodges, 2016; Lundqvist et al., 2009; White & Rickard, 2015). For example, such a reduction in dimensionality makes it difficult to draw conclusions regarding a measure’s correspondence with arousal or valence independent of the other. In summary, there are issues with stimulus selection in previous work that concern familiarity, stylistically heterogeneous, or nonecologically valid stimulus sets, and independent variables that do not fully represent an arousal-valence emotion space. These issues may have contributed to discrepant findings regarding the efficacy of widely used psychophysiological measures in the domain of music-evoked emotions. This may also have occluded any demonstration of convergence among branches of the emotion response triad. To ameliorate these issues, we introduce a more rigorous and ecologically valid level of control in line with the recent work of Gingras and colleagues (2014, 2015) and Marin and colleagues (2012; Marin & Leder, 2013).

Considerations Regarding Participant Selection

A growing body of work suggests that music training may bestow a variety of benefits to auditory processing (Kraus et al., 2014; Zhao & Kuhl, 2016) and auditory abilities (Slater et al., 2015; Tierney, Krizman, & Kraus 2015), although these effects are often subtle (Bigand & Poulin-Charronnat, 2006). Detailed examinations such as decoding the emotional content of speech prosody indicate that music training may be associated with improvements in detecting (Thompson, Schellenberg, & Husain, 2004) and processing (Strait, Kraus, Skoe,

& Ashley, 2009) emotional expressions. Given the additional evidence suggesting an overlap between the processing of emotional expressions in music and speech (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Thompson, Marin, & Stewart, 2012), perhaps music training increases a listener’s sensitivity to detecting and experiencing emotion in music as well. Indeed, if one wanted to demonstrate converging evidence for music-evoked emotions in line with the emotional response triad (Scherer, 2004, 2005), musicians may be an ideal population to study. However, the majority of physiological studies of music-evoked emotions ignore this potential aptitude (Grewe et al., 2007; Khalfa et al., 2008; Lundqvist et al., 2009; Rickard, 2004; Van den Bosch et al., 2013; Witvliet & Vrana, 2007). Additionally, drawing from a participant pool of highly trained musicians might yield a more homogeneous sample and limit the variability associated with recruiting from the general population. Indeed, two recent investigations suggest musicians may be a useful target population for investigating physiological sensitivity to musical emotions because they exhibit reliable physiological and behavioral responses to expressive performances (Egermann, Fernando, Chuen, & McAdams, 2015; Vieillard et al., 2012).

The Present Study

We aimed to examine music-evoked emotions with respect to all aspects of the emotion response triad (Scherer, 2004, 2005) via self-reports in Experiment 1 (subjective feeling) and a combination of electrodermal activity, cardiovascular measurements (physiological changes), and facial electromyography (overt expression) in Experiment 2. To maximize the likelihood of detecting the emotion response triad’s influence on music-evoked emotions, we examined musically trained participants who ought to be most sensitive to the emotional nuances of the presented musical excerpts (Vieillard et al., 2012). To control for stylistic variations and differences in timbre and orchestration across historical periods, our stimuli were selected from the Romantic piano repertoire, which is known for extremes of emotional expression (Schubert, 2004). This stylistically controlled stimulus set is refined in Experiment 1 to arrive at a final set of ecologically valid musical excerpts that still allows arousal and valence to largely vary independently. What is more, these determinations in stimulus selection were made in light of comparisons between the semantic labels of valence and pleasantness to ensure that the core features of the underlying dimension of emotion are captured. Finally, by collecting subjective

ratings and physiological measures from separate groups of listeners, we were able to mitigate the influence of subjective rating tasks on physiological responses. This issue has been discussed as a shortcoming of previous psychophysiological studies of music and emotion, and some authors have suggested that concurrently performing a task might alter the physiological results (Gingras et al., 2015; White & Rickard, 2015) in addition to behavioral demand characteristics. To offset some of the potential issues introduced by this between-subjects design, we took care to recruit participants who were similar in many demographic and music-experience related variables in Experiment 2.

The aim of Experiment 2 was to find a combination of physiological measures that distinguishes the dimensions of subjective arousal and valence among the excerpts selected in Experiment 1. The measures we employed have been reliable predictors of emotion in more general psychological contexts, but the results among studies of music-evoked emotions have often been inconsistent with this wider literature. Subjects in this experiment simply listened to the stimuli and were not instructed to monitor their own emotional experiences. Our electrodermal measure captured overall skin conductance level. Cardiovascular measures included global heart rate and amplitude of blood volume pulses, as well as several measures of heart rate variability. Heart rate variability measures included the standard deviation of normal-to-normal beat intervals, the square root of the mean squared differences of successive normal-to-normal beat intervals, the number of pairs of successive normal-to-normal interbeat intervals that differ by more than 50 ms, low and high frequency components of heart rate variability derived from the frequency spectrum, as well as the ratio of the low to high frequency components of heart rate variability. Finally, somatovisceral measures included recordings of zygomaticus major and corrugator supercilii muscle activity via electromyography. If emotions evoked by music are similar to everyday emotions, we expect skin conductance level (Boussein et al., 2012; Dawson et al., 2007), heart rate, and most heart rate variability measures to increase in response to higher emotional arousal, whereas blood volume pulse amplitude and high-frequency heart rate variability are expected to decrease (Berntson et al., 1997, 2007; Bradley & Lang, 2007). Similarly, zygomaticus and corrugator electromyography measures are expected to correspond to the positive and negative ends of the valence dimension, respectively (Bradley & Lang, 2007; Larsen et al., 2003; Tassinary & Cacioppo, 1992; Tassinary et al., 2007).

Experiment 1

METHOD

Participants. Thirty members (15 female) of the Montreal community recruited through the Schulich School of Music at McGill University participated in the experiment. Participants were required to have at least five years of formal study on a musical instrument, and all but one reported at least eight years of formal training. Participants were excluded if they reported a history of emotion or anxiety disorder, or were unwilling to abstain from coffee, alcohol, and drugs on the day of the experiment. Ages ranged from 18 to 30 years ($M = 23.0$, $SD = 3.5$). A self-developed questionnaire was administered to assess music preferences and training. On average, the participants had 12.6 years of study on a musical instrument ($SD = 4.3$), and all participants indicated that they enjoyed listening to classical music. A standard audiogram was administered before the experiment to confirm that hearing thresholds were below 20 dB HL (ISO 398-8, 2004; Martin & Champlin, 2000). All participants gave informed consent. The study was certified for ethical compliance by McGill University's Research Ethics Board II.

Materials. The stimuli consisted of 40 excerpts of Romantic piano music that were chosen on the basis of their form, duration, and emotional content by a musicologist (author MM, see Appendix). This selection process was similar to Marin and colleagues (2012, 2013). An effort was made to minimize familiarity with the excerpts by avoiding solo piano works from well-known nineteenth-century composers. The excerpts lasted between 50 and 90 s and were selected such that an equal number would potentially occupy each quadrant of the arousal-valence space. Care was taken to ensure that stimuli were consistent in their emotion category assignment throughout their duration. To limit any effects resulting from differences in the form of each excerpt, all excerpts were in small ternary form (ABA' or AABA'A'), which consists of three main sections: an exposition (A), a contrasting middle (B), and a recapitulation of the material from the exposition (A') (Caplin, 1998).

Apparatus. The experiment was conducted on a Macintosh G5 PowerPC (Apple Computer, Cupertino, CA) in a double-walled IAC Model 1203 sound-isolation booth (IAC Acoustics, Bronx, NY). The stimuli were reproduced using an M-Audio Audiophile 192 sound card (Avid, Irwindale, CA), converted to analog using a Grace Design m904 monitor system (Grace Design, Boulder, CO), and presented over a pair of Sennheiser HD 280

Pro headphones (Sennheiser Electronics, GmbH, Wedemark, Germany). Stimuli were presented at a level of 65 dB SPL on average as measured with a Bruel & Kjaer (Holte, Denmark) Type 2250 sound level meter and a Type 4157 artificial ear to which the headphone was coupled. The experimental program, participant interface, and data collection were programmed using the Max/MSP environment from Cycling 74' (San Francisco, CA) controlled by the PsiExp software environment (Smith, 1995).

Procedure. Upon arriving, each participant completed a consent form, a medical survey, and a music experience survey. The participant was then directed into the audiometric testing booth and the audiogram was administered. During the experimental session, participants were presented with a randomized set of 40 excerpts and asked to rate their subjective emotional experience on a set of 7-point Likert scales that assessed their familiarity with the music (1 = *unfamiliar* and 7 = *familiar*), the valence of their emotional response (1 = *negative* and 7 = *positive*), their experienced arousal/excitement (1 = *not excited* and 7 = *very excited*), their liking of the musical excerpt (1 = *not at all* and 7 = *very much*), and finally the intensity (1 = *very low* and 7 = *very intense*) and pleasantness of their emotional experience (1 = *very unpleasant* and 7 = *very pleasant*). It was emphasized that the participant should respond based on the emotion they *felt*, not the emotion they recognized in the music. Participants responded to all six scales before moving on to the next trial. No time constraints were placed on responses. Scales were presented in the order above. This order was chosen arbitrarily, because previous research has suggested that the order in which response scales of this type are presented does not influence responses (Marin et al., 2012). To prevent trial order effects, all excerpts were preceded by 9.84 s of bird song, which has been used previously to decrease orienting responses (Guhn et al., 2007).

RESULTS

To justify averaging subjective ratings over participants, we first examined the rating scales for consistency, with the excerpt as the unit of measurement. The familiarity, valence, arousal, and intensity ratings proved to be internally consistent (Cronbach's α : familiarity = .77, valence = .71, arousal = .92, intensity = .76), based on the widely used criterion threshold of .70 (Bland & Altman, 1997), which justifies averaging across participants' ratings per excerpt. The liking and pleasantness rating scales, however, did not meet this criterion (Cronbach's α : liking = .59, pleasantness = .61). We

TABLE 1. Summary of Familiarity Ratings by Stimulus

Stimulus	Mean	SD
101	3.53	2.43
102	2.13	1.53
103	3.23	2.10
104	2.80	1.79
105	2.27	1.44
106	2.93	2.07
107	2.77	2.05
108	2.80	2.02
109	2.30	1.86
110	2.30	1.51
201	2.70	2.04
202	3.47	2.29
203	2.37	1.67
204	2.73	1.96
205	2.63	1.90
206	2.97	2.08
207	2.47	2.06
208	2.90	2.11
209	4.87	2.40
210	2.53	2.03
301	2.27	1.70
302	2.77	2.14
303	3.10	2.14
304	2.37	1.79
305	2.83	1.86
306	2.23	1.45
307	2.37	1.71
308	2.17	1.60
309	2.13	1.76
310	2.93	2.03
401	3.07	2.10
402	2.33	1.69
403	3.13	2.26
404	3.30	2.34
405	3.17	2.25
406	3.43	2.31
407	2.77	2.18
408	2.27	1.89
409	3.73	2.52
410	3.17	2.00

chose to examine both valence and pleasantness ratings because of the widespread use of both in the literature. However, due to low consistency among pleasantness ratings, we elected to base further stimulus selection on valence. We include the below analyses in regards to pleasantness simply to provide more information for future researchers regarding the differences between the semantic labels of pleasantness and valence.

Continuing with the excerpt as the unit of analysis, the familiarity ratings were typically low ($M = 2.81$, $SD = 0.54$; for mean familiarity ratings by excerpt, see Table 1). Spearman rank-order correlations between the individual scales are provided in Table 2. Familiarity was weakly to moderately correlated with all of the other

TABLE 2 *Spearman Rank-order Correlations Between Averaged Behavioral Response Measures in Experiment 1*

Scale	1	2	3	4	5	6
1. Familiarity	–	.38*	.54**	.44**	.34*	.34*
2. Valence		–	.17	.43**	.00	.50**
3. Arousal			–	.26	.62**	.14
4. Liking				–	.66**	.90**
5. Intensity					–	.47**
6. Pleasantness						–

Note: $N = 40$. * $p < .05$; ** $p < .01$

TABLE 3 *Partial Spearman Rank-order Correlations Between Averaged Behavioral Response Measures in Experiment 1 Controlled for Familiarity*

Scale	1	2	3	4	5
1. Valence	–	–.05	.31	–.15	.43**
2. Arousal		–	.03	.55**	–.06
3. Liking			–	.60**	.89**
4. Intensity				–	.40*
5. Pleasantness					–

Note: $N = 40$. * $p < .05$; ** $p < .01$

rating scales (r_s range: .34–.54, all $p < .04$). Valence and arousal were not significantly correlated with one another ($r_s = .17$, $p = .30$), and both measures exhibited the fewest significant correlations with the other measures. The strongest correlation observed was between liking and pleasantness ($r_s = .90$, $p < .01$), suggesting that these scales are redundant and most related to hedonic evaluations. Because familiarity was correlated with all of the other scales, we calculated partial Spearman correlations to control for familiarity effects (see Table 3). This resulted in only a slight reduction in most correlation coefficients, but the overall pattern of results remained the same. Valence and arousal continued to exhibit no correlation after controlling for familiarity ($r_s = -.05$, $p = .76$).

Next, we conducted analyses of arousal and valence that used each subject as the unit of analysis. To evaluate the relation between our initial arousal-valence classifications and the participants' responses, these rating scales were submitted to a 2×2 repeated-measures ANOVA with the a priori arousal and valence classifications as within-subjects factors (hereafter referred to as Aclass and Vclass to distinguish these independent variables from the arousal and valence ratings as dependent variables). Post hoc tests were calculated with Bonferroni correction using pairwise t -tests between all six pairs of arousal-valence quadrants (critical $\alpha = .0083$).

As expected, these ANOVAs generally confirmed our initial classifications in the arousal-valence space. Arousal ratings yielded significantly higher values for high Aclass excerpts, $F(1, 29) = 43.54$, $p < .01$, $\eta_p^2 = .60$, and negative Vclass excerpts, $F(1, 29) = 19.22$, $p < .01$, $\eta_p^2 = .40$. A significant Aclass \times Vclass interaction for arousal ratings, $F(1, 29) = 7.04$, $p = .01$, $\eta_p^2 = .20$, was driven by significant differences between all quadrants ($p < .01$), except positive and negative quadrants within the same arousal categories ($p > .05$). Valence ratings yielded significantly higher values for positive Vclass excerpts, $F(1, 29) = 9.39$, $p < .01$, $\eta_p^2 = .24$, as well as higher values for high Aclass excerpts, $F(1, 29) = 4.43$, $p = .04$, $\eta_p^2 = .13$, but the interaction was not significant, $F(1, 29) = 3.20$, $p = .08$, $\eta_p^2 = .10$. Finally, a 2×2 ANOVA for the pleasantness ratings revealed a significant interaction between Aclass and Vclass, $F(1, 29) = 6.52$, $p = .02$, $\eta_p^2 = .18$, but no main effects of Vclass, $F < 1$, or Aclass, $F(1, 29) = 1.43$, $p = .24$, $\eta_p^2 = .05$. Also, no post-hoc tests were significant (all $p > .05$), suggesting that pleasantness ratings did not differentiate the excerpts among the a priori categories.

We next sought to determine whether stimulus ratings created clusters that were representative of the arousal and valence quadrants of interest, and to identify which stimuli best represented each arousal and valence quadrant. This was accomplished using k-means clustering analyses (Bishop, 1995) of the averaged ratings of the 40 individual stimuli with four clusters specified. Ratings were standardized to a continuous scale between -1 and 1 because arousal ratings were found to exhibit a larger range than the valence ratings (range = 2.83 for arousal, range = 1.77 for valence, range = 1.27 for pleasantness). As shown in Figure 1, the clustering solution for the arousal and valence scales corresponds to our a priori classifications with 75% accuracy. By comparison, the clustering solution for the arousal and pleasantness ratings only resulted in a classification accuracy of 50% compared to our musicologically guided a priori Aclass and Vclass groupings (see Figure 2).

To select excerpts that best exemplified each quadrant of the arousal-valence space, the five stimuli within each quadrant of the arousal and valence clustering solution were ranked according to the shortest distance to the quadrant extremes. Using this method, the following excerpts (see list in Appendix) were selected: high arousal, positive valence–209, 406, 206, 204, and 210; high arousal, negative valence–407, 401, 408, 402, and 404; low arousal, negative valence–306, 305, 309, 304, and 301; low arousal, positive valence–108, 308, 101, 107, and 109.

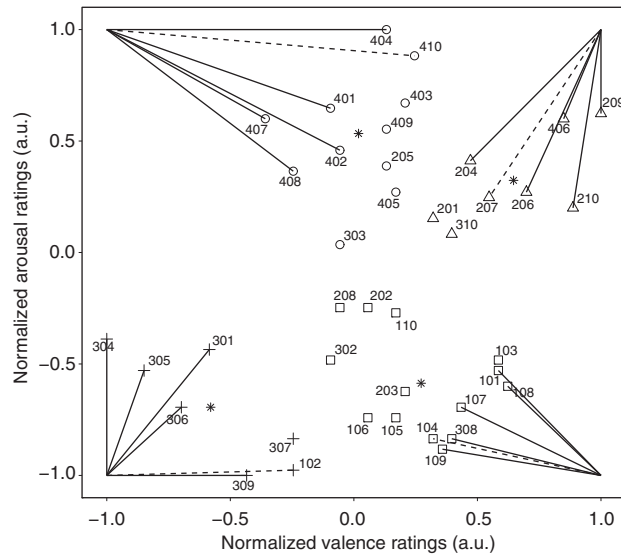


FIGURE 1. K-means clustering analysis of normalized arousal and valence ratings. Asterisks represent cluster centers. Points are coded with different symbols based on cluster membership. Solid lines indicate the five excerpts in each quadrant closest to the arousal-valence extremes. Dashed lines indicate excerpts that did not strictly meet this criterion, but were included in the physiological study to achieve a balance among musicological concerns. Stimulus codes denote the a priori arousal/valence categorization: 100 = positive valence, low arousal; 200 = positive valence, high arousal; 300 = negative valence, low arousal; 400 = negative valence, high arousal.

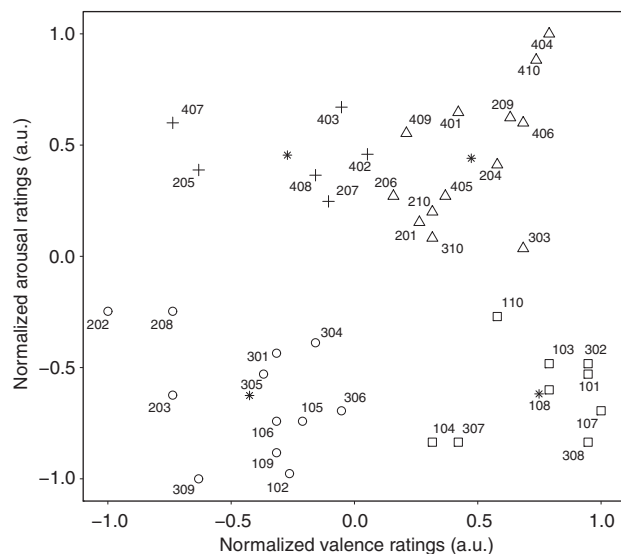


FIGURE 2. K-means clustering analysis on normalized arousal and pleasantness ratings. Asterisks represent cluster centers. Points are coded with different symbols based on cluster membership. Stimulus codes denote the a priori arousal/valence categorization (see Figure 1 caption).

In a final analysis, the arousal/valence k-means clustering solution was reviewed to confirm that the cluster analysis selections within each quadrant were controlled with respect to their rated familiarity and other musical characteristics. Among the selections identified by the clustering analysis, two were rated highest among all 40 stimuli on familiarity (209 and 101), so these excerpts were replaced with the next closest excerpts to the corresponding quadrant extreme (207 and 104, respectively). There was also a large difference in dynamic range between the two A sections of excerpt 301 compared to our other selections, so this excerpt was replaced with excerpt 102. Lastly, to maintain a balance of two ABA' and three AABA'A' forms among stimuli in each quadrant; excerpt 401 was substituted with 410 (see Appendix).

Follow-up repeated-measures ANOVAs were run on the arousal and valence ratings of the 20 selected excerpts to confirm these selections. The ANOVA on valence ratings revealed a significant effect of Aclass, $F(1, 29) = 8.23, p < .01, \eta_p^2 = .22$, and Vclass, $F(1, 29) = 39.40, p < .01, \eta_p^2 = .57$, but no interaction, $F(1, 29) = 2.20, p = .15, \eta_p^2 = .07$. The ANOVA on arousal ratings revealed a significant effect of Aclass, $F(1, 29) = 48.65, p < .01, \eta_p^2 = .63$, and Vclass, $F(1, 29) = 6.33, p = .02, \eta_p^2 = .18$, and an interaction, $F(1, 29) = 4.56, p = .04, \eta_p^2 = .14$.

DISCUSSION

We selected 20 excerpts for further study in Experiment 2. These excerpts best exemplified the crossing of high and low arousal with positive and negative valence based on the behavioral ratings of musically experienced participants. Our findings indicate that ratings for this stimulus set were reliable, largely correspond to our initial classifications, and could be well classified along the dimensions of arousal and valence. This interpretation is supported by large effect sizes for Aclass and Vclass on arousal and valence ratings, respectively, in addition to good accuracy and well-defined clusters in the k-means clustering analysis. These stimuli should maximize the sensitivity of the physiological measures examined in Experiment 2, and allow us to examine the influence of each dimension on these physiological measures as independently as possible.

Pleasantness ratings were found to be less internally consistent than most of our other measures, which is in line with some previous work (Marin, Lampatz, Wandl, & Leder, 2016). Additionally, pleasantness ratings did not correspond with the a priori categorizations (Aclass and Vclass) of the arousal and valence dimensions determined by a musicological analysis (classification

accuracy = 50%). Pleasantness ratings exhibited a high correlation with ratings of liking a complicated interaction between arousal and valence, and no main effects of Aclass or Vclass. Taken together, these findings suggest that valence and pleasantness ratings assess relatively distinct aspects of emotion, and that valence ratings were more relevant to the current investigation. Next, we explored the psychophysiological components of these emotional experiences.

Experiment 2

METHOD

Participants. Twenty musicians (ten females) with over eight years of formal music training took part in the experiment. The average age of the participants was 21.8 years ($SD = 2.5$), and the average number of years of music training was 11.0 ($SD = 2.6$). On the day of the experiment, male participants were required to shave, and all participants were required not to wear makeup and agreed to abstain from coffee, alcohol, and drugs. As in Experiment 1, all participants had normal hearing as confirmed by an audiogram.

Participants were recruited and screened in the same manner as Experiment 1, but with some additional measures that were relevant to this experiment. To minimize familiarity with the stimuli, all of which were solo piano compositions, participants were also screened to ensure that they had no more than four years of amateur piano lessons in the past and were not taking lessons currently. Percussionists were also excluded to further control for familiarity, as they might be more likely to have piano experience than other instrumentalists. All participants gave informed consent. This study was certified for ethical compliance by McGill University's Research Ethics Board II.

Stimuli. The stimuli consisted of the 20 excerpts chosen from Experiment 1. However, due to a programming error, one excerpt was never presented (408, see Appendix), and in its place the preceding stimulus in the randomized sequence was presented a second time. This second iteration was not analyzed. This confined our analysis to 19 stimuli. All quadrants had five excerpts except for the negative-valence, high-arousal quadrant, which had four.

Apparatus. The apparatus was identical to that of the preliminary study, with the addition of psychophysiological equipment. All physiological measurements were recorded using the Procomp Infiniti biometric acquisition system (Thought Technology Ltd., Montreal, QC) at a sampling rate of 256 Hz. Skin conductance was

measured using electrodes (SA9309M) on the distal phalange of the index and ring fingers of the left hand. Blood volume pulse was measured using a photoplethysmograph (SA9308M) on the palmar side of the distal phalange of the middle finger of the left hand. Activation of muscles during facial expression was measured using two Myoscan-Pro electromyography electrodes (SA9401M-60) placed over and in line with the corrugator supercilii and zygomaticus major muscles on the right side of the face, which are active during frowning and smiling, respectively. Plots of the physiological signals were visually monitored using MATLAB (The Mathworks, Natick, MA).

Procedure. Participants completed a biographical questionnaire to provide more details about their music training and listening habits, and a brief medical survey to indicate any medical conditions that might affect the results. Participants also filled out a Profile of Mood States questionnaire (McNair, Lorr, & Droppleman, 1971) before and after the experiment to identify any significant changes in mood that might occur during the experimental session. They were then directed into the sound isolation booth and given a short audiometric exam to ensure that their hearing was suitable for the study. Only one participant was asked not to continue for this reason and was not included among the 20 participants.

After the hearing test, the electrode placement sites were cleaned with alcohol, and electrodes were attached to the participant's face and hand. During pilot testing, these electrodes were found to be sensitive to light sources in the testing booth, thus the experiment was conducted with the overhead lights switched off, leaving only the light from the computer screen. Once the sensors were attached, the participant was instructed to choose a comfortable sitting position facing away from the screen and was asked to remain still for the duration of the session, as movement would introduce artifacts into the recordings. Stimulus presentation was identical to Experiment 1.

The session began with a two-minute silent baseline, which was followed by a brief (9.84 s) excerpt of bird sounds, and then the first of 20 randomly ordered excerpts. After each excerpt, there was a silent inter-stimulus interval of 45 s, followed by the bird song recording to elicit an orienting response prior to the next musical excerpt (Guhn et al., 2007). After the session was completed, participants completed another Profile of Mood States questionnaire. They were then compensated with ten dollars, thanked, and debriefed.

Data analysis. Continuous physiological data were processed in MATLAB (The Mathworks, Inc., Natick, MA) using custom scripts written by the second author. Cardiovascular features to index emotional arousal (both sympathetic and parasympathetic activation, except where noted) consisted of heart rate in beats per minute (HR), blood volume pulse amplitude (BVPamp, sympathetic activation), as well as time-domain and frequency-domain measures of heart rate variability (see feature descriptions below). Electrodermal activity, or skin conductance level (SCL), was also used to index emotional arousal (sympathetic activation). Somatovisceral features to index emotional valence (potential for overt facial expression) consisted of the electromyographic (EMG) signals for the corrugator and zygomaticus muscles.

To remove extraneous information, physiological signals were first filtered with a zero-phase fourth-order Butterworth filter. Skin conductance signals were low-pass filtered with a cutoff frequency of .3 Hz (Boucsein et al., 2012). Blood volume pulse data were low-pass filtered with a cutoff frequency of 3 Hz (Bernston et al., 2007). The electromyography signals were high-pass filtered with a cutoff of 20 Hz, which also served to eliminate noise and movement artifacts (Van Boxtel, 2001), and were then full-wave rectified (Fridlund & Cacioppo, 1986; Tassinari et al., 2007).

To obtain a measure of blood volume pulse amplitude (which decreases with increased sympathetic activation), the upper and lower amplitude envelopes of the blood volume pulse signal were obtained by interpolating between local maxima and minima, respectively. The blood volume pulse amplitude measure was then calculated by extracting the absolute difference between the upper and lower envelopes. Any outliers more than four standard deviations from the mean were replaced using spline interpolation. To obtain measures of heart rate and heart rate variability, inter-beat intervals were calculated by identifying the intervals between local maxima (Jennings et al., 1981). Any outliers in the resulting beat period time series lying more than four standard deviations from the mean were replaced using cubic spline interpolation. Heart rate was then calculated from the final beat period series in beats per minute (Jennings et al., 1981). Three time-domain measures of heart rate variability were derived from the inter-beat interval series: the standard deviation (SDNN), the root mean square of successive differences between adjacent intervals (RMSSD), and the number of successive intervals differing by more than 50 ms (NN50) (Bernston et al., 2007).

In the frequency domain, low-frequency heart-rate variability (LF) in the range of 0.04 – 0.15 Hz has been

shown to be sensitive to both sympathetic and parasympathetic activation (Bernston et al., 2007; Harmat et al., 2011; Iwanaga, Kobayashi, & Kawasaki, 2005), with sympathetic innervation perhaps being dominant (da Silva et al., 2014). High-frequency heart-rate variability (HF) 0.15 – 0.4 Hz reflects parasympathetic control (Bernston et al., 2007; Iwanaga et al., 2005). Thus, to recover sympathetic effects on the cardiovascular system, a ratio of low- to high-frequency heart rate variability is often analyzed (Nakahara, Furuya, Obata, Masuko, & Kinoshita, 2009).

To obtain measures of heart rate variability in the frequency domain, a power spectrum density estimate was calculated for the inter-beat interval series using Welch's fast Fourier transform-based periodogram, which divides the signal into overlapping segments, and averages the spectra across each segment. First, each inter-beat interval series was converted to an equidistantly sampled series by cubic spline interpolation sampled at 4 Hz. Next, a Welch's fast Fourier transform was calculated to obtain absolute power estimates (ms^2) for each frequency band described above. Finally, the absolute power values for each frequency band were obtained by integrating the spectrum over the band limits.

Each pre-stimulus birdsong presentation was removed from the physiological signals, as it was merely included to induce orienting responses prior to each stimulus onset, and thus was not of interest in the analysis. Next, the mean of each 45-s silent baseline signal was subtracted from the subsequent stimulus signal. However, some investigators have pointed out that for electromyographic signals, a muscle at rest exhibits electrical silence (i.e., 0 μV) (Fridlund & Cacioppo, 1986; Gratton, 2007). Thus, it may not be necessary to subtract stimulus measurements from a preceding baseline recording. In light of this claim, we elected to examine the electromyographic signals both with and without reference to the preceding baseline. Finally, to remove inter-individual differences in physiological activity, all extracted features were z-normalized by participant across the experimental session.

RESULTS

Participants in Experiments 1 and 2 did not differ in demographic variables such as age, $t(48) = 1.23, p = .22$, or years of music training, $t(48) = 1.53, p = .13$. The results of the Profile of Mood States questionnaire revealed one outlier on the Total Mood Disturbance outcome measure and this participant was removed from subsequent analyses (Eerola & Vuoskoski, 2011). The removal of this outlier did not substantially affect

the pattern of results reported below. A series of paired-samples *t*-tests comparing participant ratings measured before and after the experiment revealed that participants scored lower on the subscales for tension, $t(18) = 3.33$, $p < .01$, and vigor, $t(18) = 2.41$, $p = .03$, after the experiment. However, there was no change in the total mood disturbance scale, $t(18) = -.43$, $p = .67$. This result therefore indicated that the participants were feeling slightly more relaxed following the experiment, but their overall mood was largely unchanged.

Before examining the physiological measures using parametric models, normality assumptions were assessed for each physiological measure using Shapiro-Wilk tests. The majority of the Shapiro-Wilk tests were nonsignificant ($p > .05$). However, two measures did reach significance: high-frequency heart rate variability: $S-W(20) = .97$, $p = .04$; zygomaticus electromyography without baseline adjustment: $S-W(20) = .97$, $p = .04$. Visual inspection did not reveal common outliers, and the histogram plots of these measures did not appear to deviate extremely from normality (high-frequency heart rate variability: Skewness = -0.54 , Kurtosis = 2.74 ; zygomaticus electromyography without baseline adjustment: Skewness = 0.21 , Kurtosis = 2.06). Thus, analysis commenced with parametric tests. Results regarding psychophysiological measures were analyzed using a 2×2 repeated-measures ANOVA with a priori arousal (low, high) and valence (positive, negative) as categorical, within-subjects factors as determined in Experiment 1.

We first sought to test whether electrodermal activity increased in response to our high-arousal stimuli. As predicted, RM-ANOVAs on the skin conductance level signals revealed a significant main effect of arousal (see Figure 3a), $F(1, 18) = 4.90$, $p = .04$, $\eta_p^2 = .21$, which was driven by higher responses to excerpts rated high in arousal in Experiment 1. There was no main effect of valence, $F(1, 18) = 1.16$, $p = .30$, and no interaction between arousal and valence, $F(1, 18) = 1.71$, $p = .21$ for skin conductance level.

We next examined whether increases in cardiovascular activity correspond to high arousal stimuli. For the cardiovascular measures, RM-ANOVAs revealed a significant main effect of arousal for both heart rate, $F(1, 18) = 32.54$, $p < .01$, $\eta_p^2 = .64$, and blood volume pulse amplitude, $F(1, 18) = 21.27$, $p < .01$, $\eta_p^2 = .54$, but no main effect of valence (both $F < 1$) or interaction was observed for either measure: heart rate: $F < 1$; blood volume pulse: $F(1, 18) = 2.37$, $p = .14$. Excerpts from the high-arousal quadrants elicited significant increases in heart rate (Figure 3b) and significant decreases in blood volume pulse amplitude (Figure 3c), the latter

of which is consistent with increased vasoconstriction following increased sympathetic activation.

The heart rate variability measures suggested some sensitivity to the valence characteristics of the stimuli. A marginally significant effect of valence was observed for low-frequency heart rate variability (Figure 3d), $F(1, 18) = 3.89$, $p = .06$, $\eta_p^2 = .18$, as well for the number of successive heart beat intervals differing by more than 50 ms (Figure 3e), $F(1, 18) = 4.23$, $p = .06$, $\eta_p^2 = .19$, with both measures exhibiting higher responses for positively valenced stimuli. However, there was no effect of arousal, and no interaction for either measure (both $F < 1$). No significant effects were observed for the other heart rate variability measures (the standard deviation and root-mean square of inter-beat intervals, high-frequency heart rate variability, and the ratio of low- to high-frequency heart rate variability).

To examine physiological sensitivity to valence, we examined whether facial electromyographic activity responds to stimuli rated as positively or negatively valenced by measuring activity in the zygomaticus major and corrugator supercilii muscles, respectively (see Figure 4a and 4b). RM-ANOVAs relative to the preceding 45 s inter-stimulus baseline revealed a main effect of arousal for the zygomaticus major, $F(1, 18) = 4.83$, $p = .04$, $\eta_p^2 = .21$, with an increased response to high arousal stimuli (Figure 4a). We also observed a main effect of valence for the corrugator supercilii, $F(1, 18) = 5.28$, $p = .03$, $\eta_p^2 = .23$, with an increased response to negative valence stimuli (Figure 4b). No other significant effects were found for these measures.

The results for the facial electromyography signals changed when analyzed without reference to the preceding baseline (see Figure 4c and 4d). RM-ANOVAs for the zygomaticus muscle activity revealed significant main effects of arousal, $F(1, 18) = 5.69$, $p = .03$, $\eta_p^2 = .24$, and valence, $F(1, 18) = 17.88$, $p < .01$, $\eta_p^2 = .50$, but no interaction between the two, $F < 1$, indicating that this measure was sensitive to high-arousal excerpts and positive-valence excerpts (Figure 4c). A significant main effect of valence also indicated that corrugator muscle activity was sensitive to excerpts in the negative-valence category, $F(1, 18) = 6.42$, $p = .02$, $\eta_p^2 = .26$, with larger activity in response to negative-valence ratings (Figure 4d). Corrugator electromyography showed no effect of arousal and no interaction, $F < 1$, in both cases.

DISCUSSION

Based on findings outside of the music literature, we predicted that the electrodermal and cardiovascular measures would be sensitive to arousal or activation

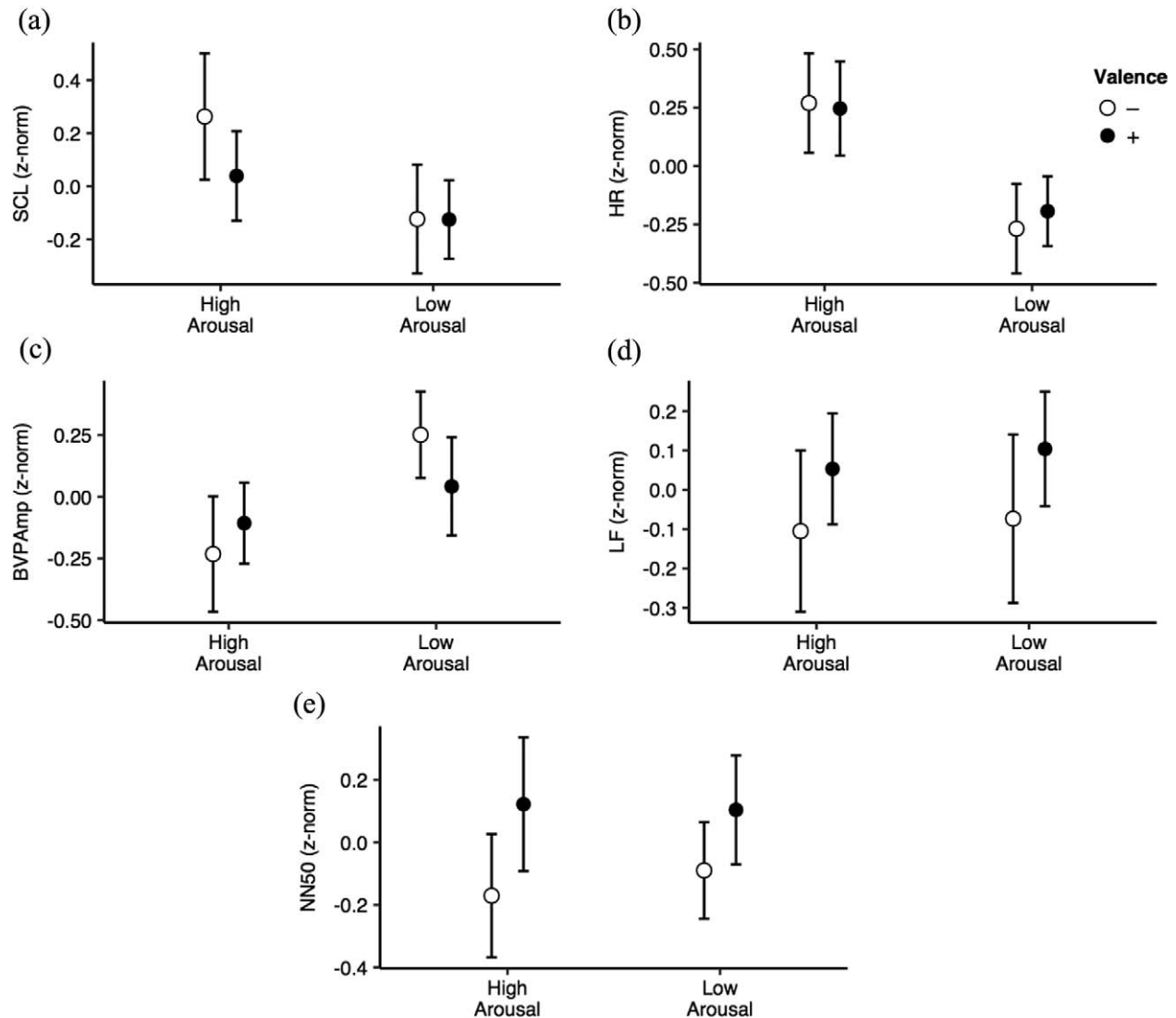


FIGURE 3. Psychophysiological measures of emotional arousal (electrodermal and cardiovascular) for the four quadrants of the arousal-valence space: (a) Normalized skin conductance levels (SCL), (b) heart rate (HR), (c) blood volume pulse amplitude (BVPAmplitude), (d) low-frequency heart rate variability (LF), and (e) number of successive intervals differing by more than 50 ms (NN50). Bars represent 95% confidence intervals.

aspects of the emotional experience, whereas measures of somatic muscle activity would be sensitive to the valence of emotional reactions. These hypotheses were largely confirmed, with many measures exhibiting strong main effects of arousal or valence and with few interactions, validating this set of psychophysiological tools for use in future investigations of music-evoked emotions. Moreover, the observed effects align closely with results obtained for emotional responses to non-musical stimuli (Berntson et al., 1997, 2007; Bouscein et al., 2012; Bradley & Lang, 2007; Chuen et al., 2016; Dawson et al., 2007; Gerger et al., 2014; Larsen et al., 2003; Tassinari et al., 2007). This finding suggests that

these responses are common to both musical and everyday emotional experiences, thus contributing to the ongoing discussion about the nature of musical emotions.

Among our electrodermal measures, musicians exhibited higher skin conductance levels for highly arousing stimuli, suggesting that skin conductance level is a reliable indicator of sympathetic arousal during music listening (Bouscein et al., 2012). This finding deviates from the results of some investigations (Blood & Zatorre, 2001; Grewe et al., 2007; White & Rickard, 2015). However, the stimulus sets employed in these studies were especially varied, incorporating participant-selected excerpts

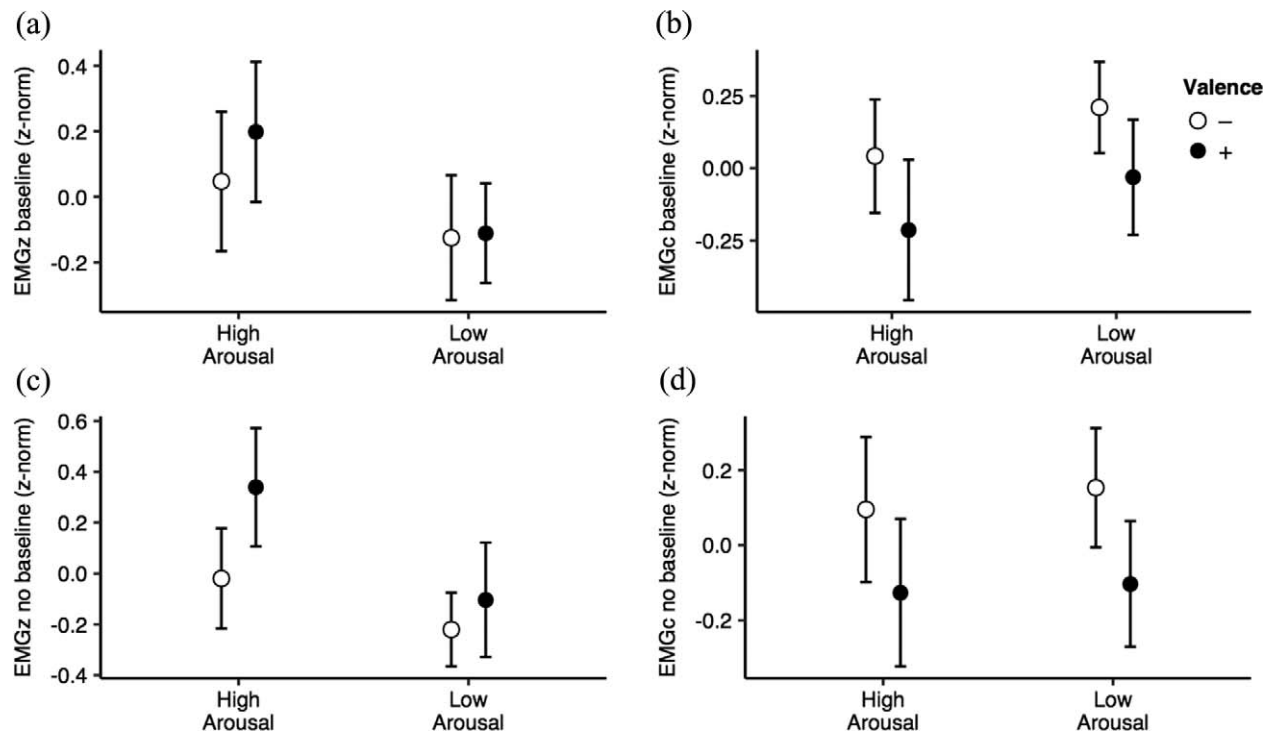


FIGURE 4. Psychophysiological measures of emotional valence (electromyography) for the four quadrants of the arousal-valence space: (a) Normalized electromyography signals of the zygomaticus major (EMGz) referenced to baseline, (b) corrugator supercillii (EMGc) referenced to baseline, (c) zygomaticus major (EMGz) without baseline reference, and (d) corrugator supercillii (EMGc) without baseline reference. Bars represent 95% confidence intervals.

or selections from multiple genres. In some cases, this led to an under-representation of low-arousal and negatively valenced stimuli, which perhaps resulted in a lack of differentiation among various levels of arousal in electrodermal analysis (Grewe et al., 2007).

As expected, our cardiovascular measures of heart rate and blood volume pulse amplitude also exhibited sensitivity to emotional arousal. In the case of blood volume pulses, lowered amplitudes result from increased sympathetic activity (Berntson et al., 1997, 2007), which is consistent with our electrodermal results. However, our findings conflict with some previous work, which found no significant differences in heart rate responses to musical stimuli (Lundqvist et al., 2009) or cardiovascular responses that decreased during exposure to arousing music (White & Rickard, 2015). These investigations only utilized stimuli that were “happy” or “sad” and in some cases were experimenter-composed songs for voice and guitar, thus collapsing two potentially distinct emotional dimensions into a single-dimensional space. Importantly, the present study was able to examine arousal and valence independently by selecting stimuli that varied along both dimensions. This allowed for

a closer investigation of the effect of arousal on cardiovascular activity based primarily on changes in the music, unrelated to linguistic effects from the presence of lyrics. Indeed, this approach led to heart rate and blood volume pulse amplitude effects being some of the strongest observed in our data.

Some previous research has shown heart rate and heart rate variability to be responsive to the pleasantness of musical excerpts (Nakahara et al., 2009; Orini et al., 2010). The heart rate measures observed here were largely sensitive to arousal except for the number of inter-beat intervals exceeding 50 ms and the low-frequency heart rate variability measures, which approached significance for valence. In the frequency domain, the trend toward a valence effect observed in the low-frequency region differs from some previous findings (Orini et al., 2010), where high-frequency heart rate variability was found to be most responsive to pleasant stimuli. It should be noted that these previous studies rigorously controlled for respiratory effects, which are known to greatly affect high-frequency heart rate variability. Respiratory measures were not recorded in the present study, so their effect on high-frequency

heart rate variability could not be considered. In the temporal domain, contrary to other studies, only a marginal effect of the number of inter-beat intervals exceeding 50 ms was observed. One explanation for this divergence might be that previous studies using temporal heart rate variability measures have been based on physical activity or performance, which could in turn affect variability (da Silva et al. 2014; Harmat et al., 2011; Nakahara et al., 2009). Further work should explore how these heart rate variability measures might correspond to emotions related to preparatory actions of performance, as well as anticipatory responses to upcoming familiar passages as neither of these issues were examined in the present study.

Our findings from electromyography of the corrugator and zygomaticus muscles highlight several methodological issues. Most studies using these measures in music research offset responding during stimulus presentation by a baseline measurement, although the utility of applying baseline subtraction has been questioned in the psychophysiology community (Fridlund & Cacioppo, 1986; Gratton, 2007), because there is not the same level of tonic activity to be accounted for as in electrodermal or cardiovascular measures when the muscle is not engaged. When corrected to a prior baseline, zygomaticus muscle activity exhibited a significant response to high-arousal stimuli, and corrugator muscle activity exhibited a larger response to negatively valenced stimuli. Without baseline correction, strong effects of positive valence in the zygomaticus measure also became apparent, indicating that such processing choices can have effects on these measures. Analyses without baseline correction are consistent with previous work demonstrating complementary valence effects in both of these facial muscle groups (Larsen et al., 2003). In further agreement with previous studies, we found that the corrugator measure was a more pure indicator of valence than the zygomaticus measure (Vieillard et al., 2012), because the zygomaticus showed additional activity during highly arousing stimuli (Witvliet & Vrana, 2007). However, it should also be noted that the valence effect observed in the zygomaticus measure (and not the arousal effect) was among the highest effect sizes reported in this study.

It should also be noted that the electromyography measures were constrained by a relatively narrow frequency bandwidth of 20–128 Hz. Due to limitations in the signal recording and stimulus presentation software, these signals were sampled at 256 Hz, thus making the highest frequency captured 128 Hz. Guidelines for this type of signal typically suggest capturing a bandwidth between 20–500 Hz, and thus, sample rates as high as

1000 Hz are often encouraged (Tassinari et al., 2007; Van Boxtel, 2001). The current findings are informative, however, given that most of the pertinent frequency content that occurs during emotion elicitation in these signals is present in the narrow bandwidth captured here (Van Boxtel, 2001). Thus, additional electromyography bandwidth may not be necessary to capture electrophysiological correlates of emotional valence.

General Discussion

The present investigation examined a wide variety of traditional psychophysiological measures of electrodermal, cardiovascular, and somatic muscle activity to assess emotional reactions to a set of stylistically homogeneous musical excerpts of Romantic piano music in musicians, utilizing a converging-methods approach that relied on all aspects of the emotion response triad (Bradley & Lang, 2000; Izard, 1992; Scherer, 2004). The controls used in selecting our ecologically valid musical stimuli also allowed us to examine the effects of arousal and valence on these psychophysiological measures as independently as possible. Importantly, we used separate groups of listeners to explore the relationship between subjective ratings and physiological responses, which prevented one task from explicitly affecting the other (Gingras et al., 2015; White & Rickard, 2015). Our results demonstrate a convergence of all aspects of the emotion response triad on the music-evoked responses. Behavioral ratings in Experiment 1 demonstrated subjective emotional responses, whereas the psychophysiological recordings in Experiment 2 illuminated corresponding changes in physiological responses and overt expression.

The range of controls implemented in our study and the lack of consensus regarding these measures among previous studies underscore the complexity of studying emotional responses to music (Juslin & Sloboda, 2010a). Such responses intersect with an individual's tastes, personality, and autobiographical experiences, as well as with many physiological responses to the acoustic stimulus itself (Juslin, Barradas, & Eerola, 2015; Juslin, Harmat, & Eerola, 2014). Indeed, the results of Experiment 1 emphasize this through several instances where averaged ratings disagreed with our a priori musicological predictions. Such a result validates our choice to make stimulus selections based on empirical data from many individuals, rather than preselecting stimuli that fit our assumptions. The stimulus control and validation methods we implemented in turn allowed us to isolate some of the qualities of the emotional experience of music listening.

Great care was taken to properly represent both dimensions of arousal and valence in our experiments. The results from Experiment 1 indicate that we were largely successful, but some biases among high-arousal stimuli were still present. Nonetheless, these effects are consistent with other investigations (Grewe et al., 2007; Khalfa et al., 2008; Krumhansl, 1997), including those that used similar stimulus-selection procedures and controls in this genre (Gingras et al., 2014, 2015; Marin & Leder, 2013; Marin et al., 2012). These previous findings could be due to systematic trends in how listeners perceive this particular style of music or to larger issues regarding incomplete orthogonality between arousal and valence in emotion more generally (Kuppens, Tuerlinckx, Russell, & Barrett, 2012). It is possible that this bias in the high-arousing, positive-valence excerpts influenced the arousal effects observed in the zygomaticus activity.

In Experiment 1, we took many steps to limit familiarity with our stimulus set. The relatively low overall familiarity ratings for stimuli used in Experiment 2 suggest that these specific excerpts were unlikely to be associated with past emotional experiences. Moreover, we controlled for much of the heterogeneity present in previous studies, such as variable instrumentation, variable levels of music training among participants, stylistic differences among stimuli, and different musical forms. However, the decision to select stimuli from the Romantic piano repertoire, which was motivated by a desire to achieve a high level of validity and control, may nevertheless limit the generalizability of the findings reported here. Future work should seek to replicate and expand these findings by examining emotional responses to similarly controlled stimulus sets in other, perhaps more contemporary musical styles, as well as to compare responses between styles. Other informative follow-up work could examine how effects of familiarity and personality interact with different genres during emotion induction (Eerola & Vuoskoski, 2011; Kuppens et al., 2012; Vuoskoski, Thompson, McIlwain, & Eerola, 2012).

Our choice to investigate musicians also seems to have been effective in indexing physiological emotional responses to music. In line with previous research, musicians exhibited a wide array of music-evoked physiological responses that were sensitive to the affective characteristics of our stimuli (Vieillard et al., 2012). These findings suggest that the null effects found for some physiological measures in previous studies may have been due to the use of heterogeneous participant populations (Grewe et al., 2007; Khalfa et al., 2008; Lundqvist et al., 2009). It could be that the extensive

level of training among our subjects—who on average reported 11 years of explicit training on a musical instrument—led to increased emotional sensitivity to our stimuli, which featured particularly expressive performances. This would be in line with the behavioral and corrugator electromyography findings of Vieillard and colleagues (2012), as well as with other findings that musicians are more perceptive to emotive cues in other auditory signals like speech (Thompson et al., 2004). It should also be kept in mind that—as with all studies involving musicians or effects of musical experience—any processing advantages of musicians may be the result of either a pre-existing sensitivity to music or the intensive training they have received. This is a much larger issue in the field, and we leave it to future investigations to untangle these two causes. Also, since our study only examined musicians, an interesting future study might examine these physiological responses as a function of music training or other aspects of musical sophistication (Müllensiefen, Gingras, Musil, & Stewart, 2014) in a heterogeneous participant sample.

Some exciting recent work has focused on better understanding the mechanisms through which music influences emotion (Juslin et al., 2014). Our results also align with the physiological results of Juslin et al. (2015), who attempted to isolate individual mechanisms of emotional induction via their stimulus set, with skin conductance being sensitive to the more highly arousing responses, and the corrugator muscle activity showing sensitivity to negative valence. Interestingly, Juslin and colleagues' (2015) zygomaticus results also show some sensitivity to arousal (or brainstem response emotional induction) in addition to positive valence. Our results complement their findings in that some mechanisms, such as episodic memory, were relatively controlled in our study because the music was unfamiliar, while still yielding a similar pattern of physiological results. However, other factors from their framework could have contributed to our physiological results, such as emotional contagion, reflex activity, and expectancies. An interesting possibility may be that the arousal and valence qualities of an emotional reaction are agnostic to their mechanism of induction, but that certain mechanisms are particularly well suited to altering arousal or valence.

Finally, the results of this study resonate with larger questions concerning the nature of musical emotions. The approach used here incorporated all aspects of the emotion response triad: subjective feeling in Experiment 1, plus physiological changes (electrodermal and cardiovascular measures), and overt expression (electromyography) in Experiment 2 among two groups of

participants that were shown to be quite homogeneous with respect to their intense involvement with the Western art music repertoire and were thus likely to be receptive to the expressive qualities in the performances (Vieillard et al., 2012). Additionally, by controlling for familiarity, style, instrumentation, and even musical form, the present findings lend strength to the view that emotional responses to music are felt by the listener rather than merely perceived (Kivy, 1990; Konečni, 2008), and our physiological findings support the view that music can induce emotions like those experienced in everyday situations (Juslin & Västfjäll, 2008). This interpretation is supported by our use of measurement criteria derived from emotion methodologies in non-musical domains of psychology and that possible issues introduced by stimulus familiarity and stylistic variations were controlled for. Although it could be argued that during the preliminary study, respondents might have appraised a given stimulus based on what was being *expressed* rather than what was *felt* (despite explicit instructions to the contrary), in the physiological study, respondents were not asked to consciously monitor their emotional experiences. The physiological and nonexplicit nature of these responses also indicates that the emotional experience was embodied. Taken together, these results support Cochrane's (2010) claim that a demonstration of the elicitation of physiological changes purely from the music itself might satisfy

philosophical scrutiny over the nature of music-evoked emotions.

Author Note

Mattson Ogg is currently at the Department of Psychology, University of Maryland, College Park and Program in Neuroscience and Cognitive Science, University of Maryland, College Park.

This research was conducted at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Funding was provided by a University of Maryland Flagship Fellowship and Dean's Fellowship awarded to Mattson Ogg, by a Richard H. Tomlinson fellowship and a Quebec Merit Fellowship for Foreign Doctoral Students awarded to David Sears, and by Canadian Natural Sciences and Engineering Research Council (RGPIN 213887-10), and Social Sciences and Humanities Research Council (410-2009-2201) grants, and a Canada Research Chair awarded to Stephen McAdams. Part of Manuela Marin's contribution was completed while at the University of Vienna. The authors would like to thank Bennett K. Smith for the design of the interfaces used in this experiment.

Correspondence concerning this article should be addressed to Stephen McAdams, Schulich School of Music, 555 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 1E3. E-mail: stephen.mcadams@mcgill.ca

References

- BERNTSON, G. G., BIGGER, J. T., ECKBERG, D. L., GROSSMAN, P., KAUFMANN, P. G., MALIK, M., ET AL. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, 34, 623-648.
- BERNTSON, G. G., QUIGLEY, K. S., & LOZANO, D. (2007). Cardiovascular psychophysiology. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 182-210). Cambridge, UK: Cambridge University Press.
- BIGAND, E., & POULIN-CHARRONNAT, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100, 100-130.
- BISHOP, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.
- BLAND, J. M., & ALTMAN, D. G. (1997). Cronbach's alpha. *BMJ: British Medical Journal*, 314, 572.
- BLOOD, A. J., & ZATORRE, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98, 11818-11823.
- BOUCSEIN, W., FOWLES, D. C., GRIMNES, S., BEN-SHAKHAR, G., ROTH, W. T., DAWSON, M. E., & FILION, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49, 1017-1034.
- BRADLEY, M. M., & LANG, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 242-276). New York: Oxford University Press.
- BRADLEY, M. M., & LANG, P. J. (2007). Emotion and motivation. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 581-607). Cambridge, UK: Cambridge University Press.
- BROWN, S. L., & SCHWARTZ, G. E. (1980). Relationships between facial electromyography and subjective experience during affective imagery. *Biological Psychology*, 11, 49-62.
- CAPLIN, W. E. (1998). *Classical form: A theory of formal functions for the instrumental music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.
- CHUEN, L., SEARS, D., & MCADAMS, S. (2016). Psychophysiological responses to auditory change. *Psychophysiology*, 53, 891-904.

- COCHRANE, T. (2010). Music, emotions and the influence of the cognitive sciences. *Philosophy Compass*, 5, 978-988.
- COLOMBETTI, G. (2005). Appraising valence. *Journal of Consciousness Studies*, 12, 103-126.
- DA SILVA, S. A. F., GUIDA, H. L., DOS SANTOS ANTONIO, A. M., DE ABREU, L. C., MONTEIRO, C. B. M., FERREIRA, C., ET AL. (2014). Acute auditory stimulation with different styles of music influences cardiac autonomic regulation in men. *International Cardiovascular Research Journal*, 8, 105-110.
- DAWSON, M. E., SCHELL, A. M., & FILION, D. L. (2007). The electrodermal system. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159-181). Cambridge, UK: Cambridge University Press.
- DILLMAN CARPENTIER, F. R., & POTTER, R. F. (2007). Effects of music on physiological arousal: Explorations into tempo and genre. *Media Psychology*, 10, 339-363.
- DOWLING, W. J. (1989). Simplicity and complexity in music and cognition. *Contemporary Music Review*, 4, 247-253.
- EEROLA, T., & VUOSKOSKI, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39, 18-49.
- EGERMANN, H., FERNANDO, N., CHUEN, L., & MCADAMS, S. (2015). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology*, 6, 1341.
- EGERMANN, H., PEARCE, M. T., WIGGINS, G. A., & MCADAMS, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, and Behavioral Neuroscience*, 13, 533-553.
- EKMANN, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169-200.
- FRIDLUND, A. J., & CACIOPPO, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23, 567-589.
- GERGER, G., LEDER, H., & KREMER, A. (2014). Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. *Acta Psychologica*, 151, 174-183.
- GINGRAS, B., MARIN, M. M., & FITCH, W. T. (2014). Beyond intensity: Spectral features effectively predict music-induced subjective arousal. *The Quarterly Journal of Experimental Psychology*, 67, 1428-1446.
- GINGRAS, B., MARIN, M. M., PUIG-WALDMÜLLER, E., & FITCH, W. T. (2015). The eye is listening: Music-induced arousal and individual differences predict pupillary responses. *Frontiers in Human Neuroscience*, 9, 619.
- GOMEZ, P., & DANUSER, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, 53, 91-103.
- GRATTON, G. (2007). Biosignal processing. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 834-858). Cambridge, UK: Cambridge University Press.
- GREWE, O., KOPIEZ, R., & ALTENMÜLLER, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, 27, 61-74.
- GREWE, O., NAGEL, F., KOPIEZ, R., & ALTENMÜLLER, E. (2007). Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion*, 7, 774-788.
- GUHN, M., HAMM, A., & ZENTNER, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception*, 24, 473-484.
- HAILSTONE, J. C., OMAR, R., HENLEY, S. M., 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. *The Quarterly Journal of Experimental Psychology*, 62, 2141-2155.
- HARMAT, L., ULLÉN, F., DE MANZANO, Ö., OLSSON, E., ELOFSSON, U., VON SCHÉELE, B., & THEORELL, T. (2011). Heart rate variability during piano playing: A case study of three professional solo pianists playing a self-selected and a difficult prima vista piece. *Music and Medicine*, 3, 102-107.
- HEVNER, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48, 246-268.
- HODGES, D. (2010). Psychophysiological measures. In P. N. Juslin & J. A. Sloboda (Eds.) *Handbook of music and emotion: Theory, research, applications* (pp. 279-312). New York: Oxford University Press.
- HODGES, D. (2016). Bodily responses to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (2nd ed., pp. 183-196). Oxford, UK: Oxford University Press.
- 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.
- IWANAGA, M., KOBAYASHI, A., & KAWASAKI, C. (2005). Heart rate variability with repetitive exposure to music. *Biological Psychology*, 70, 61-66.
- IZARD, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, 99, 561-565.
- JENNINGS, J. R., BERG, W. K., HUTCHESON, J. S., OBRIST, P., PORGES, S., & TURPIN, G. (1981). Publication guidelines for heart rate studies in man. *Psychophysiology*, 18, 226-231.
- JUSLIN, P. N., BARRADAS, G., & EEROLA, T. (2015). From sound to significance: Exploring the mechanisms underlying emotional reactions to music. *The American Journal of Psychology*, 128, 281-304.
- JUSLIN, P. N., HARMAT, L., & EEROLA, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, 42, 599-623.
- 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., & SLOBODA, J. A. (Eds.) (2010a). *Handbook of music and emotion: Theory, research, applications*. New York: Oxford University Press.
- JUSLIN, P. N., & SLOBODA, J. A. (2010b). At the interface between the inner and outer world: Psychological perspectives. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 279-312). New York: Oxford University Press.
- JUSLIN, P. N., & VÄSTFJÄLL, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559-575.
- KHALFA, S., ROY, M., RAINVILLE, P., DALLA BELLA, S., & PERETZ, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, 68, 17-26.
- KIVY, P. (1990). *Music alone: Philosophical reflections on the purely musical experience*. Ithaca, NY: Cornell University Press.
- KONEČNÍ, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 115-139.
- KRAUS, N., SLATER, J., THOMPSON, E. C., HORNICKEL, J., STRAIT, D. L., NICOL, T., & WHITE-SCHWOCH, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *The Journal of Neuroscience*, 34, 11913-11918.
- KRUMHANS, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, 51, 336-352.
- KUPPENS, P., TUERLINCKX, F., RUSSELL, J. A., & BARRETT, L. F. (2012). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, 139, 917-940.
- LARSEN, J. T., NORRIS, C. J., & CACIOPPO, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40, 776-785.
- LUNDQVIST, L., CARLSSON, F., HILMERSSON, P., & JUSLIN, P. N. (2009). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, 37, 61-90.
- MARIN, M. M., & LEDER, H. (2013). Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS ONE*, 8, e72412.
- MARIN, M. M., GINGRAS, B., & BHATTACHARYA, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion*, 12, 618-631.
- MARIN, M. M., LAMPATZ, A., WANDL, M., & LEDER, H. (2016). Berlyne revisited: Evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers in Human Neuroscience*, 10, 536.
- MARTIN, F. N., & CHAMPLIN, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, 11, 64-66.
- McNAIR, D. M., LORR, M., & DROPPLEMAN, L. F. (1971) *The Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- MÜLLENSIEFEN, D., GINGRAS, B., MUSIL, J., & STEWART, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9, e89642.
- MULLIGAN, K., & SCHERER, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, 4, 345-357.
- NAGEL, F., KOPIEZ, R., GREWE, O., & ALTENMÜLLER, E. (2007). EMUJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, 39, 283-290.
- NAGEL, F., KOPIEZ, R., GREWE, O., & ALTENMÜLLER, E. (2008). Psychoacoustical correlates of musically induced chills. *Musicae Scientiae*, 12, 101-113.
- NAKAHARA, H., FURUYA, S., OBATA, S., MASUKO, T., & KINOSHITA, H. (2009). Emotion-related changes in heart rate and its variability during performance and perception of music. *Annals of the New York Academy of Sciences*, 1169, 359-362.
- NATER, U. M., ABBRUZZESE, E., KREBS, M., & EHLERT, U. (2006). Sex differences in emotional and psychophysiological responses to music stimuli. *International Journal of Psychophysiology*, 62, 300-308.
- ORINI, M., BAILÓN, R., ENK, R., KOELSCH, S., MAINARDI, L., & LAGUNA, P. (2010). A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis. *Medical and Biological Engineering and Computing*, 48, 423-433.
- PANKSEPP, J. (1995). The emotional sources of "chills" induced by music. *Music Perception*, 13, 171-207.
- PARNCUTT, R., & MARIN, M. M. (2006). Emotions and associations evoked by unfamiliar music. In H. Gottesdiener & J.-C. Vilatte (Eds.), *Culture and communication: Proceedings of the International Association of Empirical Aesthetics* (pp. 725-729). Avignon, France: IAEA.
- RICKARD, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, 32, 371-388.
- RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- SALIMPOOR, V. N., BENOVOY, M., LONGO, G., COOPERSTOCK, J. R., & ZATORRE, R. J., (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE*, 4, e7487.
- SCHERER, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, 33, 239-251.
- SCHERER, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695-729.
- SCHUBERT, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, 21, 561-585.

- SLATER, J., SKOE, E., STRAIT, D. L., O'CONNELL, S., THOMPSON, E., & KRAUS, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, 291, 244-252.
- SMITH, B. K. (1995). *PsiExp: An environment for psychoacoustic experimentation using the IRCAM Musical Workstation*. Paper presented at the Society for Music Perception and Cognition Conference, University of California, Berkeley.
- STEINBEIS, N., KOELSCH, S., & SLOBODA, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, 18, 1380-1393.
- STRAIT, D. L., KRAUS, N., SKOE, E., & ASHLEY, R. (2009). Musical experience and neural efficiency-effects of training on sub-cortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29, 661-668.
- TASSINARY, L. G., & CACIOPPO, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science*, 3, 28-33.
- TASSINARY, L. G., CACIOPPO, J. T., VANMAN, E. J. (2007) The skeletomotor system: Surface electromyography. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 267-299). Cambridge, UK: Cambridge University Press.
- TIERNEY, A. T., KRIZMAN, J., & KRAUS, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, 112, 10062-10067.
- THOMPSON, W. F., MARIN, M. M., & STEWART, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, 109, 19027-19032.
- THOMPSON, W. F., SCHELLENBERG, E. G., & HUSAIN, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4, 46-64.
- VAN BOXTEL, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, 38, 22-34.
- VAN DEN BOSCH, I., SALIMPOOR, V. N., & ZATORRE, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Frontiers in Human Neuroscience*, 7, 534.
- VIEILLARD, S., ROY, M., & PERETZ, I. (2012). Expressiveness in musical emotions. *Psychological Research*, 76, 641-653.
- VUOSKOSKI, J. K., THOMPSON, W. F., MCILWAIN, D., & EEROLA, T. (2012). Who enjoys listening to sad music and why? *Music Perception*, 29, 311-317.
- WHITE, E. L., & RICKARD, N. S. (2015). Emotion response and regulation to "happy" and "sad" music stimuli: Partial synchronization of subjective and physiological responses. *Musicae Scientiae*, 20, 11-25.
- WITVLIET, C. V., & VRANA, S. R. (2007). Play it again Sam: Repeated exposure to emotionally evocative music polarizes liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition and Emotion*, 21, 3-25.
- WUNDT, W. (1912). *An introduction to psychology* (R. Pinter, Trans.). London, UK: George Allen.
- ZATORRE, R. J., & SALIMPOOR, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences*, 110, 10430-10437.
- ZHAO, T. C., & KUHLM, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, 113, 5212-5217.

Appendix

List of Stimuli. Stimulus codes denote the a priori category to which the excerpt was assigned. Excerpts coded as 100 correspond to the positive-valence, low-arousal category, 200 to the positive-valence, high-arousal category, 300 to the negative-valence, low-arousal category, and 400 to the negative-valence, high-arousal category. In cases where no time window is specified, the excerpt starts at the beginning of the piece. Some excerpts have

a short fade out (500 ms) because of an abrupt ending. Excerpts selected as a quadrant extreme in K-means clustering analysis are denoted with an asterisk (*). Stimuli selected for inclusion in the physiological study are indicated by two additional superscript characters denoting the arousal-valence category in which they were used via the following convention: high arousal (^H), low arousal (^L), positive valence (⁺), negative valence (⁻).

Stimulus Code	Composer	Piece & CD	Form	Duration
101*	Fryderyk Chopin	2 Nocturnes, Op. 32, Nocturne No. 1 in B major EMI Classics, 7243 5 56196 2 4 Byron Janis, 1996	AABA'A'	1:19 B: 0:30.40 A: 0:48.68

(continued)

Appendix. (continued)

Stimulus Code	Composer	Piece & CD	Form	Duration
102 ^{L-}	Carl Maria von Weber	Piano Sonata 1 in C major, Op. 42, Mvt. 2, Adagio Naxos, 8.550988 Alexander Paley, 1994	ABA'	0:49 B: 0:15.49 A: 0:34.32
103	Robert Schumann	Albumblätter, Op. 124, Mvt. 16, Schlummerlied McGill Records, MR 2006-01-2 Kyoko Hashimoto, 2006	AABA'A'	0:50 0:18–1:08 B: 0:16.07 A: 0:34.57
104 ^{L+}	Robert Schumann	Humoreske in B flat, Op. 20, Einfach Naxos, 8.550469 Wolf Harden, 1990	AABA'A'	1:22 fade out B: 0:23.96 A: 1:01.13
105	Robert Schumann	3 Fantasiestücke, Op. 111, No. 2 in A flat major, Piu toso lento- Un poco piu mosso – Tempo 1 CPO, 999 431-2 Volker Banfield, 1997	ABA'	0:53 0:15–1:08 B: 0:13.82 A: 0:40.09
106	Franz Schubert	Piano Sonata 19 in C minor, D. 958, Mvt. 2, Adagio Naxos, 8.550475 Jenö Jandó, 1991	ABA'	0:52 B: 0:18.23 A: 0:39.15
107 ^{L+*}	Felix Mendelssohn	Sonata in G minor, Op. 105, Adagio: Cantabile et Lento Naxos, 8.553358 Benjamin Frith, 1998	ABA'	0:55 4:40–5:35 B: 0:15.10 A: 0:31.96
108 ^{L+*}	Edvard Grieg	Lyriske Stykker, Vol. 1, Op. 12, No. 1, Arietta Victoria VCD 19025 Geir Henning Braaten, 1990	AABA'A'	0:52 fade out B: 0:11.75 A: 0:41.14
109 ^{L+*}	Johannes Brahms	3 Intermezzi, Op. 117, No. 1 in E flat major Nimbus Records, NI 5376 Martin Jones, 1992	AABA'A'	0:56 fade out B: 0:18.14 A: 0:38.50
110	Antonín Dvořák	Silhouettes, Op. 8, No. 6, Poco sostenuto Supraphon, SU 3290-2111 Radoslav Kvapil, 1997	ABA'	1:15 fade out B: 0:22.84 A: 0:50.13
201	Felix Mendelssohn	Piano Sonata 1 in B flat major, Op. 106, Mvt. 1, Allegro Vivace Naxos, 8.553186 Benjamin Frith, 1997	ABA'	0:47 0:08–0:55 fade out B: 0:10.87 A: 0:35.24
202	Robert Schumann	Kinderszenen, Op. 15, Wichtige Begebenheit Tacet, 153 Evgeni Koroliov, 2006	AABA'A'	0:51 B: 0:12.87 A: 0:37.80
203	Edvard Grieg	23 short pieces for piano (EG 104), Scherzo Victoria, VCD 19034 Geir Henning Braaten, 1993	ABA'	0:50 0:13–1:03 B: 0:13.17 A: 0:36.44
204 ^{H+*}	P. I. Tchaikovsky	The Seasons, Op. 37, February: Carnaval Virgin Classics, 7243 5 45042 2 8 Mikhail Pletnev, 1994	AABA'A'	1:11 B: 0:22.48 A: 0:49.86
205	Robert Schumann	Davidsbündlertänze, Op. 6, No. 3, Mit Humor Naxos, 8.550493 Benjamin Frith, 1991	AABA'A'	1:05 0:08–1:13 B: 0:17.35 A: 0:49.41

(continued)

Appendix. (continued)

Stimulus Code	Composer	Piece & CD	Form	Duration
206 ^{H+*}	Franz Liszt	Schubert Transcriptions, Franz Schuberts Märsche für das Pianoforte übertragen; Grande Marche caractéristique Hyperion, CDA66953 Leslie Howard, 1995	ABA'	0:55 fade out B: 0:17.92 A: 0:36.48
207 ^{H+}	Felix Mendelssohn	Seven Characteristic Pieces, Op. 7, No. 7, Leicht und luftig Naxos, 8.553541 Benjamin Frith, 1995	AABA'A'	1:04 fade out B: 0:18.39 A: 0:47.15
208	Franz Schubert	Sonata in B flat major, D 960, No. 4, Allegro, ma non troppo Harmonic Records, H/CD 8610 Paul Badura-Skoda, 1986	ABA'	1:29 3:30–4:59 fade out B: 0:18.36 A: 1:10.00
209*	Fryderyk Chopin	Polonaise in A flat major, "Heroic," Op. 53 Brilliant Classics, 99802/2 Folke Nauta, 1998	AABA'A'	0:58 0:30–1:30 fade out B: 0:19.00 A: 0:38.12
210 ^{H+*}	Edvard Grieg	Lyriske Stykker, Op. 57, No. 5, Hun danser BIS-CD-105 Eva Knardahl, 1987	ABA'	1:11 0:18–1:29 fade out B: 0:15.32 A: 0:55.11
301*	Robert Schumann	Bunte Blätter, Op. 99, Albumblätter, No. 4, Sehr langsam Tacet, 153 Evgeni Koroliov, 2006	ABA'	0:57 0:13–1:10 B: 0:11.49 A: 0:33.70
302	P. I. Tchaikovsky	Romance in F minor, Op. 5 Naxos, 8.550504 Ilona Prunyi, 1991	AABA'A'	1:12 fade out B: 0:27.15 A: 0:57.16
303	Fanny Mendelssohn-Hensel	Das Jahr, 12 Charakterstücke, No. 3, März Sony Classical, 88697030162 Lauma Skirde, 2007	ABA'	1:00 fade out B: 0:20.23 A: 0:38.31
304 ^{L-*}	Robert Schumann	Etudes symphoniques, Op. 13, Thema EMI Classics, 5 69521 2 Dimitri Alexeev, 1996	ABA'	1:03 0:22–1:25 B: 0:18.72 A: 0:40.30
305 ^{L-*}	Johannes Brahms	4 Balladen, Op. 10, Ballade No. 1 in D minor Nimbus Records, 5372 Martin Jones, 1992	AABA'A'	1:09 0:03–1:12 B: 0:27.84 A: 0:46.12
306 ^{L-*}	Franz Liszt	Alleluja et Ave Maria (Arcadelt) Mvt. 2, Ave Maria Naxos, 8.553073 Philip Thomson, 1997	AABA'A'	1:12 3:14–4:24 fade out B: 0:18.16 A: 0:53.33
307	Johannes Brahms	Variations on a theme by Robert Schumann in F sharp minor, Op. 9 Nimbus Records, 5372 Martin Jones, 1992	AABA'A'	1:15 B: 0:25.55 A: 0:50.16
308 ^{L+*}	Franz Schubert	Sonata in B flat major, D 960, No. 2, Andante Sostenuto Harmonic Records, H/CD 8610 Paul Badura-Skoda, 1986	ABA'	1:18 0:51–2:11 B: 0:17.49 A: 1:04.34

(continued)

Appendix. (continued)

Stimulus Code	Composer	Piece & CD	Form	Duration
309 ^{L-*}	Robert Schumann	Waldszenen, Op. 82, No. 3, Einsame Blume MIREs 29-017 Denise Trudel, 2006	AABA'A'	1:22 B: 0:34.22 A: 0:48.94
310	Fryderyk Chopin	Mazurka in E minor, Op. 17, No. 2 Brilliant Classics, 99802/8 Cor de Groot, 1998	ABA'	1:24 0:21–1:45 fade out B: 0:21.56 A: 1:05.03
401*	Felix Mendelssohn	Zwei Klavierstücke, No. 2, Presto agitato in G minor Naxos, 8.553358 Benjamin Frith, 1998	AABA'A'	1:00 fade out B: 0:18.87 A: 0:41.16
402 ^{H-*}	Johannes Brahms	8 Piano Pieces, Op. 76, No. 5, Capriccio in C sharp minor, Agitato, ma non troppo presto London Records, 430 053-2 Julius Katchen, 1990	AABA'A'	1:17 fade out B: 0:28.03 A: 0:53.79
403	Franz Schubert	Six Moments Musicaux, Op. 94, D. 780 Harmonia Mundi, 2908205 Alain Planès, 2000/2006	ABA'	0:54 fade out B: 0:16.06 A: 0:36.21
404 ^{H-*}	Edvard Grieg	Lyriske Stykker, Op. 71, Nr. 3, Kobold Deutsche Grammophon, 419 749-2 Emil Giles, 1996	ABA'	0:53 B: 0:13.52 A: 0:41.26
405	Felix Mendelssohn	Etude in F Minor, Op. 104 Naxos, 8.553358 Benjamin Frith, 1998	ABA'	0:57 0:16–1:13 fade out B: 0:18.43 A: 0:43.94
406 ^{H+*}	Franz Schubert	Drei Klavierstücke, D 946, No. 1 in E flat minor Harmonic Records, H/CD 8610 Paul Badura-Skoda, 1986	AABA'A'	0:52 B: 0:11.49 A: 0:41.48
407 ^{H-*}	Fryderyk Chopin	Scherzo No. 1 in B minor, Op. 20 Brilliant Classics, 99802/6 Alwin Bär, 1998	AABA'A'	0:58 0:06–1:04 B: 0:21.74 A: 0:38.12
408 ^{H-*}	Robert Schumann	Etudes symphoniques, Op. 13, Etude IV (Variation III) EMI Classics, 7234 573277 2 5 Tzimon Barto, 1990	AABA'A'	0:55 B: 0:27.42 A: 0:34.09
409	Johannes Brahms	6 Klavierstücke, Op. 118, Mvt. 3, Ballade in G minor London Records, 430 058-2 Julius Katchen, 1990	AABA'A'	0:50 fade out B: 0:13.13 A: 0:29.51
410 ^{H-}	Franz Liszt	Ungarische Rhapsodien, Vol. 2, No. 15 in A minor Naxos 8.554481 Jeno Jandó, 1998	ABA'	0:52 0:39–1:31 B: 0:15.21 A: 0:37.21