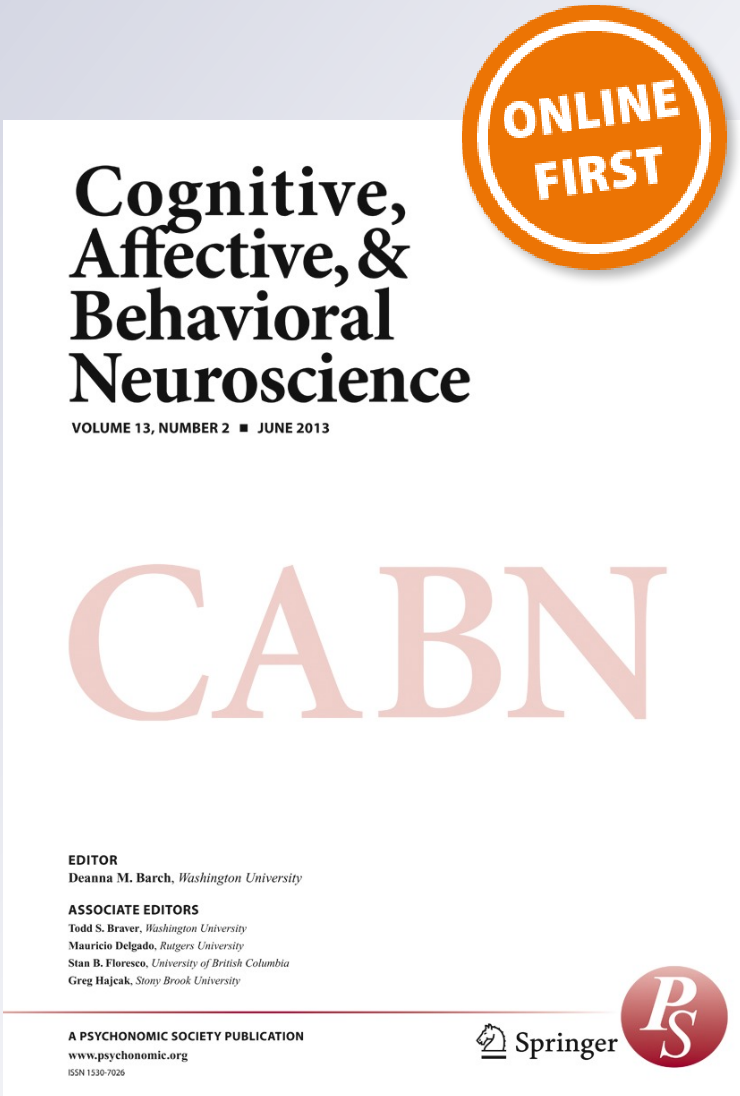
Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music

Hauke Egermann, Marcus T. Pearce, Geraint A. Wiggins & Stephen McAdams

Cognitive, Affective, & Behavioral Neuroscience

ISSN 1530-7026

Cogn Affect Behav Neurosci
DOI 10.3758/s13415-013-0161-y


ONLINE FIRST

Cognitive, Affective, & Behavioral Neuroscience
VOLUME 13, NUMBER 2 ■ JUNE 2013

CABN

EDITOR
Deanna M. Barch, *Washington University*

ASSOCIATE EDITORS
Todd S. Braver, *Washington University*
Mauricio Delgado, *Rutgers University*
Stan B. Floresco, *University of British Columbia*
Greg Hajcak, *Stony Brook University*

A PSYCHONOMIC SOCIETY PUBLICATION
www.psychonomic.org
ISSN 1530-7026

Springer

Springer

Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com”.

Springer

Author's personal copy

Cogn Affect Behav Neurosci
DOI 10.3758/s13415-013-0161-y

Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music

Hauke Egermann · Marcus T. Pearce · Geraint A. Wiggins · Stephen McAdams

© Psychonomic Society, Inc. 2013

Abstract We present the results of a study testing the often-theorized role of musical expectations in inducing listeners' emotions in a live flute concert experiment with 50 participants. Using an audience response system developed for this purpose, we measured subjective experience and peripheral psychophysiological changes continuously. To confirm the existence of the link between expectation and emotion, we used a threefold approach. (1) On the basis of an information-theoretic cognitive model, melodic pitch expectations were predicted by analyzing the musical stimuli used (six pieces of solo flute music). (2) A continuous rating scale was used by half of the audience to measure their experience of unexpectedness toward the music heard. (3) Emotional reactions were measured using a multicomponent approach: subjective feeling (valence and arousal rated continuously by the other half of the audience members), expressive behavior (facial EMG), and peripheral arousal (the latter two being measured in all 50 participants). Results confirmed the predicted relationship between high-information-content musical events, the violation of musical expectations (in corresponding ratings), and emotional reactions (psychologically and physiologically). Musical structures leading to expectation reactions were manifested in emotional reactions at different emotion component levels (increases in subjective arousal and autonomic nervous system activations). These results emphasize the role of musical structure in emotion induction, leading to a further understanding of the frequently experienced emotional effects of music.

Keywords Emotion · Music · Expectation · Statistical learning · Computational modeling · Psychophysiology

Music has been shown to induce emotional reactions that are accompanied by activations in several reaction components: subjective feelings, psychophysiological activations, and expressive behavior (Juslin & Västfjäll, 2008). However, most previous experimental research has been rather exploratory, showing that music induces emotion, but not providing theoretically founded explanations for the phenomena observed. More than a decade ago, Scherer and Zentner (2001) noted: "This is a bad omen for future research, since it is to be feared that additional, isolated research efforts with little or no theoretical underpinnings are more likely to add to the current confusion than to the insight to which the researchers aspire" (p. 382). However, beginning with Scherer and Zenter's paper, several theoretical attempts have been made to explain the underlying mental processes that are involved in creating emotional responses to music. Scherer and Zentner formulated "production rules" describing in detail several mental mechanisms that could be used to explain emotional responses to music. A few years later, Juslin and Västfjäll continued this idea and presented a seminal review paper, positing seven possible ways to explain the observed effects of music. Here, they summarized previous ideas about emotion induction mechanisms in general and those specific to music: cognitive appraisal of music and the listening situation, brain stem reflexes to acoustic characteristics, visual imagery induced through sound, evaluative conditioning from pairing music with another emotion-inducing stimulus, emotional episodic memory associated with the music, emotional contagion through emotional expressions in the music, and musical

H. Egermann · S. McAdams
McGill University, Montreal, Canada

H. Egermann (✉)
Audio Communication Group, Technische Universität Berlin,
Sekr. EN-8, Einsteinufer 17c,
10587 Berlin, Germany
e-mail: hauke@egermann.net

M. T. Pearce · G. A. Wiggins
Queen Mary University of London, London, UK

Published online: 20 April 2013

Author's personal copy

expectation. The last mechanism will be the focus of the study presented here. There have been many theoretical and empirical attempts to link musical structures and expectations, but empirical evidence explicitly investigating the connection between expectation and emotion in music is limited. Therefore, we conducted an experiment in which we tested whether statistical properties of composed musical structures violate or confirm subjective expectations, and also whether they lead to emotional reactions in subjective, expressive, and physiological response components. In order to maintain a naturalistic research paradigm, we conducted the experiment in a live concert setting, using an audience response system developed in-house that measured participants' reactions in real time. Advancing the theory on the underlying mechanisms of musical expectation, we furthermore used a computational machine-learning algorithm to analyze the music presented and predict human expectations and emotions.

Musical expectations

As early as the 1950s, Leonard B. Meyer (1956, 1957) began to theorize about the relationships between musical structures and listener's expectations (which may be confirmed or violated). Emphasizing the role of cultural learning through exposure to the syntactical properties of music, he engendered a great deal of scholarship and empirical research, describing and testing which musical properties create which expectations. Reviewing this work, Huron (2006) suggested that there are four different types of expectation associated with music and created by different auditory memory modules. *Veridical expectations* are derived from episodic memory and contain knowledge of the progression of a specific piece. *Schematic expectations* arise from being exposed to certain musical styles and contain information about general event patterns of different musical styles and music in general (based on semantic memory). *Dynamic expectations* are built up through knowledge stored in short-term memory about a specific piece that one is currently listening to and are updated in real time through listening. Finally, Huron also described *conscious expectations* that contain listeners' explicit thoughts about how the music will sound.

Expectations brought up by melodic, harmonic, and rhythmic features have been studied and researched extensively. However, because it remains difficult to experimentally differentiate between the different forms of expectation (Huron, 2006), research has focused mostly on the role of different musical features in evoking expectations. Rhythmic periodicities of music have been shown to create preattentive hierarchical expectations of beat and meter in listeners (Ladinig, Honing, Háden, & Winkler, 2009), which have psychophysiological correlates in high-frequency gamma-band activity in the auditory cortex (Zanto, Snyder, & Large, 2006). Harmonic expectations have been investigated by comparing responses to chords varying in harmonic relatedness to the context with more distantly related chords (assumed to be less expected) leading to longer reaction times in priming tasks (Bharucha & Stoeckig, 1986, with chord distance quantified by the number of shared parent keys), delayed and lower completion/expectation ratings (Bigand & Pineau, 1997; Schmuckler & Bolz, 1994), and several specific event-related brain potentials, such as the P300 (Carrión & Bly, 2008; Janata, 1995). Concerning melodic expectations, several theoretical models making expectation-related predictions have been suggested (Larson, 2004; Margulis, 2005; Ockelford, 2006) with partial empirical support. Music theorist Eugene Narmour (1990, 1992) was among the most popular, proposing several melodic principles in his implication-realization theory, which are intended to describe the expected melodic continuation of implicative intervals. Some of those principles, such as the principle of pitch proximity, have been confirmed in experimental testing (Cuddy & Lunney, 1995; Schellenberg, 1996, Thompson & Stainton, 1998). Unlike Meyer (1956), Narmour conceived some of his melodic organization principles as universal, innate, and bottom-up processes, similar to Gestalt principles of perception. However, recent theories of auditory statistical learning, also supported by evidence reported by Huron (2006) and Pearce and Wiggins (2006), propose that melodic expectations do not rely on underlying patterns of universal bottom-up principles but have merely been formed through exposure to syntactic relationships within musical structures of a given culture (Abdallah & Plumbley, 2009; Pearce & Wiggins, 2006).

Furthermore, computational simulations of this learning process have yielded robust predictions of perceptual expectations, outperforming other rule-based models like Narmour's (1990, 1992). For that reason, our experiment uses a computational model of auditory expectation to specify precise, quantitative measures of structural predictability for each note in a melody (*the information dynamics of music model* [IDyOM]). The model itself has been presented (Pearce 2005; Pearce, Conklin, & Wiggins, 2005; Pearce & Wiggins, 2004) and evaluated (Pearce, 2005; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010; Pearce & Wiggins, 2006) elsewhere, so here we just provide a brief overview.

The central feature of the model is that it learns about sequential dependencies between notes in an unsupervised manner through exposure to melodies. At any given point in processing a melody, the model generates a probability distribution governing some property of the next note (e.g., its pitch or onset time). This probability distribution reflects the prior experience of the model and represents its expectations about the next note in the melody. The learning and generation of probabilities is achieved using a Markov or n-gram model (Manning & Schütze, 1999), which computes the conditional probability of a note given the n−1

Author's personal copy

preceding notes in the melody. The quantity $n-1$ is called the order of the model. In IDyOM, basic Markov modeling is extended in three ways.

First, the model is of variable order, incorporating an interpolated smoothing strategy to combine probabilities from models of different order. This allows the system to benefit from both the structural specificity of longer (but relatively rare) contexts and the statistical power of more frequent (but less specific) low-order contexts. Second, the model is configured with two subcomponents: a long-term model (LTM), which is exposed to an entire corpus (modeling learning based on a listener's long-term exposure to music), and a short-term model (STM), which is exposed only to the current musical material (modeling local learning of the structure and statistics in the current listening episode). The full model (BOTH) generates probability distributions by combining those generated by the LTM and STM.

Third, the system has the ability to use a combination of different features, or *viewpoints*, to predict the properties of notes. We do not use this aspect of the system in the present research but refer the interested reader to the literature on *multiple viewpoint systems* (Conklin & Witten, 1995; Pearce et al., 2005).

The use of this system as a cognitive model of auditory expectation is motivated by empirical evidence of implicit learning of statistical regularities in musical melody and other sequences of pitched events (Oram & Cuddy, 1995; Saffran, Johnson, Aslin, & Newport, 1999). Consistent with an approach based on statistical learning, melodic pitch expectations vary between musical styles (Krumhansl et al., 2000) and cultures (Carlsen, 1981; Castellano, Bharucha, & Krumhansl, 1984; Eerola, 2004; Kessler, Hansen, & Shepard, 1984, Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999), throughout development (Schellenberg, Adachi, Purdy, & McKinnon, 2002), and across degrees of musical training and familiarity (Krumhansl et al., 2000; Pearce, Ruiz, et al., 2010). The use of LTMs and STMs is motivated by evidence that pitch expectations are informed both by long-term exposure to music (Krumhansl, 1990) and by the encoding of regularities in the immediate context (Oram & Cuddy, 1995). Tillmann and colleagues have shown that target chords are processed more accurately and quickly when they are related both to the local and to the global harmonic contexts (previous chord and prior context of six chords, respectively; Tillmann, Bigand, & Pineau, 1998) and that these effects can be explained by a mechanism of implicit statistical learning of sequential harmonic patterns in music (Tillmann, Bharucha, & Bigand, 2000).

With regard to melodic expectation, the model summarized above has been tested by comparing its pitch expectations with those of human listeners (Omigie, Pearce, & Stewart, 2012; Pearce, Ruiz, et al., 2010; Pearce & Wiggins, 2006). In a series of reanalyses of existing behavioral data (Cuddy & Lunney, 1995; Manzara, Witten, & James, 1992; Schellenberg, 1997), it was shown that this model predicts listener's expectations better than do existing models of melodic expectation based on innate principles (Narmour, 1990; Schellenberg, 1997). Using a novel visual cueing paradigm for eliciting auditory expectations without pausing playback, Pearce, Ruiz, et al. (2010) confirmed that the model predicts listeners' expectations in melodies without explicit rhythmic structure.

Music and emotion

Here, the term *emotion* is used in the sense of the *component process model* presented by Scherer (2004, 2005). According to this model, an emotion episode consists of coordinated changes in three major reaction components: (1) physiological arousal, (2) motor expression, and (3) subjective feelings. There are two major theoretical positions concerning emotional effects of music. The *cognitivist* position states that music is capable only of *representing* emotion, and not of *inducing* emotions similar to those occurring in everyday life with synchronized reaction components and object focus (e.g., being angry about something that presents an obstacle to reaching one's personal goals; Kivy, 1990; Konečni, 2008). According to the *emotivist* view, music does indeed induce emotions similar to those induced by other events in everyday life, often demonstrated by citing the research that shows emotional reactions in all components (Juslin & Västfjall, 2008). For example, Lundqvist, Carlsson, Hilmersson, and Juslin (2008) showed that music-induced feelings of happiness or sadness were associated with activations of the autonomic nervous system (measured through skin conductance) and activations of expressive facial muscles. In addition, Grewe, Kopiez, and Altenmüller (2009) showed that strong emotional responses to music, like the *chill* response (experience of shivers or goose bumps), have been accompanied by increases in felt emotional intensity, skin conductance, and heart rate (HR). Finally, it has also been shown recently that those strong music-induced emotions are manifested neurochemically by dopamine release in the reward system in the human brain, in a similar manner to other pleasurable stimulations such as food intake, sex, or drugs (Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011). However, cognitivists often argue that all this empirical evidence does not demonstrate that the music itself stimulated these emotional responses, because external emotional objects might also have been associated with the music, making it appear to induce emotion (Konečni, 2008). In order to prove that music is able to induce emotions on its own, one would have to show that musical structures by themselves generate emotional responses in listeners without external reference (Cochrane, 2010).

Author's personal copy

Linking expectation and emotion

Musical expectation is a good candidate for demonstrating this emotional induction by the music itself without the help of any external association. Huron (2006) proposed the ITPRA theory of expectation giving a detailed account of five different response types (grouped into two pre- and three postevent responses). The imagination response accounts for emotional reactions to imaginative processes before the occurrence of a musical event. The tension response functions as physiological preparation for an anticipated event by adjusting the needed arousal. After this musical event has occurred, the prediction and reaction responses happen simultaneously. Here, the accuracy of the prediction is rewarded or punished (prediction response), and the pleasantness of the outcome itself is evaluated in a fast and less accurate way (reaction response) and in a slower and more elaborated way leading to the appraisal response. Thus, there are several affective phenomena associated with expectation that are more or less related to everyday emotions. Imagination responses may be very subtle and difficult to separate in measurement from responses to the event itself. Tension responses to music have been researched widely (Krumhansl, 1996, 2002) and are likely to lack a coordinated event-related onset. The three postevent responses described by Huron are more likely to create reactions that are synchronized across emotional components and measurable in an experimental research design. These responses might create surprise (potentially leading to strong emotions like chills), pleasure from making correct predictions or appraising false predictions as not harmful, and also displeasure from making wrong predictions (Huron, 2006). However, expectation might also influence musical experience in another way. It may be necessary to differentiate event-related emotions from the perceptual qualities that arise from statistical properties of musical structures like the scale degree qualia in tonality or rhythm that, according to Huron, produce feelings of closure, tendency, and pleasure. Those qualities may be different from emotions, at least as we have defined them, in being undetectable consciously (Margulis & Levine, 2006) and too weak to measure in a real-time listening context.

Some of the first empirical evidence for a link between expectation and emotion was presented by Sloboda (1991), who reported that musical structures like unexpected harmonies can induce strong emotions. However, this finding must be viewed as merely suggestive, because it is based only on retrospective reports in a survey. To our knowledge, there are only three published empirical studies explicitly linking musical expectation to emotional responses (Koelsch, Fritz, & Schlaug, 2008; Koelsch, Kilches, Steinbeis, & Schelinski, 2008; Steinbeis, Koelsch, & Sloboda, 2006). Steinbeis et al. and Koelsch, Kilches, et al. showed that harmonies that contravene the principles of Western tonal harmony (presumably violating listeners' expectations) lead to increases in retrospective emotion ratings and continuous tension ratings, with corresponding increases in skin conductance but no correlated changes in continuous emotion rating and HR. Employing a similar research paradigm, Koelsch, Fritz, and Schlaug further demonstrated that irregular chord sequences ending on a Neapolitan sixth chord instead of the tonic chord, thus being presumably less expected, lead to bilateral activations of the amygdala (associated with negative emotional processing) and are also rated as being less pleasant. However, both of these studies are limited in their external validity, because only the effects of listening to intensively repeated and artificially recomposed chord progression endings were measured and participants provided no ratings of subjective expectation.

Aims of the study

While the idea that expectation confirmation and violation in musical listening can induce affective responses has a venerable history (e.g., Meyer, 1956), quantitative empirical evidence for this impact has not yet been established. We aim to provide such evidence, using both a computational model of auditory expectation (Pearce, 2005) and subjective ratings to quantify the expectedness of events in a live performance of solo flute music and to relate these measures quantitatively to the psychophysiological emotional state of the audience.

In doing so, we address the identified limitations of previous work by using real, naturally performed compositions and by gathering subjective unexpectedness ratings during listening. In order to increase ecological validity, the study was conducted during a live concert using an audience response system developed by the research team. Previous research suggests that continuous subjective experience ratings in similar settings can be successfully employed to assess the emotional effects of large-scale music structures (McAdams, Vines, Vieillard, Smith, & Reynolds, 2004) or audience response to dance (Stevens et al., 2009). However, in the present study, additional assessment of physiological indicators of emotional experience was added. Furthermore, in contrast to previous research on expectation and emotion, we also predict listeners' expectations using a computational model that is theoretically grounded in statistical auditory learning (Pearce, 2005). We also focus in this study on expectations generated by complex melodic styles that have not previously been investigated in this context.

We predicted that musical events with low conditional probability, as compared with those with high probability, would be experienced as unexpected and would, at the same time, induce emotions that are manifest as changes in the activity of all three response components measured: increased

Author's personal copy

autonomic arousal, expressive facial muscle activity, and subjective feeling. By identifying highly unexpected and expected musical events using a computational model of auditory cognition, we ensure that we include events whose expectedness is based on implicit schematic memory and is therefore not available to conscious introspection (Huron, 2006). However, the cognitive model can capture only the effects of statistical learning and memory from the local musical context and global schematic musical context. It does not, for example, account for the effects of veridical knowledge or audiovisual performance cues on expectations. Therefore, in a second part of our analyses, we used the continuous unexpectedness ratings of participants to identify events in the entire performance that were highly unexpected and tested whether they were also accompanied by emotional reactions.

Method

Participants

Participants were recruited via several e-mail lists. They were screened with the help of an online questionnaire before taking part to ensure that they had some familiarity with and preference for classical music, had normal hearing, would show willingness to be filmed, and were willing to wear no makeup and to shave (due to facial electrode placement for females and males, respectively). Fifty participants were selected (21 female), with an average age of 23 years (*SD* = 6 years). With the exception of two nonmusicians, all were recruited as amateur (n = 32) or professional (including university music students; n = 16) musicians. We made this decision because, in general, musicians were assumed to be more practiced at listening to music analytically and, thus, were presumably better at the continuous unexpectedness rating task. They have also been previously shown to react more emotionally to music than nonmusicians (Grewe et al., 2009; Grewe, Nagel, Kopiez, & Altenmüller, 2007), increasing the probability of finding expectation-induced emotional responses. Participants were randomly assigned to two different continuous rating tasks. One half continuously rated their subjective feelings listening to the music, and the other continuously rated the unexpectedness of the musical events presented. All were paid $10 Canadian as compensation.

Stimuli description and analyses

The music was selected to represent different musical styles, chosen from the performing musician's current repertoire. Table 1 presents all six pieces used in this concert. First, two recorded flute pieces were presented to participants to familiarize them with the continuous rating task. Subsequently, a highly recommended flute performance student played the other four pieces live on stage.

The computational model used to analyze this music was set up as follows. On the basis of the composed MIDI representation of the music, the pitch of each note in the six pieces was predicted using a variable-order context and a simple pitch viewpoint (i.e., no derived viewpoints such as pitch interval, contour, or scale degree were used in predicting pitch; cf. Pearce et al., 2005). The model was configured using a combination of the LTM and STM (i.e., a BOTH configuration; see description above). The LTM is intended to reflect the schematic effects of long-term exposure to music on expectations, whereas the STM is intended to reflect the effects of online learning of repeated structures within each individual composition. The LTM was trained on a corpus of 903 melodies from Western folk songs and hymns as used by Pearce and Wiggins (2006); therefore, the expectations encoded by the LTM are for tonal music.

For each note in each melody, the model returns an estimate of the conditional probability of the note's pitch given the pitches appearing previously in the melody. These probabilities are converted into information content (IC), the negative logarithm to the base 2 of the probability, which is a lower bound on the number of bits required to encode an event in context (Mackay, 2003). The IC represents the model's estimate of how unexpected the pitch of each note is. Thus, for every piece, a time series of one IC value per note was generated and used for further analysis.

Table 1 Music stimuli presented

Order of presentation	Title	Composer	Presentation mode	Duration (min:s)
1.	Acht Stücke Für Flöte Allein: VI. Lied, Leicht Bewegt	Paul Hindemith	recorded	0:38
2.	Un Joueur de Flûte Berce les Ruines	Francis Poulenc	recorded	1:18
3.	Density 21.5	Edgar Varèse	live	3:30
4.	Syrinx	Claude Debussy	live	2:35
5.	Solo Partitas for Flute, A-minor: 2nd Movement “Corrente”	Johann S. Bach	live	1:53
6.	Solo Partitas for Flute, A-minor: 3rd Movement “Sarabande”	Johann S. Bach	live	2:11

Author's personal copy

Measurements

All participants were equipped with an iPod Touch (Apple Inc., Cupertino, CA) that was fixed on the thigh of the dominant leg (assessed by self-reported handedness) with the help of a Velcro strip. Both groups of participants were asked to keep their finger on the iPod surface during the presentation of each complete piece.

Continuous rating of emotion

For one half of the participants, the iPod displayed an emotion space, based on the two-dimensional emotion model with vertical arousal and horizontal valence dimensions (Russell, 1980). The heuristic value of the two-dimensional emotion space has been confirmed in numerous previous studies measuring emotional expressions and inductions through music (e.g., Egermann, Grewe, Kopiez, & Altenmüller, 2009; Egermann, Nagel, Altenmüller, & Kopiez, 2009; Nagel, Kopiez, Grewe, & Altenmüller, 2007; Schubert, 1999) and, as a consequence, was also adopted in the present study in order to capture participants' emotional responses. By moving the index finger of their dominant hand from left to right, participants were instructed to indicate how pleasant the effect of the music was (left = negative and unpleasant; right = positive and pleasant). We followed Russell's original definition of valence as "pleasantness of the induced feeling," instead of the also commonly used "emotion valence," where participants have to rate the valence of this emotion as it would occur in everyday life (Colombetti, 2005). By moving their finger from top to bottom, participants indicated the degree of their emotional arousal while listening to the music (top = excited; bottom = calm). Participants were instructed to rate their current emotional state on both dimensions simultaneously, with the finger position at each moment reflecting their emotional response to the piece as they were listening. They were also asked not to rate emotions recognized, but only their own emotional response. In order to help participants to scale their ratings, the extremes of the rating scales were defined to represent the extremes of participants' emotional reactions to music in general in everyday life.

Continuous rating of unexpectedness

For the other half of the participants, the iPod displayed a one-dimensional vertical unexpectedness rating scale, which was developed in an internal pretest ($n = 9$) comparing four different interface designs that could be used to continuously capture expectation: (1) continuous assessment of the fit of the current musical event to previous context (similar to Krumhansl et al., 2000), (2) feeling of surprise (Huron, 2006), (3) continuous rating of unexpectedness of musical events (Pearce, Ruiz, et al., 2010), and (4) buttonpresses indicating unexpected musical events. Interface (3) was finally chosen for the concert experiment on the basis of participants' evaluations concerning ease of use and comprehensibility of instructions. In the concert experiment, participants were instructed to rate continuously with their index finger during the music presentation the unexpectedness of the musical events. Both rating interfaces and instructions employed are presented in the Appendix.

Psychophysiological measurements

Physiological measurements were recorded through 50 ProComp Infiniti (Thought Technology Ltd., Montreal, Canada) units that were taped to the back of the participants' seats. Each ProComp Infiniti was connected with four others via an optical cable and an optical-to-USB converter to a functionally expanded Asus router. This device converted incoming signals into TCP/IP packets that were sent via network cables to several switches that were all connected to one Mac Pro workstation (Apple Inc., Cupertino, CA). Here, a custom program received all data packets and stored them on an internal hard disk. Respiration was measured using a belt with a stretch sensor attached around the chest. Blood volume pulse (BVP) was measured using a photoplethysmograph on the palmar side of the distal phalange of the middle finger of the nondominant hand. Skin conductance was measured using electrodes on the distal phalanges of the index and middle fingers of the nondominant hand. Expressive muscle activations were measured using two electromyography (EMG) electrodes (MyoScan-Pro surface EMG sensors) placed on the corrugator supercilii (associated with frowning) and zygomaticus major (associated with smiling) muscles (Cacioppo, Petty, Losch, & Kim, 1986). EMG electrodes were placed on the side of the face contralateral to the dominant hand (with positive and negative electrodes aligned with the respective muscles and the reference electrodes placed on the cheek bone/forehead).

Questionnaires

Participants completed questionnaires on a clipboard after every piece, including a question about their familiarity with the piece (rated on a 7-point scale from 1 = *unfamiliar* to 7 = *familiar*). At the end of the concert, participants also filled in a general questionnaire including background variables about socio-demographic characteristics, musical training, and music preferences.

Audiovisual recordings and analyses

Three different HD video cameras (Sony XDCAM) recorded the entire experiment with synchronized time code. One faced the performer, and the other two each faced a different half of the audience in order to be able to monitor participants' behavior during the experiment. Two pairs of microphones

Author's personal copy

attached to two of the three cameras recorded the music performed: one about 2 m away from the flute performer (DPA 4011), and the other one binaurally using a dummy head placed in the middle of the concert hall (Neumann KU100 Kunstkopf). The signal recorded with the DPA 4011 microphones was also recorded in a low sample-rate version on the computer together with the physiological signals in order to synchronize behavioral and physiological data with the audio recording, using corresponding time codes and visual identification of the time lag between the two types of recordings. The audio signal from the two DPA 4011 microphones was also recorded in high quality on a MacBook Pro with an external sound card; this high-quality recording was then used to detect the onset times of the notes played during the concert. In this way, all MIDI versions created from the composed score were visually overlaid on the peak pitch display of all six audio recordings using Sonic Visualizer (Cannam, Landone, & Sandler, 2010). Subsequently, MIDI note events were manually aligned with the performed notes creating a MIDI reproduction of the performance, used to align analyses of event-related responses of participants.

Procedure

The experiment was conducted in Tanna Schulich Recital Hall at McGill University starting at 7:00 p.m. Participants were asked to come in 1 hour earlier to allow enough time for seating and sensor placement. At the entrance, they were handed written instructions with questionnaires and the respiration belt. They were shown how to fix it on their own and were handed skin cleaning tissues to clean their face and finger tips. Subsequently, they were assigned a seat number in the first eight rows of seats (alternating empty rows with seated rows to allow access to participants). The two different rating groups were placed on alternate seats (seat 1 = emotion rating, seat 2 = unexpectedness rating, seat 3 = emotion rating, . . .) in order to reduce visibility of the subjective response interface between members of the same group. After sitting down, participants read written instructions and provided their informed consent. Then a team of 10 assistants attached the electrodes and visually tested sensor placement on the recording computer's live display of incoming signals. Afterward, the experimenter gave a talk repeating every detail of the written instructions (approximately 10 min long), allowing participants to ask questions about the procedure and instructions. Subsequently, the six pieces of music were presented to participants in the following manner. Before every piece, we recorded 45 s of physiological baseline activity without any stimulation (for live performed pieces without performer on stage). Then the music was presented, and after each piece, participants filled out the associated form. Finally, participants filled in the final form, were detached from the sensor cables, returned their iPods, and received their compensation. During baseline recording and stimulus presentation, participants were instructed not to move, so as to reduce movement artifacts in physiological recordings.

Data analyses

Physiology

Preprocessing of all continuous signals recorded with a sample rate of 256 Hz was done in MATLAB (Mathworks, Version 7.14.0.739). Due to technical malfunction, physiological recordings from 2 participants could not be used. Visual inspection of all other recordings revealed that there were no other significant measurement errors. However, due to some scattered sample loss in transmission to the recording Mac Pro workstation (in the range of 12–72 dropped samples), all signals were linearly interpolated at the original sample rate first. This posed no problems to the analyses, since all physiological signals recorded here are known to change on a much longer timescale. Subsequently, BVP (low pass 2 Hz), respiration activity (low pass 1 Hz), and skin conductance (low pass 0.3 Hz) were filtered in order to remove extraneous information, using a linear phase filter based on the convolution of a 4th-order Butterworth filter impulse response (also convolved with itself in reverse time in order to avoid phase shifting). Creating a measure for skin conductance response (SCR), the phasic component of the skin conductance signal (Boucsein, 2001), we performed linear detrending on the corresponding recording, also in order to remove any negative trends over time with breakpoints every 60 s (which are caused by an accumulation of charge over time between the skin and sensor; see Salimpor, Benovoy, Longo, Cooperstock, & Zatorre, 2009). We extracted continuously interpolated HR and respiration rate (RespR) in beats per minute from the BVP and respiration signals by inversing the interbeat period (detected by identifying adjacent minima). The MyoScan-Pro EMG sensors automatically converted their signal to a root mean square signal (after an internal analog rectification), which was therefore not preprocessed any further (capturing EMG activity at frequencies up to 500 Hz). By subtracting from the filtered and extracted signals the mean baseline activity in the silent 40 s preceding each stimulus presentation, we finally removed any linear trends over the course of the concert and individual differences in baseline physiological activity (baseline normalization).

Continuous iPod ratings

Due to technical malfunctioning, data were missing for one iPod emotional rating. Since only rating changes and their corresponding time points were recorded as iPod data, we first programmed a stepwise interpolation function to sample participants' ratings at a rate of 256 Hz. We subsequently checked

Author's personal copy

that all participants provided changing ratings for all pieces. This analysis indicated that 1 participant failed to use the rating device during piece 6, leading to the removal of the corresponding data. We then removed any individual differences in scale use by individual range normalization (dividing each participant's rating by his or her individual range of ratings over the entire concert and then subtracting each participant's resulting minimum rating value over the entire concert, creating a range for each participant from 0 to 1).

Event-related statistical response analyses

In order to test for significant event-related changes in all continuous responses, we employed a novel linear mixed-effects modeling (LMM) approach (West, Welch, & Galecki, 2007), similar to a conventional linear regression analysis, that allowed estimation of significant coefficients of predictors controlling for random sources of variance and nonindependent observations in the data set (autocorrelation). Furthermore, this procedure also allowed for significance tests with high statistical power, since the event-related response data were not averaged over conditions per participant. We included crossed random effects for participants and two items (unexpected events within music pieces), in a way suggested by Baayen, Davidson, and Bates (2008). Equation 1 illustrates the general model formulation by these authors (with random effects for participants and one item):

$$y_{ij} = X_{ij}\beta + S_i s_i + W_j w_j + \varepsilon_{ij}, \quad (1)$$

where, y_{ij} denotes the responses of subject *i* to item *j*. X_{ij} is the experimental design matrix, consisting of an initial column of ones (representing the intercept) and followed by columns representing factor contrasts and covariates. This matrix is multiplied by the population coefficients vector β. The terms $S_i s_i$ and $W_j w_j$ help to make the model's predictions more accurate for the subjects and items (pieces of music and nested events) used in the experiment. The S_i matrix (the random effects structure for subject) is a full copy of the X_{ij} matrix. It is multiplied with a vector specifying the adjustments required for subject *i*. The W_j matrix represents the random effect for item *j* and is again a copy of the design matrix X_{ij}. The vector w_j contains adjustments made to the population intercept for each item *j*. The last term is a vector of residual errors ε_{ij}, including one error for each combination of subject and item. As suggested by Baayen et al., all analyses were conducted using the software R (2.13) using the lmer function from the lme4 package (Bates, Maechler, & Bolker, 2011). Estimation of parameters was based on restricted maximum likelihood, and likelihood ratio tests were used to test the significance of random effects. Significance of fixed predictors was tested using the pamer.fnc function (Tremblay, 2011), which outputs upper- and lower-bound *p*-values based on ANOVAs with upper and lower numbers of degrees of freedom (due to the addition of random effects to the linear model; Baayen et al., 2008). However, due to the large sample size investigated in this study, *p*-values obtained with both degrees of freedom were never computationally different, and only one will be reported.

Results

Evaluation of method

Since the method of conducting a live concert experiment with psychophysiological recordings is new and may lack some control over the experimental setting, we asked participants to evaluate the experiment along several dimensions concerning their experience of the experimental setting (Table 2).

Table 2 Participants' evaluation of live concert experiment separated by continuous rating task group (n = 50)

Question	Group		*F*	*p*
	Emotion rating: Mean (*SD*)	Unexpectedness rating: Mean (*SD*)		
1. How much did you interact with others during the pieces?	1.28 (0.67)	1.32 (0.69)	0.04	.84
2. How much did you interact with others between the pieces?	1.48 (0.77)	1.5 (0.89)	0.01	.93
3. Did you feel comfortable in the listening situation?	3.52 (1.12)	3.48 (1.08)	0.02	.90
4. Were your emotional responses influenced by the other people in the room during music listening?	1.6 (0.96)	1.38 (0.64)	0.92	.34
5. Did the other music listeners distract you from music listening?	1.4 (0.76)	1.25 (0.53)	0.63	.43
6. Did the iPod rating affect your listening negatively?	2.33 (1.01)	2.21 (1.25)	0.15	.70
7. How intuitive was the iPod rating?	3.44 (0.96)	3.17 (1.17)	0.8	.37
8. Indicate the degree to which the sensors interfered with your listening to the piece.	2.36 (1.04)	2.52 (1.45)	0.2	.66

Note. Q1–Q8: one-factorial (emotion vs. unexpectedness rating group) ANOVAs. Rating scales were labeled as follows: for Q1–Q2, 0 = *very little*, 5 = *very strongly*; for Q3–Q7, 0 = *very little*, 5 = *a lot*; and for Q8, 0 = "*No interference,*", 6 = "*A great deal of interference.*"

Author's personal copy

Participants rated their own degree of interaction with each other and distraction by others to be very low. This rating was also validated through inspection of the video recordings of participants. No one had to be excluded for not following the instructions. They also indicated, on average, that they felt quite comfortable in the listening situation and that they had not been influenced by the presence of other people in the concert. Finally, they reported that continuous iPod ratings probably did not influence their experience negatively (group mean was slightly lower than the middle of the rating scale), and the rating device was rather intuitive (group mean was higher than the middle of the rating scale). Interference of sensors with listening was also rated to be low, on average. There were no significant differences on any of those rating scales between the two groups with different continuous rating tasks (emotion vs. unexpectedness rating). In summary, both groups understood their rating tasks reasonably well and indicated that their reported results were not influenced much by the experimental aspect of the musical setting.

The following section, testing the proposed hypotheses, is structured in the following stages. First, we identify musical events corresponding to outstanding peaks of IC and subjective expectation across the entire concert. Then we test whether IC at these peak events predicts changes in unexpectedness ratings. Next, we test whether those segments identified as IC and subjective unexpectedness peaks also predict changes in psychophysiological response measures of emotion.

Identifying unexpected and expected moments

In order to identify very unexpected or very expected events in the music, the single note events presented to participants were grouped into short segments to ensure that the epochs selected for our subsequent analyses would be long enough to elicit a response in participants. Therefore, a trained music theorist carried out a motivic analysis on all six music pieces, identifying coherent melodic units segmented at a level of about one to two measures per unit. She identified 193 segments that were, on average, 3.7 s (SD = 2.5 s) long. These segmentations were compared with the independent analyses of a second music theorist and showed a high similarity. We then calculated, for each segment, the mean IC of all notes within that segment and the mean of the unexpectedness ratings. In order to identify very unexpected moments in corresponding individual ratings, we first differentiated them (using the diff function in MATLAB) and then averaged across the entire group of raters (indicating segments with high increases or high decreases in corresponding ratings). Figure 1 illustrates the results of this segmentation for the third piece presented (*Density 21.5* by Edgar Varese). The first row presents the IC of each note, and the second row contains the corresponding averaged IC per segment. Rows 3 and 4 present participants' unexpectedness ratings (third row, group mean; fourth row, mean of rating change per segment). High rating values correspond to the experience of unexpected events, whereas low values correspond to expected events.

Subsequently, we identified peak segments in both the averaged IC and unexpectedness rating time series by computing percentages of corresponding distributions across the entire concert. We excluded segments that were too close to the beginning or the end of pieces to extract event-related response time windows (see below), leaving 183 segments in the data set. Since participants already showed consensual responses in continuous ratings in the first two practice pieces, we also decided to include those two in these analyses. Figure 2 presents the distributions of all segments for mean IC and mean unexpectedness ratings for all six pieces. Segments with corresponding values higher than the 90th percentile of their distribution were classified as high-IC peaks (n = 18)/very unexpected moments (n = 19). Segments with values lower than the 10th percentile of their distribution were classified as low-IC troughs (n = 19)/very expected moments (n = 18). The dashed lines in Fig. 2 illustrate the corresponding percentile thresholds. Table 3 presents a cross-tabulation of the resulting two segment variables, coding each segment as a peak, a trough, or not used for both IC and unexpectedness ratings. Several segments were identified as peaks or troughs in both the IC and the unexpectedness ratings. Pearson's chi-squared test identified that at this level of analysis, there was a significant association between IC event type and unexpectedness rating event type, $\chi^2(4) = 27.2$, $p < .001$. Table 4 presents a cross-tabulation of the resulting segment variables and piece. Peak and trough segments identified by unexpectedness ratings were evenly distributed across the six pieces of music. However, peak and trough segments identified by IC were not as equally distributed: Half of the high-IC segments came from the third piece (*Density 21.5*).

Figure 1 also includes labels identifying peak IC and subjective expectation segments. For example, segment (a) was identified only as a high-IC peak, whereas segments (d) and (e) were identified as peaks on both analyses (see also corresponding score excerpts). Segments (b) and (c) were identified in the continuous unexpectedness ratings as very expected, presumably because they included repetitions of one of the main motives (sharing a similar rhythmic and pitch structure on the first three notes). This example piece contained no segments that were identified as a low-IC trough.

Testing for IC event-related changes in unexpectedness ratings

Subsequently, we tested whether the onset of the previously identified high-IC peaks or low-IC troughs led to a change in continuous unexpectedness ratings, employing the previously described LMM approach using only a subset of response data including the identified peak and

Author's personal copy

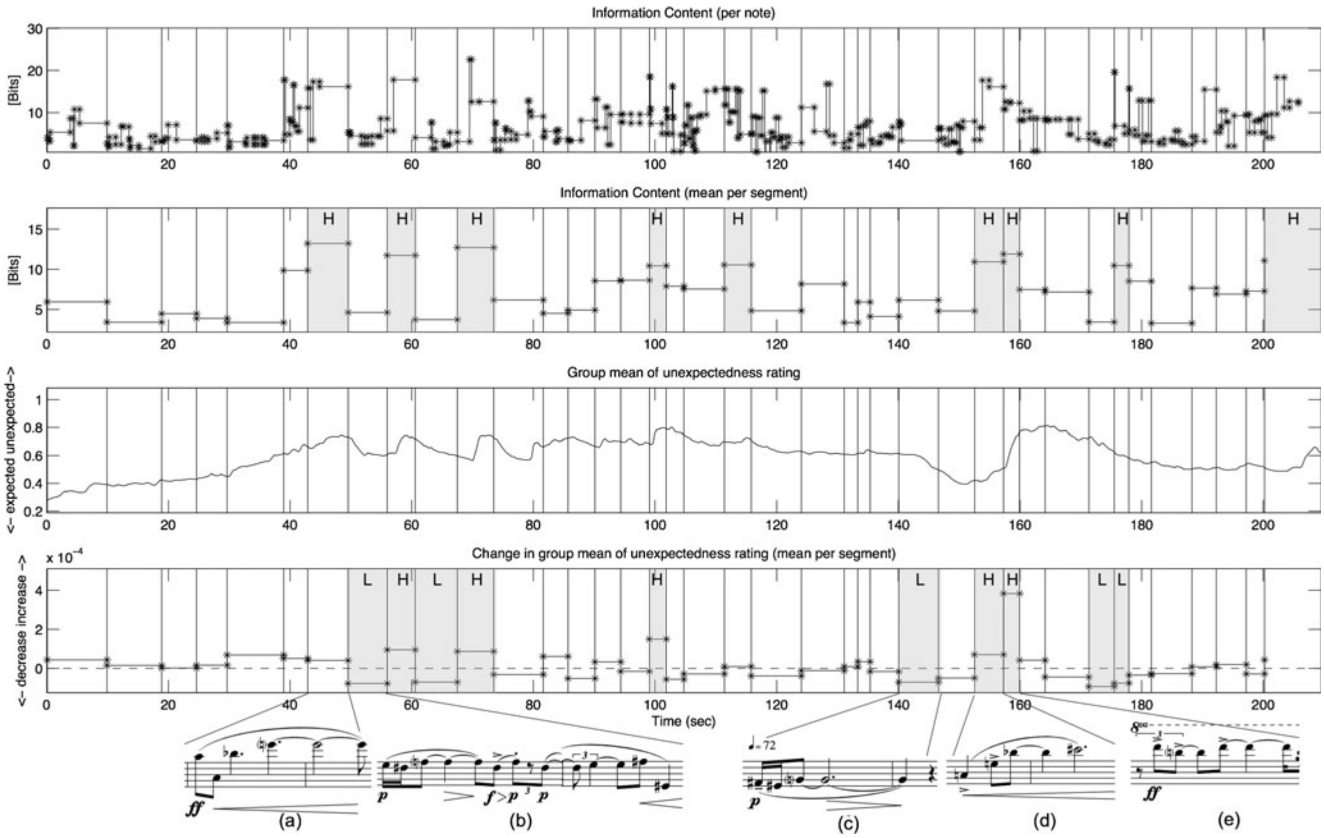


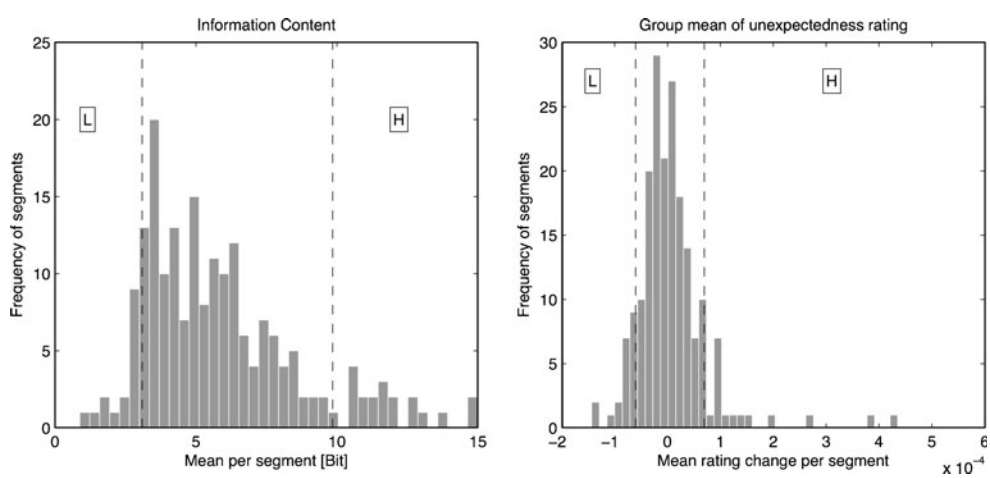
Fig. 1 Plot of information content (IC) and average unexpectedness ratings for *Density 21.5* by Edgar Varèse. Vertical lines represent segment boundaries. First row, IC per note; second row, mean IC per segment; third row, group mean of unexpectedness ratings (n = 25); fourth row, mean of change in group mean of unexpectedness rating per segment. H = high IC or unexpectedness rating peak segments, L = low IC or unexpectedness rating trough segments. (**a**), (**b**), (**c**), and (**d**) present musical score excerpts for the selected example peak/trough segments

trough segments. Therefore, we extracted all participants' individual mean unexpectedness rating for seven 1-s windows starting 1 s before the onset of the IC peak segments tested. This window size was chosen because previous research had shown that buttonpress response times to unexpected notes are about 2–3 s (Pearce, Ruiz, et al., 2010), so a 6-s-long postevent time window was assumed to be long enough to capture participants' event-related continuous rating changes. The estimated model followed Equation 2:

$$\text{response} = b_0 + b_1 \times \text{time} \times b_2 \times \text{event type} + b_3 \times \text{time} \times \text{event type} + \text{random effects.} \qquad (2)$$

Predictors were *time*, with values from 1 to 7, representing 1 s before the onset of the segments tested to 6 s after that

Fig. 2 Histograms of mean information content (IC) or average change in unexpectedness ratings per segment over the entire concert (n = 183). H = high-IC or unexpectedness rating peak segments (higher than 90 % of the frequency distribution), L = low-IC or expectedness rating trough segments (lower than 10 % of the frequency distribution). Segments between L and H were not included in further analyses



Author's personal copy

Table 3 Cross-tabulation of frequency of segments in unexpectedness rating event type separated by information content event type

	Mean information content			Total
	Low trough (<10th percentile)	High peak (>90th percentile)	Not used	
Very expected (<10th percentile)	3	2	13	18
Very unexpected (>90th percentile)	0	8	11	19
Not used	16	8	122	146
Total	19	18	146	183

segment onset, and *event type,* a dummy variable coding high-IC events (with 1) or low-IC events (with 0). The models furthermore included an intercept, an interaction term for both main effects, and several random effects, modeling the random correlation in the data set (random intercepts and slopes for each participant). Figure 3 presents group averaged unexpectedness ratings as a function of time, separated by the two predictor variables (n = 6,475). As can be seen, in addition to a main effect of event type (in general, ratings for high-IC segments were higher than ratings for low-IC segments), the onset of the high-IC event led to an increase in unexpectedness ratings, as compared with the low-IC event peaks. LMM fixed-effects coefficients were estimated as b_0 = .40 (intercept), b_1 = −.0013 (time), b_2 = .0096 (event type), and b_3 = .0093 (time × event type interaction). The significance of predictors was subsequently tested with F-tests, indicating that b_1 and b_3 were significantly different from zero [b_1, $F(1, [6328–6471]) = 8.70$, $p = .003$; b_2, $F(1, [6328–6471]) = 2.07$, $p = .15$; b_3, $F(1, [6328–6471]) = 19.87$, $p < .001$, where the numbers in square brackets represent the range of degrees of freedom due to the addition of random effects to the model]. Significant random effects were included as random intercepts for each participant, peak segment, and piece, as well as random slopes for all fixed effects within participants (based on chi-squared likelihood-ratio tests). Since we were interested only in event-related changes in these analyses, we will not interpret any main effect of event type, because these might be due to the context of the peak segments investigated, and not to responses caused by their onset. In summary, therefore, as was expected, the onset of any IC peak was modeled as leading to a slight decrease in unexpectedness ratings (as reflected in the negative b_1 coefficient and possibly the onset of low-IC troughs), and for high-IC peaks (>90th percentile), there was a significant increase (due to the significant interaction term b_3 between time and event type).

Testing emotional effects of high-IC peak versus low-IC trough segments

Following this LMM specification, we subsequently ran several similar analyses, testing for significant change in affective psychophysiological response measures after the onset of those previously identified IC peak moments. Response variables were all continuously recorded measures of emotion: arousal ratings, valence ratings, SCR, HR, RespR, and EMG from the corrugator and zygomaticus muscles. Predictors were again time (seven levels from 1 s before to 6 s after onset) and event type (a dummy variable: high-IC peak = 1, low-IC trough = 0). The models furthermore included an intercept, an interaction term for both main effects, and several random effects, modeling the random correlation in the data set (random intercepts for participants, peak segments, and pieces, plus random slopes for all predictors within participants).

Table 4 Cross-tabulation of frequency of segments separated by IC or unexpectedness rating event types and music pieces

Piece (presen-tation no.)	Frequency of segments: Analyses based on IC			Frequency of segments: Analyses based on unexpectedness ratings		
	Low trough (<10th percentile)	High peak (>90th percentile)	Not used	Very expected (<10th percentile)	Very unexpected (>90th percentile)	Not used
1.	0	2	14	3	2	11
2.	2	2	8	1	1	10
3.	0	9	29	5	5	28
4.	4	3	35	5	7	30
5.	5	0	41	1	1	44
6.	8	2	19	3	3	23
Total	19	18	146	18	19	146

Note. Total n = 183.

Author's personal copy

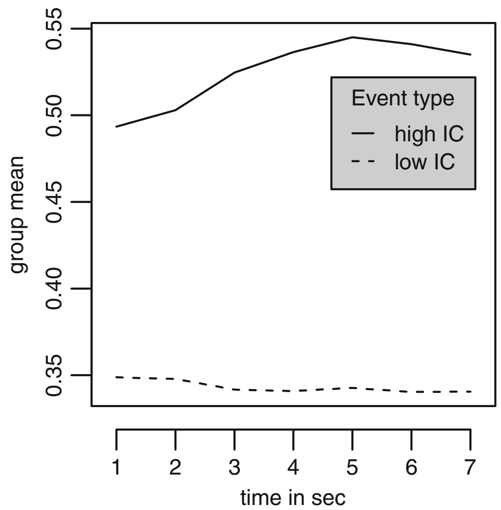


Fig. 3 Plot of mean unexpectedness ratings as a function of time, separated by event type (high information content (IC) segments, n = 18, vs. low IC segments, n = 19). Segment onset is between seconds 1 and 2

As can be seen in Fig. 4 (upper row), there was an event-related change in subjective feelings after the onset of those IC peak moments. Arousal ratings significantly increased and valence ratings significantly decreased for high-IC peak segments, as compared with low-IC trough segments (indicated by corresponding significant interaction terms between time and event type in the LLM results; Table 5, upper row). Although there were event-related changes in recordings of expressive facial movements (Fig. 4, lower row), LLM estimates show that for both EMG measures (corrugator and zygomaticus activity), no predictors were estimated as significant (Table 5, lower row).

Measures of autonomous nervous system (ANS) activity did show event-related changes corresponding to the onset of high- and low-IC segments (Fig. 5). In contrast to low-IC events, high-IC events were accompanied by increases in SCR and a decrease in HR. For SCR, there was only a significant interaction term (Table 6, upper row). For HR,

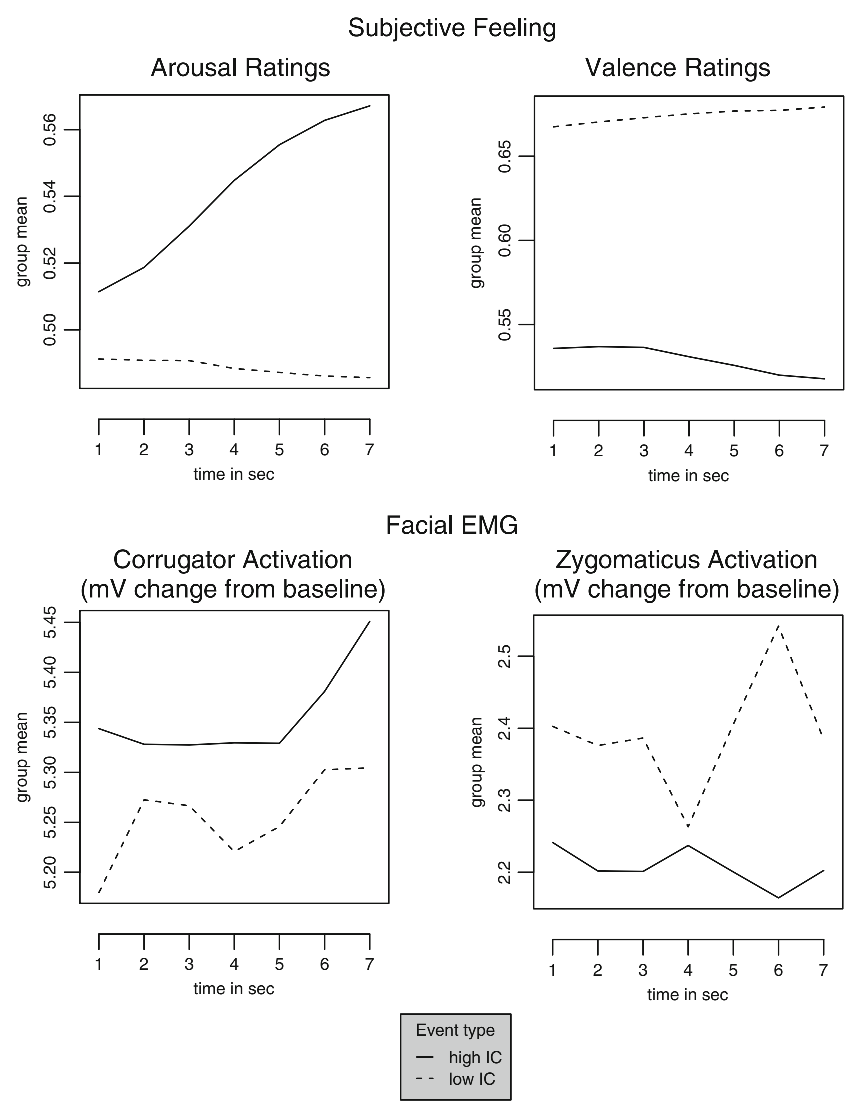


Fig. 4 Plot of subjective emotion ratings and facial EMG as a function of time, separated by event type (high information content (IC) segments, n = 18, vs. low-IC segments, n = 19). Segment onset is between seconds 1 and 2

Author's personal copy

Table 5 Linear mixed effects modeling (LMM) coefficient estimates for event-related change in subjective feeling and facial EMG predicted by IC event type (high- vs. low-IC segments), time (seconds 1–7), and their interaction

	Fixed-effects coefficients	F	Fixed-effects coefficients	F
	Subjective feeling			
	Arousal ratings		Valence ratings	
b_0 (intercept)	.49	–	.61	–
b_1 (time[1])	−.001**	8.57	.002	0.13
b_2 (event type[2])	.01	1.49	.01	0.05
b_3 (time × event type[3])	.01***	16.27	−.005*	4.88
	Facial EMG			
	Corrugator activity		Zygomaticus activity	
b_0 (intercept)	5.10	–	2.36	–
b_1 (time[1])	.02	1.44	.01	0.01
b_2 (event type[2])	−.19	0.6	−.06	0.57
b_3 (time × event type[3])	.001	0	−.02	0.39

Note. [1] Seconds 1–7. [2] Dummy variable: 1 = *high-IC peak*, 0 = *low-IC trough*. [3] Interaction term. Results of F-test (subjective feeling, df1 = 1, df2 = 6,003–6,142; EMG, df1 = 1, df2 = 12,193–12,428): $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$. The following random effects were included: (1) random intercepts for participants, pieces, and segments, (2) random slopes for time, event type, and time × event type (all within participants).

both event types showed a significant event-related decrease (Table 6). However, if only seconds 1 to 4 were evaluated, a significant negative interaction coefficient was found, indicating a difference in responding to high-IC, as compared with

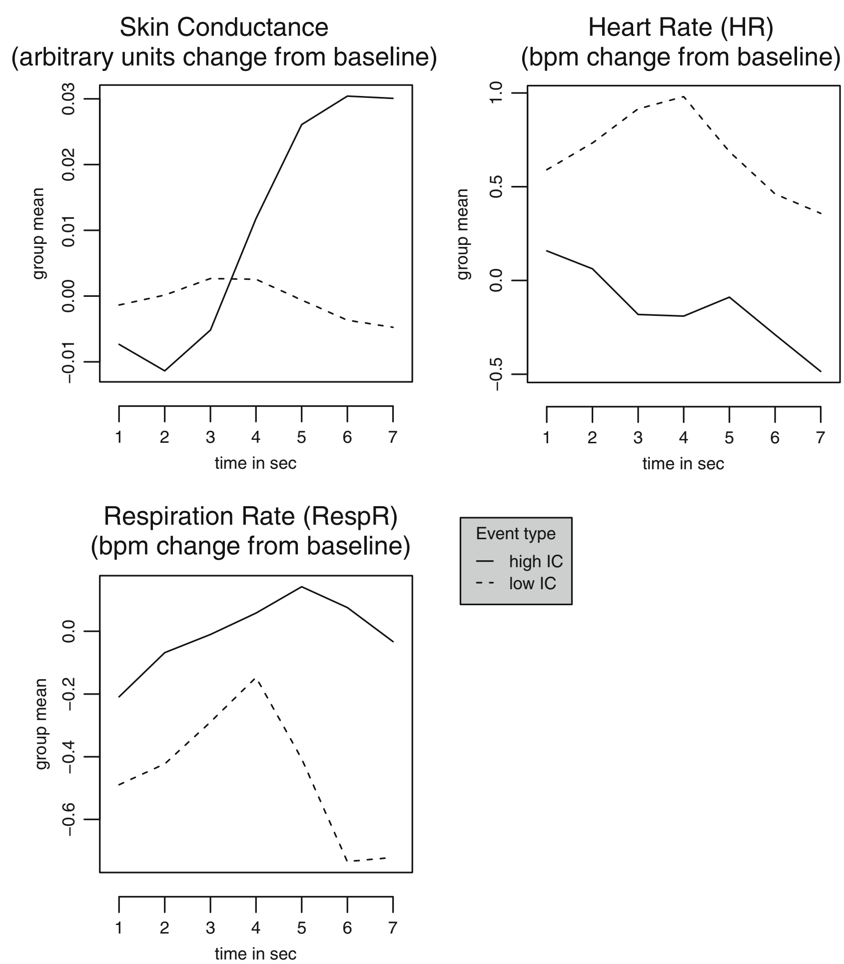


Fig. 5 Plot of mean skin conductance response, heart rate, and respiration rate measurements as a function of time, separated by event type (high information content (IC) segments, n = 18, vs. low-IC segments, n = 19). Segment onset is between seconds 1 and 2. BPM, beats per minute

Author's personal copy

Table 6 Linear mixed effects modeling (LMM) coefficient estimates for event-related change in ANS and respiration rate measures predicted by IC event type (high- vs. low-IC trough segments), time (second 1–7), and their interaction

	Fixed-effects coefficients	F	Fixed-effects coefficients	F
	Skin conductance		Heart rate (time = 1–7)	
b_0 (intercept)	.002	–	.875	–
b_1 (time[1])	−.001	0.09	−.052***	11.35
b_2 (event type[2])	−.024	0.57	−.547*	4.14
b_3 (time × event type[3])	.009***	9.38	−.038	0.94
	Respiration rate		Heart rate (time = 1–4)	
b_0 (intercept)	−.197	–	.533	–
b_1 (time[1])	−.051	0	.135	0.03
b_2 (event type[2])	.485	2.62	−.189**	6.21
b_3 (time×event type[3])	.086	1.72	−.264**	8.42

Note. [1] Second 1–7. [2] Dummy variable: 1 = *high-IC peak*, 0 = *low-IC trough*. [3] Interaction term. Results of F-test (time = 1–7, df1 = 1, df2 = 12,193–12,428; time = 1–4, df1 = 1, df2 = 6,865–7,100): *$p < .05$,**$p < .01$,***$p < .001$. The following random effects were included: (1) random intercepts for participants, pieces, and segments, (2) random slopes for time, event type, and time×event type (all within participants).

low-IC, events (Table 6, lower row). The initial decrease associated with high-IC peaks was stronger. Although RespR appeared to increase after both event types (Fig. 5), the corresponding predictors were not significant in LMM (Table 6).

Testing emotional effects of unexpected peak versus expected trough segments

We subsequently ran several LMM analyses, testing for significant change in emotional response measures after the onset of the previously identified peak moments in listeners' continuous unexpectedness ratings. The response variables were again all continuously recorded psychophysiological measures presumed to be related to affective response. Predictors were again time, event type, and their interaction. Event type was a dummy variable coding very unexpected events (1) and very expected events (0). The models furthermore included intercepts and the same random effects used above, modeling the random correlation in the data set.

As can be seen in Fig. 6 (upper row), there was an event-related increase in arousal ratings after the onset of those unexpected peak segments and a decrease in arousal after very expected events. This observation was also supported by a corresponding significant negative effect of time and a significant positive interaction term between time and unexpected events (Table 7, upper row). No event-related changes in valence and EMG activity were significant here (Table 7, lower row).

Similar to peak events identified with IC analyses, markers of ANS activity also showed a significant event-related change corresponding to the onset of very unexpected and very expected peak moments. For both types of segments, an increase of skin conductance is indicated in Fig. 7 (upper row). In the LMM, a main effect of time was significant, but not the interaction with event type (Table 8, upper row). HR also decreased after the onset of very unexpected segments (Fig. 7, upper row), and the corresponding interaction between time and event type was significant (Table 8, upper row). Finally, for RespR, a significant decrease was observed for expected events (indicted by a significant negative main effect of time), whereas for unexpected events, RespR increased (Fig. 5, lower row). This difference between unexpected and expected events was illustrated by a significant positive interaction term between time and event type (Table 8, lower row).

Discussion

The experiment tested three main hypotheses. First, it was proposed that information-theoretic analyses could predict whether participants perceived particular segments of the music presented as expected or unexpected. Second, it was predicted that, as compared with low-IC segments, high-IC peak segments would be associated with event-related changes in measurement of the three emotion components subjective feeling, peripheral arousal, and expressive behavior. Third, it was hypothesized that similar activations would be associated with segments that were identified as subjectively very unexpected, as compared with those identified as very expected. All three hypotheses were partially corroborated.

Focusing on segments with extreme IC values, IC was associated with ratings of subjective expectation on two levels of analysis. First, a cross-tabulation of segments corresponding with peaks in IC and unexpectedness rating showed a significant association between them. Second, continuous unexpectedness ratings significantly increased after the onset of high-IC peaks and decreased after the onset of low-IC troughs. Therefore, these findings confirm the validity of the cognitive

Author's personal copy

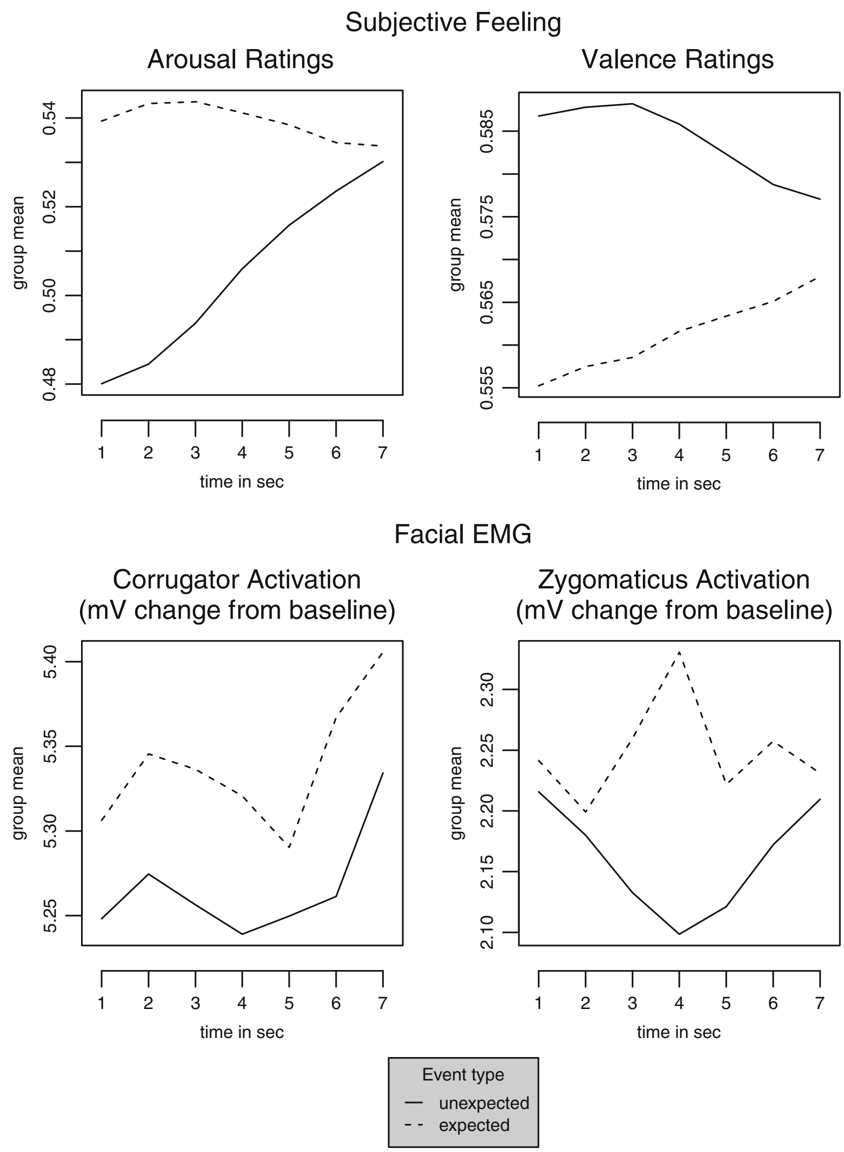


Fig. 6 Plot of subjective feeling ratings and facial EMG measurements as a function of time, separated by event type (very unexpected segments, n = 19, vs. very expected segments, n = 18). Segment onset is between seconds 1 and 2

model used to predict listeners' expectations, replicating previous studies (Omigie et al., 2012; Pearce, Ruiz, et al., 2010; Pearce & Wiggins, 2006) and confirming assumptions about the role of statistical learning in creating expectations.

The second hypothesis was partially supported, in that high-IC, as opposed to low-IC, segments were associated with changes in two components of emotion. First, arousal increased and valence decreased for high-IC segments. Second, high-IC segments were also associated with changes in peripheral arousal as indicated by increases in SCR and decreases in HR. No event-related changes were found in RespR or EMG measures of expressive activity in the corrugator and zygomaticus muscles. Considering the event-related psychophysiological responses together, their response patterns may be understood in terms of the "defense response cascade" described by Lang, Bradley, and Cuthbert (1997). This reaction pattern is usually induced when highly arousing aversive stimuli are encountered. A general increase in arousal (also indicated by an increase in skin conductance) is accompanied by an initial fast decrease in HR that represents a freezing response, functioning to allocate attention. This orienting period is then followed by an increase in HR that indicates increased sympathetic dominance preparing for adaptive fight or flight behavior (*circa-strike period*). Thus, the observed emotional effects of unexpected melodic events may be interpreted as being mediated by this affective response mechanism.

The third hypothesis, concerning the induction of emotional reactions by very unexpected segments, was also partially corroborated. Here, due to the use of continuous unexpectedness rating changes as an identifier of peaks, we were able to compare effects of very unexpected to very expected moments. Unexpected events induced a psychophysiological reaction pattern that was very similar to those of high-IC peaks, with

Author's personal copy

Table 7 Linear mixed effects modeling (LMM) coefficient estimates for event-related change in subjective feeling and facial EMG predicted by event type (unexpected vs. expected segment), time (second 1–7), and their interaction

	Fixed-effects coefficients	F	Fixed-effects coefficients	F
	Subjective feeling			
	Arousal ratings		Valence ratings	
b_0 (intercept)	.55	–	.59	
b_1 (time[1])	−.0001***	7.05	.003	0.21
b_2 (event type[2])	−.07	2.87	.03	0.32
b_3 (time × event type[3])	.01***	12.37	−.004	3.38
	Facial EMG			
	Corrugator activity		Zygomaticus activity	
b_0 (intercept)	5.11	–	2.33	–
b_1 (time[1])	.01	0.38	.002	0.02
b_2 (event type[2])	−.14	2.37	−.04	0.27
b_3 (time × event type[3])	−.002	0.02	−.003	0.02

Note. [1] Second 1–7. [2] Dummy variable: 1 = *very unexpected*, 0 = *very expected*. [3] Interaction term. Results of F-Test (subjective feeling, df1 = 1, df2 = 5,996–6,135; EMG, df1 = 1, df2 = 12,193–12,428): $*p < .05$. $**p < .01$. $***p < .001$. The following random effects were included: (1) random intercepts for participants, pieces, and segments; (2) random slopes for time, event type, and time × event type (all within participants).

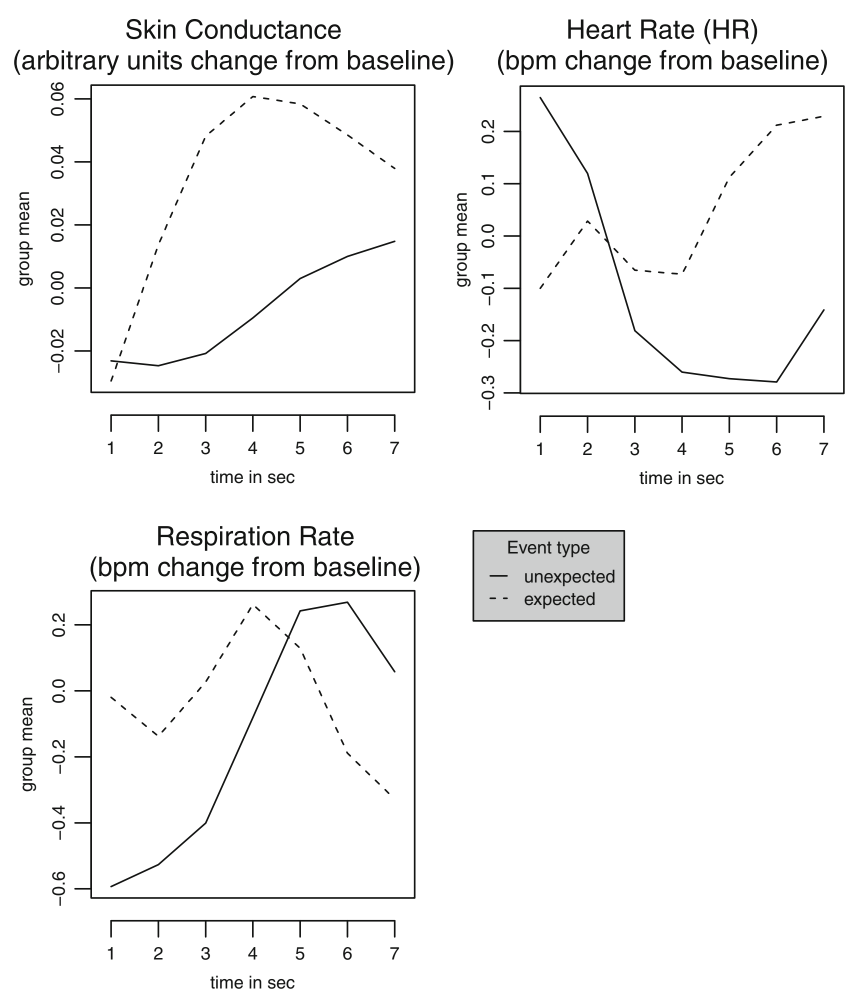


Fig. 7 Plot of mean skin conductance response, heart rate, and respiration rate measurements as a function of time, separated by event type (very unexpected segments, n = 19, vs. very expected segments, n = 18). Segment onset is between seconds 1 and 2. BPM, beats per minute

Author's personal copy

Table 8 Linear mixed effects modeling (LMM) coefficient estimates for event related change in ANS and respiration rate measures predicted by event type (unexpected vs. expected segments), time (second 1–7), and their interaction

	Fixed-effects coefficients	F	Fixed-effects coefficients	F
	Skin conductance		Heart rate	
b_0 (intercept)	−.01	–	−.03	–
b_1 (time¹)	.01***	11.61	.05	0.64
b_2 (event type²)	−.03*	5.59	.40	0.41
b_3 (time × event type³)	−.003	0.38	−.13**	10.57
	Respiration rate			
b_0 (intercept)	.15	–		
b_1 (time¹)	−.03*	5.33		
b_2 (event type²)	−.61	0.91		
b_3 (time × event type³)	.18*	5.27		

Note. ¹ Second 1–7. ² Dummy variable: 1 = *very unexpected*, 0 = *very expected*. ³ Interaction term. Results of F-test (df1 = 1, df2 = 12,193–12,428): *p < .05. **p < .01. ***p < .001. The following random effects were included: (1) random intercepts for participants, pieces, and segments; (2) random slopes for time, event type, and time × event type (all within participants).

increased arousal and SCR and decreased HR. However, different from high-IC peaks, there was no associated effect on valence ratings, and RespR also significantly increased after the onset of unexpected peak moments. Furthermore, even for very expected segments, SCR significantly increased after the event onset. For both event types, very unexpected and very expected measures of EMG showed no event-related responses. Thus, all analyses in this study failed to show any IC- or expectation-related responses in EMG recordings. However, other research has elicited affective event-related facial EMG responses to computer gaming (Ravaja, Turpeinen, Saari, Puttonen, & Keltikangas-Järvinen, 2008) and to pictures or sounds (Bradley, Moulder, & Lang, 2005). But to our knowledge, for music, only tonic effects on EMG measurements have been shown to date by testing activation measures of entire stimulus sequences (Lundqvist et al., 2008).

In summary, these results corroborate parts of the theory proposed by Huron (2006). Statistical properties of melodic events and moments of strong expectation or surprise created prediction responses that were correlated with several emotional response components. By identifying expectation-related events in two ways (with the help of statistical analyses and subjective unexpectedness ratings), we were able to show that, independently of this identification mode, similar psychophysiological responses were observed for segments that were subjectively unexpected and segments that were unpredictable according to the computational model of auditory expectation (Pearce, 2005). Thus, modeled expectations based on short- and long-term memory generate responses similar to those that are also consciously represented in participants' experiences.

These findings extend those of Steinbeis et al. (2006), Koelsch, Kilches, et al. (2008), and Koelsch, Fritz, and Schlaug (2008) to melodic stimuli heard in an ecologically valid concert setting with quantitative measures of expectedness supplied by a cognitive model (Pearce, 2005). Violations and confirmations of musically induced expectations were associated with affective psychophysiological activations in several response components. Like Steinbeis et al., we found general increases of physiological arousal for very unexpected moments, and at the same time, Koelsch, Fritz, et al. (2008) interpretation of their fMRI data, in which unexpected moments induce unpleasant feelings, was corroborated, because high-IC segments induced a decrease in valence ratings. (However, the effects of unexpected events, identified by subjective unexpectedness ratings, on valence or pleasantness were less clear, and no event-related EMG activations were found for IC or unexpectedness peaks.) The negatively valenced effects of high-IC events may also depend on the stimuli used and the population investigated. Half of the high-IC events identified here were in the piece *Density 21.5*, and thus the valence effects associated with them were more strongly associated with this piece than with the others. Different participants listening to different stimuli might interpret expectancy violations differently. As Huron (2006) notes, violations of expectations that are originally negatively valenced may be evaluated positively by subsequent appraisal responses potentially based on individual evaluation criteria, thus leading to contrastive valence.

To summarize, the findings of this study extend previous research on physiological responses to musical expectations in four ways.

First, in contrast with previous research focusing on harmonic expectations in Western tonal music, we focused on expectations in melody, which is arguably a more universal aspect of musical structure.

Second, we quantified predictability in our stimuli using a computational model of auditory expectation making probabilistic predictions based on statistical learning of music structure (Pearce, 2005). Our analysis of the effects of

Author's personal copy

expectation on psychophysiological measures of emotional response using this model were compared with an analysis using subjective measures of expectation. There was significant overlap between these two approaches because some events corresponded to expectation peaks in both analyses, and there was also a significant increase in unexpectedness ratings after the onset of high-IC events. However, the measures also allowed us to study expectation in two different ways. We were able to test, first, for the effects of the implicit statistical structure of the music presented and, second, for those expectation-related musical structures that were not captured by the computational modeling approach.

Third, our approach allowed us to test for emotional effects of very expected moments. Huron's (2006) theory of emotion and expectation describes two outcomes of the prediction response: one negative, penalizing incorrect predictions, and one positive, rewarding correct predictions. Here, for the first time, we tested whether emotional responses were induced when participants were able to make correct predictions. The results indicated divergent changes in physiological and subjective components of arousal (increase in SCR, decrease in RespR, and decrease in subjective arousal).

Finally, our experiment was conducted in a concert setting using live performance of actual compositions. To the best of our knowledge, this study is among the first to explicitly test for music-induced emotions with psychophysiological measures in a live concert situation. In doing so, it is unique in two ways. For the first time, the emotional effects of music have been investigated in an experimental field setting, employing measures of several emotion components in parallel: Subjective feelings, expressive behavior, and peripheral arousal were continuously monitored throughout the concert. Additionally, the study employed performances of actual compositions from the repertoire of Western art music spanning several styles and historical periods. This allows us to generalize results from laboratory environments often using artificial stimuli that are pale reflections of real performed music. Conducting research in natural listening contexts may be important, since previous research has shown that emotional responses to music are sensitive to the presence of other people (Egermann et al., 2011; Liljeström et al., 2012). As far as we know, there is only one previous exploratory study taking psychophysiological measures from a very limited number of participants (3) to investigate emotional effects of Leitmotifs from a Wagner opera performed in Bayreuth, Germany (Vaitl, Vehrs, & Sternnagel, 1993).

Limitations and future research suggestions

The naturalistic and ecologically valid methodological approach employed in this investigation also entailed several potential weaknesses that should be considered. First, in conducting a study within a concert context, we lacked some control over the participants' actual behavior. Although instructed not to do so, participants might have been distracted or otherwise influenced by interaction with peers during the study. We took measures against this possibility by asking them to rate their own degree of focus and monitored them visually throughout the concert; no participants had to be excluded on the basis of these criteria. Another potential weakness is that because we used actual complex compositions, the selected segments varying in degrees of expectation and IC might actually differ in other respects related to expressive performance (including performer's gestures). However, we tried to eliminate these differences by sampling participants' responses over several segments of each expectation/IC condition.

Future research might explore complementary ways of ensuring that the emotional responses are related specifically to expectations based on the musical structure, and not any other underlying covarying performance feature. This could be done through more controlled laboratory research where the music is recomposed to systematically vary the degree of expectedness of the music using IC as a quantitative indicator of expectations. One could also remove any features associated with the music performance (tempo, dynamics, or timbre) and test whether findings comparing the different event types are replicable. Computational modeling could also be improved, since only a limited number of viewpoints representing the music were employed in this study. In future research, it would be interesting to investigate the effects of derived features such as scale-degree or timing information on the quantitative modeling of expectation-related emotional responses.

In this study, analyses were based on a segmentation provided by two music theorists. In future research, segmentation might be also based on the IC itself, since phrase boundaries have been shown to be perceived before notes with very high IC (Pearce, Müllensiefen, & Wiggins, 2010). Furthermore, average IC across a segment might not be representative for all notes within one segment. Individual unexpected notes might also be effective in inducing emotional responses, as was confirmed by preliminary analysis of this data set (not presented here).

We also tested for different lengths of response windows from segment onset and decided that 6 s after segment onset provided optimal results, since segments were, on average, 3.7 s (SD = 2.5 s) long and previous research indicated that subjective and physiological response measures have time lags between 2 and 3 s (Pearce, Ruiz, et al., 2010; Ravaja et al., 2008). However, if no significant effects were reported for 7-s-long response windows, we also tested with 4-s-long windows. If the results did change, they were reported (e.g., HR analyses).

Author's personal copy

Finally, future analyses could also test for individual differences in participants' emotional responses to violations of expectations, which were beyond the scope of this study due to space limitations. Emotional reactions to music have been shown to have a high interindividual variance (Grewe, Nagel, Altenmüller, & Kopiez, 2009–2010) that may be explained by interindividual differences in music-related syntactical knowledge creating different expectations in different listeners.

Conclusions

On the basis of statistical modeling and on subjective measures of expectation, this study showed that violations of structural expectations in live music performance induced emotional reactions in listeners, with associated activations in two different response components: subjective feelings and peripheral arousal. This study extends previous research to a greater range of psychophyisological responses to melodic expectations induced in a live concert experiment. There was also limited support for two additional findings. Unexpected musical moments induced unpleasant feelings (only in IC-based analyses), and highly expected segments also produced physiological responses (changes in SCR and RespR).

These results contribute evidence to discussions concerning the ability of music to induce fully component-synchronized emotions by itself (Cochrane, 2010). Here, musical structures and their performances induced varying degrees of predictability that, in turn, had effects on several levels of emotion measurement, supporting the emotivist position described above. Finally, we note that we understand expectation as being only one of many mechanisms that might be involved in creating emotional responses to music (Juslin & Västfjäll, 2008). Therefore, we have focused on analyzing events in the music that were relevant to this mechanism, corroborating the often-predicted link between expectation and emotion in music (Meyer, 1956, 1957). Our results advance research on music-induced emotion by taking it beyond exploratory studies, like those cited in our introduction, to the next scientific level of formulating and testing theoretical mechanistic models that generate falsifiable hypotheses.

Author Notes H.E. and S.M.'s work was partially funded by the Canadian Social Sciences and Humanities Research Council through a grant to S.M. (#410-2009-2201), as well as S.M.'s Canada Research Chair. The CIRMMT Audience Response System was funded by a grant from the Canada Foundation for Innovation. M.T.P. and G.A.W.'s contribution was funded by EPSRC research grant EP/H01294X, "Information and neural dynamics in the perception of musical structure". We would like to thank all participants, members of the Music Perception and Cognition Laboratory, and the technical team of the Centre for Interdisciplinary Research in Music Media and Technology for being very supportive in carrying out this study.

Appendix

Continuous Emotion Ratings | **Continuous Unexpectedness Ratings**

Rating interfaces

Rating instructions

… By moving your finger from left to right you can indicate how pleasant the music is to you (left = negative and unpleasant; right = positive and pleasant). By moving your finger from top to bottom you can indicate your degree of emotional arousal during listening to the music (top = excited; bottom = calm). You should try to rate what your current emotional state is along both dimensions simultaneously. The position of your finger should reflect at each moment your emotional response to the piece as you are listening. …

… By moving your finger from top to bottom you can indicate how unexpected the music events you are hearing are (top= very unexpected; bottom = very expected). The position of your finger should reflect at each moment the unexpectedness of the events as you are listening. You need to constantly monitor your expectations for every musical event in order to keep your finger at the corresponding position. ….

References

Abdallah, S. A., & Plumbley, M. D. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science, 21*(2), 89–117.

Author's personal copy

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes, R package version 0.999375-39 [Computer Software].

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology. Human Perception and Performance, 12*(4), 403–410.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics, 59*(7), 1098–1107.

Boucsein, W. (2001). Physiologische Grundlagen und Meßmethoden der dermalen Aktivität [Physiological Bases and Measurement Methods for Electrodermal Activity]. In F. Rösler (Ed.), *Enzyklopädie der Psychologie, Bereich Psychophysiologie: Vol. 1. Grundlagen und Methoden der Psychophysiologie [Encyclopedia of psychology, area psychophysiology: Vol. 1. Basics and methods of psychophysiology]* (pp. 551–623). Hogrefe: Göttingen.

Bradley, M. M., Moulder, M., & Lang, P. J. (2005). When good things go bad: The reflex physiology of defense. *Psychological Science, 16,* 468–473.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*(2), 260–268.

Cannam, Landone, & Sandler, (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files [Computer Software]. *MM'10,* October 25–29, 2010, Firenze, Italy.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology, 1,* 12–29.

Carrión, R. E., & Bly, B. M. (2008). The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology, 45*(5), 759–775.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology. General, 113*(3), 394–412.

Cochrane, T. (2010). Music, emotions and the influence of the cognitive sciences. *Philosophy Compass, 11,* 978–988.

Colombetti, G. (2005). Appraising valence. *Journal of Consciousness Studies, 12*(8), 103–126.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research, 24,* 51–73.

Cuddy, L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Attention, Perception, & Psychophysics, 57*(6), 451–462.

Eerola, T. (2004). Data-driven influences on melodic expectancy: Continuations in North Sami Yoiks rated by South African traditional healers. In S. D. Libscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, 2004* (pp. 83–87). Adelaide, Australia: Causal Productions.

Egermann, H., Grewe, O., Kopiez, R., & Altenmüller, E. (2009). Social feedback influences musically induced emotions. *The Neurosciences and Music III: Disorders and plasticity: Annals of the New York Academy of Sciences, 1169,* 346–350.

Egermann, H., Nagel, F., Altenmüller, E., & Kopiez, R. (2009). Continuous measurement of musically-induced emotion: A web experiment. *International Journal of Internet Science, 4*(1), 4–20.

Egermann, H., Sutherland, M. E., Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2011). Does music listening in a social context alter experience? A physiological and psychological perspective on emotion. *Musicae Scientiae, 15*(3), 307–323.

Grewe, O., Kopiez, R., & Altenmueller, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception, 27*(1), 61–74.

Grewe, O., Nagel, F., Altenmüller, E., & Kopiez, R. (2009–2010). Individual emotional reactions towards music: Evolutionary-based universals? *Musicae Scientiae, Special Issue*, 261–287.

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*(4), 774–788.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge: MIT Press.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience, 7*(2), 153.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *The Behavioral and Brain Sciences, 31*(5), 559–575. Discussion 575–621.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception, 2*(2), 131–165.

Kivy, P. (1990). *Music alone: Philosophical reflections on the purely musical experience*. Ithaca, NY: Cornell University Press.

Koelsch, S., Fritz, T., & Schlaug, G. (2008). Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroreport, 19*(18), 1815.

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS One, 3*(7), e2631.

Konecni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts, 2*(2), 115–129.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*(3), 401–432.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11*(2), 45–50.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual hymns: Convergence of statistical, behavioral and computational approaches. *Music Perception, 17,* 151–195.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jarvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition, 76*(1), 13–58.

Ladinig, O., Honing, H., Háden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception, 26*(4), 377–386.

Lang, P. J., Bradley, M. M., & Cuthbert, M. M. (1997). Motivated attention: Affect, activation and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97–136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Larson, S. (2004). Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception, 21*(4), 457–498.

Liljeström, S., Juslin, P. N., & Vastfjall, D. (2012). Experimental evidence of the roles of music choice, social context, and listener personality in emotional reactions to music. *Psychology of Music*. doi:10.1177/0305735612440615

Lundqvist, L.-O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2008). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music, 37*(1), 61–90.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the entropy of music: An experiment with Bach chorale melodies. *Leonardo, 2,* 81–88.

Author's personal copy

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception, 22*(4), 663–714.

Margulis, E. H., & Levine, W. (2006). Timbre priming effects and expectation in melody. *Journal of New Music Research, 35*(2), 175–182.

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22*(2), 297–350.

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

Meyer, L. B. (1957). Meaning in music and information theory. *Journal of Aesthetics and Art Criticism, 15*(4), 412–424.

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods, 39*(2), 283–290.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity.* Chicago: University of Chicago Press.

Ockelford, A. (2006). Implication and expectation in music: A zygonic model. *Psychology of Music, 34*(1), 81–142.

Omigie, D., Pearce, M. T., & Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia, 50,* 1483–1493.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch- distributional information in melodic sequences. *Psychological Research, 57*(2), 103–118.

Pearce, M.T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis, London, UK: Department of Computing, City University.

Pearce, M. T., Conklin, D., & Wiggins, G. A. (2005). Methods for combining statistical models of music. In U. K. Wiil (Ed.), *Computer music modelling and retrieval* (pp. 295–312). Berlin: Springer.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*(10), 1365–1389.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50*(1), 302–313.

Pearce, M. T., & Wiggins, G. A. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research, 33*(4), 367–385.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23*(5), 377–405.

Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., & Keltikangas-Järvinen, L. (2008). The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion, 8*(1), 114–120.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience, 14,* 257–262.

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PloS One, 4*(10), e7487.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58*(1), 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception, 14,* 295–318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology. General, 131*(4), 511–537.

Scherer, K. (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*(3), 239–251.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695–729.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). Oxford: Oxford University Press.

Schmuckler, M. A., & Boltz, M. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception & Psychophysics, 56*(3), 313–325.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two dimensional emotion space. *Australian Journal of Psychology, 51,* 154–165.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19*(2), 110–120.

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*(8), 1380–1393.

Stevens, C. J., Schubert, E., Morris, R. H., Frear, M., Chen, J., Healey, S., et al. (2009). Cognition and the temporal arts: Investigating audience response to dance using PDAs that record continuous data during live performance. *International Journal of Human Computer Studies, 67*(9), 800–813.

Thompson, W. F., & Stainton, M. (1998). Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Perception, 15,* 231–252.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*(4), 885–913.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception, 16*(1), 99–117.

Tremblay, A. (2011). *A suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions.* R package version 1.6 [Computer Software].

Vaitl, D., Vehrs, W., & Sternagel, S. (1993). Prompts—leitmotif—emotion: Play it again, Richard Wagner. In N. Birnbaumer & A. Öhman (Eds.), *The structure of emotion: Psychophysiological, cognitive, and clinical aspects* (pp. 169–189). Göttingen: Hogrefe & Huber.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software*. Boca Raton: Chapman & Hall/CRC Press.

Zanto, T. P., Snyder, J. S., & Large, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology, 2*(2), 221–231.