Emotion
2012, Vol. 12, No. 3, 552–567

1528-3542/11/$12.00 DOI: 10.1037/a0023747

Emotional Response to Musical Repetition

Steven R. Livingstone and Caroline Palmer
McGill University

Emery Schubert
University of New South Wales

Two experiments examined the effects of repetition on listeners' emotional response to music. Listeners heard recordings of orchestral music that contained a large section repeated twice. The music had a symmetric phrase structure (same-length phrases) in Experiment 1 and an asymmetric phrase structure (different-length phrases) in Experiment 2, hypothesized to alter the predictability of sensitivity to musical repetition. Continuous measures of arousal and valence were compared across music that contained identical repetition, variation (related), or contrasting (unrelated) structure. Listeners' emotional arousal ratings differed most for contrasting music, moderately for variations, and least for repeating musical segments. A computational model for the detection of repeated musical segments was applied to the listeners' emotional responses. The model detected the locations of phrase boundaries from the emotional responses better than from performed tempo or physical intensity in both experiments. These findings indicate the importance of repetition in listeners' emotional response to music and in the perceptual segmentation of musical structure.

Keywords: emotional arousal, repetition, music, phrase structure, computational model

Musical repetition and variation (musical ideas that are repeated in an altered form) are fundamental elements of music; however, their effects on listeners' emotional response have not received much attention. Repetition is prevalent at multiple levels in musical forms of many cultures, from the basic units of discrete pitches and durations to the hierarchical structures of rhythm and phrase structure. Many societies have songs that consist of "short phrases repeated several or many times, with minor variations" (Nettl, 2000, p. 469). The concept of musical repetition has been formalized by several theorists as enhancing listeners' perception of structural units (Meyer, 1973; Schenker, 1906/1954). Thus, repetition in music may aid perceptual segmentation by demarcating musical units. The ubiquity of repetition for both small and large musical sections motivates us to examine its effect on listeners' emotional response to music.

Some evidence suggests that musical repetition also affects listeners' emotional responses to music. Krumhansl (1996) examined the correspondence between listeners' continuous ratings of musical tension and the hierarchical phrase structure of the music, with larger changes in tension found at points of more significant phrase boundaries. Importantly, listeners' responses to musical sections that were repeated in the work were correlated; their responses to musical variations or sections containing less repetition, however, were not contrasted. The high correspondence in participants' responses to musical repetition may arise from how listeners react to familiar musical features, which then influence their perceived emotional response to a musical work.

Emotional response to music has been differentiated in terms of "perceived" and "felt" emotion (Evans & Schubert, 2008; Gabrielsson, 2002; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009). A perceived emotional response is an evaluation of the emotion expressed by the acoustic signal and is determined by the underlying musical features (Gabrielsson & Lindström, 2001; Juslin, 2001; Livingstone, Muhlberger, Brown, & Thompson, 2010). This evaluation is not expected to change across repeated presentations when the signal remains unchanged, and the repeatability of perceived emotion is expected to be independent of the musical work. In contrast, a listener's felt emotional response is thought to arise from a more complex relationship involving musical preferences and physiological factors, and is expected to show greater variability to repeated musical phrases that may vary depending on the musical work (Thaut & Davis, 1993). We test the role of musical repetition in perceived emotional responses, which are expected to be more consistent for music (Schubert, 2007).

Listeners' emotional responses to music arise from several musical features, including musical mode, harmonic complexity, tempo and intensity (Livingstone & Thompson, 2009; Schubert, 2004). Several studies document the effects of manipulated musical features on listeners' emotional responses (Bresin & Friberg, 2000; Livingstone et al., 2010). If the relationship between sounded musical features and listeners' emotional responses is consistent, then listeners' emotional responses should be most similar for repeated phrases in which the musical features are held

---

Steven R. Livingstone and Caroline Palmer, Department of Psychology, McGill University, Montreal, Québec, Canada; Emery Schubert, School of English, Media and Performing Arts, The University of New South Wales, Sydney, New South Wales, Australia.

This research was funded in part by an ACN-Create NSERC Fellowship to the first author and by NSERC Grant 298173 to the second author. Portions of the first experiment were presented at the ISPS meeting in Auckland, 2009. The authors thank Janeen Loehr for assistance and Erik Koopmans, Frances Spidle, Rachel Brown, Max Anderson, and Pascale Lidji for their comments.

Correspondence concerning this article should be addressed to Steven R. Livingstone, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, QC H3Y 1 × 1, Canada. E-mail: steven.livingstone@mcgill.ca

constant. We hypothesize that the degree of change in listeners' emotional responses for two musical sections should vary as a function of the shared musical content between those sections.

We examine the relationship between listeners' emotional responses to repetition and their perception of repeated structural units. One method for examining this relationship is the extraction of musical phrase structure from listeners' emotional responses to repeated phrases. Musical phrases are often marked at their boundaries by changes in several musical features, such as pitch, intensity and duration (cf. Palmer, 1997; Palmer & Krumhansl, 1987; Todd, 1985; Todd, 1992), that are perceptually salient to adults and infants (Krumhansl & Jusczyk, 1990; Neuhaus, Knösche, & Friederici, 2006). Strong emotional responses are often elicited at phrase boundaries (Krumhansl, 1996), locations that typically mark the closure of higher-level musical structures, leading to a completion of musical expectations (Meyer, 1956; Narmour, 1990). Listeners' emotional responses to repeated phrases may therefore contain regions of increased consistency that yield reliable markers for the identification of musical phrase boundaries. We report tests of a computational model that automatically identifies phrase boundaries from listeners' emotional responses. The model implements a windowed cross-correlation algorithm, as this algorithm is well-equipped to identify trends between continuous time series (Boker, Rotondo, Xu, & King, 2002; Schulz & Huston, 2002). This approach differs significantly from existing detection models, which typically identify changing events in a symbolic musical score or an acoustic recording (Cambouropoulos, 2001; Cambouropoulos, 2006; Friberg, Bresin, Frydén, & Sundberg, 1998; Grachten & Widmer, 2007). The current model also implements a growing-window algorithm, which permits flexibility in adapting to repeated segments of different lengths. Previous windowed-correlation models have used a fixed-length sliding window (Boker, Rotondo, Xu, & King, 2002; Schulz & Huston, 2002; Stevens, Schubert, Wang, Kroos, & Halovic, 2009).

Listeners' emotional responses to music have been measured along several dimensions with the use of continuous response methods (Broughton, 2008; Fredrickson, 1995; Luck et al., 2008; Madsen, 1998; Nielsen, 1983; Sloboda, Lehmann, & Parncutt, 1997; for a review see Schubert, 2010). Russell (1980, 2003) provided influential support for a two-dimensional representation of emotion consisting of arousal (high-low activity or activation) and valence (pleasure-displeasure). The current study uses Schubert's (1999) implementation of Russell's two dimensional arousal-valence model of emotional response. Schubert's (1999) findings suggested that a two-dimensional model of continuous emotional response provided a reliable, data-rich alternative for measuring musical emotions. In order to identify specific locations in the music where emotional responses changed over time, Schubert (2002) introduced the use of first-order derivatives of arousal ratings, which analyze the rate of change in arousal (also referred to as affective velocity, Vines, Nuzzo, & Levitin, 2005). Rate of change measures (first-order derivatives) are sensitive to the effects of the current auditory event on listeners' emotional responses and also address issues of serial correlation (Schubert, 2002). The current study reports rate of change in both arousal and valence responses to music as measures of listeners' sensitivity to change.

Two experiments report listeners' emotional response to music that contained repeating musical segments of different lengths; a computational model is reported that identifies phrase boundaries arising from musical repetition as identified in the emotional responses. In each experiment, listeners heard a recording of orchestral music while they rated continuously the emotion they perceived the music to express on a 2-dimensional scale of arousal and valence. Each musical work contained a large section that was repeated twice in the recording. In Experiment 1, the musical work had a symmetric phrase structure (same-length subphrases), and in Experiment 2 the musical work had an asymmetric phrase structure (different-length subphrases), which we hypothesized may lower the predictability of perceiving musical repetition. Continuous measures of arousal and valence were compared for musical phrases that contained repeated (identical), variation (related), or contrasting (unrelated) structure. We expected that participants would respond most similarly for repeated sections, next most similarly for musical variations, and least similarly for contrasting sections. A computational model's performance in detecting phrase boundaries from emotional responses is compared with its performance from physical intensity and tempo measures of the musical recordings, as these acoustic features can have a strong influence on listeners' emotional responses to music (Schubert, 2004). Based on previous findings for musical features (beyond intensity) that affect perception of musical boundaries (Krumhansl & Jusczyk, 1990), we expected detection to be more accurate when based on emotional responses than on the acoustic features.

Experiment 1

Experiment 1 investigated the effects of musical repetition on listeners' perceived emotional response to music. Participants listened to a 3-min recording of an orchestral work (the Pizzicato Polka, by J. and J. Strauss) while rating continuously the emotional arousal and valence they perceived the music to express. Based on previous findings that link positive valence ratings to music in major keys (Gabrielsson & Lindström, 2001; Livingstone et al., 2010), valence was expected to remain constant in listener' responses to the Polka, which is composed in a major key, whereas arousal ratings were expected to change. Listeners' emotional responses were compared across repeated (identical), variation (related), and contrasting (unrelated) musical segments (subphrases, phrases, and sections). We hypothesized that participants' responses would differ most across contrasting segments and differ least across repeated segments. As large numbers of musical features change at musical segment boundaries (Krumhansl & Jusczyk, 1990), it was also hypothesized that participants' emotional responses may be related in systematic ways at these locations.

A computational model for the detection of musical phrase boundaries, based on a growing-window correlation algorithm that detected repetitive patterns of increased consistency, was tested on listeners' emotional responses to the Pizzicato Polka. The model was expected to perform best in detecting phrase boundaries at locations with salient changes in the emotional responses. The model's detection of phrase boundaries based on the sounded intensity and tempo of the music performance was contrasted with its performance based on listeners' emotional responses. If the model performed better on emotional responses, this would suggest that emotional responses provide cues to phrase structure beyond physical intensity and tempo.

Method

Participants

Sixty-seven participants (38 females, mean age = 30.6, *SD* = 12.3, range = 15 to 63), most of whom were undergraduate students from the University of New South Wales community. About one third had taken some coursework in music (*n* = 22). Some participants reported playing a musical instrument (*n* = 58) for different amounts of time (19 had 1–10 years, 39 had 10+ years).

Stimuli

An orchestral recording of the Pizzicato Polka was used (Strauss & Strauss II, 1870; length 2:37 min). The Pizzicato Polka is written in the key of C major, with a moderate tempo and simple repeating harmonies. The work has been characterized as “cheerful” by listeners (Schubert, 1999). The Pizzicato Polka contained the following phrase structure, shown in Figure 1: An introduction (4 metrical bars; 7 seconds), phrase A (8 bars; 12 s), B (12 bars; 16 s), A′ (8 bars; 13 s), followed by a transition (32 bars; 45 s), and a repeat of the introduction and phrases A, B, and A′, finishing with a coda (16 bars; 16 s). Phrases A, B, and A′ were further subdivided into 4-bar subphrases: a1, a2, b1, b2, b3, a′1, a′2. Temporal locations of subphrase boundaries were identified from the orchestral performance with a precision of 0.05 s. Phrases A and A′ had related melodic and harmonic content, whereas Phrase B differed from both A and A′. The two occurrences of A, B, A′ are referred to as Section 1 (0:08–0:48 s; 85 quarter-notes per minute) and Section 2 (1:34–2:14 s; 86 quarter-notes per minute). Sections 1 and 2 (musical repetition) were the focus of the analyses, and are shown in Figure 1.

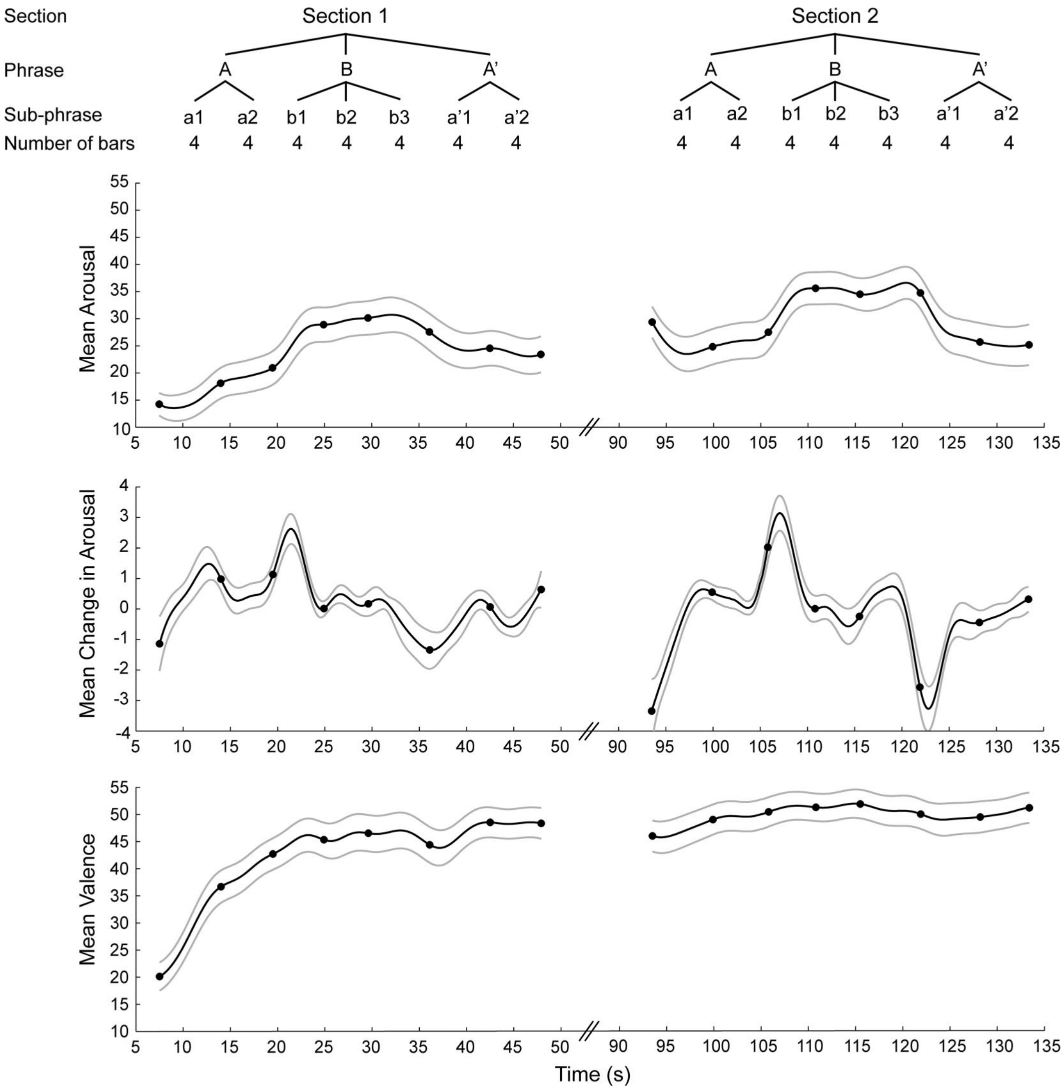

Figure 1. Phrase structure of Sections 1 and 2 of the Pizzicato Polka, with listeners’ mean arousal, change in arousal and valence values (gray lines indicate one standard error). Theoretical phrase boundaries indicated as filled circles.

Apparatus

Stimulus presentation and data collection were controlled by the Two-Dimensional Emotion Space software (2DES, Schubert, 1999). The software presents a visual square with the x- and y-axes (arousal and valence respectively) dividing the space into four equal quadrants. Each axis consists of a 201-point scale ranging from -100 to $+100$ (further details in Schubert, 1999). Positive arousal values corresponded to a high energy emotion (indicated on the 2D scale with wide-eyed faces) and negative arousal values corresponded to a low energy emotion (indicated on the 2D scale with sleepy-eyed faces). Positive valence values corresponded to a positive emotion (indicated on the 2D scale with smiling faces) and negative values to a negative emotion (indicated with frowning faces). Emotion ratings were sampled continuously at 1 Hz, with a mouse used to control the cursor position within the 2D Emotion Space. Testing was conducted on a Power Macintosh and sound recordings were presented over headphones adjusted to a comfortable volume.

Procedure

Participants were seated in front of the computer where they first completed a questionnaire about their musical background. They were then given a set of training tasks explaining the use of the 2D Emotion Space software. The training task introduced the arousal and valence dimensions individually; arousal was described as how active or passive a stimulus was and valence was described as how happy or sad a stimulus was. Participants were then presented with five emotional terms, and asked to mark where on the dimension each emotion fell. The combined 2-Dimensional Emotion Space was then introduced. During training, participants used the 2D Emotion Space to rate the emotion of words and pictures of facial expressions. During testing, participants were instructed to judge the emotion they perceived the music to be expressing, rather than the emotion they may feel in response to the music. Participants then listened to the musical work, while rating continuously the emotion they perceived the music to be expressing. Afterward, participants completed other measures that were not the focus of this study. At the end of the task, participants rated their familiarity with the musical composition using a 7-point scale, ranging from 1 (*never heard it*) to 7 (*know the piece intimately*).

Data Analysis

Arousal and valence responses were analyzed with functional data analysis techniques (Ramsay & Silverman, 2005), which turn discrete measures into a continuous function. Functional techniques enable comparisons across unequal length sections of the music by generating equivalent numbers of data points from the continuous function. Functional data analysis also enables analysis of derivatives (rates of change) after reducing sampling noise with data smoothing. Order 6 B-splines were fit to each individual's raw emotional response data, with a 1:1 ratio of knots (break points) to data samples. The data were smoothed with a constant smoothing parameter ($\lambda = 0.13809$) applied to the 2nd derivative, which minimized the mean Generalized Cross-Validation estimate across all individuals. Landmark registration at each phrase boundary was used to align participants' responses across Sections 1 and 2 which were of different durations (40.4 and 39.9 s, respectively). Smoothed functional data were interpolated to create 405 equally spaced data points per section, yielding a new sampling rate of 10 Hz. First order derivatives (velocity) of arousal and valence responses were then produced from the smoothed functional form. To enable comparisons between subphrases of variable length, the functional data were interpolated to produce 80 equally spaced data points per subphrase with a minimum sampling rate of 10 Hz.

Physical intensity values were extracted from the 44.1 kHz 16-bit stereo recording with Praat (Boersma & Weenink, 2010), utilizing a weighted average Gaussian window to resample and smooth the intensity contour, with a window size of 32 ms (minimum pitch of 100 Hz) and a time step of 0.1 ms. This yielded a new sampling rate of 10 Hz, matching that of the arousal and valence section datasets. The resampled intensity contour still contained a significant amount of fine-grained activity and, therefore, was smoothed with the same functional data analysis methods used in the analysis of arousal and valence but with a modified smoothing parameter ($\lambda = 10^{-6}$). Tempo values were extracted using spectral and waveform analysis. The tempo contour was smoothed with the same functional data analysis methods used in the analysis of arousal and valence.

Results

Listeners' Emotional Ratings

Figure 1 presents the mean arousal and valence responses with the Pizzicato Polka's phrase structure. Listeners' emotional ratings were high (arousal) and positive (valence), consistent with a description of the Pizzicato Polka as "happy" or "cheerful". Listeners' arousal responses for Section 1 were highly similar to those for Section 2, which occurred later in the musical context. The correlation across repeated sections for mean listeners' responses, shown in Figure 1, was significant (number of data points = 405, $r = .83$, $p < .001$). As well, the mean correlation across sections within individuals was significant (mean $r = .30$, $p < .001$). Figure 1 suggests that listeners' arousal measures corresponded with the phrase structure, with higher arousal during phrase B relative to phrases A and A′. In contrast, valence responses did not vary much or correspond to the musical work's phrase structure. As shown in Figure 1, valence ratings tended to increase at the beginning and then stayed stable until the end of the music. Therefore, further analyses of the emotional responses to the Pizzicato Polka piece are reported only for the arousal ratings.

Next we analyzed listeners' responses to subphrases that reoccurred within and across sections. Three categories of musical repetition were identified. Repeated subphrases ($n = 7$) were those first heard in Section 1 and later repeated (in identical compositional form) in Section 2 (e.g., Section 1 a2 and Section 2 a2; see Appendix A for a complete listing of phrase pair comparisons). Variation subphrases ($n = 36$) included remaining pairings of subphrases from within each phrase A-A′ or B; thus, Variations contained related (nonidentical) harmonic and melodic content (e.g., Section 1 a2 and Section 1 a′2). Contrasting subphrases ($n =$

48) included all remaining pairings of subphrases from across phrases A-A′ and B; thus, Contrasting subphrases contained unrelated harmonic and melodic content (e.g., Section 1 a1 and Section 1 b1). The number of emotional response samples in each subphrase was kept constant across all comparisons ($n = 80$). All possible subphrase pairings were included in the Variation and Contrasting categories to prevent a bias from any particular subphrase. Rate of change in listeners' arousal measures (first-order derivative of arousal), which captures the change in emotional state for each auditory event, was examined in terms of these categories of musical repetition. Listeners' mean change in arousal responses, shown in Figure 1, suggest consistency across the repeated Sections 1 and 2.

We tested listeners' sensitivity to degree of repetition (Repeated, Variation, and Contrasting subphrases) by computing difference scores on change in arousal responses for each pair of subphrases in the three subphrase categories (see Appendix A). Listeners' mean difference scores (absolute value) are shown in Figure 2 for pairs of Repeated, Variation, and Contrasting subphrases (number of data points per subphrase = 80). Absolute values of difference scores were computed to examine total deviations from zero in change in arousal. A one-way functional analysis of variance (Ramsay & Silverman, 2005) was conducted on the continuous functions for the three conditions shown in Figure 2. Regions of significance, indicated by the horizontal line beneath the curves (Dunn-Bonferroni adjustment of p values for the number of tests), show significant differences across the subphrase categories, with least difference in emotional response to Repeating subphrases and most difference for Contrasting subphrases. Importantly, the categories differed across the majority of the subphrase durations, suggesting listeners were influenced by repetition throughout the segments.

Next, we compared the degree of similarity in listeners' emotion ratings across the Repeated, Variation, and Contrasting subphrases. Listeners' mean change in arousal responses for Repeated subphrases, shown in Figure 1, were highly similar (mean $r = .81$, $p < .001$). As well, the correlations on listeners' mean change in arousal among Variation subphrases were significant (mean $r = .74$, $p < .001$). A significant negative correlation was found for Contrasting subphrases (mean $r = -.65$, $p < .001$), indicating an inverse emotional relationship between Contrasting and Repeating subphrases. A one-way analysis of variance (ANOVA) on individuals' correlation values for the three subphrase categories confirmed a significant main effect [Repeated, Variation, and Contrasting subphrases; $F(2, 132) = 4.08$, $MS_e = .073$, $p < .05$]. Post hoc comparisons (Tukey's HSD, $\alpha = .05$) confirmed that both emotional responses to Repeated (mean $r = .11$) and Variation (mean $r = .10$) subphrases were more highly correlated than Contrasting subphrases (mean $r = -.01$).

The relatively short subphrases contained four metrical bars; to evaluate whether emotional response to repetition were consistent over longer musical passages, listeners' mean correlations were compared for Repeated passages at the Subphrase (4-bar), Phrase (8-bar), and Section (16-bar) levels shown in Figure 2 (the number of data points was held constant across levels and increased to $n = 160$ to yield adequate resolution at the larger timescale of Phrase and Section levels). A one-way ANOVA on individuals' correlation values for musical segments at the three hierarchical levels confirmed a significant main effect of level [$F(2, 132) = 3.74$, $MS_e = .025$, $p < .05$], with greatest similarity in listeners' ratings

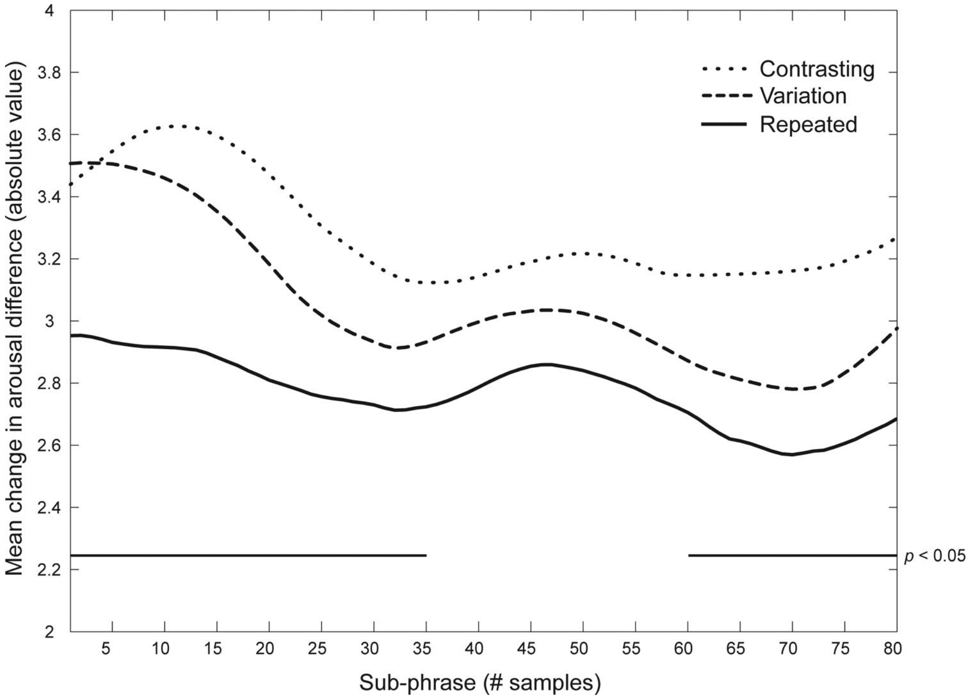

Figure 2. Listeners' mean difference scores for Repeated, Variation, and Contrasting subphrases (absolute value) in the Pizzicato Polka. Horizontal line indicates significance region at $p < .05$.

at higher levels (Section $M = .18$; Phrase $M = .14$; Subphrase $M = .11$). This finding suggests that emotional response to musical repetition was not limited to the smaller subphrase units.

We investigated whether listeners' familiarity with the musical work, musical experience, or age influenced the change in arousal responses between repeated sections. Individual listeners' correlation values across Sections 1 and 2 (which include the entire set of emotional responses) were compared with their familiarity with the musical work (which listeners ranked on the scale of 1 = *never heard it* to 7 = *know the piece intimately*), musical experience (measured in years of playing a musical instrument), and age. None of the correlations reached significance (all p's $> .05$). Thus, correspondences in emotional responses across repeating musical sections were not limited to listeners with specific or general musical experience.

Boundary Detection Model

To determine whether phrase boundaries could be identified from listeners' emotional responses, an automatic boundary detection algorithm was developed which identified peaks in correlations across listeners' change in arousal ratings. The growing-window correlation algorithm iteratively correlated listeners' mean change in arousal responses for Section 1 to responses in Section 2, by applying an increasing correlation window size to identify the highest points of similarity across sections. Window sizes began at 0.3 s (first three emotional response samples of Section 1 were correlated with the first three samples of Section 2) and grew iteratively (+0.1 s) to include the entire $t = 40.4$ s section lengths ($n = 405$). Each of the 403 correlation windows produced a single correlation value. The resulting correlation value from each window was plotted against time to form a correlation curve. Peaks in the resulting correlation curve were defined as three successive increasing values followed by three successive decreases; peaks represented regions in which emotional responses to Section 1 increased and then decreased in their similarity to responses to Section 2. To balance the number of correlations computed at each time point, the algorithm was then repeated, working backward from the end of the section toward the beginning. Peak values were then averaged across the forward and backward correlation curves. The locations of all correlation peaks ($p < .01$, adjusted for window size) were then compared with the theoretical subphrase boundary locations, shown in Figure 3.

The model's performance in identifying phrase boundaries was then compared with notated phrase boundaries in the musical score; the model's identification accuracy is reported in Table 1. Hits were defined as locations identified by the model that corresponded to those in the performance, occurring within a 4-s window (−1.5 to +2.5 s). This window was used to capture participants' change in arousal responses which often lag (+2.5 s) the auditory stimulus presentation (Krumhansl, 1996; Schubert, 2001), as well as listeners' expectations which may influence emotional responses prior to the phrase boundary (Meyer, 1956), estimated as −1.5s. False positives were defined as those locations identified by the model that were not identified in the score.

The algorithmic model was successful (100% hit rate) at detecting all phrase boundary locations in listeners' mean change in arousal responses within −1.5 to +2.5 seconds (−15 to +25 samples) of the expected boundary, and it was relatively accurate (67% hit rate) within −1.2 to 1.2 seconds. The model generated only one false positive for listeners' mean change in arousal responses. Similar findings were obtained when the model was fit to individuals' change in arousal responses; a mean of 5.1 (85.1%) boundaries and 3.6 (41.2%) false positives were found.

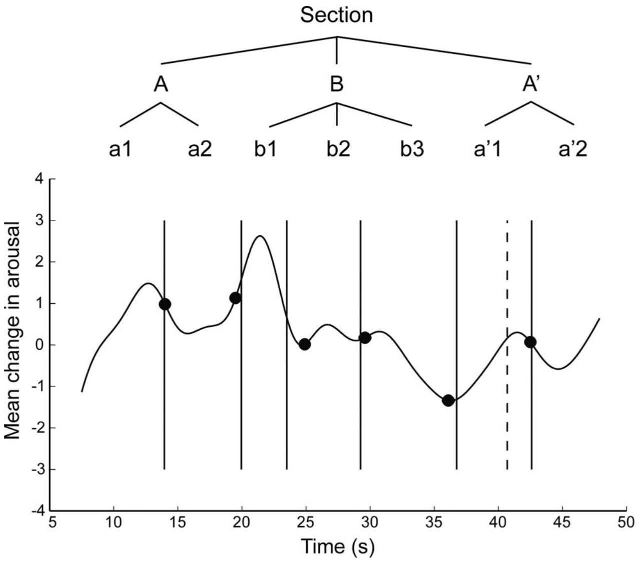

Figure 3. Listeners' mean change in arousal and boundaries detected by model (vertical lines). Filled circles indicate theoretical phrase boundaries based on information in the musical score; vertical dashed lines indicate false positives by the algorithm.

The model's accuracy on listeners' emotional responses was compared with its accuracy on the physical intensity and tempo measures. The physical intensities for Section 1 were highly similar to those of Section 2 in this performance (number of data points = 405, $r = .86$, $p < .001$). The tempo of Sections 1 ($M = 81$ beat per minute, range = 44–101) and 2 ($M = 86$ beats per minute, range = 60–112) were also similar. Although the algorithm successfully identified all phrase boundaries based on physical intensity (hits), it also identified an increased number of false positives ($n = 3$). The algorithm identified four of six phrase boundaries based on tempo (hits), and also identified an increased number of false positives ($n = 4$). Thus, listeners' mean change in emotional arousal responses provided the algorithm with more discriminating cues to phrase structure than those provided by the physical intensity or tempo of the performance.

Discussion

Listeners' emotional arousal responses differed for repeated, variation, and contrasting musical subphrases. This effect of repetition complements the finding that arousal responses were highly similar for repeated musical subphrases and highly dissimilar for musically contrasting subphrases. Both raw emotional responses and similarity metrics indicated differential sensitivity to degree of musical repetition. Furthermore, emotional response to repetition was consistent over longer musical passages (section level) as well as shorter passages (phrase and subphrase levels). These effects on emotional response were independent of the participant's musical

Table 1

Detection Algorithm Accuracy for Change in Arousal and Physical Intensity Measures From Experiment 1

Model fit	Hits[1]	False positives[2]
Mean change in arousal	6 (100%)	1 (14.3%)
Physical Intensity	6 (100%)	3 (33.3%)
Tempo	4 (66.7%)	4 (50%)

[1] (%) of hits is number of correctly detected boundaries divided by total number of boundaries determined by the musical score. Number of opportunities for hits = 240/403 samples (59.6%). [2] (%) for false positives is number of incorrectly detected boundaries divided by total number of detected boundaries. Number of opportunities for false positives = 163/410 samples (40.4%).

experience, familiarity with the Pizzicato Polka, and age, indicating that the perceived emotion was consistent across a range of listeners.

The automatic boundary detection model successfully identified all notated phrase boundaries from listeners' mean change in arousal responses, with a low number of false positives. This finding supported the hypothesis that listeners' emotional responses reflected the underlying musical phrase structure and were more consistent at phrase boundaries. The model was also successful on within-individual emotional responses, with a higher number of false positives. Listeners' change in arousal responses provided a more accurate predictor for the model's detection of phrase boundaries than did the physical intensity or tempo of the sounded performance. This suggests that listeners integrate multiple features in their emotional judgments, and that features beyond intensity or tempo are required to signal the presence of a phrase boundary. Another explanation is that listeners integrate the closure of higher level phrase structure information with low level features, such as tempo and intensity, when making emotional judgments, and that the model is therefore able to recover phrase boundary information from the emotional responses.

The findings of Experiment 1 supported the main hypothesis that change in emotional arousal is responsive to repeated musical sections. However, the Pizzicato Polka has a simple and symmetrical phrase structure, with each subphrase of equal length (four bars in length), using conventional Western harmonic structures. It is unknown if the main hypotheses would hold for more structurally complex and variable musical works. Thus, a second experiment was conducted with a different musical work.

Experiment 2

Experiment 2 applied the methods of Experiment 1 to a structurally complex musical work. The Slavonic Dance No. One in C major, Op. 46, No. One by Antonín Dvorák was chosen because the orchestral work, like the Pizzicato Polka, contains melodic repetition as well as varied and contrasting phrases. In contrast to the Pizzicato Polka, the Slavonic Dance contains an asymmetric phrase structure consisting of successive subphrases of unequal length (subphrase lengths ranged from 8 to 12 bars) and more complex harmonies. Participants' emotional responses to the musical repetition in this structurally complex piece were measured with the same methods as in Experiment 1. The effectiveness of the boundary detection model was also tested on listeners' emotional responses.

Method

Participants

The same participants from Experiment 1 participated in Experiment 2.

Stimuli

An orchestral recording of the Slavonic Dance was used (Dvorák, 1878; length 3:52 min). The Slavonic Dance is composed in the key of C major with some sections in A minor, and has more complex harmonic, rhythmic, and phrase structure than the Pizzicato Polka. The work is performed at a fast tempo and has been characterized as "exciting" and "triumphant" by listeners (Schubert, 1999). Again, previous findings that link music in major keys to positive valence ratings (Gabrielsson & Lindström, 2001; Livingstone et al., 2010) led to the prediction that arousal but not valence was expected to change in listener' responses. The Slavonic Dance has the following phrase structure: opening (1 bar, 2 seconds), phrase A (16 bars; 11 s), B (16 bars; 11 s), C (21 bars; 15 s), A′ (16 bars; 11 s) followed by a transition (90 bars; 78 s), and a repeat of the phrases A, B, C, and A′, finishing with a coda (55 bars; 55 s). Phrases A, B, C, and A′ were further subdivided into variable length subphrases: a1, a2, b1, b2, c1, c2, a′1, a′2. Phrases A and A′ contained varied (related, nonidentical) melodic and harmonic content, whereas Phrases B and C were not related to A, A′, or to each other. The two occurrences of A, B, C, A′ are referred to as Section 1 (0:03 – 0:51 s; 117 dotted half-notes per minute) and Section 2 (2:09 – 2:57 s; 115 dotted half-notes per minute). Sections 1 and 2 (musical repetition) were the focus of the analyses, and are shown in Figure 4.

Design and Procedure

The Experimental design and procedures were the same as in Experiment 1.

Data Analysis

Listeners' arousal and valence responses were analyzed with functional data analysis techniques, using the same parameter values as in Experiment 1. Landmark registration across each subphrase boundary was used to align participants' responses across Sections 1 and 2, which were of different durations (48.6 and 49.2 sec respectively). Smoothed functional data were interpolated to create 487 equally spaced data points per section, yielding a new sampling rate $\approx$ 10 Hz. First order derivatives of arousal and valence were then produced from the smoothed functional form. The functional data were interpolated to produce a second dataset with 80 equally spaced data points per subphrase. Analyses of the musical work's intensity were also conducted using the same parameters as in Experiment 1.

Results

Listeners' Emotional Ratings

Listeners' mean arousal responses for Section 1 were highly similar to those for Section 2, as shown in Figure 4. Listeners'

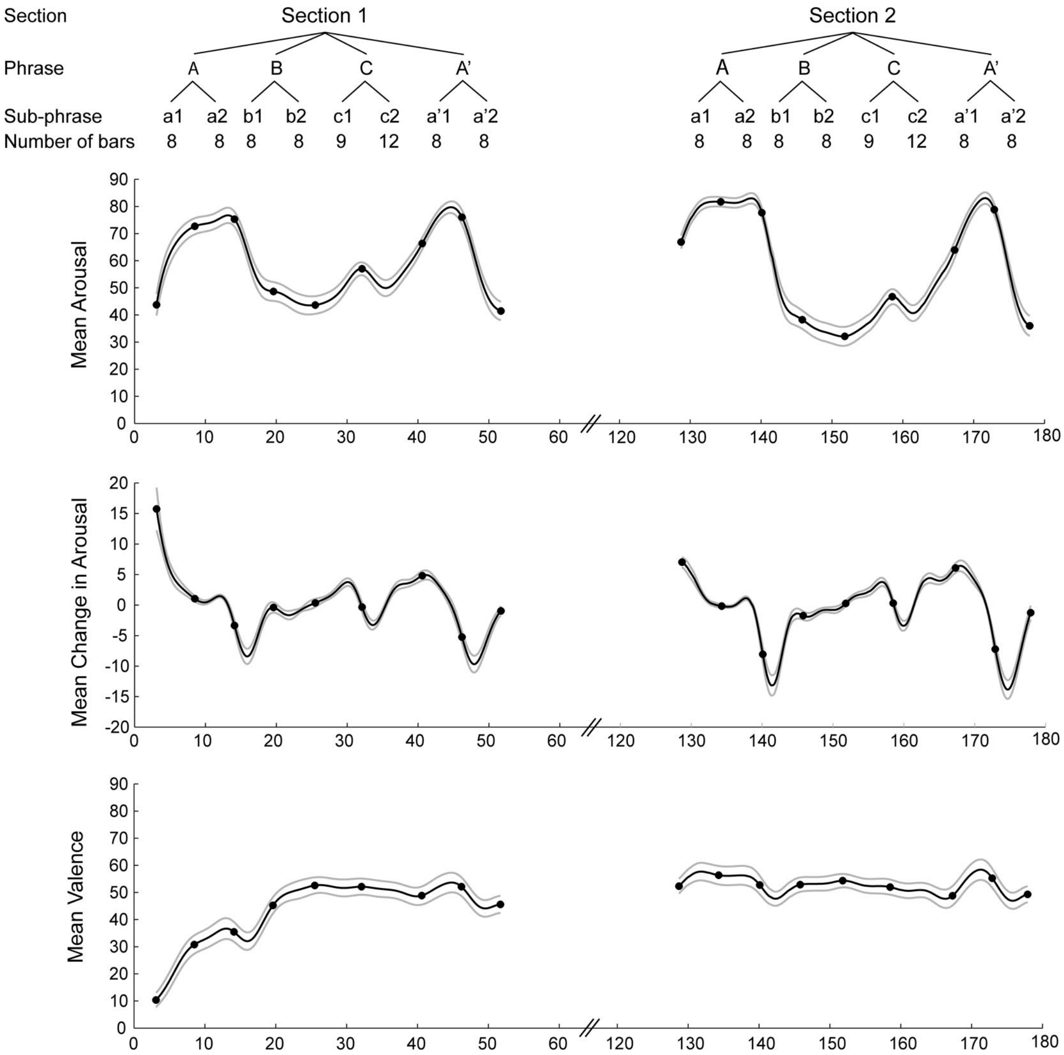

Figure 4. Phrase structure of Sections 1 and 2 of the Slavonic Dance, with listeners' mean arousal, change in arousal and valence values (gray lines indicate one standard error). Theoretical phrase boundaries indicated as filled circles.

arousal and valence ratings were high and positive, consistent with a description of the Slavonic Dance as "exciting" or "triumphant". The correlation across sections, based on the mean arousal responses, was significant (number of data points = 487, $r = .94$, $p < .001$). As well, the correlation within individuals across the two sections was significant (mean $r = .53$, $p < .001$). Figure 4 indicates that listeners' arousal measures corresponded with the phrase structure, with higher arousal during phrases A and A′ relative to phrases B and C. In contrast, listeners' mean valence responses did not appear to correspond with the work's phrase structure. As shown in Figure 4, valence ratings tended to increase at the beginning and reached a plateau: at a positive level (similar to listeners' responses in Experiment 1). Therefore, further analyses of the emotional responses to the Slavonic Dance piece were conducted only on the arousal ratings.

We examined listeners' sensitivity to musical repetition by comparing their responses to Repeated, Variation, and Contrasting subphrases. As in Experiment 1, we focus on rate of change in arousal measures, which capture changes in emotional response as the piece unfolds. The categorization of subphrase similarity was more complex than in Experiment 1 due to the increased complexity of the Slavonic Dance's phrase structure. Repeated subphrases ($n = 8$) were those first heard in Section 1 and later reheard in Section 2 (an example is Section 1 a1 and Section 2 a1). Variation subphrases ($n = 20$) included all remaining subphrase pairings that arose from within phrases A-A′, B, or C; thus, Variations contained related harmonic and melodic content (a complete list of all subphrase pairings is provided in Appendix B). Contrasting subphrases ($n = 84$) included all pairings of subphrases that arose from different phrases (A-A′, B, and C) and contained unrelated harmonic and melodic content.

We tested listeners' sensitivity to degree of repetition (Repeated, Variation, and Contrasting subphrases) by computing difference scores for each pair of subphrases in the three repetition categories. Listeners' mean difference scores (absolute value) are shown in Figure 5 for Repeated, Variation, and Contrasting subphrases (number of data points per subphrase = 80). A one-way functional ANOVA was conducted on the continuous functions by repetition category; regions of significance, indicated by the horizontal line at the bottom (Dunn-Bonferonni corrections applied to *p* values for the number of tests), show significant differences across the categories. Repeating phrases elicited the least difference in emotional responses and Contrasting phrases elicited the greatest difference. Importantly, the categories differed across the entire subphrase region, suggesting listeners were influenced by repetition throughout the segments.

Next, we compared the degree of similarity in listeners' emotion ratings across the Repeated, Variation, and Contrasting subphrases. Listeners' mean change in arousal responses, shown in Figure 4, indicate highly similar responses for Repeated subphrases (number of data points = 80, mean $r = .90$, $p < .001$). As well, the correlation across Variation subphrases was significant (mean $r = .70$, $p < .001$). The negative correlation across Contrasting subphrases was also significant (mean $r = -.32$, $p < .01$), indicating an inverse emotional relationship between Contrasting and Repeated subphrases. A one-way ANOVA on individuals' subphrase correlation values revealed a significant main effect of Similarity [Repeated, Variation, and Contrasting; $F(1, 66) = 38.40$, $MS_e = 0.063$, $p < .001$]. Post hoc comparisons (Tukey's HSD, $\alpha = .05$) confirmed that both Repeated (mean $r = .33$) and Variation (mean $r = .20$) subphrases were more highly correlated than Contrasting subphrases (mean $r = -.04$) and that Repeated subphrases were more highly correlated than Variation subphrases.

We evaluated whether emotional response to repetition differed over longer musical passages by comparing listeners' mean correlations for Repeated passages at the Subphrase, Phrase, and Section levels shown in Figure 4 (the number of data points was held constant across levels and increased to $n = 160$, to yield adequate resolution at larger timescales). A one-way ANOVA on individuals' correlation values for musical segments at the three hierarchical levels confirmed a significant effect of level [$F(2, 132) = 29.1$, $MS_e = .022$, $p < .01$], with greatest similarity in listeners' ratings at higher levels (Section $M = .49$; Phrase $M = .39$; Subphrase $M = .30$). Again, emotional response to musical repetition was highly consistent across larger segments.

We investigated whether listeners' familiarity with the musical work, musical experience, or age influenced the consistency of their change in arousal responses between sections. The correlation values across Sections 1 and 2 within listener were correlated with their familiarity with the musical work, musical experience, and age. No factor reached significance (all *p*'s > .05). Thus, correspondences in emotional responses across repeating musical sections were not influenced by specific or general musical experience.

Boundary Detection Model

The boundary detection model, based on a growing-window correlation algorithm, was applied to participants' emotional measures for the more complex Slavonic Dance. The same model parameters used in Experiment 1 were maintained in Experiment 2. Window sizes began at 0.3 s (first three emotional response samples) and grew iteratively (+0.1 s) to include the entire (48.6 s)

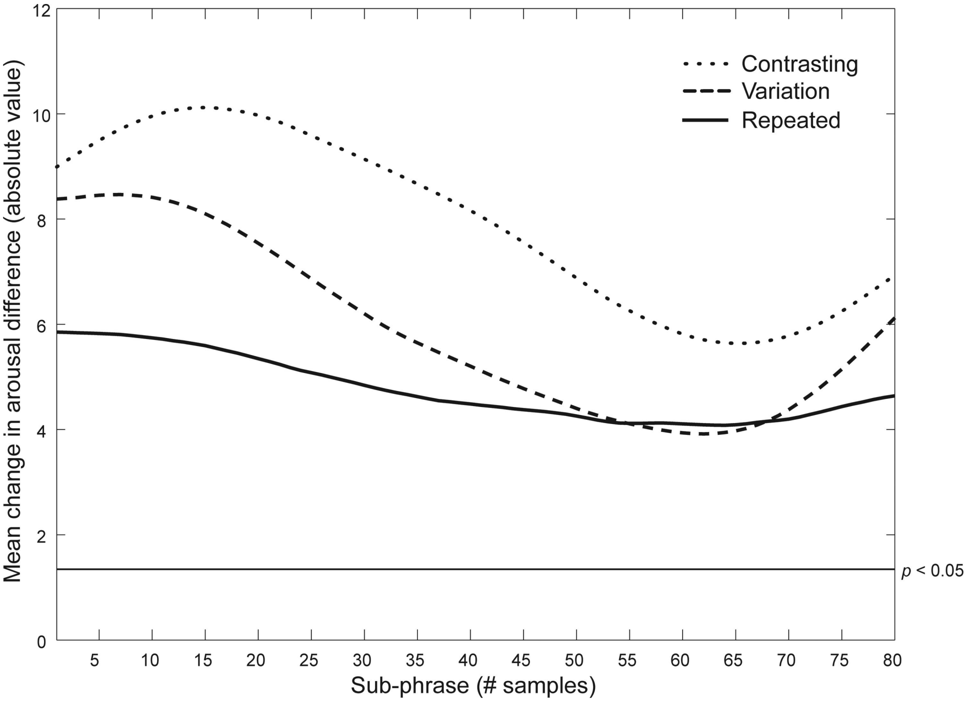

Figure 5. Listeners' mean difference scores for Repeated, Variation, and Contrasting subphrases (absolute value) in the Slavonic Dance. Horizontal line indicates significance region at $p < .05$.

section length. Peaks in the correlation curves were then compared with known subphrase boundary locations, shown in Figure 6.

The model's identification accuracy for the phrase boundaries as notated in the musical score is reported in Table 2 mean change in emotional arousal, intensity, and tempo measures. Hits were defined as locations identified by the model that corresponded to those in the musical score, occurring within a -1.5 to $+2.5$ s (-15 to $+25$ samples) time window. False positives were defined as those locations identified by the model that were not identified in the score.

The algorithmic model was accurate at detecting phrase boundary locations (hits = 100%) in the mean change in arousal response data within -1.5 to $+2.5$ seconds of the expected boundary, and relatively accurate (57% hit rate) within -1.2 to $+1.2$ seconds. The model generated three false positives (locations identified as phrase boundaries that were not present in the musical score) from listeners' mean change in arousal ratings; comparison of these locations with performance cues did not indicate any intensity or tempo changes. Fits of the model to individuals' change in arousal responses indicated similar findings, with means of 5.7 boundaries (82.1% hits) and 5.4 (48.6%) false positives.

Two of the three false positives occurred in close proximity, the first in subphrase b1, and the second in the musical variation subphrase b2. These 8-bar subphrases are musically related as the first six bars of b1 are repeated in b2. Both false positives (16.8 s, 23.6 s) occurred close to bar 4 of the two subphrases (16.9 s, 22.6 s). Thus, the first false positive detected in b1 may have reoccurred in b2 because listeners were responding to an identical musical passage. A reanalysis of the score and performance indicated that subphrases b1 and b2 could each be further subdivided into two 4-bar subphrases. Thus, these false positives may reflect phrase boundaries occurring at a lower level in the phrase structure than was assessed in the model's accuracy criteria. Using this adjusted detection criterion, the number of false positives for the mean data was reduced to 1.

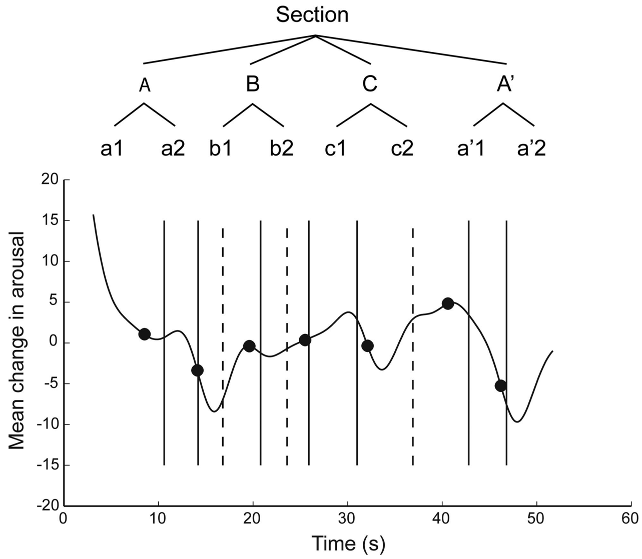

Figure 6. Listeners' mean change in arousal and boundaries detected by model (vertical lines). Filled circles indicate theoretical phrase boundaries based on information in the musical score; vertical dashed lines indicate false positives by the algorithm.

Table 2

Detection Algorithm Accuracy for Emotional Arousal and Physical Intensity Measures From Experiment 2

Model fit	Hits[1]	False positives[2]
Mean change in arousal	7 (100%)	3 (30%)
Physical Intensity	7 (100%)	6 (46.2%)
Tempo	7 (100%)	4 (36%)

[1] (%) of hits is number of correctly detected boundaries divided by total number of boundaries determined by the musical score. Number of opportunities for hits = 280/485 samples (57.7%). [2] (%) for false positives is number of incorrectly detected boundaries divided by total number of detected boundaries. Number of opportunities for false positives = 205/485 samples (42.3%).

The boundary detection model's performance on listeners' emotional ratings was compared with its performance on the sounded intensity and tempo of the performance. The physical intensities for Section 1 were highly similar to those of Section 2 in this performance (number of data points = 487, $r = .98$, $p < .001$). The tempo of Sections 1 ($M = 85$ beats per minute, range = 80–91) and 2 ($M = 83$ beats per minute, range = 74–91) were also similar, but not as consistent across sections as was intensity ($r = .65$, $p < .01$). Although the algorithm successfully identified all phrase boundaries based on intensity, it also identified many false positives ($n = 6$). Similarly, while the algorithm successfully identified all phrase boundaries based on tempo (hits), it also identified an increased number of false positives ($n = 4$). As in Experiment 1, listeners' mean change in arousal responses were more discriminating with respect to the musical phrase structure than the cues provided by the physical intensity or tempo.

Discussion

Experiment 2 replicated the main findings from Experiment 1 with a structurally more complex work. Listeners' emotional arousal responses differed for repeated, variation, and contrasting musical subphrases. In addition, listeners' arousal responses were highly similar for repeated musical subphrases, moderately similar for variations, and highly dissimilar for musically contrasting subphrases. Furthermore, emotional response to repetition was consistent over longer musical passages (section level) as well as shorter passages (phrase and subphrase levels). These results were independent of the listeners' musical experience, familiarity with the music, or age. As in Experiment 1, valence ratings exhibited activity at the beginning of the task, but displayed minimal change thereafter.

The boundary detection model successfully identified all theoretical phrase boundaries from listeners' emotional response data, replicating the result of Experiment 1. This finding supported the main hypothesis that listeners' emotional responses were sensitive to boundaries defined by repeated musical sections. As in Experiment 1, the boundary detection model was most accurate when applied to the rate of change in listeners' mean arousal responses, which achieved 100% detection accuracy with a moderate rate of false positives (detected events that were not phrase boundaries). The model also detected most boundaries when applied to individual data but with a higher rate of false positives. The location of these false positives did not coincide with large changes in intensity or tempo, but instead, coincided with locations that suggest boundaries occurring at a lower level in the phrase structure than was originally hypothesized. Thus, the boundary detection model may reveal listeners' segmentation tendencies above and beyond the locations hypothesized as boundaries. This supports the assumption that listeners integrate several musical features in their emotional responses (beyond intensity or tempo) which allow the model to recover higher level phrase structure.

General Discussion

Musical repetition influenced listeners' emotional response to music in two experiments. Listeners' emotional arousal ratings differed most for contrasting phrases and differed least for repeated phrases. Furthermore, the emotional arousal responses to repetition were highly similar across all hierarchical levels, including repeated sections, phrases, and subphrases, indicating consistency at both large and small timescales. An interesting finding was that listeners' responses to musically contrasting (unrelated) phrases were highly dissimilar, suggesting that listeners' emotional response to these phrases changed in opposite directions. The strong associations between emotional arousal and musical repetition were replicated in the symmetric musical structure (same-length phrases) of Experiment 1 and asymmetric musical structure (different-length phrases) of Experiment 2, indicating that the predictability of phrase boundaries did not constrain the effects of repetition on listeners' emotional response.

These findings support the hypothesis that listeners' emotional responses to music vary as a function of musical repetition across segments. The computational model's ability to derive segment boundaries from regions of musical repetition provides support for the assumption that a listener's emotional response to music can be influenced by the set of musical features that change near segment boundaries, and that in turn, boundary information can be recovered from listeners' emotional responses. These findings extend previous research on the emotional effects of musical repetition (Krumhansl, 1996) in three ways. First, emotional responses to musical repetition occurred at multiple hierarchical levels, from 4-bar subphrases to 69-bar musical sections, suggesting that the perception of repetition is not constrained to small segments by auditory working memory. Second, emotional responses to Repeated, Variation, and Contrasting subphrases differed throughout the musical segments, indicating that perceived differences were not simply due to changes at segment boundaries. Finally, listeners' emotional responses were independent of their musical experience, familiarity with the music, or age. This generalizability across listeners with different experience complements previous findings of stable emotional responses across a wide range of adult ages (Lima & Castro, in press).

These findings suggest that continuous emotional response may be an implicit measure of musical similarity. Previous studies have typically examined the perception of musical similarity with explicit, retrospective response tasks, in which participants rated how similar two musical excerpts were (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001; McAdams, Vieillard, Houix, & Reynolds, 2004; Schubert & Stevens, 2006). Although such responses are sufficient for short musical pieces, they are not feasible for longer musical works such as those used in the current study. As the total number of unique combinations of excerpt pairs from a musical work with 'n' excerpts is $(n^2 - \mathrm{n})/2$, an exhaustive listener response approach becomes infeasible for moderate- to larger-sized works.

One goal of the current experiments was to determine whether listeners were sensitive to large-scale repetition. Listeners' emotional responses are usually collected for short musical segments (less than 1 min duration), that may not be sufficient to enable perception of higher-level structure. In both experiments, listeners rated their emotional responses to complete 3 min orchestral works, which were long enough to allow large-scale structural comparisons to arise. Listeners' responses were more consistent at large than at small timescales, despite the fact that the orchestral performances of these sections were not acoustically identical in physical intensities and tempi. Listeners' similarity in responses to compositional repetitions was also higher than for variations or contrasting sections, despite these minor acoustic differences. Large-scale structural repetition, therefore, may be more salient than acoustic repetition in emotional response. Future work may address this hypothesis with musical works that contain compositional repetition, but vary in their acoustic similarity.

Listeners' emotional valence responses exhibited little change across the musical works. Previous examinations of the relationship between emotional response and musical features suggest that valence responses are typically affected only by the musical mode and harmonic complexity (Gabrielsson & Lindström, 2001; Livingstone et al., 2010). As these features remained relatively constant throughout the repeated sections of the musical works, listeners' valence responses might be expected to show minimal change. The small valence changes that accompanied larger changes in arousal may reflect movement co-occurrence in both dimensions of the 2-dimensional input device.

A boundary detection model was developed to examine the relationship between listeners' emotional arousal and the perception of repeated musical units, by identifying the boundaries of musical segments from listeners' emotional responses. The model was based on the assumption that listeners' emotional responses would be most similar across regions of repeated segments. These regions reflect increased consistency in the use of musical features, which we hypothesized increased their perceptual salience. The model provided a quantification of the detectable influence of boundaries on listeners' emotional responses to music; all phrase boundaries in both musical works were successfully identified. A reanalysis of the false positives reported in Experiment 2 suggested that some phrase boundaries were perceived at a lower level than was notated as theoretical phrase boundaries, which were assessed by the model. The detection model was able to identify boundaries based on the physical intensity and tempo of the performances but also identified more false positives than in the emotional responses, suggesting that intensity and tempo were less discriminating inputs for the automatic detection of musical phrase boundaries.

This research extends the field of automatic boundary detection models which rely on information in the notated musical score or acoustic recording as input (Cambouropoulos, 2001, 2006; Friberg et al., 1998; Grachten & Widmer, 2007). These models typically locate patterns of changes in note durations, which often mark a phrase boundary. Detection accuracy for these categories of models is around 65%, with a false positive rate of 31–45% (Cambouropoulos, 2001). The current model's application yielded a 100% hit rate and lower false positive rates than those previously reported (Experiment 1: false positive rate of 14%; Experiment 2: 30%). Thus, the use of emotional responses as input can provide an effective alternative for automatic detection models that use notated scores or acoustic recordings. The current model also differed from previous models in its use of a variable-length growing window algorithm, compared with other windowed-correlation models which used a fixed-length sliding window (Boker, Rotondo, Xu, & King, 2002; Schulz & Huston, 2002). The variable-length growing window permitted more flexibility and sensitivity to small- and large-scale repeated segments. Similar to existing detection models, the current model inferred the location of surface phrase boundaries but not the hierarchical structure of the musical work. Future research may develop methods for the derivation of hierarchical information from listeners' emotional responses to music.

Conclusion

In summary, listeners' emotional arousal responses to musical segments that repeated in a large musical work differed systematically with the degree of musical repetition. Listeners' judgments of emotional arousal at multiple hierarchical levels were most similar for repeating segments, moderately similar for variations, and least similar for contrasting segments. These findings were documented in musical works with both symmetric and asymmetric phrase structures, suggesting that the effects of repetition on emotional response did not depend on the predictability of the segment lengths. The automatic boundary detection model successfully identified all theoretical phrase boundaries in listeners' emotional responses, which reflected the structure of the musical work beyond the physical performance attributes. Continuous emotional response offers an effective method for examining listeners' response to musical repetition and their perception of musical structure.

References

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer program]. Version 5.1.42. Retrieved from http://www.praat.org/

Boker, S., Rotondo, J., Xu, M., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods, 7,* 338–355.

Bresin, R., & Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal, 24,* 44–63.

Broughton, M. C. (2008). *Music, movement and marimba: Solo marimbists' bodily gesture in the perception and production of expressive performance.* Unpublished PhD, University of Western Sydney, Sydney, Australia.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference* (*ICMC01*) (pp. 232–235). Havana, Cuba.

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation. *Music Perception, 23,* 249–268.

Dvorák, A. (1878). Slavonic Dance No. 1 in C major, Op. 46, No. 1 [Slovak Philharmonic Orchestra, conducted by Zdenek Kosler]. On *Discover Classical Music* [CD]. Hong Kong: Naxos.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception, 18,* 275–296.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae, 12,* 75–99.

Fredrickson, W. (1995). A comparison of perceived musical tension and aesthetic response. *Psychology of Music, 23,* 81–87.

Friberg, A., Bresin, R., Frydén, L., & Sundberg, J. (1998). Musical punctuation on the microlevel: Automatic identification and performance of small melodic units. *Journal of New Music Research, 27,* 271–292.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae, Special Issue 2001–2002,* 123–147.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: theory and research* (pp. 223–248). Oxford: Oxford University Press.

Grachten, M., & Widmer, G. (2007). Towards phrase structure reconstruction from expressive performance data. In E. Schubert, K. Buckley, R. Eliott, B. Koboroff, J. Chen & C. Stevens (Eds.), *Proceedings of the international conference on music communication science* (pp. 56–59). Sydney, Australia: University of Western Sydney.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: theory and research* (pp. 309–337). Oxford: Oxford University Press.

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

Krumhansl, C., & Jusczyk, P. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73.

Lima, C. F., & Castro, S. L. (in press). Emotion recognition in music changes across the adult life span. *Cognition and Emotion.*

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal, 34,* 41–64.

Livingstone, S. R., & Thompson, W. F. (2009). The emergence of music from the Theory of Mind. *Musicae Scientiae, Special Issue, 2009–2010,* 83–115.

Luck, G., Toiviainen, P., Erkkila, J., Lartillot, O., Riikkila, K., Makela, A. . . . Varri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music, 36,* 25–45.

Madsen, C. (1998). Emotion versus tension in Haydn's Symphony No. 104 as measured by the two-dimensional Continuous Response Digital Interface. *Journal of Research in Music Education, 46,* 546–554.

McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Perception, 22,* 207–237.

Meyer, L. B. (1956). *Emotion and Meaning in Music.* Chicago, IL: University of Chicago Press.

Meyer, L. B. (1973). *Explaining Music: Essays and Explorations.* Berkeley, CA: University of California Press.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model.* Chicago, IL: University of Chicago Press.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin & S. Brown (Eds.), *The origins of music* (pp. 463–472). Cambridge, MA: MIT Press.

Neuhaus, C., Knösche, T., & Friederici, A. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience, 18,* 472–493.

Nielsen, F. (1983). *Oplevelse af misikalsk spending [The experience of musical tension*]. Copenhagen: Akademisk Forlag.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48,* 115–138.

Palmer, C., & Krumhansl, C. (1987). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics, 41,* 505–518.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.

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

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172.

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

Schenker, H. (1906/1954). *Harmony.* Cambridge, MA: MIT Press.

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

Schubert, E. (2001). Continuous Measurement of Self-report Emotional Response to Music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: theory and research* (pp. 393–414). Oxford: Oxford University Press.

Schubert, E. (2002). Correlation analysis of continuous emotional response to music: Correcting for the effects of serial correlation. *Musicae Scientiae, Special Issue 2001–2002,* 213–236.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception, 21,* 561–585.

Schubert, E. (2007). Locus of emotion: The effect of task order and age on emotion perceived and emotion felt in response to music. *Journal of Music Therapy, 44,* 344–368.

Schubert, E. (2010). Continuous self-report methods. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 223–253). Oxford: OUP.

Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research, 35,* 161–174.

Schulz, D., & Huston, J. (2002). The sliding window correlation procedure for detecting hidden correlations: Existence of behavioral subgroups illustrated with aged rats. *Journal of Neuroscience Methods, 121,* 129–137.

Sloboda, J. A., Lehmann, A. C., & Parncutt, R. (1997). Perceiving intended emotion in concert-standard performances of Chopin's Prelude No. 4 in E-minor. In *Proceedings of the Third Triennial ESCOM Conference* (pp. 629–634). Uppsala, Sweden: ICMPC.

Stevens, C. J., Schubert, E., Wang, S., Kroos, C., & Halovic, S. (2009). Moving with and without music: Scaling and lapsing in time in the performance of contemporary dance. *Music Perception, 26,* 451–464.

Strauss, J., & Strauss, J., II (1870). Pizzicato Polka, Op. 234 [Slovak Radio Symphony Orchestra, conducted by Ondrej Lenard]. On *Discover Classical Music* [CD]. Hong Kong: Naxos.

Thaut, M., & Davis, W. (1993). The influence of subject-selected versus experimenter-chosen music on affect, anxiety, and relaxation. *Journal of Music Therapy, 30,* 210–223.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33–58.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91,* 3540–3550.

Vines, B., Nuzzo, R., & Levitin, D. (2005). Analyzing temporal dynamics in music. *Music Perception, 23,* 137–152.

Appendix A

Pizzicato Polka Sub-Phrase Pairs Included in Repeated, Variation, and Contrasting Categories

Phrase pair #	Repeated ($n = 7$)	Variation ($n = 36$)	Contrasting ($n = 48$)
1	S1A1-S2A1	S1A1-S1A2	S1A1-S1B1
2	S1A2-S2A2	S1A1-S1A′1	S1A1-S1B2
3	S1B1-S2B1	S1A1-S1A′2	S1A1-S1B3
4	S1B2-S2B2	S1A1-S2A2	S1A1-S2B1
5	S1B3-S2B3	S1A1-S2A′1	S1A1-S2B2
6	S1A′1-S2A′1	S1A1-S2A′2	S1A1-S2B3
7	S1A′2-S2A′2	S1A2-S1A′1	S1A2-S1B1
8		S1A2-S1A′2	S1A2-S1B2
9		S1A2-S2A1	S1A2-S1B3
10		S1A2-S2A′1	S1A2-S2B1
11		S1A2-S2A′2	S1A2-S2B2
12		S1B1-S1B2	S1A2-S2B3
13		S1B1-S1B3	S1B1-S1A′1
14		S1B1-S2B2	S1B1-S1A′2
15		S1B1-S2B3	S1B1-S2A1
16		S1B2-S1B3	S1B1-S2A2
17		S1B2-S2B1	S1B1-S2A′1
18		S1B2-S2B3	S1B1-S2A′2
19		S1B3-S2B1	S1B2-S1A′1
20		S1B3-S2B2	S1B2-S1A′2
21		S1A′1-S1A′2	S1B2-S2A1
22		S1A′1-S2A1	S1B2-S2A2
23		S1A′1-S2A2	S1B2-S2A′1
24		S1A′1-S2A′2	S1B2-S2A′2
25		S1A′2-S2A1	S1B3-S1A′1
26		S1A′2-S2A2	S1B3-S1A′2
27		S1A′2-S2A′1	S1B3-S2A1
28		S2A1-S2A2	S1B3-S2A2
29		S2A1-S2A′1	S1B3-S2A′1
30		S2A1-S2A′2	S1B3-S2A′2
31		S2A2-S2A′1	S1A′1-S2B1
32		S2A2-S2A′2	S1A′1-S2B2
33		S2B1-S2B2	S1A′1-S2B3
34		S2B1-S2B3	S1A′2-S2B1
35		S2B2-S2B3	S1A′2-S2B2
36		S2A′1-S2A′2	S1A′2-S2B3
37			S2A1-S2B1
38			S2A1-S2B2
39			S2A1-S2B3
40			S2A2-S2B1
41			S2A2-S2B2
42			S2A2-S2B3
43			S2B1-S2A′1
44			S2B1-S2A′2
45			S2B2-S2A′1
46			S2B2-S2A′2
47			S2B3-S2A′1
48			S2B3-S2A′2

(*Appendices continue*)

Appendix B

Slavonic Dance Sub-Phrase Pairs Included in Repeated, Variation, and Contrasting Categories

Phrase pair #	Repeated ($n = 8$)	Variation ($n = 20$)	Contrasting ($n = 84$)
1	S1A1-S2A1	S1A1-S1A2	S1A1-S1B1
2	S1A2-S2A2	S1A1-S1A′1	S1A1-S1B2
3	S1B1-S2B1	S1A1-S2A2	S1A1-S1C1
4	S1B2-S2B2	S1A1-S2A′1	S1A1-S1C2
5	S1C1-S2C1	S1A2-S1A′1	S1A1-S1A′2
6	S1C2-S2C2	S1A2-S2A1	S1A1-S2B1
7	S1A′1-S2A′1	S1A2-S2A′1	S1A1-S2B2
8	S1A′2-S2A′2	S1B1-S1B2	S1A1-S2C1
9		S1B1-S2B2	S1A1-S2C2
10		S1B2-S2B1	S1A1-S2A′2
11		S1C1-S1C2	S1A2-S1B1
12		S1C1-S2C2	S1A2-S1B2
13		S1C2-S2C1	S1A2-S1C1
14		S1A′1-S2A1	S1A2-S1C2
15		S1A′1-S2A2	S1A2-S1A′2
16		S2A1-S2A2	S1A2-S2B1
17		S2A1-S2A′1	S1A2-S2B2
18		S2A2-S2A′1	S1A2-S2C1
19		S2B1-S2B2	S1A2-S2C2
20		S2C1-S2C2	S1A2-S2A′2
21			S1B1-S1C1
22			S1B1-S1C2
23			S1B1-S1A′1
24			S1B1-S2A1
25			S1B1-S2A2
26			S1B1-S2C1
27			S1B1-S2C2
28			S1B1-S2A′1
29			S1B2-S1C1
30			S1B2-S1C2
31			S1B2-S1A′1
32			S1B2-S2A1
33			S1B2-S2A2
34			S1B2-S2C1
35			S1B2-S2C2
36			S1B2-S2A′1
37			S1C1-S1A′1
38			S1C1-S1A′2
39			S1C1-S2A1
40			S1C1-S2A2
41			S1C1-S2B1
42			S1C1-S2B2
43			S1C1-S2A′1
44			S1C1-S2A′2
45			S1C2-S1A′1
46			S1C2-S1A′2
47			S1C2-S2A1
48			S1C2-S2A2
49			S1C2-S2B1
50			S1C2-S2B2
51			S1C2-S2A′1
52			S1C2-S2A′2
53			S1A′1-S1A′2
54			S1A′1-S2B1
55			S1A′1-S2B2
56			S1A′1-S2C1
57			S1A′1-S2C2
58			S1A′1-S2A′2
59			S1A′2-S2A1

(*Appendices continue*)

Appendix B (*continued*)

Phrase pair #	Repeated ($n = 8$)	Variation ($n = 20$)	Contrasting ($n = 84$)
60			S1A′2-S2A2
61			S1A′2-S2C1
62			S1A′2-S2C2
63			S1A′2-S2A′1
64			S2A1-S2B1
65			S2A1-S2B2
66			S2A1-S2C1
67			S2A1-S2C2
68			S2A1-S2A′2
69			S2A2-S2B1
70			S2A2-S2B2
71			S2A2-S2C1
72			S2A2-S2C2
73			S2A2-S2A′2
74			S2B1-S2C1
75			S2B1-S2C2
76			S2B1-S2A′1
77			S2B2-S2C1
78			S2B2-S2C2
79			S2B2-S2A′1
80			S2C1-S2A′1
81			S2C1-S2A′2
82			S2C2-S2A′1
83			S2C2-S2A′2
84			S2A′1-S2A′2

Received September 17, 2010
Revision received February 25, 2011
Accepted March 10, 2011 ■