International Symposium on Performance Science
ISBN 978-94-90306-01-4

Emotional arousal and the automatic detection of musical phrase boundaries

Steven R. Livingstone[1], Emery Schubert[2], Janeen D. Loehr[1], and Caroline Palmer[1]

[1] Department of Psychology, McGill University, Canada
[2] School of English, Media, and Performing Arts,
University of New South Wales, Australia

This study investigated correspondences between listeners' emotional responses to an orchestral work and the underlying phrase structure. An algorithm for the automatic detection of musical phrase boundaries was developed, based on listeners' continuous ratings of perceived emotion with a two-dimensional tool (valence and arousal). Rates of change in arousal (velocity) and change in change in arousal (acceleration) for each musical phrase were combined into a single metric. Similarity in listeners' responses for musically related phrases was higher than those for musically contrasting phrases. Based on listeners' responses to identical sections, a recursive algorithm identified all phrase boundaries. These findings indicate novel measures of change in listeners' emotional arousal that correspond with musical phrase structure.

Keywords: emotional response; arousal; phrase structure; automatic boundary detection; similarity metric

Emotional response to music is influenced by contributions of both the composition and performance (Kendall and Carterette 1990, Livingstone and Thompson 2009). Although significant research has investigated the role of musical cues such as tempo and articulation on emotional response (Gabrielsson and Lindstrom 2001), fewer studies have investigated the role of phrase structure. Recent work (Krumhansl 2002, Vines *et al.* 2005) found that listeners' tension ratings coincided with the phrase structure, suggesting that listeners' emotional responses could be used for the computational (automatic) detection of phrase structure (Cambouropoulos 2001).

We investigated listeners' emotional responses to an orchestral work to identify response patterns across phrases that varied in melodic and harmonic context. Two measures of emotional response were analysed: velocity (first derivative) and acceleration (second derivative) of arousal (amount of emotional activity). These instantaneous measures reduce effects of the preceding musical context. The velocity-acceleration arousal responses were compared with the theoretical phrase structure. Based on the recurrent nature of listeners' responses, a detection algorithm was developed to identify segment boundaries, which confirmed that listeners' emotional arousal reflects musical phrase structure.

METHOD

Participants and materials

Sixty-seven adult listeners with varied amounts of musical training (9 had no training, 19 had 1-10 years, 39 had 10+ years), most of whom were undergraduate music education students participated (further details in Schubert 1999). An orchestral recording of the *Pizzicato Polka*, Op.234 (J. and J. Strauss), length 2:37 mins, was used, which contains the following phrase structure: introduction (4 bars), phrase A (8 bars), B (12 bars), A' (8 bars), followed by a middle section (32 bars), and a repeat of the introduction, phrases A, B, and A', followed by a coda. The A, B, A' section that occurred twice within the composition was the focus of the analyses; the two occurrences are referred to as section 1 and section 2. Phrases A and A' had related melodic and harmonic content, whereas Phrase B was not related to A or A'. Sections 1 and 2 were further subdivided into 4-bar sub-phrases: a1, a2, b1, b2, a'1, a'2, as shown in Figure 1.

Procedure and analysis

Listeners continuously reported their emotional response using EmotionSpace Lab, a two-dimensional tool that records valence from positive (100) to negative (-100) and arousal (amount of emotional activity) from strong (100) to weak (-100), sampled at 1 Hz (Schubert 1999). Listeners' valence responses did not correspond to the phrase structure, and therefore we focus on arousal responses. Responses were smoothed with *functional data analysis* (Ramsay and Silverman 2005). Listeners' mean arousal responses and their velocity (first derivative) and acceleration (second derivative) values are shown in Figure 1. Phase-plane plots that combined velocity and acceleration were generated for each sub-phase (for an example, see inset of Figure 1). Similarities

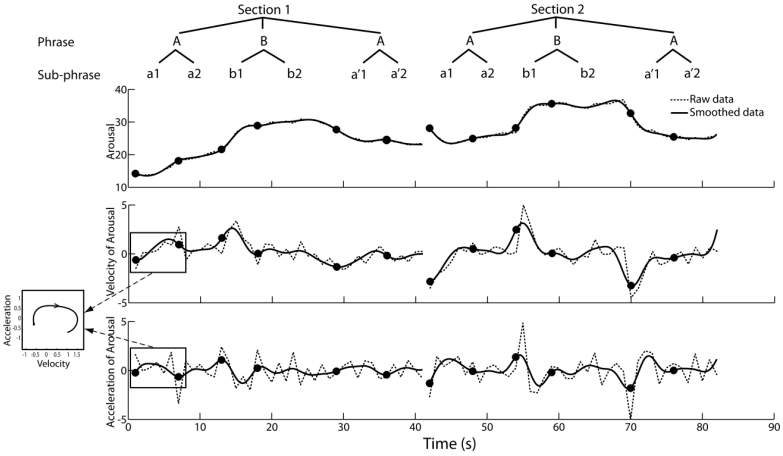

Figure 1. Phrase structure of sections 1 and 2 of the *Pizzicato Polka*, with listeners' mean arousal, velocity, and acceleration values. Theoretical phrase boundaries indicated as dots. Phase-plane plot for sub-phrase a1 in section 1 (inset) shows velocity and acceleration values and begins at the filled square.

between phase-plane plots, which reflect the 2-dimensional responses to arousal, were evaluated with procrustes analyses (2-dimensional measure of similarity).

Listeners' responses to different sub-phrases were divided into three groups: repetition, within-category, and between-category. The repetition group contained pairs of responses to the same sub-phrase from section 1 and section 2 (n=6). The within-category group compared all musically related sub-phrases (same letter, e.g. a1 to a2) in and across both sections, excluding repetition pairs (n=28). The between-category comparisons contained pairs of responses to musically unrelated sub-phrases (different letters, e.g., a1 to b1) in and across both sections (n=38).

RESULTS

Listeners' mean emotional arousal values across the entire section 1 were highly similar to those for section 2, which occurred later in the musical context (mean arousal: $r=0.84$, $p<0.01$, mean velocity: $r=0.70$, $p<0.01$, mean acceleration: $r=.69$, $p<0.01$). Next we measured response similarity among the 2D phase-plane plots, shown in Figure 2, at the level of individual sub-phrases. The procrustes similarity values for repetition pairs (Table 1, first column) indicated that listeners' emotional responses to individual sub-

Table 1. Procrustes similarity values for repetition, within-category, and between-category sub-phrase comparisons for original and new phrase structures. All p<0.01.

	Original phrase structure	*New phrase structure (with B3)*
Repetition	0.75	0.81
Within-category	0.69	0.75
Between-category	0.63	0.77

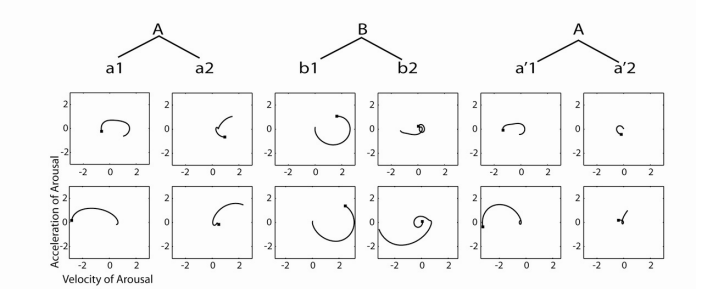

Figure 2. Phase plane plots for sub-phrases sections 1 (top row) and 2 (bottom) of the *Pizzicato Polka*. Sub-phrases begin at the filled squares and possess a cyclical structure.

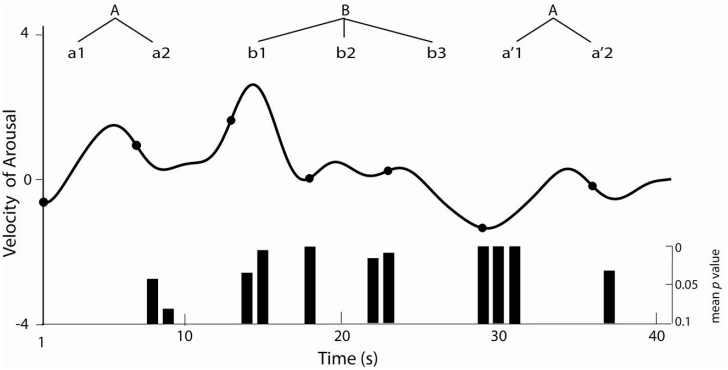

Figure 3. Automatic phrase detection outcomes aligned with listeners' mean arousal responses (velocity) for section 1. Filled circles indicate theoretical phrase boundaries. Bars indicate points in time at which a phrase boundary was algorithmically detected (taller bars=smaller p-values).

phrases in section 1 reheard in section 2 were highly similar. Furthermore, emotional responses for within-category sub-phrases were perceived as next most similar and responses to between-category sub-phrases as least similar.

The phase-plane plots in Figure 2 show a cyclical structure to listeners' emotional responses; each sub-phrase generated semi-elliptical responses that shifted clockwise around the midpoint. However, b2 sub-phrase responses possessed an unusual double swirl. These phrases were also considerably longer in duration than other sub-phrases (see Figure 1). Therefore, b2 sub-phrases were subdivided into 2 sub-phrases (b2 and b3) and procrustes analyses were recalculated for the new phrase structure (Table 1, second column). The similarity values based on the new phrase structure were higher than those based on the original phrase structure, suggesting that the new phrase structure is a more accurate representation of listeners' emotional responses to the *Pizzicato Polka*.

The high similarity in listeners' emotional responses to repeated phrases suggests that this information may benefit the automatic detection of phrase boundaries in this piece, which has a simple harmonic structure. A recursive algorithm was developed which takes as input the listeners' velocity responses for sections 1 and 2. Beginning with a window size of 2 s, responses for sections 1 and 2 were correlated and the window length grew iteratively (+1 s) up to the entire 41 s response. The resulting correlation function was used to identify peaks of most similarity, for comparison with locations of sub-phrase boundaries. Peaks in the correlation function were defined by two successive increasing values followed by two successive decreases. The algorithm was then reiterated once at each value, with the new window onset time equal to the sequential location of the peak. To evenly distribute the effect of sequence beginnings and endings on window lengths, the algorithm was run in both sequence directions (forward/backward) and the outcomes averaged. Figure 3 shows the p-values associated with each detected boundary. All six boundaries were correctly identified automatically within two seconds of the hypothesized phrase boundaries, and no false positives.

DISCUSSION

Listeners' emotional arousal responses to repeated musical sections were highly consistent. Similarity analyses revealed that the first and second derivatives of emotional arousal were more similar for repeated musical sub-phrases than for musically related or less related sub-phrases. These findings indicate that listeners had distinct arousal responses at the 4-bar sub-phrase level in the *Pizzicato Polka* and these patterns recurred during musical repetitions.

The phrase boundary detection algorithm used the similarity in listeners' responses to repeated musical sections to identify all phrase boundaries suc-

cessfully. The algorithm operates under the premise that listeners' responses to identical musical sections are most similar at phrase boundaries. Phrase boundaries are points of perceptual salience, and listeners' emotional responses may accelerate at these locations. Use of the first and second derivatives of emotional arousal helped to provide a more precise temporal marker of change in listeners' responses. This research demonstrates novel analytic techniques for measuring listeners' emotional responses and their correspondence with the underlying phrase structure of music.

Acknowledgments

This work was supported by the Canada Research Chairs program and NSERC Grant 298173 to the fourth author. We thank Erik Koopmans and Frances Spidle.

Address for correspondence

Steven Livingstone or Caroline Palmer, Department of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal H3A 1B1, Canada; *Email:* steven.livingstone@mcgill.ca or caroline.palmer@mcgill.ca

References

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

Gabrielsson A. and Lindstrom E. (2001). The influence of musical structure on emotional expression. In P.N. Juslin and J. S. Sloboda (eds.), *Music and Emotion*. Oxford: Oxford University Press.

Kendall R. A. and Carterette. E. C (1990). The communication of musical expression. *Music Perception, 8,* pp. 129-164.

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

Livingstone S. R. and Thompson W. F. (2009). The emergence of music from the theory of mind. *Musicae Scientiae, 10,* pp. 83-115.

Ramsay J. O. and Silverman B. W. (2005). *Functional Data Analysis*. New York: Springer.

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

Vines B.W., Nuzzo R. L., and Levitin D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics and functional data analysis techniques. *Music Perception, 23*, pp. 137-152.