Nonmusicians Express Emotions in Musical Productions Using Conventional Cues

Haley E. Kragness¹ and Laurel J. Trainor¹,²

Abstract
Expert musicians use a number of expressive cues to communicate specific emotions in musical performance. In turn, listeners readily identify the intended emotions. Previous studies of cue utilization have studied the performances of expert or highly trained musicians, limiting the generalizability of the results. Here, we use a musical self-pacing paradigm to investigate expressive cue use by non-expert individuals with varying levels of formal music training. Participants controlled the onset and offset of each chord in a musical sequence by repeatedly pressing and lifting a single key on a MIDI piano, controlling tempo and articulation. In addition, the velocity with which they pressed the key controlled the sound level (dynamics). Participants were asked to “perform” the music to express basic emotions that were (1) positively or negatively valenced and (2) high- or low-arousal (joy, sadness, peacefulness, and anger). Nonmusicians’ expressive cue use was consistent with patterns of cue use by professional musicians described in the literature. In a secondary analysis, we explored whether formal training affected how tempo, articulation, dynamics, rhythm, and phrasing were employed to express the target emotions. We observed that the patterns of cue use were strikingly consistent across groups with differing levels of formal musical training. Future work could investigate whether expertise is implicated in the expression of more complex emotions and/or in the expression of more complex musical structures, as well as explore the role of emotional intelligence and informal musical experiences in expressive performance.

Keywords
Emotion, expression, music, performance, training

Submission date: 18 June 2018; Acceptance date: 9 February 2019

Introduction
Music is pervasive in our everyday lives. One particularly compelling aspect of music is that it can be used as a non-verbal medium for emotional communication. Caregivers use affective songs and song-like speech in interactions with infants (e.g., Fernald et al., 1989; Ilari, 2003; Trehub & Trainor, 1998; Young, 2008) and adults commonly report using music to regulate their own emotions (Lonsdale & North, 2011). Surveys indicate that affective value is a primary motivation for listening to music (Juslin & Laukka, 2004; Sloboda & O’Neill, 2001). Likewise, musicians attest that communicating emotions is a central goal in their performances (Lindström, Juslin, Bresin, & Williamson, 2003).

It is widely accepted that there is a distinction between emotions that are induced by music (“felt emotions”) and emotions that are communicated by music (“perceived emotions”; Gabrielsson, 2001). A number of studies have found that listeners’ felt and perceived emotional responses to music can be different (for a review, see Schubert, 2013). For example, many report experiencing pleasant feelings when listening to sad music, despite being able to identify the music as sad-sounding (e.g., Garrido & Schubert, 2011; Huron, 2011; Kawakami, Furukawa, Katahira, & Okanoya, 2011).
2013). Even when the felt and perceived emotions are the same, listeners tend to rate perceived emotions as more intense than felt emotions (Schubert, 2013). Such contrasts suggest that musical emotions are not identified by simply reflecting on felt emotions; rather, listeners make an appraisal about performers’ intended emotions based on acoustic cues (Juslin, 1997; Juslin & Laukka, 2003).

In general, adults agree on the emotion communicated by musical excerpts (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Juslin & Laukka, 2003; Mohn, Argstatter, & Wilker, 2011; Vieillard et al., 2008). In some circumstances, more complex emotions such as longing, humor, and awe have also been identified by listeners (Huron, 2006; Laukka et al., 2013; Senju & Ohgushi, 1987). Such findings have led to questions about the emotional “code” used by composers, performers, and listeners in expressing and interpreting musical emotions. Amplitude envelope, rhythm, pitch contour, melodic range, mode, loudness, variations in loudness, articulation, pitch level, tempo, and timbre are among the musical cues that have been implicated in emotional communication (see Gabrielson & Lindström, 2010 and Juslin & Timmers, 2010 for reviews).

Cues for emotional expression are embedded to some extent in the score by the composer. For example, complex rhythms are often perceived to express joy or anger, and simple, regular rhythms to convey sadness or boredom (e.g., Scherer & Oshinsky, 1977; Thompson & Robbitalle, 1992). However, studies of performance using the standard content paradigm (Gabrielson & Juslin, 1996; Seashore, 1947) have demonstrated that performers also have control over communicating emotions. In such experiments, performers are asked to perform a prescribed score with the goal of communicating different emotions and listeners are asked to identify the intended emotion. The extent to which the intended and judged emotions match is considered a measure of the communicative success. Studies using this paradigm with opera singers, electric guitarists, violinists, trumpeters, percussionists, and pianists (among others) have found that both musically trained and untrained listeners can indeed decode expressive intent in such performances (Behrens & Green, 1993; Juslin, 1997; Juslin & Laukka, 2000; Juslin & Madison, 1999; Kotlyar & Morozov, 1976; Laukka & Gabrielson, 2000).

Researchers have considered how various cues relate to different dimensions of emotions. Russell’s (1980) circumplex model of emotions, which characterizes emotions on the dimensions of valence (negative to positive) and arousal (low to high), has been widely used in this regard. Performers seem to have particular control over cues that distinguish between high- and low-arousal emotions. For example, performances conveying high-arousal emotions, such as happiness and anger, typically incorporate fast tempi, low tempo variability, disconnected articulations, and high sound levels. In contrast, performances of low-arousal emotions (such as sadness and peacefulness) involve slow tempi with final ritardandi, high tempo variability, connected articulations, and low sound levels (Juslin & Timmers, 2010). These cues are redundant to some extent—it is not necessary to utilize all cues to communicate an emotion, but using multiple cues increases the chances of correct identification (Juslin & Laukka, 2000).

Though these studies have been very informative, it is interesting to note that “encoders” and “decoders” are usually mismatched—encoders must be musically experienced in order to perform the music, but decoders may have had no experience with formally producing music at all. It is not clear that formal music training discernably improves emotional decoding of professional performances (Bigand et al., 2005; Juslin, 1997), and these results are consistent with research demonstrating that day-to-day exposure to music confers a degree of sensitivity to many aspects of music, resulting in “musically experienced listeners” (for a review, see Bigand & Poulin-Charronrat, 2006). Nevertheless, there is reason to believe that the situation is different for musical production. While many societies have a participatory music culture in which non-experts are regularly active in music production and performance, Western culture is largely presentation, with production activities limited to highly trained individuals (Turino, 2008). In one study, Juslin and Laukka (2000) noted that a novice guitarist (one year of experience) communicated happiness, sadness, anger, and fear less accurately than more experienced guitarists (up to 15 years of experience). Experience was not explicitly investigated, however, and it is impossible to determine whether this difference in accuracy was due to differences in expressive intentions, differences in technical proficiency, or some other factor.

Technological advances have made it possible to examine musical production without using musical instruments. Several experiments have employed an apparatus that includes a series of sliders, each controlling a different musical feature. In the first such study, Bresin and Friberg (2011) allowed participants to systematically vary different musical features (including tempo, sound level, articulation, phrasing, register, timbre, and attack speed) to communicate happy, sad, peaceful, or scary expressions. They found that participants’ use of expressive cues to communicate each emotion were broadly consistent with those that have been found in previous studies, confirming that using a non-instrumental apparatus is sufficient for expression.
Although the apparatus did not require musical expertise to use, only musical experts were recruited.

Two additional studies have reported using a similar apparatus with participants who were not musically trained. Saarikallio, Vuokoski, and Luck (2014) examined adolescents’ emotional communication of happy, sad, and angry excerpts. While approximately half of the sample reported taking music lessons at some point, the extent of their formal musical training was not quantified. Sievers, Polansky, Casey, and Wheatley (2013) also used a slider apparatus to examine similarities between nonmusicians’ cross-modal expressions of emotion (music and movement) in different cultural contexts. Interestingly, they did not include a number of performer cues that are typically examined in performance analysis, such as articulation, rhythm, and loudness, presumably because these cues do not have clear analogues in the movement modality.

In the present study, we used a simple self-pacing apparatus to examine the use of expressive cues in musical production across non-expert undergraduate performers with different levels of formal music training. A similar apparatus has been used in previous self-pacing studies to examine expressive timing and musical phrase structure in nonmusicians and preschool children (Kragness & Trainor, 2016, 2018). In previous studies, participants controlled the onset of each successive chord in a prescribed musical sequence. In the present study, this setup was adapted such that participants additionally controlled the offset and sound level of each chord. Thus, they could control the timing and loudness of each chord in the sequence. We predicted that nonmusician participants would use similar patterns of expressive cues as expert participants in previous studies, based on shared representations of musical expression from years of music listening, and that expressive cues differentiating emotions would be enhanced in participants in our sample who had relatively high levels of training.

Method

Participants

This research was approved by the university’s research ethics board. Twenty-four undergraduates ($M_{age} = 18.65$ years, $SD = 1.22$ years, ranging from 18 to 22 years) were recruited through the participant pool. Participants were granted course credit as compensation. Because a pilot study ($N = 30$) conducted with a slightly different methodology and apparatus (using a computer keyboard instead of a MIDI keyboard) demonstrated large differences in participants’ tempo and articulation use across the target emotions used here, a sample size of 24 was deemed to be sufficient in the present study. First, participants completed a questionnaire that asked about handedness, as well as experiences with languages, music, and dance. They additionally filled out the self-report inventory portion of the Goldsmith Music Sophistication Index, v1.0 (“Gold-MSI,” Müllensiefen, Gingras, Stewart, & Musil, 2014). When asked to report “years of formal training on a musical instrument (including voice)” (question 36), participants’ selections included: 0 years ($n = 8$), 0.5 years ($n = 3$), 1 year ($n = 1$), 2 years ($n = 2$), 3–5 years ($n = 3$), 6–9 years ($n = 3$), and 10 or more years ($n = 4$).

Stimuli

We selected four excerpts from the chorales of J. S. Bach. Each excerpt contained three sub-phrases that were each eight chords in length, for a total of 24 chords, and began with an anacrusis (or “pick up” chord). Two of the excerpts (“Maj1” and “Maj2”) were originally composed in the major mode and two of the excerpts (“Min3” and “Min4”) were originally composed in the minor mode. Each excerpt was transposed to the key of F and small alterations were made to eliminate passing tones and ornamentations between each quarter-note-length chord. For each excerpt, a second version was created in the parallel major or minor mode (“Min1,” “Min2,” “Maj3,” and “Maj4”).² Thus, there were eight excerpts in total.

Each participant “performed” the emotions (joy, sadness, peacefulness, and anger) using four of the eight excerpts: two major and two minor (except one participant, who played only one major excerpt due to a technical error). Thus, each participant produced 16 “performances.” Of the two major and two minor excerpts, one was original and the other was modally altered. In a single experimental session, at no time did a participant perform both an excerpt and its modal alteration. Major and minor excerpts always alternated. The possible orders of excerpts can be seen in Table 1.

| Table 1. Order number. |
|------------------------|
| Excerpt 1 | Excerpt 2 | Excerpt 3 | Excerpt 4 |
| Order 1 | Maj1 | Min2 | Maj3 | Min4 |
| Order 2 | Maj2 | Min1 | Maj4 | Min3 |
| Order 3 | Min3 | Maj4 | Min1 | Maj2 |
| Order 4 | Min4 | Maj3 | Min2 | Maj1 |

Apparatus

During the experiment, each chord was generated online in the default piano timbre in Max MSP (version 5). The interface used by participants was an M-AUDIO Oxygen-49 MIDI keyboard, and the sounds were presented through a pair of external speakers located to the left and right of the participant in the sound booth (WestSun Jackson Sound, model JSI P63 SN 0005). Participants were seated in a chair in front of a monitor at a distance of approximately 3.5 feet. In front of the participant was a desk on which the MIDI keyboard sat.
Using the MIDI keyboard, participants could control the onset and the offset of each chord in succession by pressing and releasing the key. Each chord was sustained until the key was released, which terminated the chord. Only the middle C key elicited a chord; the other keys did not respond if pressed. The middle C key was marked with an orange sticker to remind participants which key to use. Additionally, participants could control the sound level of each chord by pressing with more or less velocity (greater velocity produced louder chords).

Procedure

Training phase. After completing the questionnaires, participants were trained to use the MIDI keyboard apparatus. They were informed that only the key marked with a sticker (middle C) would elicit a note or chord, and instructed to use the index finger of their dominant hand to “perform” music. They were told to continue to press the key for each chord until pressing the key elicited no sound, which was an indication that the excerpt was complete. They were told that they did not control what chord was being played, but when, how loudly, and for how long each chord was played. First, participants were asked to practice “performing” an excerpt, which was a C major scale (C4 to C5), without any instructions about how to play it. Next, they were asked to perform the same excerpt in six different ways: with short notes, with long notes, with loud notes, with soft notes, with fast notes, and with slow notes. This training session was done to give participants familiarity with the breadth of expressive options available to them during the testing phase.

Testing phase. Prior to the testing phase, we asked participants to jot down words or pictures to remind themselves of a time they felt each of the target emotions: joy, sadness, peacefulness, and anger. These emotions were selected because they each fall into a different quadrant of a two-dimensional valence-by-arousal circumplex (Russell, 1980). The purpose of this activity was not to act as an emotion “induction,” since, as previously noted, there is no evidence that “feeling” an emotion is necessary for communicating it. Rather, we wanted the participants to have a consistent reference point to consider for their performances throughout the experiment.

Next, there were four separate blocks, each containing a listening session, a mechanical performance, and four alternating practice and performance sessions (one practice and one performance session for each emotion). The experimenter informed the participant that they would be asked to perform each excerpt with the four different emotions, and that they should do their best to communicate each emotion, because a new set of participants would later be asked to guess their intended emotion. Then, the experimenter left the sound booth and they were guided for the rest of the experiment by instructions on the computer monitor.

In each block, participants first heard the excerpt that they would be performing. The excerpt was played in the default piano timbre at an inter-onset interval of 450 ms (approximately 133 onsets per minute). This rate was chosen because it is well within the range of adults’ spontaneous motor tempo and was therefore likely to be around the chosen tempi (Drake, Jones, & Baruch, 2000). After listening, participants were instructed via the monitor to play the excerpt “mechanically, without emotion.” Next, one of the four target emotions was displayed and they were instructed to practice performing the excerpt with that emotion one time. Finally, they were instructed to undertake the “performance” for that emotion. The practice and performance sessions were repeated for each target emotion. The listening, mechanical, and practice sessions were all included to give participants an opportunity to familiarize themselves with the chord sequences and plan their performances to the best of their ability. The order in which participants were asked to play the different emotions was chosen randomly for each participant for each block. Participants were instructed to exit the sound booth and take a short break when the screen indicated the block was complete. The experimenter manually started the next block when participants indicated they were ready to continue and had reentered the sound booth. The experimenter monitored the participants’ progress via a computer screen mirroring the instructions the participants were viewing, but the experimenter could not hear the performances inside the booth.

After the testing phase was complete, participants were given a questionnaire. They were asked to indicate how easy they found it to play each of the emotions (1 = very easy to 5 = very difficult), as well as the extent to which they felt the emotion while they were playing it (1 = not at all to 5 = very much). Finally, they were asked to report in a free response format whether they had used any strategies to try to convey different emotions.

Results

We investigated three expressive cues: tempo (time between consecutive onsets), key velocity (which was experienced as loudness), and articulation (for each interval between onsets, the proportion of that interval in which the chord was played). We additionally analyzed durational variability using the normalized pairwise variability index (nPVI; Grabe & Low, 2002). This measure of variability describes the degree of contrast between pairs of consecutive durations (e.g., Hannon, Lévêque, Nave, & Trehub, 2016, Huron & Ollen, 2003; Patel & Daniele, 2003; Quinto, Thompson, & Keating, 2013). Though rhythmic patterning is usually considered to be a compositional cue rather than expressive cue, in the current study participants could use rhythmic patterns to communicate emotions if they desired.

Because planned emotional performances were of particular interest, the “mechanical” trial and the practice trials
were excluded in the present analyses. For each cue, a separate $2 \times 2$ within-subjects analysis of variance (ANOVA) was performed with factors arousal (high and low) and valence (positive and negative). All participants experienced all four performance blocks, except for one participant who completed only three performance blocks due to technical malfunction. Archived data are available at the link provided in the Supplemental Material.

**Tempo**
Tempo was defined as the average number of onsets per minute. The ANOVA indicated a significant main effect of arousal ($F(1,23) = 81.620, p < .0001, \eta_g^2 = .459, M = 76.071$ beats per minute), demonstrating that across valence conditions, high-arousal emotions were played at a higher rate of onsets per minute than low-arousal emotions (Figure 1(a)). No significant main effect of valence ($F(1,23) = 4.064, p = .056, \eta_g^2 = .015$) or interaction effect ($F(1,23) = 1.287, p = .268, \eta_g^2 = .009$) were observed.

**Rhythmic variation**
Normalized pairwise variability was calculated using the formula developed by Grabe and Low (2002) for speech analysis and subsequently used in the context of music by Patel and Daniele (2003):

$$nPVI = 100 \times \left[ \sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{2} \right] / (m - 1)$$

where $m$ is the number of intervals in the excerpt and $d$ is the duration of the $k$th interval. The nPVI values range from 0 to 200, such that 0 indicates perfect isochrony (no durational contrast between pairs at all) and 200 represents

![Figure 1. Use of expressive cues for each emotion. (a) Average tempo used to convey each emotion. Higher values indicate more onsets per minute and faster tempi. (b) Average nPVI used to convey each emotion. Higher values indicate more pairwise durational contrast. (c) Average velocity used to convey each emotion. Higher values indicate greater velocity (and higher sound level). (d) Average articulation used to convey each emotion. Higher values indicate more connected chords.](image)
maximal durational contrast (Figure 1(b)). We observed no significant main effects for arousal ($F(1,23) = 2.940$, $p = .10$, $\eta_G^2 = .015$) or valence ($F(1,23) = 0.099$, $p = .757$, $\eta_G^2 = .0005$), but there was a significant interaction ($F(1,23) = 6.238$, $p = .020$, $\eta_G^2 = .020$). Further analysis revealed that the significant interaction was driven by greater nPVI for low-arousal than high-arousal emotion in the negative valence condition ($t(23) = -2.964$, $p = .007$, 95% CI = $-17.352$ to $-3.088$), but not in the positive valence condition ($t(23) = 0.170$, $p = .867$, 95% CI = $-0.691$ to $8.137$).

**Sound level (key velocity)**

Sound level was measured by the velocity with which the chord was pressed (recorded as MIDI velocity) on a scale of 0 (minimum velocity) to 125 (maximum velocity). The ANOVA revealed main effects of both valence ($F(1,23) = 62.970$, $p < .0001$, $\eta_G^2 = .141$) and arousal ($F(1,23) = 137.68$, $p < .0001$, $\eta_G^2 = .597$), as well as a significant interaction ($F(1,23) = 45.621$, $p < .0001$, $\eta_G^2 = .142$). Further analyses demonstrated that the significant interaction was driven by the fact that high-arousal emotions were played with greater velocity than low-arousal emotions overall, and that while anger was played with significantly greater velocity (higher sound level) than joy ($t(23) = 8.456$, $p < .0001$, 95% CI = $18.803$–$30.981$), there was no significant difference in the velocity participants used for sadness and peace ($t(23) = -0.033$, $p = .974$, 95% CI = $-3.674$ to $3.558$).

**Articulation**

Articulation was considered to be the proportion of the inter-onset interval in which the chord was played (e.g., if a chord was played for 300 ms and the next onset was initiated 300 ms later, the articulation value would be .5; if the next onset was initiated 900 ms later, the articulation value would be .25). The ANOVA revealed main effects of both valence ($F(1,23) = 19.457$, $p < .001$, $\eta_G^2 = .042$) and arousal ($F(1,23) = 85.021$, $p < .0001$, $\eta_G^2 = .358$), such that low-arousal emotions were played with more connected chords than high-arousal emotions, and negatively valenced emotions were played with more connected chords than positively valenced emotions (Figure 1(d)). No significant interaction was observed ($F(1,23) = 2.584$, $p = .121$, $\eta_G^2 = .013$).

**Correlations between expressive cues**

To examine relationships between expressive cues, Spearman’s rho was calculated for each pair of expressive cues (Table 2). Because six correlations were examined, the Bonferroni-corrected significance cut-off of $p = .008$ was used (.05/6). Tempo was correlated with sound level and articulation, such that faster tempi were correlated with greater sound level ($r_S = .697$, $p < .0001$) and less connected notes ($r_S = -.791$, $p < .0001$). Articulation was also significantly correlated with loudness ($r_S = -.526$, $p < .0001$), such that more connected notes (higher articulation values) were played more softly (lower sound levels). No correlations with nPVI were observed at the Bonferroni-corrected significance threshold (tempo, $r_S = -.225$, $p = .028$; articulation, $r_S = .239$, $p = .019$; sound level, $r_S = -.090$, $p = .385$).

### Table 2. Correlation coefficients between expressive cues.

|                 | Tempo | nPVI | Sound level |
|-----------------|-------|------|-------------|
| nPVI            | -.225 | –    | –           |
| Sound level     | .697* | -.090| –           |
| Articulation    | -.791*| .239 | -.526*      |

Note. Correlation coefficients represent Spearman's rho. Asterisks indicate statistical significance: *$p < .008$, †$p < .05$.

**The role of musical training**

Although all participants were undergraduates and non-experts musically, they represented a wide range of musical experiences, as revealed by their responses to the self-report inventory of the Gold-MSI. Specifically, eight participants reported no formal music training at all (0 years), while others reported upwards of 10 years of training. In order to examine whether those with formal music training used cues differently from those with no training, we separated participants into three equal-sized groups based on their scores on the Formal Musical Training subscale of the Gold-MSI (see Table 3), which combines information about years of formal music lessons, music practice, music theory training, and more. Participants in the “no training” group reported 0 years of formal lessons ($n = 8$), participants in the “low training” group reported 0.5–9 years of formal lessons (mean = 2.25 years, $n = 8$), and participants in the “moderate/high training” group reported 3–10 and 10 or more years of formal lessons (mean = 7.56 years or higher, $n = 8$). One-way ANOVAs performed on participants’ scores on the Active Musical Engagement and Sophisticated Emotional Engagement subscales of the Gold-MSI suggested that the groups did not significantly differ on these measures ($F(2,21) = 1.456$, $p = .256$ and $F(2,21) = 0.783$, $p = .470$, respectively).

Next, all of the previous ANOVAs were rerun with the additional between-subjects variable of musical training (none, low, moderate/high). However, there were no significant main effects or interactions with musical training in any condition (see Supplemental Material for ANOVA tables).

**Phrasing.** Although no differences were found between levels of music training for any of the expressive cues above, it was possible that differences would emerge in analyses of
longer-term sequential aspects of the performances, such as occur at the phrase level. This idea was motivated by previous findings that musicians are more sensitive to longer-term musical structures than nonmusicians (Chiappe & Schmuckler, 1997; Drake, Penel, & Bigand, 2000). To examine the degree of expressive speeding and slowing in phrases, a quadratic equation was fit to the inter-onset intervals of each eight-chord phrase. Because the end of each trial was initiated by the offset of the final chord, no inter-onset interval was available for the final chord. Thus, the final phrase (chords 17–24) of each excerpt was discarded for this analysis. The quadratic coefficient (curvature) was compared across emotions. A larger quadratic coefficient indicates greater curvature (and thus more exaggerated demarcation of phrasing by time variation). An ANOVA with within-subject factors arousal and valence and between-subject factor music training found no significant effects or interactions (Figure 2(a)), though it is worth noting that the main effect of arousal approached significance ($F(1,21) = 3.927, p = .061, \eta^2_G = .064$) and may merit future investigation.

To examine whether participants used timing variation at the phrase level as part of their expressive performances at all, the quadratic coefficients across all performance conditions were collapsed and compared to the quadratic coefficients in mechanical conditions. An ANOVA with within-subject factor expressive condition (mechanical or performance) and between-subject factor music training (Figure 2(b)) revealed no significant interaction ($p = .295$) or main effect of training ($p = .836$), but there was a main effect of expressive condition, such that participants’ phrasing was more exaggerated for performance excerpts than mechanical excerpts ($F(1,21) = 21.669, p = .0001, \eta^2_G = .261$).

**Table 3.** Participant demographics (Gold-MSI self-report questionnaire, percentile scores).

| No training | Low training | Moderate/high training |
|-------------|--------------|------------------------|
| Age         | 18.83 years (1.60) | 18.40 years (0.89) | 18.67 years (1.21) |
| Gender      | 5 F, 1 M     | 5 F, 0 M   | 5 F, 1 M  |
| Active musical engagement | 23.63 (18.14) | 28.81 (11.02) | 41.06 (23.52) |
| Perceptual abilities | 18.06 (14.55) | 26.63 (20.30) | 37.44 (33.82) |
| Musical training | 13.38 (7.70) | 42.88 (10.85) | 68.44 (7.42) |
| Singing abilities | 26.25 (11.99) | 26.63 (19.76) | 53.00 (33.60) |
| Sophisticated emotional engagement | 30.44 (19.75) | 44.38 (29.18) | 46.56 (28.09) |
| General musical sophistication | 17.19 (8.08) | 31.13 (17.49) | 52.44 (19.27) |

**Figure 2.** Participants’ use of phrasing, operationalized by the curvature of a quadratic fit to participants’ inter-onset intervals across each eight-chord phrase. Higher values indicate greater curvature, and more exaggerating phrasing. (a) Average phrasing used to convey each emotion. (b) Average phrasing across all performance trials (collapsed across emotion) compared to phrasing in mechanical trials.

**Associations between practice and performance.** Because we were primarily interested in examining participants’ planned output, only performances were included in the previous analyses. As a post-hoc exploratory analysis, we
examined participants’ degree of consistency for each expressive cue between practice trial and subsequent performance trial pairs. Pearson’s $r$ was calculated for the practice–performance pairs for each expressive cue. Each participant performed four excerpts four times, resulting in 16 pairs of trials per participant, except for one participant who had performed only three excerpts, and thus had 12 pairs of trials.

In general, participants were highly consistent. One participant had correlations that were greater than two standard deviations away from the mean for each expressive cue, and was therefore excluded in the analyses (see archived data for details).

The correlation data for each expressive cue (tempo, nPVI, sound level, and articulation) were submitted to an ANOVA with music training as a between-subjects factor and expressive cue as a within-subjects factor (see Figure 3). No significant main effect of music training was observed ($F(2,20) = 0.089, p = .968$), but there was a significant main effect of expressive cue ($F(3,60) = 32.070, p < .0001, \eta_G^2 = .516$), as well as a significant interaction ($F(6,60) = 3.035, p = .012, \eta_G^2 = .168$). The interaction was not significant after a Greenhouse–Geiser correction due to violations of sphericity ($F(2.492, 24.918) = 3.035, p = .056$), but the main effect of expressive cue remained significant ($F(1.246, 24.918) = 32.070, p < .0001$).

A Dunnett’s T3 test of pairwise comparisons found that participants’ consistency in using nPVI was lower than all other expressive cues (all adjusted $p < .01$), and that participants’ consistency in using articulation was lower than consistency in tempo and sound level (all adjusted $p < .001$).

**Ratings**

After participants completed all four performance blocks of the experiment, they were asked to rate the difficulty of conveying each emotion (Figure 4(a)). Because a Shapiro–Wilks test indicated that the data were non-normally distributed ($W = .852, p < .0001$), a Friedman rank sum test was used, revealing significant differences in difficulty ratings among emotions (Friedman chi-squared = 8.521, $N = 24, p = .036$). However, post-hoc pairwise comparisons using a Nemenyi multiple comparison test revealed no significant differences between emotions.

Participants were also asked to rate the extent to which they felt the intended emotion while playing it (Figure 4(b)). Because the data were non-normally distributed ($W = .874, p < .0001$), a Friedman rank sum test was used to evaluate whether there were differences in difficulty ratings. The Friedman ranked sum test did not reveal any significant differences (Friedman chi-squared = 7.723, $p = .065$).

**Discussion**

Although many studies have investigated how performers use expressive cues to communicate emotions in music, few previous experiments have examined nonmusicians’ expressive productions. This study demonstrated that those with little to no musical training use timing and loudness cues to differentiate musical emotions in a production setting. Moreover, musically untrained participants used the available cues in ways that were nearly identical to those used by musically trained individuals in our sample. Consistent with previous studies using highly trained musicians, differences in expression were most strongly associated with differences in the arousal of the intended emotion, such that performances of joy and anger were played with a faster tempo, more loudly, and with more disconnected chords than peacefulness and sadness (e.g., Gabrielsson & Juslin, 1996; Juslin, 1997). Valence was also represented to a lesser extent - anger was played more loudly than joy, and within each arousal level (anger vs. joy; sadness vs. peacefulness) negatively valenced emotions were played with more connected articulation than positively valenced emotions.

Finally, durational contrast as measured by the nPVI was used to portray sadness more than other emotions. This is broadly consistent with Quinto, Thompson, and Keating’s (2013) previous finding that high-level musicians used greater nPVI values in brief compositions intended to portray sadness than other emotions, though this contrast was not significant in their analysis. The finding is somewhat inconsistent, however, with previous reports that rhythms with durational contrast are often perceived to convey positive emotions (Keller & Schubert, 2011; Thompson & Robitaille, 1992). One possible explanation for this apparent inconsistency is that the chord sequences were drawn from chorales by J. S. Bach in which the chords are primarily equally spaced. Also, the task itself tended to encourage an isochronous interpretation, so rhythmic...
opportunities for the production of different categories of note length (e.g., eighth notes and quarter notes, the latter being twice as long as the former) were constrained. If this is the case, larger nPVI values may have been observed in sadness and peace than in anger and joy simply because participants intended to maintain isochrony and found it more difficult to do so when the overall tempo was slow than when it was fast. This interpretation is tentatively supported by the observation that participants were less consistent in their use of nPVI between practice and performance trials compared to their consistency when using other expressive cues. Overall, the role of rhythm in emotional communication has received little attention, and warrants future investigation.

Although this is among the first studies to examine non-expert participants, multiple levels of music training were represented. Thus, as a secondary analysis, we explored whether participants with formal training used cues in a different way from musically untrained participants. The profile of expressive cues used by those with no formal training was strikingly similar to those with training, in pattern and in magnitude. This result could have been obtained if participants simply used the maximum and minimum values of each cue (e.g., playing anger and joy as loudly as possible and sadness and peace as quietly as possible). However, the consistent use of intermediate values (especially for loudness and articulation) suggests that similarities across training levels were not simply due to use of extreme values. We further considered whether musically trained participants differed in a more global measure of expression by comparing the extent of participants’ timing variation at the phrase level across emotions and training. Overall, participants incorporated more exaggerated timing variation in their phrasing in performance trials than mechanical trials, but there was no evidence that this effect was strengthened by musical training, nor modulated by arousal and valence.

Interestingly, a previous observational study of music lessons found that instructors spend surprisingly little time on expressive instruction, tending to focus more on technique (Karlsson & Juslin, 2008), despite the widespread belief that expressivity is central to music performance (Juslin & Laukka, 2004; Lindström et al., 2003). It has been proposed that this is because teachers often conceptualize expressivity as instinctual and difficult to verbally communicate (Hoffren, 1964; Lindström et al., 2003). This lack of expressive instruction is especially interesting considering that feedback from music teachers can improve emotional communication accuracy (Juslin, Karlsson, Lindström, Friberg, & Schoonderwaldt, 2006). In combination with the present findings, these studies suggest that formal training can enhance expression of musical emotions, but that formal music curricula likely do not utilize expression instruction to its fullest potential.

The present results are consistent with previous perceptual studies that have found that nonmusicians are equally adept as musicians at identifying emotions in music (Bigand et al., 2005; Juslin, 1997). In the present study, participants mainly controlled performer cues (timing and loudness), rather than composer cues, such as mode. Indeed, performer cues are typically found to be the most consistently decoded across ages and cultures (Dalla Bella et al., 2001; Quinto, Thompson, & Taylor, 2013; Thompson & Balkwill, 2010). Whether awareness of these cues rests on a “universal affect code” shared across domains including vocal expression (e.g., Juslin, 1997, 2001, 2013; Juslin & Laukka, 2003) or other mechanisms such as cultural transmission is unknown, but the present study shows that nonmusicians use performance cues in production tasks as well as in perceptual tasks.
We asked participants to report any strategies that they implemented while completing the task using a free response format (see Supplemental Material). Though no statistical analyses were conducted on these responses, they offer several interesting insights. First, it is clear that regardless of level of training, participants were often explicitly aware of the cues that they intended to use. For example, S22 (moderate/high training) reported, “I sustained the notes more to match the peaceful and sad emotion. I cut the notes short for angry,” and S09 (no training) wrote, “sad = slower/louder, angry = faster/louder, peaceful = slow/quiet, happy = loud/fast.” A number of strategies were reported, including “imagining movie scenes (as well as their soundtracks) to fit with each emotion,” and “thinking about what music provokes these emotions.” Though participants’ free responses were not explicitly analyzed, they suggest that participants have some capacity to introspect about expressive cues and emotional musical production. Exploring the relationship between participants’ introspections and their expressive productions would be an interesting future direction.

Many participants expressed that one primary strategy was to reflect on memories that incorporated the target emotions, as they were instructed. This is consistent with past studies with musicians, many of whom reported that they believed that feeling the intended emotion is important for expressive performance (Lindström et al., 2003) and that they used recalling emotional memories as one strategy for inducing the intended mood (Persson, 2001). Interestingly, although differences in “difficulty” and “feeling” ratings were not significant, participants indicated greater mood induction for joy and sadness, which were also reported to be the easiest to communicate. Peace was the least-felt of the four target emotions, and also rated as the most difficult to communicate. This pattern suggests that there may indeed have been an association between feeling and communicating emotions in this context, although whether feeling the emotion resulted in more accurate communication is unknown.

The present results should not be taken as evidence that music training has no effect on expressivity. It is important to note that the target emotions each belonged to different quadrants of the two-factor model (Russell, 1980) and had been previously observed to elicit relatively high agreement across listeners in emotion recognition tasks (e.g., Gabrielsson & Juslin, 2003; Juslin & Laukka, 2003; Vieillard et al., 2008). Mixed emotions and aesthetic emotions (such as longing, love, awe, and humor), which generally have lower levels of agreement, are likely to be more difficult to express. Future studies could investigate whether music training alters expressive cues in the context of complex emotions. It is possible that there were features that differed across emotions or participants that were not captured by the measures included in the present analyses. Archived data are openly available for further exploratory analyses at a link available in the Supplemental Material.

Additionally, though our “no training” participants had no formal lessons at all, a post-hoc one-way ANOVA did not offer any evidence for differences between the three formal training groups in the Active Musical Engagement subscale of the Gold-MSI. Given that expressivity is not typically a focus in music lessons, it is possible that active engagement with music is more important for expressive production than formal training. The participants also did not differ significantly on their self-reported emotional engagement in music. However, one previous study has reported that people who score higher on a measure of emotional intelligence are better at recognizing emotions in music (Rescinow, Salovey, & Repp, 2004). Recent work has shifted toward more sophisticated and multidimensional conceptions of musical experience using subscales and composite measures of overall engagement (e.g., Chin & Rickard, 2012; Müllensiefen et al., 2014). Informal musicianship and emotional engagement with music may be important contributors to expressive production, although they were not investigated in the present study.

Overall, the present study demonstrated that participants with a variety of musical backgrounds can use the self-pacing apparatus to express emotions in music. Future work could use similar paradigms to investigate questions of expressive timing in a variety of other populations. The self-pacing method provides a complementary technique to the slider apparatus described in previous studies (Bresin & Friberg, 2011; Saarikallio et al., 2014; Sievers et al., 2013). Though the slider apparatus is relatively simple to use, Sievers and colleagues (2013) reported that participants in rural Cambodia found the continuous sliders to be uncomfortable to use, leading to decision paralysis. The self-pacing apparatus emulates many of the properties of instruments used world-wide, allowing intuitive control of onsets, offsets, and loudness. Using our self-pacing apparatus, it would be possible to examine amateur and expert performers’ expressive tendencies while performing either music congruent with their own musical system or music of a foreign musical system, which could illuminate questions about the universality of expressive cues for basic emotions. Similarly, whether the slider technique would be conducive to testing young children’s musical expression is an open question, but the self-pacing apparatus can be used with children as young as three years old. Previous studies examining children’s expressive productions have been limited to analyzing singing (Adachi & Trehub, 2000; Adachi, Trehub, & Abe, 2004), and we are presently using this apparatus to investigate children’s expressive musical production.

**Author contribution**

HEK and LJT jointly conceived and designed the study. HEK was responsible for participant recruitment, data analysis, and drafting the manuscript. Both authors revised the manuscript and approved the final version.
Acknowledgments
The authors thank Dr. Matthew Woolhouse for his guidance in stimulus creation and Dave Thompson for technical assistance. We additionally thank Farriyan Hossain and Mrinalini Sharma for their assistance running participants.

Declaration of conflicting interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grants to LJT from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR).

ORCID iD
Haley E. Kragness https://orcid.org/0000-0003-3137-7655

Supplemental material
Supplemental material for this article is available online.

Notes
1. As noted by Juslin (2013), “expression” in this context does not suggest that performers are necessarily experiencing the emotion they intend to convey, but rather that they are performing a symbolic representation of that emotion (as would an actor in a play).
2. Alterations and transpositions were done with the support and advice of an assistant professor of music theory with significant experience in Baroque-era harmony.
3. Years of training were determined based on participants’ responses to question 36 of the self-report questionnaire of the Gold-MSI. In instances in which a response represented a range of values (e.g., “4–6 years”), individuals’ response was coded as the middle value of the range to compute the average years of training for each group (i.e., “4–6 years” was coded as 5 years). The highest possible response, “10 or more” years, was coded as 10 years. Because of this, the actual average years of training for the “moderate/high training” group may be higher than 7.56.

Peer review
Jonna Vuoskoski, University of Oslo, Department of Musicology & Department of Psychology.
Jean-Julien Aucouturier, IRCAM/CNRS/Sorbonne-Université, Science & Technology of Music and Sound Lab.

References
Adachi, M., & Trehub, S. E. (2000). Decoding the expressive intentions in children’s songs. *Music Perception: An Interdisciplinary Journal, 18*, 213–224. doi:10.2307/40285909
Adachi, M., Trehub, S. E., & Abe, J. (2004). Perceiving emotion in children’s songs across age and culture. *Japanese Psychological Research, 46*, 322–336. doi:10.1111/j.1468-5584.2004.00264.x
Balkwill, L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*, 43–64. doi:10.2307/40285811
Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46*, 337–349. doi:10.1111/j.1468-5584.2004.00265.x
Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*, 20–33. doi:10.1177/030573569302100102
Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*, 100–130. doi:10.1016/j.cognition.2005.11.007
Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Daquet, A. (2005). Multidimensional scaling of emotional response to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*, 1113–1139. doi:10.1080/02699930500204250
Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex, 47*, 1068–1081. doi:10.1016/j.cortex.2011.05.009
Chappe, P., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review, 4*, 254–259. doi:10.3758/BF03209402
Chin, T., & Rickard, N. S. (2012). The music use (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception: An Interdisciplinary Journal, 29*, 429–446. doi:10.1080/02699930500204250
Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. *Tutorial in Quantitative Methods for Psychology, 1*, 4–45. doi:10.20982/tqmp.01.1.p042
Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion, 12*, 399–413. doi:10.1007/bf00992362
Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*, 1584–1595. doi:10.1016/s0010-0277(00)00136-0
Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music, 18*, 87–98. doi:10.1177/03057356901801007
Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attention, referent period, focal attending. *Cognition, 77*, 251–288. doi:10.1016/s0010-0277(00)00106-2
Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception: An Interdisciplinary Journal, 18*, 1–23. doi:10.2307/40285899
Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *Journal of Child Language*, 16, 477–501. doi:10.1017/S0305000900010679

Fritz, T., Jentschke, S., Gossellin, N., Sammler, D., Perezit, I., Turner, R., & ... Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19, 573–576. doi:10.1016/j.cub.2009.02.058

Gabrielsson, A. (2001). Perceived emotion and felt emotion: Same or different? *Musicae Scientiae*, 5, 123–148. doi:10.1177/10298649020050s105

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer’s intention and the listeners’ experience. *Psychology of Music*, 24, 68–91. doi:10.1080/0305735696241007

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In K. R. J. Davidson (Ed.), *Handbook of affective sciences* (pp. 503–534). New York, NY: Oxford University Press.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford: Oxford University Press.

Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception: An Interdisciplinary Journal*, 28, 279–296. doi:10.1525/mp.2011.28.3.279

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gustenhouse & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter.

Hannon, E. E., Lévêque, Y., Nave, K. M., & Trehub, S. E. (2016). Exaggeration of language-specific rhythms in English and French children’s songs. *Frontiers in Psychology*, 7, 939. doi:10.3389/fpsyg.2016.00939

Hoffren, J. (1964). A test of musical expression. *Bulletin of the Council for Research in Music Education*, 2, 32–35.

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

Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae*, 15, 146–158. doi:10.1177/1029864911401171

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception: An Interdisciplinary Journal*, 21, 267–271. doi:10.1080/102986491101300202

Ilari, B. (2003). Singing and listening to music with infants: An interview study with Canadian mothers. *Abstracts of the 5th Triennial ESCOM Conference*, 364.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception: An Interdisciplinary Journal*, 14, 383–418. doi:10.2307/40285731

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). New York, NY: Oxford University Press.

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00596

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientiae*, 4, 151–183. doi:10.1177/10298649000400202

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33, 217–238. doi:10.1080/0929821042000317813

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, 17, 197–221. doi:10.2307/40285891

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 453–489). Oxford: Oxford University Press.

Juslin, P., Karlsson, J., Lindstrom, E., Friberg, A., & Schoonderwaldt, E. (2006). Play it again with feeling: Computer feedback in musical communication of emotions. *Journal of Experimental Psychology: Applied*, 12, 79–95. doi:10.1037/1076-898x.12.2.79

Karlsson, J., & Juslin, P. N. (2008). Musical expression: An observational study of instrumental teaching. *Psychology of Music*, 36, 309–334. doi:10.1080/0305735607086040

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, 8, 189–201. doi:10.2307/40285296

Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00311

Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology*, 7, 142–156. doi:10.2478/v10053-008-0094-0

Kotlyar, G. M., & Morozov, V. P. (1976). Acoustic correlates of the emotional content of vocalized speech. *Soviet Physics Acoustics*, 22, 208–211.

Kragness, H. E., & Trainor, L. J. (2016). Listeners lengthen phrase boundaries in self-paced music. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1676–1686. doi:10.1037/xhp0000245
Kragness, H. E., & Trainor, L. J. (2018). Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues. *Developmental Psychology, 54*, 842–856. doi:10.1037/dev0000405

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion, 13*, 434–449. doi:10.1037/a0031388

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music, 28*, 181–189. doi:10.1073/p0035735600282007

Lindström, E., Justlin, P., N., Bresin, R., & Williamson, A. (2003). “Expressivity comes from within your soul”: A questionnaire study of music students’ perspectives on expressivity. *Research Studies in Music Education, 20*, 23–47. doi:10.1177/1321103X030200010201

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology, 102*, 108–134. doi:10.1348/000712610X506831

Mohn, C., Argstatter, H., & Wilker, F. (2011). Perception of six basic emotions in music. *Psychology of Music, 39*, 503–517. doi:10.1073/p0035735610378183

Mote, J. (2011). The effects of tempo and familiarity on children’s affective interpretation of music. *Emotion, 11*, 618–622. doi:10.1037/a0022573

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. (2014). The musicality of non-musicians: An index for measuring musical sophistication in the general population. *PLoS One, 9*, e89642. doi:10.1371/journal.pone.0089642

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music, 31*, 75–92. doi:10.1073/p00305735603031001325

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*, B35–B45. doi:10.1016/s0010-0277(02)00187-7

Persson, R. (2001). The subjective world of the performer. In P. N. Justlin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 275–289). Oxford: Oxford University Press.

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology, 4*, 184. doi:10.3389/fpsyg.2013.00184

Quinto, L., Thompson, W. F., & Taylor, A. (2013). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music, 42*, 503–524. doi:10.1073/p00305735613482023

Rescino, J. E., Salovey, P., & Repp, B. H. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception, 22*, 145–158. doi:10.1525/mp.2004.22.1.145

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178. doi:10.1037/h0077714

Saarikko, S., Vuokskoski, J., & Luck, G. (2014). Adolescents’ expression and perception of emotion in music reflects their broader abilities of emotional communication. *Psychology of Well-Being, 4*, doi:10.1186/s13612-014-0021-8

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotional attribution from auditory stimuli. *Motivation and Emotion, 1*, 331–346. doi:10.1007/bf00992539

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology, 4*, doi:10.3389/fpsyg.2013.00837

Seashore, C. E. (1947). In search of beauty in music: A scientific approach to musical aesthetics. Westport, CT: Greenwood Press.

Senju, M., & Ohgushi, K. (1987). How are the player’s ideas conveyed to the audience? *Music Perception, 4*, 311–323. doi:10.2307/4028577

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences, 110*, 70–75. doi:10.1073/pnas.1209023110

Sloboda, J. A., & O’Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Justlin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 415–429). Oxford: Oxford University Press.

Terwogt, M. M., & Van Grinsven, F. (1988). Recognition of emotions in music by children and adults. *Perceptual and Motor Skills, 67*, 697–698. doi:10.2466/pms.1988.67.3.697

Thompson, W. F., & Balkwill, L. L. (2010). Cross-cultural similarities and differences. In P. N. Justlin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 755–788). Oxford: Oxford University Press.

Thompson, W. F., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts, 10*, 79–89. doi:10.2190/nbny-awdb-gw58-mtel

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research, 12*, 43–77.

Turino, T. (2008). *Music as social life: The politics of participation*. London: University of Chicago Press.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical emotions in music performance an aspect of emotional intelligence? *Research Studies in Music Education, 26*, 33–46. doi:10.1177/0255761407085648