Received: 22 August 2019 | Revised: 15 April 2020 | Accepted: 17 April 2020

DOI: 10.1111/desc.12982

PAPER

Developmental Science WILEY

Evidence for early arousal-based differentiation of emotions in children's musical production

Haley E. Kragness[1] | Matthew J. Eitel[1] | Ammaarah M. Baksh[1] | Laurel J. Trainor[1,2,3]

[1]Department of Psychology, McMaster University, Hamilton, ON, Canada

[2]McMaster Institute for Music and the Mind, McMaster University, Hamilton, ON, Canada

[3]Rotman Research Institute, Baycrest Hospital, Toronto, ON, Canada

Correspondence
Laurel J. Trainor, Psychology Building, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada.
Email: ljt@mcmaster.ca

Present address
Haley E. Kragness, Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada and Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada
Matthew J. Eitel, Department of Psychology, Goldsmiths, University of London, London, UK

Funding information
Canadian Institute for Advanced Research; Canadian Institutes of Health Research, Grant/Award Number: CIHR Operating Grant MOP 42554; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: NSERC Discovery Grant RGPIN-2014-0470

Abstract

Accurate perception and production of emotional states is important for successful social interactions across the lifespan. Previous research has shown that when identifying emotion in faces, preschool children are more likely to confuse emotions that share valence, but differ in arousal (e.g. sadness and anger) than emotions that share arousal, but differ on valence (e.g. anger and joy). Here, we examined the influence of valence and arousal on children's production of emotion in music. Three-, 5- and 7-year-old children recruited from the greater Hamilton area (N = 74) 'performed' music to produce emotions using a self-pacing paradigm, in which participants controlled the onset and offset of each chord in a musical sequence by repeatedly pressing and lifting the same key on a MIDI piano. Key press velocity controlled the loudness of each chord. Results showed that (a) differentiation of emotions by 5-year-old children was mainly driven by arousal of the target emotion, with differentiation based on both valence and arousal at 7 years and (b) tempo and loudness were used to differentiate emotions earlier in development than articulation. The results indicate that the developmental trajectory of emotion understanding in music may differ from the developmental trajectory in other domains.

KEYWORDS

articulation, emotion, loudness, music, production, tempo

1 | INTRODUCTION

From the newborn period, infants are exposed to music in their environment, such as songs sung by their caregivers (Trehub & Trainor, 1998). Infants respond differently to music that adults perceive to be calming compared to arousing (Rock, Trainor, & Addison, 1999; Tsang & Conrad, 2010), and caregiver's songs can prevent or delay infant stress, as well as modify arousal levels (Cirelli, Jurewicz, & Trehub, 2019; Corbeil, Trehub, & Peretz, 2005; Shenfield, Trehub, & Nakata, 2003). In adulthood, both music listeners and performers identify music-related emotions as among the primary motivations for engaging with music (Juslin & Laukka, 2004; Lindström, Juslin, Bresin, & Williamon, 2003). Indeed, this aspect of music has been proposed as one possible explanation for the observation that music is found across human cultures (e.g. Trainor, 2015). Thus, the association between music and emotion appears to be present in early infancy and persists throughout the lifetime.

The ability to produce and recognize emotional signals is vital for successful social interactions in general. Gestures, facial expressions, language, and music can all contribute to the communication of emotions between humans. Though much of the emotional intent in spoken language is derived from the content of a sentence, emotional intent is also conveyed through the speaker's use of expressive cues, or *prosody* (Morton & Trehub, 2001; Nygaard & Queen, 2008; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006). For example, higher speech rates and intensity are often associated with communicating happiness and anger, and lower speech rates with tenderness and sadness (for a review, see Juslin & Laukka, 2003).

Emotive prosody can arise from physiological changes in arousal that affect the vocal tract (Scherer, 1986). In addition, individuals can intentionally manipulate prosodic cues to convey different affective states (Banse & Scherer, 1996; Juslin & Laukka, 2003).

Similarly, emotional expression in music is influenced both by the content of the notated music (or 'score') and the expressive cues used by the performers. Both musically trained and untrained adults can identify intended emotions in musical performances, with the most widespread agreement for so-called 'basic' emotions, including *happiness*, *sadness*, *anger*, *fear*, and *tenderness* (Balkwill, Thompson, & Matsunaga, 2004; Fritz et al., 2009; Juslin & Laukka, 2003; Mohn, Argstatter, & Wilker, 2010). There is some evidence that these basic emotions, and even some more complex emotions, are recognized above chance levels cross-culturally, although cultural congruency between the performer and listener appears to enhance recognition (Laukka, Eerola, Thingujam, Yamasaki, & Gregory, 2013).

Though there is no definitive consensus on the best way to model emotions in music, a simple two-dimensional approach including the dimensions valence (negative to positive) and arousal (low to high activity) has been widely used (Russell, 1980; Schubert, 1999; Vieillard et al., 2008; Vuoskoski & Eerola, 2011). This conceptualization is based on the observation that musical features often map onto either *valence* (positive/negative) or *arousal* (high/low activity level). In general, *valence* tends to be modulated by musical features that are controlled by the composer—for example, mode (Gagnon & Peretz, 2003; Hevner, 1935; Lindström, 2006; Peretz, Gagnon, & Bouchard, 1998) and pitch register (Hevner, 1937; Scherer & Oshinsky, 1977). While *arousal* can also be conveyed by compositional structure, it is mainly communicated through cues that can be additionally modulated by the performer, such as speed, loudness, and connectedness of the notes (Behrens & Green, 1993; Gabrielsson & Juslin, 1996; Gagnon & Peretz, 2003; Ilie & Thompson, 2006; Juslin, 1997, 2000; Laukka & Gabrielsson, 2000).

It has been proposed that the dimensions of valence and arousal are also critical in the development of children's emotion concepts in domains other than music. Widen and Russell (2008) describe a model in which early emotion understanding is based on valence alone, with increasing differentiation throughout development. For example, infants' willingness to cross a visual cliff is influenced to a greater extent by the valence of their mother's facial expression (positive versus negative) than by the arousal of the facial expression within a valence category (e.g. fear versus sadness; Sorce, Emde, Campos, & Klinnert, 1985). Similarly, 2- and 3-year-old children appear to have trouble differentiating between facial expressions of anger and sadness (Widen & Russell, 2002). And throughout early childhood, when children make errors in naming facial expressions, their errors tend to be of the same valence as the target, rather than the same arousal (Bullock & Russell, 1984; Widen & Russell, 2004).

An important question is how children recognize emotions in music. Overall, it appears that by 5 or 6 years old, children tend to agree with adults about basic emotions expressed in orchestral music, including *happiness*, *sadness*, *anger*, and *fear*, although the latter two are often confused (Cunningham & Sterling, 1988; Giomo, 1993; Kratus, 1993; Terwogt & van Grinsven, 1991; Vidas, Dingle, & Nelson, 2018). Other studies have reported successful recognition of *happiness*, *sadness*, and *anger* by children as young as 3 or 4 years old (Franco, Chew, & Swaine, 2016; Gentile, 1998), and several studies found that infants as young as 5 months old could discriminate between excerpts that had been selected previously by adults and pre-schoolers as representative of *happy* and *sad* (Flom, Gentile, & Pick, 2008; Flom & Pick, 2012). However, from these studies, it is impossible to know which expressive cues pre-schoolers used in their judgments, and whether infant discrimination was based on differences in perceived affect per se or simply on differences in certain salient acoustic cues (for example, tempo differences between the *happy* and *sad* excerpts).

Research Highlights

- In a simple music production task, young children (5 years) differentiate emotions based on arousal more than valence.
- Musical differentiation based on both arousal and valence becomes apparent by 7 years of age.
- Manipulation of tempo and loudness for emotional expression is observed as early as 5 years, and manipulation of articulation is incorporated by 7 years.

To our knowledge, only two studies have examined the relative contribution of different expressive cues to children's recognition of emotions by manipulating features in music. Dalla Bella et al. (2001) systematically manipulated the tempo and mode (major or minor) of musical passages and asked children to judge whether each passage sounded happy (by selecting a cartoon happy face) or sad (by selecting a cartoon sad face). Children as young as 5 years rated fast passages as happier than slow passages, and by 6 years, children used mode (major, minor) as well as tempo to identify valence. Mote (2011) asked children to perform the same task with musical passages that varied in tempo and familiarity, and found associations between emotion and tempo as young as 4 years, but no associations with familiarity. However, neither of these studies examined children's understanding of musical cues from a dimensional approach—in other words, it is not possible to know whether children associated fast tempi with happiness due to its high arousal level or to its positive valence.

Though studies with adults have examined both the production and recognition of emotions in music, studies with children have been largely limited to recognition—children usually are not able to perform music at a high level. Only one study to our knowledge has investigated the development of musical production of emotion (Adachi & Trehub, 1998). In this study, 4- to 12-year-old children were asked to sing an experimenter-selected, but familiar, song with the goal of making the experimenter feel happy or sad. In subsequent studies (Adachi & Trehub, 2000; Adachi, Trehub, & Abe, 2004), same-age children and adults were asked to guess

TABLE 1 Demographic characteristics

	3-year-olds (n = 24)	5-year-olds (n = 27)	7-year-olds (n = 23)
Gender distribution	11 boys, 13 girls	13 boys, 14 girls	12 boys, 11 girls
Age (years)	3.56 years, *SD* = 0.12	5.66 years, *SD* = 0.13	7.54 years, *SD* = 0.07
Number reporting >10% exposure to non-English	2	0	5
Languages exposed to	Punjabi (1), Spanish (1)	NA	French (5)
Number in formal music training (past or current)	0	0	1 year of piano lessons (5)
Number in formal dance training (past or current)	Predance (4), gymnastics (1), ballet (2)	Acro/ballet (1), tap/ballet/jazz (1), ballet (3), jazz (1)	Tap/jazz (1), ballet/jazz (1), mixed (1), kinder-dance(1)

the children's intended emotions. Children tended to sing faster, louder, and at a higher pitch when trying to invoke happiness compared to sadness. Older children manipulated tempo to a greater extent than younger children between the happy and sad conditions. At all ages, children primarily used cues found in both music and speech (e.g. tempo/speed, dynamics/loudness), and used cues that are primarily music-specific (e.g. articulation) relatively infrequently.

Although singing studies can be useful, there are notable limitations. First, it is necessary to use familiar songs, which likely have pre-existing emotional associations and/or expressive performance characteristics. A second drawback is reliance on the voice as the instrument, as singing range and accuracy increase significantly throughout childhood (Hedden, 2012). In the present study, we used a self-pacing paradigm to test young children's expressive musical productions. Children pressed a single key on a MIDI piano (marked with a sticker) to self-pace through a musical excerpt several times, portraying a different emotion in each performance. They were able to control the onset, offset, and loudness of each chord in sequence. Specifically, children were asked to use the piano apparatus to play a music game. In this game, children played chord sequences to accompany vignettes and expressive faces that conveyed four different emotions, each of which was in a different quadrant of a valence-by-arousal circumplex. The simplicity of the self-pacing task removes the requirement of musical training.

We expected to replicate findings from previous studies that children's 'happy' song renditions were faster and louder than their 'sad' song renditions (Adachi & Trehub, 1998). However, whether children's additional renditions of 'angry' and 'peace' would be more similar to their shared-valence or shared-arousal counterparts was not known.

The first aim of the study was to investigate whether differentiation of the valence dimension dominates children's early production of musical emotions, as in children's early recognition of facial expressions. The second aim was to examine how children associate conventional expressive cues—tempo, loudness, and articulation—with the valence and arousal dimensions of musical emotions. We expected to replicate previous results showing children's production and perception of 'happy/excited' emotions to be faster and louder than 'sad' emotions, but whether these differences would generalize to 'peace' and 'anger' based on either arousal or valence was unknown.

2 | METHOD

2.1 | Participants

Eighty-six children were recruited to participate. Of those, 10 were excluded from analysis due to expressed desire to withdraw (n = 4), reporting greater than 2 years of formal musical training (n = 4), or parent-reported developmental delays including gross motor delays (n = 1) and receptive language barrier (n = 1). Two additional participants scored below the third percentile in receptive vocabulary (PPVT-4) and were excluded due to concerns about their ability to understand the task.

Thus, in the final sample, 74 children served as participants, including 24 3-year-olds (11 boys, 13 girls; M_{age} = 3.56, *SD* = 0.12 years), 27 5-year-olds (13 boys, 14 girls; M_{age} = 5.66, *SD* = 0.13 years), and 23 7-year-olds (12 boys, 11 girls; M_{age} = 7.54, *SD* = 0.07 years). See Table 1 for participant demographic details. Participants were recruited from the Developmental Studies Database at McMaster University, which includes families who expressed interest in participating in future scientific studies while in the maternity ward at hospitals in the greater Hamilton area. Children received a certificate and a small prize of their choice as a thank you. Each lab visit was approximately 1 hr. All procedures were approved by the McMaster University Research Ethics Board.

2.2 | Stimuli

2.2.1 | Music selection

As noted above, prior research has shown that mode (major versus minor) may influence children's recognition of the emotional valence of an excerpt. Therefore, we selected four excerpts from 4-voice Bach chorales as stimuli such that two that were originally composed in the major mode and two that were originally composed in the minor mode. It is also possible that the particular key of a composition might affect perceived valence, so with the guidance of an expert in Baroque music theory, each excerpt was transposed to the key of *F* (major or minor) and small alterations were made so that each excerpt consisted of 24 chords of quarter-note length. Because compositions written in major versus minor keys might tend also to differ in other compositional aspects, such as melodic contour or harmonic structure, we created four additional musical excerpts by transforming the excerpts originally in a major mode to F minor, and by transforming the minor mode excerpts to F major, again under the guidance of an expert in Baroque music theory. Though each participant only produced music using two of the eight possible excerpts, this procedure ensured that across the experiment, any expressive tendencies that were correlated with compositional aspects of a specific stimulus sequence (such as original mode, melodic contour, harmonic structure) were counter-balanced across participants. These chord sequences were used in a previous study with adults (Kragness & Trainor, 2019).

2.2.2 | Vignette selection

Seven vignettes were selected from previous studies examining children's understanding of emotions in stories (see Table 2). Two vignettes were selected for each emotion, except for *peaceful*, for which only a single vignette could be found in the previous literature. An additional *peaceful* vignette was composed by the authors, for a total of eight. Some of the vignettes were slightly adapted such that each story consisted of a character responding to a specific situation and the explicit statement of the emotion associated with the situation. In case 3-year-old children might have difficulty understanding the words 'joy' and 'peaceful', we used the phrases 'happy and excited' (joy) and 'happy and peaceful'/'happy and sleepy' (peaceful), respectively.

2.2.3 | Facial expression selection

Sixteen faces were selected from the NimStim set of facial expressions (Tottenham et al., 2009; see Table 3) to accompany the vignettes. This was done to reinforce the verbal instructions. Two male faces and two female faces were selected for each of the four emotions. For happy/excited, we selected NimStim pictures that were 'happy, open-mouth, exuberant', and for peaceful, we used the 'happy, closed-mouth' pictures. No individual actor was represented in more than one image across the set.

2.3 | Procedure

After a brief explanation of the task, informed consent was obtained from the parent and verbal consent from the child, who were informed that they could choose to stop at any time. The child was told that they would play through a musical excerpt by pressing middle C (indicated with a sticker) on a MIDI piano keyboard. Pressing the key would initiate the onset of each successive chord in the musical excerpt and releasing the key would cause the offset of the chord. Participants could alter the loudness by changing the level of force used to press the key.

TABLE 2 Vignettes

Emotion	Story	Reference
Joy (positive, high-arousal)	'It was X's birthday. All of his/her friends came to his/her birthday party and gave her presents. X jumped up and down and clapped her hands. How do you think he/she felt? He/she felt happy and excited'	Based on Widen and Russell (2010)
	'X wanted his/her friends to come over and play. So he/she asked them, and they came to play at his/her house. He/she smiled and jumped up and down. How do you think he/she felt? He/she felt happy and excited'	Based on Ribordy, Camras, Stefani, and Spaccerelli (1988)
Sadness (negative, low-arousal)	'X went to feed his/her pet goldfish. But it was not swimming. It was not even in the fish tank. X's fish had died. X walked over to a chair and sat down. Tears came to his/her eyes. He/she didn't want to talk to anyone. How do you think he/she felt? He/she felt sad'	Based on Widen and Russell (2002)
	'X's friend, who he/she really liked to play with, moved away. He/she couldn't play with his/her friend anymore. How do you think he/she felt? He/she felt sad'.	Based on Ribordy et al. (1988)
Anger (negative, high-arousal)	'X was waiting in line. Then a boy cut in line in front of her. He didn't even ask. X shoved him out of line and yelled at him. How do you think he/she felt? He/she felt angry'	Based on Widen and Russell (2010)
	'X's little brother broke his/her favorite toy on purpose. He/she clenched his/her fists and stomped his/her feet. How do you think he/she felt? He/she felt angry'	Based on Ribordy et al. (1988)
Peace (positive, low-arousal)	'X drew a picture and showed it to his/her father. His/her father really liked it and said X did a good job. That made him/her feel very happy and peaceful, and he/she gave his/her dad a hug'	Based on Ribordy et al. (1988)
	'X was feeling very tired and likes to sleep with his/her favorite stuffed animal. He/she searched around his/her whole room for the stuffed animal. When he/she found his/her stuffed animal, how do you think he/she felt? He/she felt happy and sleepy. He/she hugged his/her stuffed animal, smiled, and went to bed'	Original

TABLE 3 Emotional faces used in the study

Emotion	NimStim faces—female	NimStim faces—male
Joy (positive, high-arousal)	01F_HA_X	30M_HA_X
	09F_HA_X	36M_HA_X
Sadness (negative, low-arousal)	02F_SA_C	22M_SA_C
	07F_SA_C	34M_SA_C
Anger (negative, high-arousal)	10F_AN_C	20M_AN_C
	05F_AN_C	23M_AN_C
Peace (positive, low-arousal)	03F_HA_C	28M_HA_C
	08F_HA_C	32M_HA_C

2.3.1 | Practice phase

A practice phase was used to familiarize participants with the task and to allow participants to practice manipulating the musical cues. Prior to the testing blocks, the experimenter asked the child to perform a short practice excerpt (a C Major scale) with different musical features (quickly, slowly, with short notes, with long notes, loudly, and quietly) to ensure that they were aware of the expressive cues available to them.

2.3.2 | Testing phase

Each participant experienced two testing blocks. One block consisted of a musical excerpt in a major mode and the other block consisted of a different musical excerpt in a minor mode. The order in which participants experienced the major and minor mode excerpts was counterbalanced across participants. Participants were told that they were going to play a musical game, and that they should play the music in a manner that matched the feeling of the story. At the beginning of each testing block, the experimenter directed the participant to continually press the designated piano key to 'listen to the music you are going to play'. They were asked to keep pressing the key until no more music played, which meant that the music was over. Due to experimenter error, there were three instances in which the participants were not exposed to the music before the emotion trials in the second block (2 five-year-olds, 1 seven-year-old). Only the first block was included in analyses for these participants.

Next, the experimenter read a short vignette to the participant, while showing the participant a picture of a face that matched the emotion portrayed in the vignette, and asked the participant how the character in the vignette felt. After the participant verbally provided an answer, the experimenter indicated whether the answer was correct or incorrect by explicitly stating the target emotion (e.g. 'Yes, the character felt angry' or 'Actually, the character felt angry'). The experimenter then asked the participant to play the musical excerpt in a way that sounded like the target emotion (e.g. 'Can you play the music in a way that sounds angry?'). On each trial, participants were directed to continue to play until no more music played. Therefore, each performance trial lasted for 25 taps (24 chords plus an additional tap to end the trial). The average duration of playing the 24 taps was 12.96 s (approximately 540 ms per tap), which is well within the range of spontaneous motor tempo reported for young children previously (e.g. Provasi & Bobin-Bègue, 2003). The experimenters observed that once the children heard the excerpt several times, they sometimes forgot to initiate the final (25th) tap to end the trial. Other children appeared to treat the final tap differently from the first 24, and pressed it especially slowly or especially hard to end the trial. Because this resulted in an incorrect value for the articulation measure on the 24th chord, and because we could not be sure whether children intended for the final chord to be part of their performance, we used only chords 1–23 for all analyses for consistency across cues.

This process was repeated for each of the four emotions being tested for each of the two excerpts, for a total of 8 test trials. The order in which the emotions were presented was randomized within each block. The stories and faces used in the first block were not repeated in the second testing block. The participant either viewed only male or only female faces from the face set throughout the two testing blocks, and the gender of the faces was counterbalanced with the gender of the participants in each age group. Two children (both 5-year-olds) stopped their performance to have a conversation with the experimenter during at least one test trial and were asked to 'play the music again [emotion], without stopping'. In these cases, only the second performance was used for analyses. This portion of the experiment took 10–15 min.

2.3.3 | Additional tests

After the two testing blocks had been completed, the experimenter administered the Test of Emotional Comprehension (TEC, Pons & Harris, 2000) to obtain a measure of each participant's understanding of emotions in general. One participant (3-year-old) was not given a score for the TEC due to parent interference during the test, and this participant was excluded from analyses using the TEC. Finally, the Peabody Picture Vocabulary Test (PPVT-4, Dunn & Dunn, 2007) was administered to examine each participant's receptive vocabulary skills, which could affect their ability to understand the task instructions.

3 | RESULTS

3.1 | Developmental patterns in three-dimensional space

We conceptualized performances as data points in three-dimensional space, with each cue (tempo, loudness, articulation) representing a dimension (see a similar conceptualization of performance space in Sievers, Polansky, Casey, & Wheatley, 2013). First, we calculated each participant's average performance for each emotion, resulting in an average tempo, loudness, and articulation for each

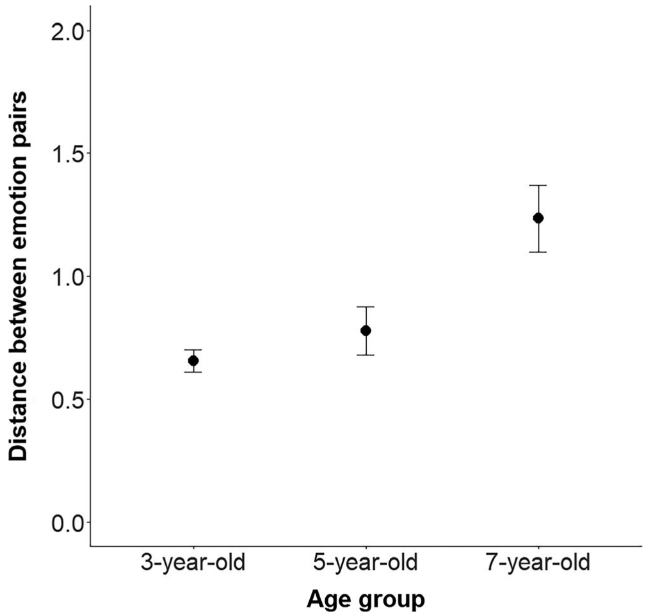

FIGURE 1 Average distance between emotion pairs for participants in each age group. The average distance used by 7-year-old participants was significantly greater than 5- or 3-year-old participants. Error bars represent standard error of the mean (*SEM*)

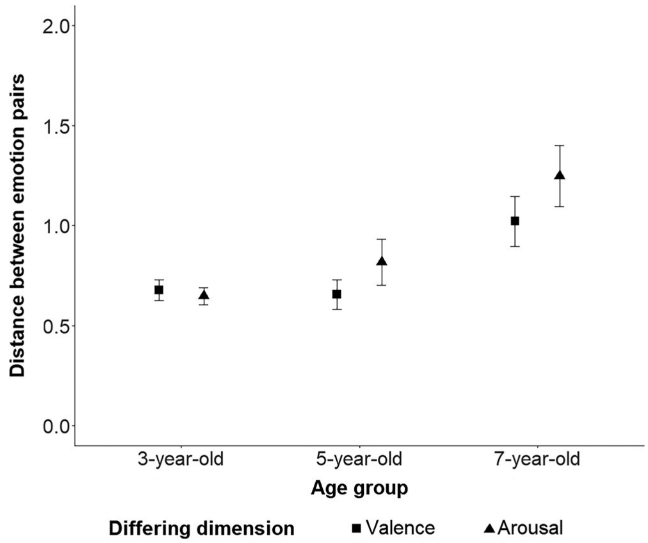

FIGURE 2 Depicted in squares is the average distance between emotion pairs that differed on valence, but not arousal (anger/joy, sad/peace). Depicted in triangles is the average distance between emotion pairs that differed on arousal, but not valence (anger/sad, joy/peace). Error bars represent *SEM*

of the four target emotions. Next, we *z*-normalized each cue to control for differences in scale between cues. Finally, we calculated the Euclidean distance between each emotion pair.

We examined whether there were age-related changes in how distant participants' emotion representations were from each other in three-dimensional space (Figure 1). We calculated the average distance between emotion pairs for each participant. Wilcoxon ranked-sum tests showed that the average distance between emotion pairs was larger for 7-year-olds than for the other age groups (5-year-olds: $W = 187$, $p = .016$; 3-year-olds: $W = 143$, $p = .004$), but that there was no significant difference in the distances used by 3- and 5-year-olds ($W = 292$, $p = .555$).

Next, we asked whether, within each age group, larger differentiation between groups was associated with higher scores on emotion understanding (raw scores on the Test of Emotion Comprehension) or receptive vocabulary (percentile scores on the (PPVT-4). No significant correlations were observed for either scale in any age group (see Table 4).

Next, we asked whether participants' three-dimensional representations of the emotions were more distant between emotions that differ on dimensions of valence or arousal. For each participant, we calculated the average Euclidean distance between *emotions that differed on valence but not arousal* (joy/anger, peace/sad) and *emotions that differed on arousal but not valence* (joy/peace, anger/sad). Paired *t*-tests were used to evaluate whether there was greater distance between emotions that differed on arousal compared to those that differed on valence (Figure 2). For 3-year-old participants, there was no significant difference for emotions that differed on arousal versus valence ($t(23) = -0.633$, $p = .533$, $d_z = -0.129$). For 5-year-old participants, emotions that differed on arousal were more distant than those that differed on valence ($t(26) = 2.398$, $p = .024$, $d_z = 0.461$). For 7-year-old participants, the difference was not significant, although there was still a trend for emotions differing on arousal to be more distant than those differing on valence ($t(22) = 1.940$, $p = .065$).

3.2 | Differences in cue use

3.2.1 | Tempo

We first examined tempo (speed) differences across emotions (see Figure 3a). Tempo was defined as chord onsets per minute, such

TABLE 4 Correlations between test scores and average distance between emotions

	3-year-olds (*n* = 24)		5-year-olds (*n* = 27)		7-year-olds (*n* = 23)	
	R	95% CI	*R*	95% CI	*R*	95% CI
TEC score	−.099	−0.491 to 0.327	.029	−0.355 to 0.405	−.093	−0.486 to 0.332
PPVT score	.162	−0.258 to 0.531	.354	−0.030 to 0.647	.276	−0.153 to 0.618

Note:: Due to parent interference, one 3-year-old participants' TEC score was omitted from analysis.

that higher values indicate faster speeds. The data were submitted to a mixed ANOVA with within-subject factors valence (positive, negative) and arousal (high, low), and the between-subject factor age group (3-year-old, 5-year-old, 7-year-old). The omnibus ANOVA revealed that age significantly interacted with both valence ($F(2,71) = 5.258$, $p = .007$, $\eta_G^2 = 0.018$, $\eta_p^2 = 0.129$) and arousal ($F(2,71) = 8.457$, $p < .001$, $\eta_G^2 = 0.025$, $\eta_p^2 = 0.192$), so further analyses were performed within each age group separately. For 3-year-olds, there were no significant main effects (p's > .07) but the valence-by-arousal interaction trended toward significance ($F(1,23) = 3.729$, $p = .066$, $\eta_G^2 = 0.006$, $\eta_p^2 = 0.139$). Post hoc analyses suggested that within the positive-valence condition, *joy* was played faster than *peace* ($t(23) = 2.098$, $p = .047$, $d_z = 0.428$, average difference of 10 chord onsets per minute), but within the negative-valence condition there was no difference ($t(23) = 0.117$, $p = .908$, average difference of .48 chord onsets per minute). For 5-year-olds, there were significant main effects of both valence ($F(1,26) = 18.270$, $p < .001$, $\eta_G^2 = 0.051$, $\eta_p^2 = 0.413$) and arousal ($F(1,26) = 11.419$, $p = .002$, $\eta_G^2 = 0.036$, $\eta_p^2 = 0.305$). Within each arousal level, positively valenced emotions were played faster than negatively valenced emotions (by 23 chord onsets per minute on average), and high-arousal emotions were played faster than low-arousal emotions (by 20 chord onsets per minute on average). The same pattern was observed for 7-year-olds (valence; $F(1,22) = 15.657$, $p < .001$, $\eta_G^2 = 0.128$, $\eta_p^2 = 0.416$, and arousal; $F(1,22) = 27.753$, $p < .001$, $\eta_G^2 = 0.157$, $\eta_p^2 = 0.558$). Within each arousal level, positively valenced emotions were played faster than negatively valenced emotions (approximately 35 chords per minute faster) and high-arousal emotions were played faster than low-arousal emotions (approximately 39 chords per minute faster).

3.2.2 | Loudness

Participants varied the loudness of each chord by varying the velocity of their key presses. Here, we used the MIDI velocity information recorded by the piano apparatus as a proxy for loudness. MIDI velocity measures range from 1 to 127, with 1 representing the minimum level (experienced as the quietest sound possible on the piano) and 127 representing the maximum level (experienced as the loudest sound possible on the piano; Figure 3b). Because there was a significant interaction between age group and arousal ($F(2,71) = 11.328$, $p < .001$, $\eta_G^2 = 0.044$, $\eta_p^2 = 0.340$), age groups were analyzed separately. For 3-year-olds, no significant interaction or main effects were observed (all $ps > 0.130$). For 5-year-olds, there was a significant valence-by-arousal interaction ($F(1,26) = 7.580$, $p = .011$, $\eta_G^2 = 0.024$, $\eta_p^2 = 0.256$). Participants played high-arousal emotions with greater velocity (louder) than low-arousal emotions in both the positive-valence condition ($t(26) = 3.362$, $p = .002$, $d_z = 0.647$) and negative-valence condition ($t(26) = 4.256$, $p < .001$, $d_z = 0.819$), but the difference was larger in the negative condition (20 MIDI velocity points) than the positive condition (8 MIDI velocity points). There was also a significant valence-by-arousal interaction in the 7-year-old group ($F(1,22) = 15.533$, $p < .001$, $\eta_G^2 = 0.091$, $\eta_p^2 = 0.414$). Similar to the 5-year-olds, participants played high-arousal emotions with greater velocity than low-arousal emotions across positive ($t(22) = 4.464$, $p < .001$, $d_z = 0.931$) and negative ($t(22) = 5.656$, $p < .001$, $d_z = 1.179$) conditions, but again, the difference was larger in the negative condition (34 MIDI velocity points) than the positive condition (12 MIDI velocity points).

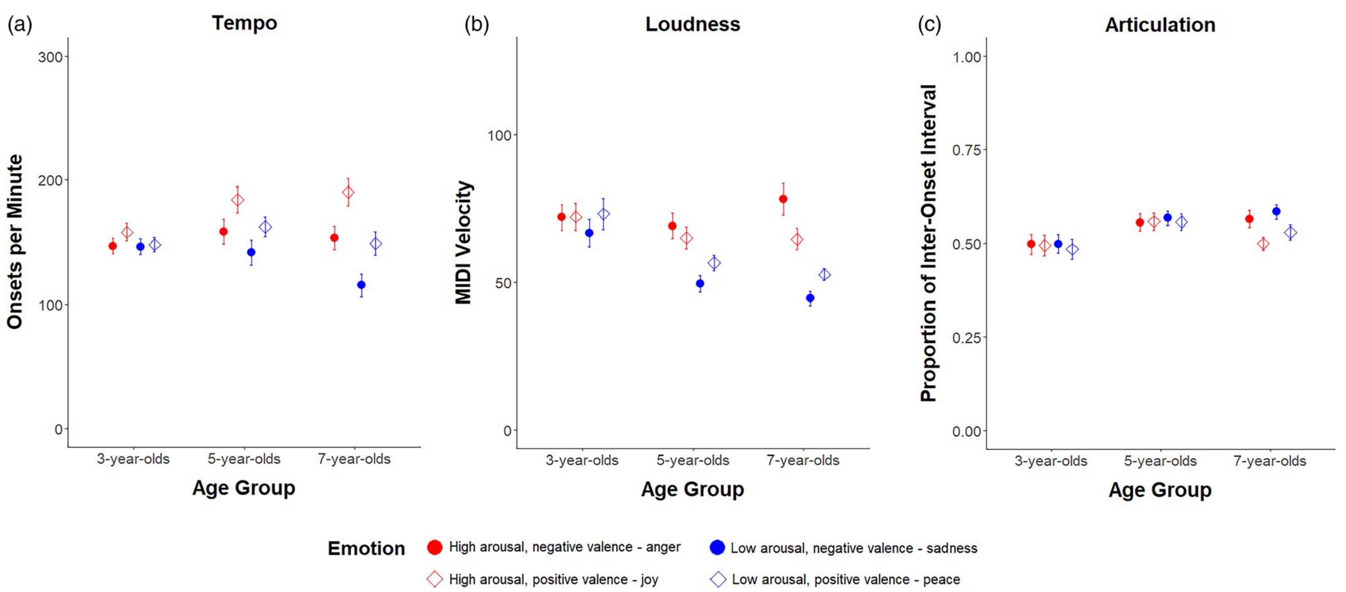

FIGURE 3 Average value for each emotion used by participants in each age group. Color denotes arousal level (*high*—red, *low*—blue) and shape denotes valence category (*positive*—diamond, *negative*—circle). (a) Average tempo, characterized as onsets per minute. Higher values represent faster tempi. (b) Loudness, characterized by MIDI velocity with which they pressed the key. Higher values represent greater MIDI velocity, experienced by the participant as loudness. (c) Articulation, characterized by the proportion of each inter-onset interval that was filled by the key press. Higher values represent greater connectedness. Error bars represent *SEM*

3.2.3 | Articulation

In music, *articulation* refers to the connectedness between notes, ranging from smooth and connected ('legato') to disconnected, with silent gaps between notes ('staccato'). In the present study, this was measured by calculating the proportion of the onset-to-onset interval in which the chord was sounded. For example, if a chord was sounded for 500 ms of a 1,000-ms interval, the articulation value would be 0.5. If a chord was sounded for 200 ms of a 1000-ms interval, the value would be 0.2. Thus, values closer to 0 represent more disconnected chords, while values closer to 1 represent more connected chords (Figure 3c). The mixed ANOVA revealed a significant interaction between age group and valence ($F(2,71) = 6.859$, $p = .002$, $\eta_G^2 = 0.012$, $\eta_p^2 = 0.162$), motivating separate valence-by-arousal ANOVAs for each age group. For 3- and 5-year-olds, there were no significant main effects or interactions (all $ps > .432$), but for 7-year-olds there was a significant main effect of valence ($F(1,22) = 16.539$, $p < .001$, $\eta_G^2 = 0.089$, $\eta_p^2 = 0.429$), such that negatively valenced emotions were played with more connected articulation than positively valenced emotions. On average, the note durations used by 7-year-olds in negatively valenced emotion contexts filled approximately 6.8% more of the inter-onset interval than note durations used in positively valenced emotion contexts.

3.3 | Vignette responses

To investigate whether children understood the intended emotions in the vignettes, in a post hoc analysis, children's spontaneous responses to each vignette were transcribed from video (six 3-year-olds, three 5-year-olds, and two 7-year-olds were not recorded due to technical issues and/or experimenter error, and recordings for several additional children's sessions were cut short due to camera battery failure).

For negatively valenced emotions, the majority of responses given were correct. For low-arousal, negative vignettes, 'sad' was given 29/36 times for 3-year-olds, 46/47 times for 5-year-olds, and 42/42 for 7-year-olds. For high-arousal, negative vignettes, 'angry' or 'mad' was given 30/36 times for 3-year-olds, 42/48 times for 5-year-olds, and 41/41 times for 7-year-olds. There were very few instances in which children gave fully incorrect responses (2 instances of 'happy' from 5-year-olds, 2 instances of 'sad' from 5-year-olds, and 2 instances of 'sad' from 3-year-olds), and several children's responses were inaudible in the recording (2 instances in 3-year-olds and 1 instance in 5-year-olds). Finally, a few children gave responses that were essentially correct, but were ambiguous with regard to arousal level, for example, 'sad and mad', 'bad', 'not happy', 'not good' (see Figure S1 in Supplemental Material). For positively valenced emotions, the response 'happy' was most common for both low-arousal vignettes (28/36, 43/48, and 33/42 for 3-, 5-, and 7-year-olds, respectively) and high-arousal vignettes (30/36, 41/47, and 25/42, respectively). Similar to the negative vignettes, very few responses were clearly incorrect (2 instances of 'sad' for low-arousal and 1 instance of 'sad' for high-arousal). Again, several ambiguous responses in addition to 'happy' were used, such as 'proud' (high-arousal), 'loud' (high-arousal), and 'good' (low-arousal). Interestingly, 'sleepy/tired/peaceful' were given 8 times in the low-arousal condition (2, 3, and 3 times by respective age groups) but 0 times in the high-arousal condition. Likewise, 'joy/joyful/excited' were given 20 times in the high-arousal condition (1, 3, and 16 times by respective age groups) but only three times in the low-arousal condition (all given by 7-year-olds). Overall, results indicate that despite the open-ended nature of the task, incorrect responses were rare, and children overwhelmingly gave 'correct' or 'essentially correct' responses across all three age groups. Regardless of children's response, we re-iterated the target emotion directly prior to their performance ('Yes, he/she felt…' or 'Actually, he/she felt…').

4 | DISCUSSION

The purpose of the present study was to examine how children use musical features to produce the valence and arousal dimensions of a target emotion. Results of previous studies have suggested that children differentiate emotion categories first based on valence, and then gradually incorporate differentiation based on arousal. The results of the present study showed that when producing emotion in music, at 5 years of age, children differentiated emotions based on arousal to a greater extent than valence. By 7 years of age, they incorporated differences in their performances based on both valence and arousal, though arousal differences were still numerically larger. The present findings therefore suggest that developmental trajectories for emotion understanding may differ across modalities.

Previous studies of children's emotional production in music have been limited to singing and have exclusively investigated the emotions 'happy' and 'sad'. Using a simple self-pacing apparatus, we found evidence that 5- and 7-year-old children express emotional arousal using tempo and loudness, with patterns of cue use mirroring those observed in adults. With respect to tempo, both 5-year-old and 7-year-old children played faster to differentiate high-arousal emotions from low-arousal emotions, and within each arousal condition, to distinguish positively valenced from negatively valenced emotions. Even for 3-year-old children, we observed a trend for *joy* to be played faster than *peace*. This is consistent with previous research using emotion identification tasks that found that tempo is among the earliest cues used by children when evaluating emotions in music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Kratus, 1993; Mote, 2011). The present results reinforce and extend these findings to the production domain.

Loudness was also used to differentiate between high- and low-arousal emotions for children as young as 5 years. To our knowledge, no previous studies have examined the role of loudness in children's emotion judgments, although Adachi and Trehub (1998) found that children as young as 4 years manipulated loudness to differentiate between sadness and happiness in vocal music. The present study extends this observation by demonstrating that both 5-year-old and 7-year-old children played high-arousal emotions louder than low-arousal emotions in both positive- and negative-valence conditions, similarly to adults. This suggests that loudness may be a particularly salient cue for children when judging musical emotions, and should be considered in future studies of children's emotional judgments.

To our knowledge, the present study is the first to examine children's emotional productions with regard to articulation. Seven-year-old children, but not 3- or 5-year-old children, performed negatively valenced emotions with more connected chords than positively valenced emotions across arousal conditions. Why does manipulation of articulation to convey emotion occur later in development than tempo and loudness? One possible explanation is that articulation is more difficult to manipulate than tempo or loudness, requiring children to control both the onset *and* the offset of each note. An alternative possibility is that articulation is not a particularly salient cue for emotional expression for children. Results from one study (Kratus, 1993) suggest that children as young as six years do rely on articulation to make emotion judgments, but that study did not examine potential correlations among various cues, so it is difficult to know the contribution of articulation specifically. Interestingly, articulation is considered to be more specific to the music domain than other cues such as loudness and speed that are used prevalently in both speech and music (Juslin & Laukka, 2003). From this perspective, we might predict that sensitivity to articulation as an emotional cue would develop later than other, more domain-general cues.

An interesting question to be explored in future research is the role of formal music training in childhood. Past research in adults suggests that music training does not significantly enhance identification of emotions in music (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Juslin, 1997; Song, Dixon, Pearce, & Halpern, 2016; but see Castro & Lima, 2014). This was corroborated by recent work using the self-pacing paradigm of the present paper, which found that adults with no music training use tempo, loudness, and articulation to express emotions similarly to those with training (Kragness & Trainor, 2019). This suggests that the level of informal experiences with music accumulated by adulthood is sufficient for recognizing and expressing musical emotions. Following that idea, we might expect to see effects of formal training differences during development, when less exposure has typically yet been acquired. To our knowledge, only one previous study has examined the effect of formal training on children's emotional recognition of excerpts and found no differences between 12-year-olds with 6 years of formal training and 6-year-olds without training (Kratus, 1993). In this case, the participants were asked to make dichotomous decisions about the valence and arousal of excerpts, so the null result may have resulted from universally high performance across participants. Effects of musical training might be observed in a more difficult task, such as the performance task in the present paper, or in a rating task with a larger gradient to choose from.

An enduring question in this literature relates to the specific music stimuli that are used. While many studies have employed orchestral music (e.g. Cunningham & Sterling, 1988; Giomo, 1993; Nawrot, 2003), others have found superior performance when using culturally familiar music that is intended for children (Mote, 2011; Schellenberg, Nakata, Hunter, & Tamoto, 2007). In the present study, we used modified chord sequences from Bach chorales, which are neither child-directed nor likely to be overly familiar (though they are likely to be culturally familiar). We selected these sequences as stimuli for several reasons. First, the sequences have a reliable phrase structure, such that each sequence is made of three 8-chord sub-phrases. Using these stimuli also enabled direct comparison with previous work on adults using the same stimuli (Kragness & Trainor, 2019). Finally, these stimuli are unlikely to be familiar to children, thus avoiding previously held associations with recognized pieces. However, it may be the case that selecting child-directed stimuli would lead to greater task engagement, potentially resulting in evidence of emotional differentiation in performances at younger ages.

In the present work, we present a new method for testing children's expressive musical production that does not rely on the vocal apparatus or the ability to play a musical instrument. This method is easy to explain to young children, highly flexible, and can be used to examine children's production of familiar music or unfamiliar music. In Western society, singing in a social setting can be intimidating. Though singing is useful for studying children's musical development, the task used in the present study can corroborate and extend those findings as it does not depend on singing ability.

Previous work has suggested that valence differentiation precedes arousal differentiation in emotional development (Widen & Russell, 2003, 2010). In the present task, children were asked to produce music expressing different emotions that were described to them through a combination of visual facial expressions, verbal stories, and emotion labels. In contrast to previous studies with non-musical materials, arousal differentiation preceded valence differentiation. This suggests that developmental trajectories for different dimensions of emotion differentiation are not generalizable across domains. It may be the case that certain modalities (e.g. visual, auditory) are more effective for communicating one or other dimension, or that production tasks ('play the music') are more likely to elicit differentiation based on arousal than labeling tasks ('name the emotion'). Future research should examine how early emotion identification is facilitated by information in different modalities and contexts across development.

ACKNOWLEDGMENTS

This research was supported by an Ontario Trillium Scholarship to HEK, and grants to LJT from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN-2014-0470), the Canadian Institute for Advanced Research (CIFAR Azrieli Program in Brain, Mind & Consciousness), and the Canadian Institutes for Health Research (CIHR Operating Grant MOP 42554). We are grateful to the children and families for their participation, and to Dr. Matthew Woolhouse for his assistance creating the stimuli.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

HEK, AMB, and LJT conceived and designed the experiment. HEK, AMB, and MJE collected and analyzed data. All authors contributed to and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

ORCID

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

REFERENCES

Adachi, M., & Trehub, S. E. (1998). Children's expression of emotion in song. *Psychology of Music*, *26*, 133–153. https://doi.org/10.1177/0305735698262003

Adachi, M., & Trehub, S. E. (2000). Decoding the expressive intentions in children's songs. *Music Perception: An Interdisciplinary Journal*, *18*, 213–224. https://doi.org/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. https://doi.org/10.1111/j.1468-5584.2004.00264.x

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. https://doi.org/10.1111/j.1468-5584.2004.00265.x

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636. https://doi.org/10.1037/0022-3514.70.3.614

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. https://doi.org/10.1177/030573569302100102

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, 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. https://doi.org/10.1080/02699930500204250

Bullock, M., & Russell, J. A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioural Development*, *7*, 193–214. https://doi.org/10.1177/016502548400700207

Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception: An Interdisciplinary Journal*, *32*, 125–142. https://doi.org/10.1525/mp.2014.32.2.125

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2019). Effects of maternal singing style on mother-infant arousal and behavior. *Journal of Cognitive Neuroscience*, *26*, 1–8. https://doi.org/10.1162/jocn_a_01402

Corbeil, M., Trehub, S. E., & Peretz, I. (2005). Singing delays the onset of infant distress. *Infancy*, *21*, 373–391. https://doi.org/10.1111/infa.12114

Cunningham, J., & Sterling, R. (1988). Developmental changes in the understanding of affective meaning in music. *Motivation and Emotion*, *12*, 399–413. https://doi.org/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*, B1–B10. https://doi.org/10.1016/s0010-0277(00)00136-0

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

Flom, R., Gentile, D. A., & Pick, A. D. (2008). Infants' discrimination of happy and sad music. *Infant Behavior and Development*, *31*, 716–728. https://doi.org/10.1016/j.infbeh.2008.04.004

Flom, R., & Pick, A. D. (2012). Dynamics of infant habituation: Infants' discrimination of musical excerpts. *Infant Behavior and Development*, *35*, 697–704. https://doi.org/10.1016/j.infbeh.2012.07.022

Franco, F., Chew, M., & Swaine, J. S. (2016). Preschoolers' attribution of affect to music: A comparison between vocal and instrumental performance. *Psychology of Music*, *45*, 131–149. https://doi.org/10.1177/0305735616652954

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

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68–91. https://doi.org/10.1177/0305735696241007

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to happy-sad judgements in equitone melodies. *Cognition & Emotion*, *17*, 25–40. https://doi.org/10.1080/02699930302279

Gentile, D. A. (1998). An ecological approach to the development of perception of emotion in music. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *59*, 2454.

Giomo, C. J. (1993). An experimental study of children's sensitivity to mood in music. *Psychology of Music*, *21*, 141–162. https://doi.org/10.1177/030573569302100204

Hedden, D. (2012). An overview of existing research about children's singing and the implications for teaching children to sing. *Update: Applications of Research in Music Education*, *30*, 52–62. https://doi.org/10.1177/8755123312438516

Hevner, K. (1935). The affective character of the major and minor mode in music. *American Journal of Psychology*, *47*, 103–118. https://doi.org/10.2307/1416710

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*, 621–630. https://doi.org/10.2307/1416385

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*, 319–329. https://doi.org/10.1525/mp.2006.23.4.319

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

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1797–1813. https://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. https://doi.org/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. https://doi.org/10.1080/0929821042000317813

Kragness, H. E., & Trainor, L. J. (2019). Nonmusicians express emotions in musical productions using conventional cues. *Music and Science*, *2*, 1–13. https://doi.org/10.1177/2059204319834943

Kratus, J. (1993). A developmental study of children's interpretation of emotion in music. *Psychology of Music*, *21*, 3–19. https://doi.org/10.1177/030573569302100101

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and cultural-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*, 432–449. 101.1037.a0031388

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, *28*, 181–189. https://doi.org/10.1177/0305735600282007

Lindström, E. (2006). Impact of melodic organization on perceived structure and emotional expression in music. *Musicae Scientiae*, *10*, 85–117. https://doi.org/10.1177/102986490601000105

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, 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. https://doi.org/10.1177/1321103x030200010201

Mohn, C., Argstatter, H., & Wilker, F. (2010). Perception of six basic emotions in music. *Psychology of Music*, *39*, 503–517. https://doi.org/10.1177/0305735610378183

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*, 834–843. https://doi.org/10.1111/1467-8624.00318

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*, 818–622. https://doi.org/10.1037/a0022573

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music*, *31*, 75–92. https://doi.org/10.1177/0305735603031001325

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1017–1030. https://doi.org/10.1037/0096-1523.34.4.1017

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*, 111–141. https://doi.org/10.1016/S0010-0277(98)00043-2

Pons, F., & Harris, P. (2000). *Test of emotion comprehension—TEC*. Oxford, UK: University of Oxford.

Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2–1/2 and 4-year-old children. *International Journal of Behavioral Development*, *27*(3), 220–231. https://doi.org/10.1080/01650250244000290

Ribordy, S. C., Camras, L. A., Stefani, R., & Spaccerelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *Journal of Clinical Child Psychology*, *17*, 322–325. https://doi.org/10.1207/s15374424jccp1704_4

Rock, A. M., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental Psychology*, *35*, 527–534. https://doi.org/10.1037/0012-1649.35.2.527

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

Schellenberg, E. G., Nakata, T., Hunter, P. G., & Tamoto, S. (2007). Exposure to music and cognitive performance: Tests of children and adults. *Psychology of Music*, *35*, 5–19. https://doi.org/10.1177/0305735607068885

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*, 76–92. https://doi.org/10.1177/0022022101032001009

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*, 331–346. https://doi.org/10.1107/BF00992539

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, *51*, 154–165. https://doi.org/10.1080/00049539908255353

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music*, *31*, 365–375. https://doi.org/10.1177/03057356030314002

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 of the United States of America*, *110*, 70–75. https://doi.org/10.1073/pnas.1209023110

Song, Y., Dixon, S., Pearce, M. T., & Halpern, A. R. (2016). Perceived and induced emotion responses to popular music: Categorical and dimensional models. *Music Perception*, *33*, 472–492. https://doi.org/10.1525/mp.2016.33.4.472

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195–200. https://doi.org/10.1037/0012-1649.21.1.195

Terwogt, M. M., & van Grinsven, F. (1991). Musical expression of mood states. *Psychology of Music*, *19*, 99–109. https://doi.org/10.1177/0305724591192001

Thompson, W. F., & Balkwill, L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*, 407–424. https://doi.org/10.1515/sem.2006.017

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B*, *370*, 20140089. https://doi.org/10.1098/rstb.2014.0089

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

Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior and Development*, *33*, 96–100. https://doi.org/10.1016/j.infbeh.2009.11.006

Vidas, D., Dingle, G. A., & Nelson, N. L. (2018). Children's recognition of emotion in music and speech. *Music & Science*, *1*, 1–10. https://doi.org/10.1177/2059204318762650

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary, and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*, 720–752. https://doi.org/10.1080/02699930701503567

Vuoskoski, J. K., & Eerola, T. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merill-Palmer Quarterly*, *48*, 248–262. https://doi.org/10.1353/mpq.2002.0013

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*, 114–128. https://doi.org/10.1037/0012-1649.39.1.114

Widen, S. C., & Russell, J. A. (2004). The relative power of an emotion's facial expression, label, and behavioral consequences to evoke preschoolers' knowledge of its cause. *Cognitive Development*, *19*, 111–125. https://doi.org/10.1016/j.cogdev.2003.11.004

Widen, S. C., & Russell, J. A. (2008). Young children's understanding of other's emotions. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 348–363). New York, NY, US: Guildford Press.

Widen, S. C., & Russell, J. A. (2010). Children's scripts for social emotions: Causes and consequences are more central than facial expressions. *British Journal of Developmental Psychology*, *28*, 565–581. https://doi.org/10.1348/026151009X457550d

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Kragness HE, Eitel MJ, Baksh AM, Trainor LJ. Evidence for early arousal-based differentiation of emotions in children's musical production. *Dev Sci*. 2021;24:e12982. https://doi.org/10.1111/desc.12982