*Article*

# Children's recognition of emotion in music and speech

## Dianna Vidas⊙, Genevieve A. Dingle and Nicole L. Nelson

## Abstract
The acoustic cues that convey emotion in speech are similar to those that convey emotion in music, and recognition of emotion in both of these types of cue recruits overlapping networks in the brain. Given the similarities between music and speech prosody, developmental research is uniquely positioned to determine whether recognition of these cues develops in parallel. In the present study, we asked 60 children aged 6 to 11 years, and 51 university students, to judge the emotions of 10 musical excerpts, 10 inflected speech clips, and 10 affect burst clips. We presented stimuli intended to convey happiness, sadness, anger, fear, and pride. Each emotion was presented twice per type of stimulus. We found that recognition of emotions in music and speech developed in parallel, and adult-levels of recognition develop later for these stimuli than for affect bursts. We also found that sad stimuli were most easily recognised, followed by happiness, fear, and then anger. In addition, we found that recognition of emotion in speech and affect bursts can predict emotion recognition in music stimuli independently of age and musical training. Finally, although proud speech and affect bursts were not well recognised, children aged eight years and older showed adult-like responses in recognition of proud music.

Music is found in every known human culture (McDermott & Hauser, 2005). While some argue that music is a by-product of other systems, and serves no adaptive role in human evolution, others take a position that music making evolved, at least partially, as a tool for experiencing shared intentionality, and maintaining social cohesion (Cross, 2016; Kirschner & Tomasello, 2010; Patel, 2008). Research in this area has focused on the overlapping neurophysiological, cognitive, and perceptual processes in common between music and language (Jentschke, 2016; Patel, 2008).

Both historically and today, individuals have marvelled that certain combinations of notes, both instrumental and vocal, can give us feelings of exhilaration or melancholy, among other emotions (Juslin & Sloboda, 2010; Spencer, 1881). Music is one of many channels through which emotion can be expressed, although how emotions in music are conveyed and understood is less researched than how we express and recognise emotions in facial expressions, body movements, non-verbal vocalisations, and speech prosody, particularly in developmental literature. Speech prosody is defined as the sounds, rhythms, and tones that accompany language. Although comparatively little is known about

how people recognise and interpret emotion from music, how people understand emotions in speech prosody has been widely researched, and it is theorised that expression of emotion in prosody and music are related evolutionarily (Juslin & Laukka, 2003; Scherer, 1995). As such, investigation into recognition of emotion in speech prosody may inform our knowledge of emotion recognition in music.

## Music and speech

The evolutionary origins of music are much debated, but a link between speech prosody and music is widely considered. Music has been linked with the evolution of language, with the suggestion that music and language co-evolved as

School of Psychology, The University of Queensland, Queensland, Australia

**Corresponding author:**
Nicole L. Nelson, School of Psychology, University of Queensland, St. Lucia, QLD, Queensland 4072, Australia.
Email: n.nelson@uq.edu.au

complementary methods of human communication (Cross, 2016). Other authors link the evolution of music with the evolution of primitive affective vocalisations (Scherer, 1995), or suggest that pre-linguistic human communication consisted of calls not unlike the "song" of gibbons (Masataka, 2009). Regardless of the true evolutionary origins of music, it is plausible that speech and music are related, and crucially, both speech and music have the ability to represent, express, and elicit emotions (Juslin & Sloboda, 2010).

Further solidifying the link between music and speech, adults with musical training have been found to perform better than non-musically trained adults in identifying emotions from speech. However, children with piano training perform equivalently on such tasks with children who have drama training, but better than those with no training in either domain (Thompson, Schellenberg, & Husain, 2004), suggesting that music training may have some benefit for children's ability to recognise emotions in speech.

The acoustic cues used to convey emotion in speech are hypothesised to be similar to those used to convey emotion in music, as reported in a meta-analysis by Juslin and Laukka (2003). For example, both angry speech and angry music are characterised by fast speech rate/tempo, high volume, and high pitch level (Juslin & Laukka, 2003). Alternatively, sad speech and sad music may be characterised by slow speech rate/tempo, low volume, and low pitch level. In both music and prosody, tempo (speed of speech or music), pitch (the relative highness or lowness of a sound), and volume are the most common acoustic cues examined, due to their emotive capabilities (Juslin & Laukka, 2003; Kamien, 2004). With music, mode (the specific pitches used in a piece of music, where a major mode typically sounds happy, and a minor mode typically sounds sad) may also be used to determine emotion.

Interestingly, the characteristics of speech prosody in both tonal (e.g., Chinese, Thai, and Vietnamese) and non-tonal (e.g., English, French, and German) languages closely mirror the tonal characteristics in music in these different cultures (Han, Sundararajan, Bowling, Lake, & Purves, 2011). Additionally, music and speech prosody recruit overlapping networks in the brain, indicating that similar mechanisms are used to infer emotion from both music and speech (Escoffier, Zhong, Schirmer, & Qiu, 2013). Thus, it should come as no surprise that the acoustic cues that convey emotion are similar in both types of stimuli.

## Children's recognition of vocal cues of emotion

### Speech prosody

Although much of the focus of emotion recognition research has been on facial expressions, studies have examined children's and adults' capacity to understand other cues, finding that understanding of speech prosody may develop later than understanding expressions in other domains. Pre-

school children can recognise happiness, sadness, anger, and fear in facial expressions and postural cues. However, while emotions presented only in voice are well-recognised by adults, three–five-year-old children's recognition is poor, with the majority failing to choose the correct label (Nelson & Russell, 2011). Similarly, while recognition of emotion in faces reaches adult levels by 11 years, emotion recognition in voices continues to develop past this age (Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015).

Other research looking solely at voice supports the conclusion that children find it difficult to interpret emotions from voice prosody. When asked to label a speaker's emotion, where the speaker emotively says a nonsense word, five-year-olds' performance is at chance levels, while levels for nine- and 13-year-olds are better than chance, although not at adult levels (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013). Additionally, recognition between five years and 10 years improves for emotions such as pride, anger, contentment, disgust, and sadness (Sauter, Panattoni, & Happé, 2013), although when hearing a foreign language or low-pass filtered speech, children as young as four years can distinguish happy and sad emotional prosody (Morton & Trehub, 2001). This suggests that understanding prosody continues to develop throughout childhood.

### Affect bursts

Although children have difficulty judging emotions from speech prosody, children as young as five years old are proficient at judging emotions from non-verbal vocalisations, or affect bursts (Sauter et al., 2013). Affect bursts can include laughs, sighs, grunts, sniffles, and other emotive human sounds, and have been used to convey emotions such as pride, happiness, surprise, anger, disgust, fear, and sadness. With such vocalisations, little improvement between the ages of five and 10 years is seen, suggesting the relative ease with which even younger children understand such cues. This contrasts with the developmental improvement seen with prosody in inflected speech stimuli. That children recognise emotions from affective vocalisations earlier than emotions from speech prosody indicates that they are sensitive to auditory expressions of emotion, and may suggest an earlier evolutionary development of vocalisations before verbal communication and speech prosody (Jentschke, 2016; Juslin & Laukka, 2003; Scherer, 1995).

## Musical cues of emotion

Similar to their understanding of speech prosody, research suggests that children's recognition of emotions in music improves with age. When presented with music written specifically for children, three- to five-year-old children can identify happiness and sadness, with increases in ability throughout this age group (Franco, Chew, & Swaine, 2016). When played instrumental versions of children's

songs, children as young as four and five years of age rate fast songs as happier than slow songs, although three-year-olds do not do so (Mote, 2011). Additionally, children aged four, seven, and nine years show improvement with age in recognition of emotions in music, regardless of whether melodies were played on a musical instrument, or sung with non-sense syllables (Dolgin & Adelson, 1990). Older children performed better than younger children overall, although not as well as adults; seven- and nine-year-old children were on par with adults for recognition of happiness, but not for sadness, anger, or fear (Dolgin & Adelson, 1990).

This developmental trend is similar regardless of whether music is child-directed or adult-directed. When presented with pieces of Western classical music, three- and four-year-old children were unable to label happiness and sadness based on tempo or mode, while five-year-olds used only tempo to determine emotion, and six- to eight-year-olds, like adults, used both tempo and mode to determine emotion in a piece of music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). With this adult-directed music, only children aged five years and older could recognise the target emotion (Dalla Bella et al., 2001), similar to findings that children as young as four can match musical pieces with facial expressions, and children respond similarly to adults even when generating their own verbal labels (Nawrot, 2003). Five- to seven-year-olds also recognise emotions at above chance levels (Stachó, Saarikallio, Van Zijl, Huotilainen, & Toiviainen, 2013). Other research with Western classical music has found that five- to ten-year-old children recognise emotion from classical music excerpts better than from affect bursts, and this ability increases with age (Allgood & Heaton, 2015). Interestingly, this study identified a correlation between recognition of emotions in music and in affect bursts. Finally, when words and acoustic signals conflict, children younger than six years old judge emotion from the content of speech or lyrics, rather than prosody or music (Morton & Trehub, 2001, 2007; Morton, Trehub, & Zelazo, 2003).

The majority of current research on emotion recognition in music (including both studies of emotion recognition and felt emotion), when considering emotion categories tends to focus on the categories of happiness and sadness, although fear and anger have also been examined in a minority of studies (Eerola & Vuoskoski, 2013). This limited range of emotions ignores the potential differences in recognition between emotions, especially as happiness and sadness are typically the easiest to recognise across a range of emotive stimuli (Calligeros, Vidas, Nelson, & Dingle, 2018; Dolgin & Adelson, 1990; Stachó, Saarikallio, Van Zijl, Huotilainen, & Toiviainen, 2013; Widen, 2013).

## The current experiment

Given the similarities between music and speech prosody, developmental research is uniquely positioned to determine whether recognition of these cues develops in parallel.

Specifically, past research suggests that children's patterns of understanding emotions in music are related to their understanding of speech prosody. Thus, we examined children's and adults' recognition of emotion in Western classical music, speech prosody, and affect bursts. We included the commonly examined emotions of happiness, sadness, anger, and fear, as well as pride, which has been examined developmentally in speech and voice cues, but not in musical cues. We asked participants to label the emotion of the stimuli in a forced-choice task, using words instead of matching to faces, as emotion labels may result in more accurate categorisation than facial expressions (Russell & Widen, 2002).

We hypothesised that children's recognition of emotions in music and speech would develop in parallel. We also expected that overall, recognition would improve on all types of stimuli with age. We further hypothesised that children's recognition of the stimuli would depend on the emotions conveyed, as well as the type of cue provided.

## Method

### Participants

Participants were 60 children and 51 adults. Children were divided into three age groups. The younger age group ranged from six to seven years ($N = 22$, 31.8% female, $M_{age} = 6.6$ years), the middle age group ranged from eight to nine years ($N = 20$, 55% female, $M_{age} = 8.6$ years) and the older age group ranged from 10 to 11 years ($N = 18$, 44.4% female, $M_{age} = 10.5$ years). Adults ranged between 17 and 47 years ($N = 51$, 82.4% female, $M_{age} = 19.5$ years). Children were recruited while visiting a local museum, as well as from a database of families who volunteered to participate in developmental research at the university. Adults were undergraduate psychology students participating for course credit.

Of the 111 participants, more than half (56.8%) had some musical training beyond school classroom music. For six- to seven-year-olds, 27.3% had some musical training ($M = 0.4$ years; range = 1 month–3 years). For eight- to nine-year-olds, 55% had some musical training ($M = 1.3$ years; range = 1–6 years). For 10–11-year-olds, 72% had some musical training ($M = 2.4$ years; range = 1–6 years). Finally, for adults, 64.7% had some musical training ($M = 3.1$ years; range = 1–13 years).

A pilot group of adults was also used to evaluate the emotional content of musical excerpts. They were undergraduate psychology students participating for course credit, or were recruited via word of mouth. Adults ranged between 18 and 52 years ($N = 27$; 70.4% female, $M_{age} = 21.37$ years). Adults in the pilot group did not participate again in the main experiment.

### Materials

*Music.* The musical excerpts, lasting between 18 and 22 seconds, were drawn from pieces of Western classical

music, depicting five emotion categories; happiness, sadness, anger, fear, and pride. We selected musical pieces that fit descriptions of common acoustic features of emotions (Juslin & Laukka, 2003), then collected pilot ratings from adults. For pride, pilot ratings were particularly necessary. A total of 17 excerpts were tested, and we selected the two best-recognised excerpts for each emotion, resulting in a total of 10 excerpts (excerpts used in this project are available at: osf.io/v7msh).

*Speech prosody.* We selected 10 previously validated inflected speech stimuli, which depicted the same five emotion categories we included in the music excerpts: happiness, sadness, anger, fear, and pride (Sauter et al., 2013). There were two stimuli per emotion, one conveyed by a male voice, and one conveyed by a female voice. Each stimulus was between 2 and 3 seconds long, and consisted of spoken three-digit numbers.

*Affect bursts.* The 10 non-verbal vocalisations conveyed the same five emotion categories: happiness, sadness, anger, fear, and pride (Sauter et al., 2013). Again, there were two stimuli per emotion, one voiced by a male, and one voiced by a female. Each stimulus was approximately 1 second long, and included sounds such as laughing, sniffling, and grunting.

## Procedure

Child participants completed the survey on an iPad, and listened to stimuli using Sennheiser headphones. To ensure children understood the task, the experimenter read children the instructions and response options for each question. Adults completed the survey on a computer, with headphones. The stimuli were presented in three blocks. The first two blocks contained the variables of primary interest to our research – music and speech prosody – and were counterbalanced. Because previous literature indicates that children's recognition of affect bursts is more advanced than of other verbal cues, the third block always contained the affect bursts, to ensure that children's recognition of these vocalisations did not influence their ratings of the music and inflected speech cues. In addition, parents provided demographic details, such as age, sex, and whether their child had previous musical instruction.

*Music.* The music block featured the 10 excerpts, presented in a random order. Participants were asked to listen to each song clip, then answer the question "what is the mood of this song?", by selecting either happy, sad, angry, scared, or proud.

*Speech.* In the speech block, participants listened to 10 stimuli, presented in a random order. Participants answered the question "what emotion do you think this person is feeling?", by selecting either happy, sad, angry, scared, or proud.

**Table 1.** Averaged percentage of responses given for music stimuli.

| | | Emotion of stimulus | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Responses | *Happy* | *Sad* | *Angry* | *Scared* | *Proud* |
| 6–7-year-olds | Happy | **68.2** | 4.5 | 11.4 | 4.6 | 27.3 |
| | Sad | 11.4 | **75.0** | 2.3 | 9.1 | 2.3 |
| | Angry | 4.6 | 0.0 | **45.5** | 22.8 | 13.7 |
| | Scared | 2.3 | 9.1 | 18.2 | **54.6** | 0.0 |
| | Proud | 13.7 | 11.4 | 22.8 | 9.1 | **56.8** |
| 8–9-year-olds | Happy | **62.5** | 5.0 | 10.0 | 5.0 | 17.5 |
| | Sad | 7.5 | **80.0** | 0.0 | 2.5 | 2.5 |
| | Angry | 5.0 | 0.0 | **47.5** | 17.5 | 0.0 |
| | Scared | 5.0 | 10.0 | 30.0 | **72.5** | 0.0 |
| | Proud | 20.0 | 5.0 | 12.5 | 2.5 | **80.0** |
| 10–11-year-olds | Happy | **77.8** | 0.0 | 8.4 | 5.6 | 11.2 |
| | Sad | 2.8 | **97.2** | 0.0 | 0.0 | 0.0 |
| | Angry | 0.0 | 0.0 | **63.9** | 8.4 | 2.8 |
| | Scared | 0.0 | 2.8 | 19.5 | **83.4** | 0.0 |
| | Proud | 19.5 | 0.0 | 8.4 | 2.8 | **86.1** |
| Adults | Happy | **97.1** | 0.0 | 3.0 | 0.0 | 10.8 |
| | Sad | 0.0 | **89.2** | 0.0 | 0.0 | 0.0 |
| | Angry | 0.0 | 0.0 | **69.6** | 3.0 | 0.0 |
| | Scared | 0.0 | 8.9 | 13.8 | **96.1** | 0.0 |
| | Proud | 3.0 | 2.0 | 13.7 | 1.0 | **89.3** |
| Average correct | | 82.0 | 86.0 | 59.9 | 81.5 | 80.6 |

*Note.* Bolded results indicate the correct answer for each column.

*Affect bursts.* In the final block of questions, participants listened to 10 non-verbal vocalisations, presented in a random order. Participants answered the question "what emotion do you think this person is feeling?", by selecting either happy, sad, angry, scared, or proud.

## Results

### Descriptives

Recognition scores were calculated by averaging across the two stimuli presented for each emotion for each stimulus type. Table 1 outlines the percentage of participants in each age group who responded with each of the five possible answers for the five averaged music stimuli. Similarly, Table 2 shows the percentage of participants who responded with each of the five possible answers for the averaged inflected speech stimuli. Finally, Table 3 shows participants' responses to the five emotions conveyed by the affect bursts.

For most stimuli, adults chose the target emotion as expected, choosing the "correct" answer at above chance levels, with chance levels set at .20, and *p*-values Bonferroni corrected (.05/15) to be .003, $ts(50) \geq 9.46$, $ps < .001$, $ds \geq 1.32$. One exception was the proud affect bursts, for which adults were more likely to select "happy", and

**Table 2.** Averaged percentage of responses given for speech stimuli.

| | Responses | Emotion of stimulus | | | | |
|---|---|---|---|---|---|---|
| | | *Happy* | *Sad* | *Angry* | *Scared* | *Proud* |
| 6–7-year-olds | Happy | **70.5** | 2.3 | 11.4 | 11.4 | 27.3 |
| | Sad | 6.8 | **70.5** | 2.3 | 41.0 | 4.5 |
| | Angry | 4.5 | 2.3 | **72.8** | 0.0 | 29.5 |
| | Scared | 2.3 | 18.2 | 2.3 | **43.2** | 9.1 |
| | Proud | 15.9 | 6.8 | 11.4 | 4.6 | **29.6** |
| 8–9-year-olds | Happy | **77.5** | 0.0 | 7.5 | 2.5 | 25.0 |
| | Sad | 0.0 | **77.5** | 0.0 | 32.5 | 2.5 |
| | Angry | 0.0 | 5.0 | **60.0** | 0.0 | 27.5 |
| | Scared | 0.0 | 12.5 | 5.0 | **60.0** | 12.5 |
| | Proud | 22.5 | 5.0 | 27.5 | 5.0 | **32.5** |
| 10–11-year-olds | Happy | **77.8** | 0.0 | 0.0 | 0.0 | 27.8 |
| | Sad | 5.6 | **86.1** | 0.0 | 25.0 | 2.8 |
| | Angry | 0.0 | 2.8 | **66.7** | 0.0 | 19.5 |
| | Scared | 5.5 | 11.1 | 5.6 | **75.0** | 8.4 |
| | Proud | 11.1 | 0.0 | 27.8 | 0.0 | **41.7** |
| Adults | Happy | **89.2** | 0.0 | 0.0 | 6.9 | 12.8 |
| | Sad | 0.0 | **96.1** | 1.0 | 11.8 | 4.0 |
| | Angry | 0.0 | 0.0 | **84.4** | 0.0 | 20.6 |
| | Scared | 1.0 | 4.0 | 2.0 | **79.4** | 3.0 |
| | Proud | 9.8 | 0.0 | 12.7 | 2.0 | **59.8** |
| Average correct | | 81.5 | 86.0 | 74.8 | 68.0 | 46.0 |

*Note.* Bolded results indicate the correct answer for each column.

**Table 3.** Averaged percentage of responses given for affect burst stimuli.

| | Responses | Emotion of stimulus | | | | |
|---|---|---|---|---|---|---|
| | | *Happy* | *Sad* | *Angry* | *Scared* | *Proud* |
| 6–7-year-olds | Happy | **75.0** | 2.3 | 0.0 | 2.3 | 47.8 |
| | Sad | 11.4 | **93.2** | 4.5 | 18.5 | 6.8 |
| | Angry | 6.8 | 0.0 | **75.0** | 2.3 | 2.3 |
| | Scared | 0.0 | 2.3 | 20.5 | **74.7** | 15.9 |
| | Proud | 6.8 | 2.3 | 0.0 | 2.3 | **27.3** |
| 8–9-year-olds | Happy | **87.5** | 2.5 | 2.5 | 2.5 | 45.0 |
| | Sad | 7.5 | **82.5** | 5.0 | 20.0 | 10.0 |
| | Angry | 2.5 | 7.5 | **75.0** | 0.0 | 2.5 |
| | Scared | 0.0 | 2.5 | 17.5 | **77.5** | 12.5 |
| | Proud | 2.5 | 5.0 | 0.0 | 0.0 | **30.0** |
| 10–11-year-olds | Happy | **77.8** | 0.0 | 0.0 | 2.8 | 69.5 |
| | Sad | 11.1 | **91.7** | 5.6 | 5.6 | 8.4 |
| | Angry | 0.0 | 0.0 | **80.6** | 5.6 | 0.0 |
| | Scared | 0.0 | 0.0 | 13.9 | **86.1** | 8.4 |
| | Proud | 11.2 | 8.4 | 0.0 | 0.0 | **13.9** |
| Adults | Happy | **85.3** | 0.0 | 0.0 | 0.0 | 60.8 |
| | Sad | 9.8 | **99.0** | 7.9 | 3.0 | 2.0 |
| | Angry | 0.0 | 0.0 | **87.3** | 5.9 | 0.0 |
| | Scared | 2.0 | 1.0 | 4.0 | **91.2** | 9.8 |
| | Proud | 3.0 | 0.0 | 1.0 | 0.0 | **27.5** |
| Average correct | | 82.4 | 93.7 | 81.5 | 84.2 | 25.7 |

*Note.* Bolded results indicate the correct answer for each column.

selected "proud" at levels not different from chance, $t(50) = 1.58$, $p = .120$, $d = 0.22$. Across the three child age groups, recognition scores for all stimuli were above chance levels (*p*-values ranged from < .001 to .002) with a few exceptions. For proud speech and proud affect bursts, all three child age groups selected "proud" at levels not different from chance responding (*p*s ≥ .019). Finally, six-to seven-year-olds selected the target emotion for angry music ($p = .012$) and scared speech ($p = .006$) at levels not different from chance.

### Proud music

As adults did not choose the expected label for proud affect bursts, and demonstrated comparatively low recognition scores for proud speech, pride stimuli were analysed separately from the other emotion categories. As can be seen in Table 1, pride in music was well recognised, with 80.6% of participants choosing the correct label. In contrast, Tables 2 and 3 show that recognition was low for inflected speech (46.0%), and affect bursts (25.7%). For affect bursts, even adults judged the clips as sounding proud in only 27.4% of trials.

We conducted a series of one-way ANOVAs to examine the influence of age group (six–seven-year-olds, eight–nine-year-olds, 10–11-year-olds, and adults) on each

stimulus type (music, inflected speech, and affect bursts) for proud stimuli. For proud affect bursts, there were no differences between age groups, $F(3, 107) = .95$, $p = .420$, $\eta^2 = .026$. This error pattern can be seen in Table 3, where the general pattern shows that participants appeared to answer with "happy" rather than "proud". However, age influenced recognition of proud music, $F(3, 107) = 7.12$, $p < .001$, $\eta^2 = .166$, and proud speech, $F(3, 107) = 6.32$, $p = .001$, $\eta^2 = .150$. For proud music, six- to seven-year-olds ($M = 0.57$) were less likely to choose the target emotion than 10–11-year-olds ($M = 0.86$) and adults, ($M = 0.89$), *p*s ≤ .008, while all other age groups scored similarly, *p*s ≥ .052. For proud speech, the younger groups of children, (six–seven-year-olds, $M = 0.30$; eight–nine-year-olds, $M = 0.33$) were less likely to choose the target emotion than adults ($M = 0.60$), *p*s ≤ .010, and were all similar to 10–11-year-olds ($M = 0.42$), *p*s ≥ .255.

### Main analysis

To understand the differences in responses to emotions and types of stimuli across age groups, we conducted a mixed ANOVA. Age group (six–seven-year-olds, eight–nine-year-olds, 10–11-year-olds, and adults) was the between-groups variable, while the emotion of the stimulus (happy,

**Table 4.** Average percentage scores for each age group by type of stimulus.

| Age group | Music | Speech | Affect bursts |
|---|---|---|---|
| 6–7-year-olds | 60.8 | 64.2 | 79.0 |
| 8–9-year-olds | 65.6 | 68.8 | 80.6 |
| 10–11-year-olds | 80.6 | 76.4 | 84.0 |
| Adults | 88.0 | 87.3 | 90.7 |

*Note.* Excluding pride.

sad, angry, and scared) and type of stimulus (music, inflected speech, and affect bursts) were within-groups variables. The dependent variable was participants' recognition of the emotion in the stimulus.[1] We followed up all significant effects with Bonferroni-corrected post hoc *t*-tests; all post hoc test *p*-values presented are Bonferroni corrected.

A main effect of age group was found, $F(3, 107) = 26.07$, $p < .001$, $\eta^2_p = .42$. Post hoc tests revealed that adults ($M = 0.89$) were more accurate than all children, $ps \leq .025$. The two older groups of children (10–11-year-olds, $M = 0.80$; eight–nine-year-olds, $M = 0.72$) scored similarly, $p = .068$, and six–seven-year-olds ($M = 0.68$) scored similarly to eight–nine-year-olds, $p = 1.0$; however, they were less accurate than 10–11-year-olds, $p = .002$.

A main effect of the type of stimulus was found, $F(2, 214) = 19.11$, $p < .001$, $\eta^2_p = .15$. As predicted, there was no significant difference between scores on music ($M = 0.77$) and speech ($M = 0.78$), $p = 1.0$. In contrast, scores on affect bursts ($M = 0.85$) were significantly higher than for both music and speech, $ps < .001$.

The age by type interaction, $F(6, 214) = 3.01$, $p = .008$, $\eta^2_p = .08$, indicated that for music stimuli, six–seven-year-olds ($M = 0.61$) and eight–nine-year-olds ($M = 0.66$), eight–nine-year-olds and 10–11-year-olds ($M = 0.81$), and 10–11-year-olds and adults ($M = 0.88$) scored similarly, $ps \geq .107$. Both 10–11-year-olds and adults scored higher than the youngest group, $ps \leq .002$, and adults scored higher than eight–nine-year-olds, $p < .001$. For the inflected speech stimuli, six–seven-year-olds ($M = 0.64$) scored similarly to eight–nine-year-olds ($M = .69$), $p = 0.552$; 10–11-year-olds ($M = 0.76$) scored similarly to all groups, $ps \geq .424$; and adults ($M = .88$) scored higher on speech than the two younger groups of children, $ps \leq .001$. Finally, for the affect bursts, all age groups scored similarly, $ps \geq .108$. Table 4 summarises these results.

The age by type interaction also revealed that for six–seven-year-olds, scores on music ($M = 0.61$) and speech ($M = 0.64$) were similar, $p = 1.0$, while scores on affect bursts ($M = 0.79$) were significantly higher than scores on music and speech, $ps \leq .001$. For children aged eight to nine years, scores on speech ($M = 0.69$) were similar to music scores ($M = 0.66$) and affect burst scores ($M = 0.81$), $ps \geq .172$, while children scored higher on music compared with affect bursts, $p = .010$. In contrast, there was no

difference between scores on the three types of stimuli for either 10–11-year-olds or adults, $ps \geq .064$.

A main effect of the emotion of the stimulus was also found, $F(3, 321) = 18.22$, $p < .001$, $\eta^2_p = .15$. Participants scored higher on sad stimuli ($M = 0.89$) than on any other emotion, $ps \leq .022$. Scores on happiness ($M = 0.82$) and fear ($M = 0.78$) were similar, $p = .440$, while participants scored lower on anger ($M = 0.72$) than all emotions, $ps \leq .001$, except fear, $p = .059$.

Finally, a significant stimulus type by emotion interaction was found, $F(6, 642) = 5.72$, $p < .001$, $\eta^2_p = .05$. For the music stimuli, participants scored similarly for happy ($M = 0.82$), sad ($M = 0.86$), and scared music ($M = 0.82$), $ps = 1.0$, but lower for angry music ($M = 0.60$), $ps \leq .001$. For the inflected speech stimuli, participants scored similarly for happy speech ($M = 0.82$), sad speech ($M = 0.86$), and angry speech ($M = 0.75$), $ps \geq .065$. Participants scored significantly lower on scared speech ($M = 0.68$) than all other emotions, $ps \leq .005$, except angry speech, $p = 1.0$. Finally, for the affect burst stimuli, participants scored higher on sad affect bursts ($M = 0.94$) than angry affect bursts ($M = 0.82$), $p = .025$. Participants scored similarly on all other emotions, including happy ($M = 0.82$) and scared ($M = 0.84$), $ps \geq .065$.

The emotion by age group interaction was non-significant, $F(9, 321) = 1.60$, $p = .113$, $\eta^2_p = .04$, as was the three-way interaction of type of stimulus, emotion, and age group, $F(18, 642) = 1.27$, $p = .198$, $\eta^2_p = .03$.

## Additional analyses

We next examined the relationship between children's recognition of emotion stimuli and their music training experience. Recognition scores were calculated for each stimulus type by summing children's correct responses for the stimuli, excluding pride. Correlations between variables are presented in Table 5. Age, in years, was correlated with all variables; musical training, total music score, total speech, and total affect bursts. As expected, recognition scores for music and speech were positively correlated. Scores for music were also correlated with scores for affect bursts, as were scores for speech.

To determine whether scores on the music variable were related to scores of speech and affect burst variables, independent of participants' age and musical training, a hierarchical multiple regression was conducted. The predictors Age and Years of Musical Training were entered into the model at Step 1, while Speech and Affect Bursts recognition scores were entered at Step 2.[2]

The combined effects of age and musical training at Step 1 explained 29.9% of variance in music scores, $F_{ch}(2, 108) = 23.02$, $p < .001$. Inspection of the squared semi-partial correlations revealed that age uniquely explained 23% of the variance, $\beta = .49$, $p < .001$, while musical training explained a non-significant 2% of the variance, $\beta = .15$, $p = .072$.

**Table 5.** Inter-correlations between variables.

| Variables | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Age | | | | |
| 2. Years musical training | .223* | | | |
| 3. Total correct music | .527** | .260** | | |
| 4. Total correct speech | .402** | .246** | .538** | |
| 5. Total correct affect bursts | .354** | .068 | .371** | .434** |

*Note.* The music, speech, and affect bursts variables were calculated as total correct of each stimuli, excluding pride. Age was measured continuously. Variables 3 to 5 measured as total instances of each stimuli where answer matched target emotion.
\* $p < .05$. \*\* $p < .01$.

At Step 2, participant scores for speech and affect bursts stimuli were included in the model, explaining a total of 42% of the variance, $F_{ch}(2, 106) = 11.04$, $p < .001$. Scores for speech and affect bursts thus accounted for 12.1% of variance in music scores once age and musical training were controlled for. For this step, we found that speech scores uniquely explained 8% of the variance, $\beta = .34$, $p < .001$, while affect bursts uniquely explained only 1% of the variance, $\beta = .10$, $p = .236$.

## Discussion

The aim of the present study was to examine whether recognition of emotion in music and speech prosody develops in parallel. We examined children's and adults' recognition of emotion in music, speech, and affect bursts. Additionally, we included the emotions happiness, sadness, anger, and fear, as well as pride, which had yet to be examined in music.

Previous literature has hypothesised that the acoustic cues used to convey emotion in speech and music are related (Juslin & Laukka, 2003), and that emotional inferences from vocalisations and music recruit overlapping networks in the brain (Escoffier et al., 2013). Our data suggest that children's recognition of emotions in music and speech develop in parallel, with both music and speech scores for children aged six to nine years being similar, but for the youngest children, lower than their scores on affect bursts. By 10 years old, children were as likely to select the target emotion as adults for music and speech, while all age groups still scored similarly for affect bursts. The similar pattern of music and speech recognition may be due to the similar performance cues reported by Juslin and Laukka (2003), such as tempo, loudness, and timbre. For example, in the speech stimuli conveying fear, the actors spoke quickly and at a high pitch, while the music stimuli conveying fear were characterised by a fast tempo and instruments playing at a high pitch. The affect bursts conveyed fear through a high-pitched scream, arguably a more primitive vocalisation (Scherer, 1995).

Children's recognition did depend on the type of stimulus, with music and speech stimuli being overall less well recognised than affect bursts. That children's recognition was higher for affect bursts than speech is consistent with previous research (e.g., Sauter et al., 2013). In contrast, our results conflict with findings that children recognise emotions better in music than in affect bursts (Allgood & Heaton, 2015). That children recognise emotions from affect bursts earlier than from speech prosody may also reflect an earlier evolutionary development of vocalisation before verbal communication and speech prosody (Jentschke, 2016; Juslin & Laukka, 2003; Scherer, 1995).

As hypothesised, recognition on all types of stimuli improved with age. We found that six–seven-year-olds and eight–nine-year-olds performed similarly, as did eight–nine-year-olds and 10–11-year-olds. All children were less likely to select target emotions than adults. This suggests that, as found in previous research, children's recognition of emotions generally improves throughout childhood (Aguert et al., 2013; Nelson & Russell, 2011; Widen, 2013). A variety of developmental aspects likely influence children's increased emotion recognition skills, including increased cognitive capacity (Izard, 2011), emotion vocabulary (Nook, Sasse, Lambert, McLaughlin, & Somerville, 2017), social experiences (Widen, 2013) and theory of mind skills (Wellman & Liu, 2004).

We further hypothesised that children's recognition of the emotions conveyed would vary with the type of cue provided. For music, happiness, sadness, and fear were equally recognisable, while anger was more difficult, partially echoing previous research, where anger or fear are typically least recognised (Dolgin & Adelson, 1990; Gaŝpar et al., 2011). For inflected speech, happiness and sadness were the most recognised, followed by anger, while fear was the least recognised. Some previous research with other types of stimuli have identified similar patterns (e.g., Nelson & Russell, 2011; Widen, 2013); however, our findings differed from earlier research using the same stimuli (Sauter et al., 2013). This may be due to our differing method – children in our study matched each clip to one of five emotion words. In the original study, children chose from four faces to match to the clip, where one face was correct, and one each was displayed of a positive, negative, and neutral category. We chose to use emotion labels, as previous research shows labels are recognised earlier than facial expressions (Russell & Widen, 2002), although future research might examine children's performance in matching music stimuli with facial expressions. Finally, for affect bursts, sadness was more recognised than anger, while scores for all other emotions were still high.

In addition, we explored whether musical training impacted children's recognition, as previous research suggests that both adults and children with musical training recognise emotions in speech better than their non-musical peers (Thompson et al., 2004). Although musical training was related to correct labelling of both music and speech stimuli, it explained only a small part of music recognition, suggesting only limited benefits in emotion recognition in

music. Our results may reflect differences in the measurement of musical training, or differential effects of musical training across ages; however, further investigation into this relationship is required to solidify this link.

This study is the first to show that people perceive pride in music, with the proud music recognised in approximately 81% of trials. As with the other emotions, a developmental pattern emerged: children's recognition of pride improved with age, such that by eight to nine years, children scored similarly to adults. Although we were unable to include pride in our main analysis due to adult performance for the proud affect bursts, our data do suggest that there is room in the music literature to branch out beyond using happy and sad music (Eerola & Vuoskoski, 2013), to include other emotions such as fear, anger, and pride. Many of the characteristics of proud music may be similar to those of happy music, such as fast tempo and high vocal intensity (Juslin & Laukka, 2003), so it is unclear why participants almost unanimously agreed that these pieces conveyed pride. We may speculate that research using a dimensional model, of valence, arousal, and dominance, may allow clearer separation of happiness and pride in music (e.g., Russell & Mehrabian, 1977). We also speculate that participants answered "proud" due to iconic coding; that is, due to the similarities between proud speech and proud music, although without further research, we cannot conclude that these answers were not due to associative coding, as the types of excerpts chosen may have been paired with pride-inducing events (Juslin, 2013).

The present study had limitations. Firstly, our study did not include a "none-of-the-above" response option for participants to choose, which can artificially inflate emotion recognition scores (e.g., Nelson & Russell, 2016; Russell, 1994). As this was the first study to compare children's recognition across these three types of auditory stimuli, we wanted to determine whether children were able to match the stimuli to their intended emotion labels. Future research should include a none-of-the-above response to further investigate the influence of response options on emotion judgments of auditory emotion. Secondly, due to limitations in children's attention, the number of stimuli per emotion in each condition was limited to two, which may limit the generalisability of our results. Future research might consider including more stimuli. Thirdly, due to the nature of the stimulus types, the durations of the three types of clips were different (approximately 20 seconds for music, 2 seconds for prosody, and 1 second for affect bursts). It could have been the case that the longer length of the music stimuli made recognition easier for participants vs. the other stimulus types. However, recognition of music and speech was similar, and the shortest stimuli – affect bursts – elicited the highest levels of recognition, suggesting that clip duration did not influence recognition. Future research could examine whether shorter music clips produce similar results. Additionally, we presented affect bursts to participants as the last block, because children recognise affect bursts better than speech or music cues (Calligeros et al., 2018; Sauter et al., 2013). We made this choice so that children's responses to the affect burst stimuli wouldn't influence their responses to the speech and music stimuli; however, future research should counterbalance the three stimulus types to determine whether affect burst stimuli influence responding to other stimuli. Finally, our sample size included 20 children per age group, and although observed power was greater than .80 for all analyses (range = .85–1.0), with an alpha of .05, future research might include a greater sample size of children.

It has been suggested that music making may have evolved as a tool for experiencing shared intentionality, and maintaining social cohesion (Cross, 2016; Kirschner & Tomasello, 2010; Patel, 2008). It is possible that given the importance of auditory communication in social environments, individuals' understanding of the emotions conveyed in music and speech is linked with shared intentionality or social cohesion, especially for an emotion like pride, which speculatively could assist in rallying teams and tribes before battle or in victory. Indeed, music and language may have dependent evolutionary and developmental origins (Jentschke, 2016). Additionally, the brain structures used in identifying emotions in music and speech have been implicated in theory of mind (Escoffier et al., 2013), and the development of emotion competence, including emotion recognition, is related to social competence (Denham et al., 2003). Given these results, and our finding that recognition of emotion in music and speech develops in parallel, future research could more directly examine the relationship between social cohesion, and children's understanding of emotions in music and speech.

In the present study, we found that recognition of emotions in music and speech develop in parallel, and later than emotion recognition in affect bursts, or non-verbal vocal stimuli. We examined children's and adults' understanding of emotion across these stimuli for the emotions happiness, sadness, anger, fear, and pride. Pride, which had yet to be examined for perceived emotion in music, showed similar scores for children aged eight years old and over and adults.

## Author note

Dianna Vidas is now studying at Goldsmiths, University of London.

## Contributorship

DV and NN conceived the study, DV researched literature, tested participants, analysed data and wrote the first draft. All authors approved and edited the final version of the manuscript.

## ORCID iD

Dianna Vidas  http://orcid.org/0000-0003-0574-5610

## Peer review

Petri Laukka, Stockholm University, Department of Psychology. Sebastian Jentschke, Universitetet i Bergen Det Psykologiske Fakultet, Institutt for biologisk og medisinsk psykologi.

## Supplemental material

The supplemental material is available online with the article.

## Notes

1. An analysis using unbiased hit rates as the dependent variable produced results identical to those using raw scores.
2. For all regression analyses, measures of multicollinearity were acceptable. Tolerance values ranged between .714–.950, and variance inflation factor values ranged between 1.052–1.400.

## References

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Le Bigot, L. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE*, 8(12), 1–8. doi:10.1371/journal.pone.0083657

Allgood, R., & Heaton, P. (2015). Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *British Journal of Developmental Psychology*, 33(3), 398–403. doi:10.1111/bjdp.12097

Calligeros, R., Vidas, D., Nelson, N. L., & Dingle, G. A. (2018). Development of emotion recognition in music and vocal bursts. Manuscript in preparation.

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, 33(2), 218–236. doi:10.1111/bjdp.12075

Cross, I. (2016). The nature of music and its evolution. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 3–17). Oxford, England: Oxford University Press. doi:10.1093/oxfordhb/9780198722946.001.0001

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(3), 1–10. doi:10.1016/S0010-0277(00)00136-0

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development*, 74(1), 238–256. doi:10.1111/1467-8624.00533

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(1), 87–98.

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, 30(3), 307–340. doi:10.1525/MP.2012.30.3.307

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, 34(8), 1796–1810. doi:10.1002/hbm.22029

Franco, F., Chew, M., & Swaine, J. S. (2016). Preschoolers attribution of affect to music: A comparison between vocal and instrumental performance. *Psychology of Music*. Advance online publication. doi:10.1177/0305735616652954

Gaŝpar, T., Labor, M., Jurić, I., Dumančić, D., Ilakovac, V., & Heffer, M. (2011). Comparison of emotion recognition from facial expression and music. *Collegium Antropologicum*, 35(Suppl 1), 163–167. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21648329

Han, S., Sundararajan, J., Bowling, D. L., Lake, J., & Purves, D. (2011). Co-variation of tonality in the music and speech of different cultures. *PLoS ONE*, 6(5). doi:10.1371/journal.pone.0020160

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, 3(4), 371–378. doi:10.1177/1754073911410737

Jentschke, S. (2016). The relationship between music and language. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 343–355). Oxford, England: Oxford University Press. doi:10.1093/oxfordhb/9780198722946.001.0001

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4(SEP), 1–14. doi:10.3389/fpsyg.2013.00596

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Sloboda, J. (2010). Introduction: Aims, organization, and terminology. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion* (1st ed., pp. 3–12). Oxford, England: Oxford University Press.

Kamien, R. (2004). Sound: Pitch, dynamics, and tone color. In *Music: An Appreciation* (8th ed., pp. 6–13). New York, NY: McGraw-Hill.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5), 354–364. doi:10.1016/j.evolhumbehav.2010.04.004

Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews*, 6(1), 11–22. doi:10.1016/j.plrev.2008.08.003

McDermott, J. H., & Hauser, M. D. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, 23(1), 29–59. doi:10.1525/mp.2005.23.1.29

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Emotion*, *72*(3), 834–843. doi:10.1111/1467-8624.00318

Morton, J. B., & Trehub, S. E. (2007). Children's judgements of emotion in song. *Psychology of Music*, *35*, 629–639. doi:10.1177/0305735607076445

Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development*, *74*(6), 1857–1868. doi:10.1046/j.1467-8624.2003.00642.x

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*(3), 618–622. doi:10.1037/a0022573

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music*, *31*(1), 75–92. doi:10.1177/0305735603031001325

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52–61. doi:10.1016/j.jecp.2011.03.014

Nelson, N. L., & Russell, J. A. (2016). A facial expression of pax: Assessing children's "recognition" of emotion from faces. *Journal of Experimental Child Psychology*, *141*, 49–64. doi:10.1016/j.jecp.2015.07.016

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881–889. doi:10.1038/s41562-017-0238-7

Patel, A. D. (2008). Evolution. In *Music, language, and the brain* (pp. 355–415). New York, NY: Oxford University Press.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? *A review of the cross-cultural studies.* Psychological Bulletin, *115*(1), 102–141. doi:10.1037/0033-2909.115.1.102

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294. doi:10.1016/0092-6566(77)90037-X

Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development*, *11*(1), 30–52. doi:10.1111/1467-9507.00185

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97–113. doi:10.1111/j.2044-835X.2012.02081.x

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248. doi:10.1016/S0892-1997(05)80231-0

Spencer, H. (1881). Mr. Herbert Spencer on the origin and function of music. *The Musical Standard*, *21*(909), 427.

Stachó, L., Saarikallio, S., Van Zijl, A., Huotilainen, M., & Toiviainen, P. (2013). Perception of emotional content in musical performances by 3–7-year-old children. *Musicae Scientiae*, *17*(4), 495–512. doi:10.1177/1029864913497617

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)*, *4*(1), 46–64. doi:10.1037/1528-3542.4.1.46

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. doi:10.1111/j.1467-8624.2004.00691.x

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72–77. doi:10.1177/1754073912451492