UvA-DARE (Digital Academic Repository)

Children's recognition of emotions from vocal cues

Sauter, D.A.; Panattoni, C.; Happé, F.

Link to publication

Children's recognition of emotions from vocal cues

Disa A. Sauter[1][*], Charlotte Panattoni[2] and Francesca Happé[2]
[1]Department of Social Psychology, University of Amsterdam, The Netherlands
[2]Institute of Psychiatry, King's College, London, UK

Emotional cues contain important information about the intentions and feelings of others. Despite a wealth of research into children's understanding of facial signals of emotions, little research has investigated the developmental trajectory of interpreting affective cues in the voice. In this study, 48 children ranging between 5 and 10 years were tested using forced-choice tasks with non-verbal vocalizations and emotionally inflected speech expressing different positive, neutral and negative states. Children as young as 5 years were proficient in interpreting a range of emotional cues from vocal signals. Consistent with previous work, performance was found to improve with age. Furthermore, the two tasks, examining recognition of non-verbal vocalizations and emotionally inflected speech, respectively, were sensitive to individual differences, with high correspondence of performance across the tasks. From this demonstration of children's ability to recognize emotions from vocal stimuli, we also conclude that this auditory emotion recognition task is suitable for a wide age range of children, providing a novel, empirical way to investigate children's affect recognition skills.

In their interactions with others, children, just like adults, need to interpret a wide range of social signals to understand the intentions and feelings of others. In order for the child to respond appropriately, several layers of functioning must be in place. The ability to discriminate between social signals is thought to develop very early, with 5-month-old infants responding differentially to infant-directed approvals and prohibitions, even in an unfamiliar language (Fernald, 1993). These discriminations can be made on the basis of low-level differences between stimulus classes, such as pitch, whereas connecting social cues with environmental events or verbal labels requires more sophisticated cognitive capacities (Walker-Andrews, 1997). Most research investigating the development of emotion recognition has focused on children of preschool age upwards (but see Denham & Couchoud, 1990b; Widen & Russell, 2003, 2008, for studies of younger children), and research into children's (and, indeed, adults') understanding of affective cues has largely focused on the recognition of facial expressions.

---

[*] Correspondence should be addressed to Disa Sauter, Department of Social Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands (e-mail: d.a.sauter@uva.nl).

Children's recognition of emotional facial expressions

Despite considerable differences in the paradigms, age groups, and emotional states that have been examined, studies have consistently found that children from the age of 3–4 years can reliably match facial expressions to verbal labels for emotions such as anger, sadness, happiness, and surprise (Bullock & Russell, 1984; 1985; Harrigan, 1984; Reichenbach & Masters, 1983; Russell & Widen, 2002; Wellman, Phillips, & Rodriguez, 2000). The ability to associate emotional facial expressions with external causes appears to develop around 5 years (Hughes & Dunn, 1997; Pons, Harris, & de Rosnay, 2004), although 2- to 3-year olds can already perform better than chance in a two-way forced-choice task with one positive and one negative option (Wellman *et al.*, 2000). However, young children's recognition performance is substantially less accurate than adults' (Felleman, Barden, Carlson, Rosenberg, & Masters, 1983; Field & Walden, 1982), and the understanding of emotional signals improves considerably with age throughout childhood (Denham & Couchoud, 1990b; Harrigan, 1984; Pons *et al.*, 2004; Thomas, De Bellis, Graham, & LaBar, 2007). The type of task used has been shown to affect performance significantly: Children do better when choosing between labels than between facial expressions (Camras & Allison, 1985; Russell & Widen, 2002), and with forced-choice recognition compared to free labelling tasks (Harrigan, 1984; Markham & Adams, 1992). Recognition accuracy has also been found to vary by emotion, with happiness typically being the most easily identified facial expression (Camras & Allison, 1985; Denham & Couchoud, 1990a), and children finding emotions such as surprise, shame, and contempt more difficult to decode (Wiggers & van Lieshout, 1985). In sum, young children can typically reliably infer at least some emotional states from facial expressions and it is clear that performance is affected by the children's age, the task employed, and the specific emotions investigated.

Emotional cues in the voice

In contrast to the wealth of research into children's understanding of facial expressions of emotions, 'How the processing of emotional prosody develops during ontogeny is only poorly understood' (Grossmann & Johnson, 2007, p. 912). Research into early development has shown that the valence of vocal cues can affect infants' attention to objects and exploration behaviour (Moses, Baldwin, Rosicky, & Tidball, 2001; Mumme, Fernald, & Herrera, 1996; Parise, Cleveland, Costabile, & Striano, 2007; Vaish & Striano, 2004), and some studies with children have found that young children can accurately classify prosody as joy or sadness in a two-way forced-choice task (Morton & Trehub, 2001; see also Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). However, this does not address the issue of whether children can meaningfully interpret specific vocal-affective cues beyond broad valence-based judgements.

Only a handful of studies of emotion recognition in children to date have included vocal stimuli from several emotions. In an early study, children's accuracy in recognizing nine emotions from semantically neutral speech was found to improve steadily with age (Gates, 1927). Another early study found that children as young as five performed above chance levels when asked to select a stick-figure drawing to match an emotional speech segment expressing one of four emotions, although performance improved considerably with age (Dimitrovsky, 1964). The age-related improvement has also been replicated in a study comparing the perception and production of emotional prosody in children between 6 and 10 years (Nowicki & Duke, 1994). In addition, children were found to be

better at perception than production, although performance on the two sub-tasks was unrelated.

Several studies that have explicitly compared the presentation of affective information from facial and auditory cues have suggested that the auditory modality may be especially important for the communication of emotions to younger children (see also Akhtar & Gernsbacher, 2008; Baldwin & Moses, 1996). Children between the ages of 5 and 18 years were exposed to either auditory-only, visual-only, or audio-visual natural parent–child interactions (Bugental, Kaswan, Love, & Fox, 1970). For younger children (up to 8 years), the auditory channel was found to be more important, in that the visual component of a message was of less importance than the inflection of the speech and the verbal contents. In contrast, older children relied more on the visual cues (see also Stifter & Fox, 1987). A study including consistent facial and vocal cues found that children as young as 3 years perform considerably better than would be expected by chance when asked to pair one of four drawn facial expressions with a target emotion (Denham & Couchoud, 1990a). The target was presented in an ambiguous story, accompanied by facial cues and vocalizations consistent with one of the interpretations. Unfortunately, this design did not allow for the contributions of auditory and visual information to children's interpretations to be distinguished, since these cues were always congruent.

Sex differences

Sex differences have been reported in children's emotion recognition from facial cues, with females typically at an advantage (see McClure, 2000 for a meta-analysis). In addition, Baron-Cohen's 'empathizing-systemizing' theory argues that better emotion recognition should be associated with the female 'brain type' (Baron-Cohen, 2002; Goldenfeld, Baron-Cohen, & Wheelwright, 2006).

The current study

The current study aimed to investigate recognition of a wide range of vocally expressed emotions across early and middle childhood to provide empirical data on the developmental trajectory of emotion recognition from non-verbal vocalizations as well as speech prosody in typically developing children. Children ranging from 5 to 10 years were asked to select the one of four picture stimuli that best matched the emotion communicated in an auditory stimulus.

We also aimed to address a limitation in current research on emotion processing based on face stimuli. The emotions typically assessed with face stimuli are biased towards negative emotions, with most of the emotions reliably recognized from the face being negative (angry, sad, disgust, and fear), while only one positive emotion is commonly included (happiness; see Sauter, 2010 for a discussion). In the voice, however, positive as well as negative emotions can be reliably distinguished by adult listeners (e.g., Sauter & Scott, 2007; Sauter, Calder, Eisner, & Scott, 2010). The use of vocal emotion stimuli, therefore, allows for a test of whether, for example, happiness has been easy for children to identify from facial expressions in previous studies because it has been the only positive emotion included.

The study aimed to test two main hypotheses. Children were expected to perform at better-than-chance levels in the recognition of emotions from vocal cues. This prediction for vocal cues was based on the fact that children from 5 years have a good understanding of causes of emotions as well as visual-affective cues (Hughes & Dunn, 1997; Pons *et al.*,

2004). This pattern was expected to extend to the auditory domain in terms of both non-verbal vocalizations and speech prosody, particularly given earlier findings, showing that even younger children could identify happiness and several negative expressions from vocal signals, although no study to date has yet examined a wide range of positive and negative emotional vocal cues. Second, we hypothesized that recognition performance would improve with age, given findings from studies of facial expression recognition (Harrigan, 1984; Pons *et al.*, 2004) as well as emotional speech (e.g., Gates, 1927).

Based on the work by Pons *et al.* (2004), Hughes and Dunn (1997), and others, we set the lower age range for our study at 5 years, and included children up to the age of 10. Although younger children than this have been found to identify emotions accurately in some studies (e.g., Wellman *et al.*, 2000), our design involved a wider range of more subtle emotions, and a forced choice from four options, including foils of the same valence as the target. We therefore set the lower age to 5 years.

In addition, two exploratory hypotheses were considered. Girls were expected to perform better than boys, given results from studies of emotion recognition from facial expressions (McClure, 2000). Finally, it was expected that performance on the two auditory emotion tasks would be related, reflecting individual differences in underlying affect recognition, and maybe wider socio-affective skills.

Methods

Participants

The sample consisted of 48 children (25 boys) between the ages of 5 and 10 years (mean age 7 years 8 months, standard deviation 1 year 7 months). The participants were divided into two groups, according to age: 24 younger children (13 boys, mean age 6 years 3 months, range 5 years 1 month to 7 years 5 months), and 24 older children (12 boys, mean age 9 years 2 months, range 8 years 3 months to 10 years 3 months). Participants were recruited from a mainstream primary school in London, with a catchment area of mixed socio-economic status. No history of auditory, visual, or behavioural problems was reported for any participant. The project was reviewed and approved by the King's College London Psychiatry, Nursing & Midwifery Research Ethics Sub-Committee (PNM/07/08–9), and informed consent was collected from the participants and their parents.

Stimuli

The vocal stimuli were taken from previously validated affective non-verbal vocalizations and inflected speech stimuli (Sauter, 2006; Sauter *et al.*, 2010), and expressions were produced by two male and two female adult British English speakers. These stimuli are well recognized by adult listeners (Sauter, 2006; Sauter *et al.*, 2010; Sauter & Scott, 2007). The vocalizations were sounds such as laughs, sighs, and grunts. The verbal sounds consisted of affectively inflected speech in the form of spoken three-digit numbers. The mean duration of the non-verbal stimuli was 0.97 s, and for the speech stimuli 1.87 s. Each set was scaled for maximum amplitude. Ten categories of verbal and non-verbal stimuli were used, consisting of positive (Amusement, Contentment, Relief, and Achievement), negative (Anger, Disgust, Fear, and Sadness), and neutral (Surprise and Neutral) sounds.

The response picture stimuli were photos of people against a neutral background, with one picture used for each emotion category throughout the tasks. The pictures were

pilot tested with eight adults, using a forced-choice task with a larger set of candidate pictures. All adult pilot participants (100%) accurately labelled all of the pictures included in the final set. The pictures were arranged in sets of four for presentation. Of the four images, one picture matched the category of the sound, and constituted the target. The other images were from one positive, one negative, and one neutral category. The constellation of the distractor categories was varied across trials of each category within each task, with each picture being presented an approximately equal number of times. The position of the target picture on the page was counterbalanced. Under each picture was the corresponding emotion label and a letter used for participants' responses (A, B, C, or D).

Procedure

Testing was done individually in a quiet room on the school campus. Before commencing the experiment, the researcher gave a brief, standardized definition, and scenario explaining each emotion label. The child was then asked to give another example of when someone might feel each of these emotions, and asked to produce a corresponding sound. If an example or a sound was incorrect, they were not given feedback, but a second standardized scenario explaining the emotion was given. At that point, the participant was again asked to give another example and corresponding sound. All participants were able to give adequate responses following this process.

Following this procedure, each child performed the two emotion recognition tasks. Consistent with most of the previous literature, the current study employed a forced-choice paradigm (e.g., Camras & Allison, 1985; Denham & Couchoud, 1990a; Russell & Widen, 2002; Thomas *et al.*, 2007). As auditory stimuli have to be presented sequentially, the vocal cue was always the target, and the response alternatives were labels, illustrated with facial cues. In each task, the participants heard 50 sounds per task (five exemplars from each of the 10 emotion categories, with half of the stimuli produced by male and female speakers, respectively), and a total of 100 trials. The order of tasks was counterbalanced between participants. The stimuli were presented in a pseudo-random order (examples of a same emotion category had to be separated by at least three other emotion sounds) within each task, consistent across participants.

On each trial, participants were asked to choose which of the four emotion labels best matched the emotional sound. Responses were given either by saying the letter underneath the picture, or pointing to indicate the selected response, resulting in a dependent measure of the child's selected category for a given stimulus. No time limitation was used, and participants were allowed to hear a stimulus more than once. All children gave a response on all trials. The sounds were presented via standard speakers on a Hewlett Packard Pavilion dv1000 laptop (Palo Alto, CA, USA) and the pictures were presented on printed sheets. Sessions lasted approximately 20 min per child. After completion of the experiment, the child was told that they had done well, and was given a sticker to thank them for their participation.

Statistical analyses

In order to test the primary hypothesis that children perform better than chance in the recognition of emotions from vocal cues, performance for each age group was calculated for each emotion. In forced-choice tasks, performance for a particular category can be artificially inflated by the disproportionate use of that response. Unbiased hit rates,

Hu scores, control for this bias. For Hu scores, a score of zero denotes no correct responses, and one is perfect performance (Wagner, 1993). Chance scores can also be calculated separately for each condition using this method (see Wagner, 1993). Hu scores are proportional and are arcsine-transformed before use in statistical tests. Thus, to test whether the children in each age group performed better than chance for each stimulus type, arcsine-transformed Hu scores and chance scores were compared using paired-sample *t*-tests.

Our second hypothesis proposed that recognition performance improves with age. In order to test this relationship, linear regressions were performed with the children's age in months as the independent measure, and the total number of correct responses as the dependent measure, with non-verbal vocalizations and speech analysed separately. In order to examine whether performance improved with age for individual emotions, separate linear regressions were run for each emotion using the total number of correct responses for that emotion category, separately for the non-verbal vocalization and speech tasks.

In order to test our exploratory hypothesis of a sex difference in the recognition of emotion from vocal cues, a repeated-measures ANOVA was performed, using the average Hu scores on each task, minus chance, as an unbiased measure of performance.

Finally, to investigate whether children's comprehension of non-verbal vocalizations and the inflected speech were related, a correlation was performed, using each participant's average Hu score, less chance for each task. Since regression analyses showed that the performance on both tasks was to some extent related to the children's age, a partial correlation was performed, controlling for the effect of age.

## Results

Do children recognize emotions in the voice?

Non-verbal vocalizations

The children overall identified the emotional sounds with an average accuracy of 78.1% for the younger children and 83.9% for the older children (see Figure 1). This level of performance is much higher than would be expected by chance in a four-way forced-choice task (approximately 25%; note that Hu-score-calculated chance levels were used in the statistical tests). The full confusion matrices are shown in Appendix A.

The children in each age group performed better than chance for each category of non-verbal emotional vocalizations, corrected for multiple comparisons (see Appendix B), demonstrating that children as young as 5 years can reliably infer emotions from non-verbal vocal cues.

Inflected speech

The children overall identified the emotion from inflected speech with an average accuracy of 52.9% for the younger children and 72.3% for the older children (see Figure 2 for performance and Appendix A for confusion matrices). In both age groups, performance was better than chance for each emotion in the emotionally inflected speech task (see Appendix B), showing that children as young as 5 years can infer emotional states from emotionally inflected speech.
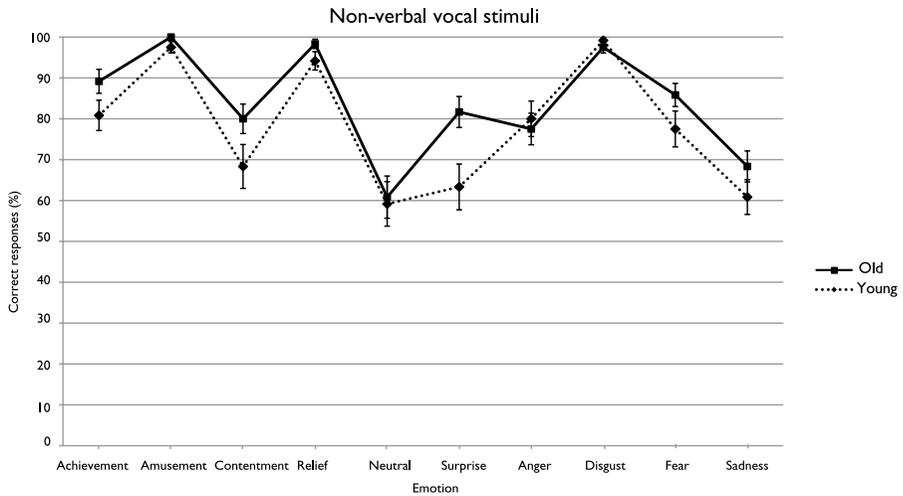
Figure 1. Percent correct in a four-way forced-choice task of 10 categories with non-verbal emotional vocalization stimuli, by younger (5- to 7-year old) and older (8- to 10-year old) children. *Note.* Chance is approximately 25% (statistical tests used Hu-score-calculated chance levels that take into account participant-level response biases, see analysis for details), error bars denote standard errors.

Does emotion recognition from vocal cues improve with age?

Non-verbal vocalizations

Age significantly predicted the children's recognition of emotions from non-verbal vocalizations (see Figure 3), accounting for just under a quarter of the variance in scores ($r^2 = .23, p < .001$). Examining individual emotions, age was found to predict the
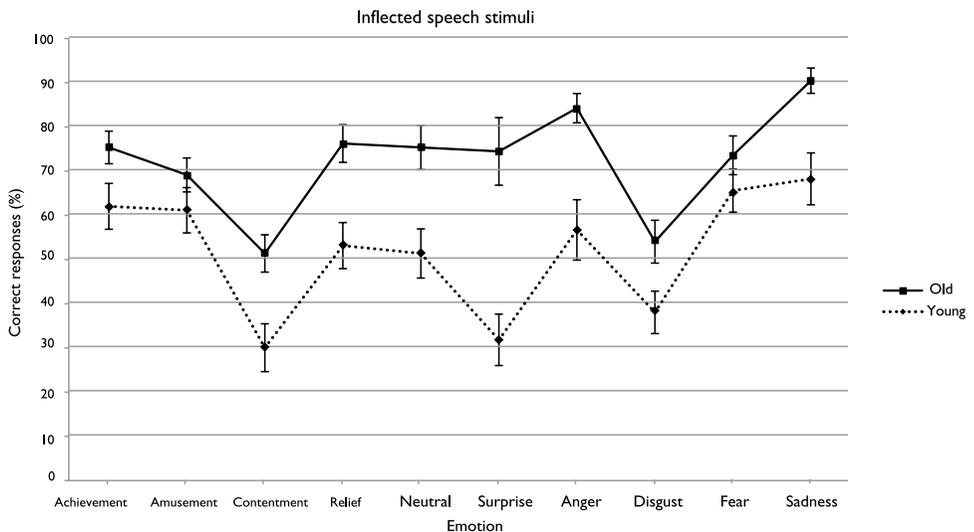
Figure 2. Percent correct in a four-way forced-choice task of 10 categories with emotionally inflected speech stimuli, by younger (5–7 years) and older (8–10 years) children. *Note.* Chance is approximately 25% (statistical tests used Hu-score-calculated chance levels that take into account participant-level response biases, see analysis for details), error bars denote standard errors.
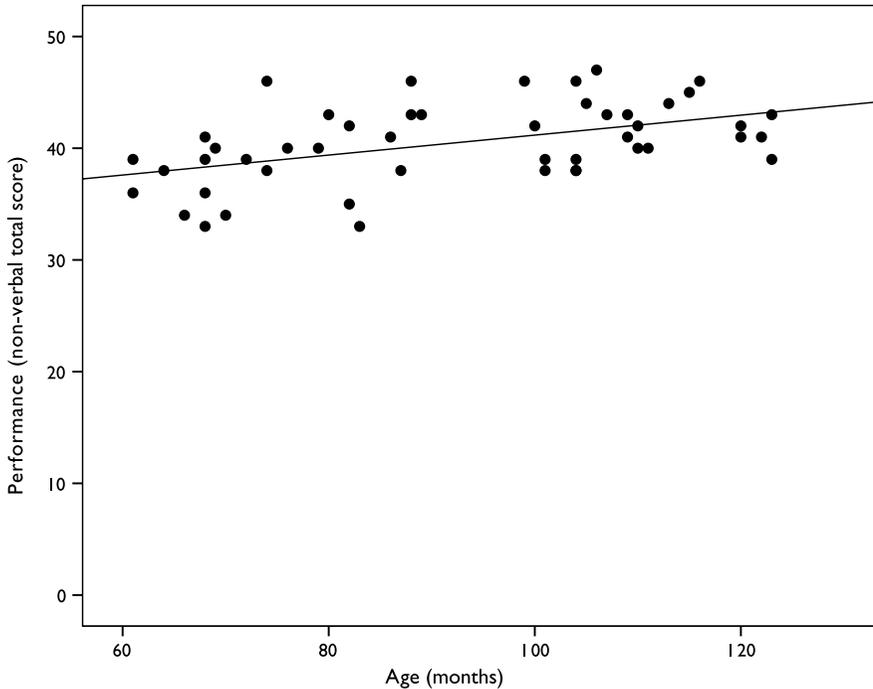
Figure 3. The relationship between children's recognition of non-verbal vocalizations and age.

children's performance significantly for surprise sounds ($r^2 = .17$ $p < .05$, Bonferroni corrected), but not for any other emotion category (all $r^2 < .10$). This indicates that, with the exception of surprise, children's recognition of emotions from non-verbal vocalizations does not improve significantly with age. This likely reflects the relatively proficient performance by even the younger children (see Figure 3), leaving little room for age-related improvement.

Inflected speech
For speech stimuli, age significantly predicted the children's overall performance, accounting for almost half of the variance in scores ($r^2 = .48, p < .001$; see Figure 4). Age predicted the children's performance for most of the emotion categories (achievement $r^2 = .18$, anger $r^2 = .26$, contentment $r^2 = .28$, disgust $r^2 = .12$, neutral $r^2 = .19$, sad $r^2 = .31$, and surprise $r^2 = .36$, all $p < .05$, Bonferroni corrected), but not for recognition of speech inflected with amusement, fear, or relief (all $r^2 < .15$). This indicates that the comprehension of many emotions from inflected speech continues to improve throughout childhood.

Is there a gender effect in children's recognition of emotion in the voice?
There was no main effect of gender ($F < 1$), and no interaction between task and gender ($F < 1$), suggesting that boys and girls did not differ systematically in their recognition of emotions from vocal cues, and that gender did not affect the relative difficulty of the tasks (average arcsine-transformed Hu scores for non-verbal vocalizations for boys: 1.02, and for girls: 1.07; for inflected speech for boys: 0.71, and for girls: 0.74).
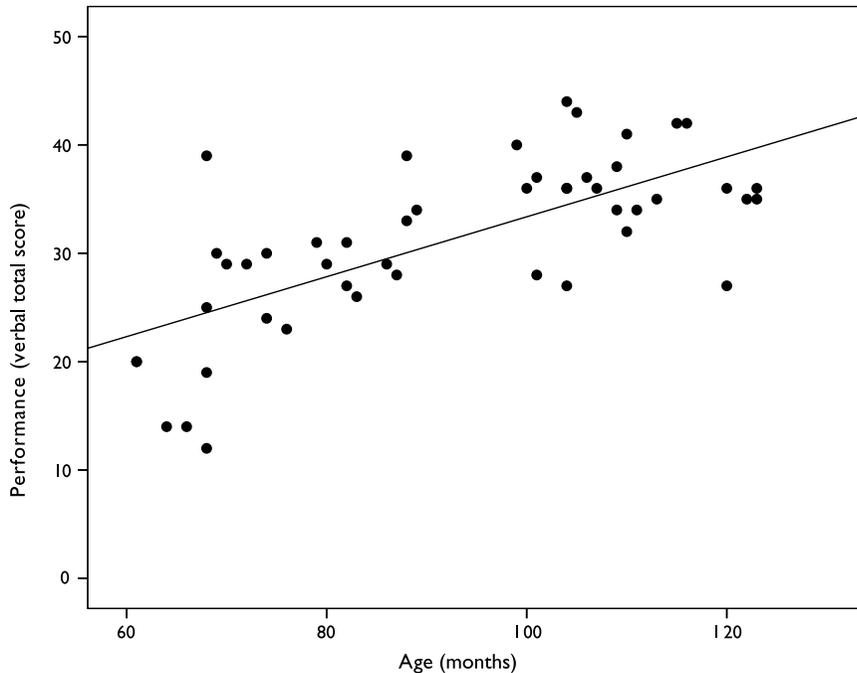
Figure 4. The relationship between children's recognition of emotionally inflected speech and age.

Are the abilities to recognize emotion in non-verbal vocalizations and inflected speech related?

The children's recognition of the two types of affective stimuli was highly correlated ($r = .53, p < .0001$), indicating that, even partialling out the effects of age, children who were better at recognizing emotions from non-verbal sounds also performed better in the recognition of emotions from speech stimuli.

Discussion

Children from 5 years of age are proficient in interpreting emotional information from vocal signals. Performance was better than would be expected by chance for all emotions, for both non-verbal vocalizations and emotionally inflected speech stimuli, demonstrating that children are sensitive to affective cues in several types of vocal signals across a range of negative, neutral, and positive states.

The confusion patterns were consistent across younger and older children and common errors included both within- and across-valence classifications (see Appendix A). For both non-verbal vocalizations and speech stimuli, performance improved with age, and compared to adult listeners' recognition of the same stimuli (e.g., Sauter & Scott, 2007) children's recognition levels appear high. However, as accuracy is affected by the number and type of response alternatives given to participants, caution should be employed in comparing across studies.

The effect of age differed both by specific emotions and stimulus type. For non-verbal vocalizations, age was found to predict performance only for the recognition of surprise sounds. In contrast, for inflected speech, age predicted performance for most of the emotion categories. This difference between the stimulus types may partly be

due to the greater difficulty of the interpretation of certain emotions from speech (e.g., Hawk, van Kleef, Fischer, & Schalk, 2009). However, the categories that became easier to recognize with age did not simply correspond to the most difficult ones (see Figure 1). Previous work has shown that young children perform better with 'simple' emotions, as proposed by Wiggers and van Lieshout (1985). Might more 'complex' categories be better recognized with age? In the current study, emotions such as fear and amusement could be considered relatively simple, whereas surprise and contentment could be considered more complex. Complexity could potentially explain why surprise sounds were the only category of non-verbal vocalizations to improve with age, since surprise may be considered a relatively complex emotion (Wiggers & van Lieshout, 1985). However, this argument does not explain the results in the speech task in the current study, where age predicted the children's performance for several simple emotions, such as anger and sadness. Some of the emotions that improved with age were among the best-recognized categories, while others were among those that children found most difficult. Nevertheless, the only category that saw an improvement with age across both tasks was surprise, consistent with Wiggers and van Lieshout's results (1985), perhaps indicating a role for emotion complexity, as understanding surprise likely involves 'theory of mind', the awareness of one's prior mistaken belief or ignorance. Surprise may also be particularly difficult for children as it is ambiguous, and can signal both a positive or negative state.

Individual differences

Our hypothesis that recognition of the two classes of stimuli would be related was confirmed by the results. This may imply that the tasks employed are tapping the underlying socio-affective skill of the children. Previous studies have found an association between children's social skills and their performance on emotion tasks: Boyatzis and Satyaprasad (1994) found that 4- to 5-year olds' non-verbal skills were correlated with peer popularity from teacher ratings. Other studies have found similar relationships, in 3- to 5-year olds, between emotion understanding and teacher ratings of social skills (Cassidy, Werner, Rourke, Zubernis, & Balaraman, 2003), and popularity (Denham, McKinley, Couchoud, & Holt, 1990), and between emotion recognition and interactions with other children or with adults (Nowicki & Mitchell, 1998). One possible mechanism for this link may be parental socialization skills, which have been shown to relate to children's emotion knowledge (Perlman, Camras, & Pelphrey, 2008). Direct investigation of these issues was beyond the scope of the current study, but our tasks nevertheless show sensitivity to individual differences, likely reflecting underlying socio-affective skills of the child.

Future work should examine the relationship between children's productive and receptive emotion abilities; some work has indicated that these may tap independent mechanisms in children (Nowicki & Duke, 1994). Unfortunately it was not possible to test for such a relationship in the current study, as no recordings were made of the vocalizations that the children produced.

Notably, we found no sex differences in auditory emotion recognition. This is counter to the predictions of, for example, Baron-Cohen's 'male brain' theory, which predicts better 'empathizing' as one marker of the female 'brain type' (Baron-Cohen, 2002; Goldenfeld *et al.*, 2006). Our study suggests that, at least between the ages of 5 and 10 years, sex differences in recognizing emotions from the voice are negligible. Although the lack of a sex difference in the task with non-verbal stimuli could possibly be due to the very high overall performance, we also found no interaction between sex and task.

Performance with the speech stimuli was more moderate and varied, but nevertheless no difference was found between boys and girls.

The recognition of affect from different auditory cues

Our developmental results are consistent with previous studies in adults (e.g., Hawk *et al.*, 2009) in showing that recognition of emotions from non-verbal vocalizations is easier than identifying emotions from inflected speech. This may reflect the fact that emotions overlaid on the speech signal are constrained by prosodic cues engaged in the communication of linguistic information, resulting in a limited range of acoustic variability (Scott, Sauter, & McGettigan, 2009). Previous research has found that when linguistic and paralinguistic cues in speech conflict, children rely more on content, whereas adults rely on inflection (Morton & Trehub, 2001), suggesting that children's understanding of the communicative functions of paralinguistic information are limited compared to their linguistic comprehension.

The current study is the first to show a developmental trajectory for the difference between non-verbal vocalizations and inflected speech. There was an especially marked effect for the younger children, with an average accuracy of 78% for non-verbal vocalizations, compared to 53% for inflected speech. The difference for the older children was less pronounced, with a mean of 84% for the non-verbal vocalizations, and 72% for the speech sounds. It would be of interest to see whether this difference would be further accentuated in even younger children, and although the number of stimuli and trials make the task used in the current study less suited to younger children, children between 2 and 3 years have verbal labels for a range of feeling states (e.g., Dunn, Bretherton, & Munn, 1987; Wellman *et al.*, 2000), and may already associate vocal signals with these labels.

Specific emotions

A number of studies investigating children's recognition of facial expressions of emotions have reported that children find happiness easiest to recognize in facial emotion tasks (see Camras & Allison, 1985; Widen & Russell, 2003). However, previous studies have typically included only one category of positive emotion, and thus have not addressed whether happiness was an easy emotion to identify because it was the only positive emotion examined, or whether expressions of happiness really were easier. In the domain of vocal-emotional signals, we were able to test several distinct positive emotions (achievement, amusement, contentment, and relief). The current study showed that the relative ease of recognition for particular emotion categories differs by stimulus type: For non-verbal vocalizations, children found amusement, relief, and disgust easiest, and neutral and sad vocalizations hardest. For speech sounds, sadness and anger were the easiest, and contentment and disgust the hardest. Widen and Russell (2003) have argued for a particular order of acquisition of emotion categories on the basis of tasks using facial expressions of emotions. According to their model, children first learn to identify happiness, then anger, then sadness, and finally fear, surprise, and disgust. That order is not reflected in the relative difficulty of categories in the current study. However, this difference could be due to a difference in methodology. Harrigan (1984) found that children recognized facial expressions of disgust well in a forced-choice task (similar to that employed in the current study), but did poorly with disgust faces in a free labelling task (also used in Widen and Russell's study). Further studies providing more direct comparisons may be informative for understanding to what extent children's development of different emotion categories is consistent across modalities and task. Acquisition may

also differ between modalities, as suggested by a recent study by Nelson and Russell (2011). Children between 3 and 5 years provided free responses to facial, vocal, and bodily signals. For the vocal stimuli, happy speech was recognized significantly less well than angry and sad speech, whereas for visual stimuli, no such difference was found.

Future considerations

Our data show that the auditory emotion recognition tasks employed in the current study are suitable for a wide age range, and can capture individual differences within typically developing children. This demonstrates that children as young as 5 years are able to make fine distinctions between vocal-emotional signals, including between emotion categories of the same valence. In addition, we hope that these tasks may be useful tools for exploration of clinical groups in future studies.

There is abundant evidence for impairments of emotion recognition in several groups of children. Reduced sensitivity to facial expressions of some emotions has been found in children from disrupted families (Reichenbach & Masters, 1983), children with psychopathic traits (Stevens, Charman, & Blair, 2001), and abused children (Pollak & Sinha, 2002; but see Smith & Walden, 1999). Furthermore, many studies have found impairments in the recognition of emotional facial expressions in children with autism (Wallace, Coleman, & Bailey, 2008; Wright *et al.*, 2008), although group differences may not be apparent when comparison groups are matched on verbal ability (e.g., Castelli, 2005; Grossman, Klin, Carter, & Volkmar, 2000).

Although there is evidence that difficulty identifying vocal emotions is associated with social avoidance and distress in healthy children (McClure & Nowicki, 2001), few studies of atypically developing children have included vocal-affective stimuli (but see Stevens *et al.*, 2001). Although five studies have investigated emotion recognition from vocal cues in children with autism spectrum disorder (ASD), the work to date has yielded equivocal findings. While some studies have found evidence for a deficit in high-functioning children with ASD (Lindner & Rosen, 2006; Mazefsky & Oswald, 2007), others have found no impairment (Jones *et al.*, 2011; Loveland *et al.*, 1997; O'Connor, 2007). Hopefully, future studies that include a wider range of emotional states will shed light on this issue.

In sum, we have shown that children as young as 5 years of age are proficient in interpreting a range of negative and positive emotional cues from vocal signals, consisting of both non-verbal vocalizations and inflected speech. Performance improves with age, and performance on the two tasks is related beyond the effect of age. We conclude that this auditory emotion recognition task is suitable for a wide age range of children, providing a novel, empirical way to investigate children's affective recognition skills in the auditory modality.

References

Akhtar, N., & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child Development Perspectives*, *2*(2), 59–65. doi:10.1111/j.1750-8606.2008.00044.x

Baldwin, D. A., & Moses, L. J. (1996). The ontogeny of social information gathering. *Child Development*, *67*(5), 1915–1939. doi:10.1111/j.1467-8624.1996.tb01835.x

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*, 248–254. doi:10.1016/S1364-6613(02)01904-6

Boyatzis, C., & Satyaprasad, C. (1994). Children's facial and gestural decoding and encoding: Relations between skills and with popularity. *Journal of Nonverbal Behavior*, *18*, 37–55. doi:10.1007/BF02169078

Bugental, D., Kaswan, J., Love, L., & Fox, M. (1970). Child versus adult perception of evaluative messages in verbal, vocal, and visual channels. *Developmental Psychology*, *2*, 367–375. doi:10.1037/h0029166

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

Bullock, M., & Russell, J. (1985). Further evidence on preschoolers' interpretation of facial expressions. *International Journal of Behavioral Development*, *8*, 15–38. doi:10.1177/016502548500800103

Camras, L., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behavior*, *9*, 84–94. doi:10.1007/BF00987140

Cassidy, K., Werner, R., Rourke, M., Zubernis, L., & Balaraman, G. (2003). The relationship between psychological understanding and positive social behaviors. *Social Development*, *12*, 198–221. doi:10.1111/1467-9507.00229

Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. *Autism*, *9*, 428–449. doi:10.1177/1362361305056082

Denham, S., & Couchoud, E. (1990a). Young preschoolers' ability to identify emotions in equivocal situations. *Child Study Journal*, *20*, 153–169.

Denham, S., & Couchoud, E. (1990b). Young preschoolers' understanding of emotions. *Child Study Journal*, *20*, 171–193.

Denham, S., McKinley, M., Couchoud, E., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development*, *61*, 1145–1152. doi:10.1111/j.1467-8624.1990.tb02848.x

Dimitrovsky, L. (1964). The ability to identify the emotional meaning of vocal expressions at successive age levels. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 69–86). New York: McGraw-Hill.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology*, *23*, 132–139. doi:10.1037/0012-1649.23.1.132

Felleman, E. S., Barden, R. C., Carlson, C., Rosenberg, L., & Masters, J. C. (1983). Children's and adults' recognition of spontaneous and posed emotional expressions in young children. *Developmental Psychology*, *19*, 405–413. doi:10.1037/0012-1649.19.3.405

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*, 657–674. doi:10.1111/j.1467-8624.1993.tb02934.x

Field, T., & Walden, T. (1982). Production and discrimination of facial expressions by preschool children. *Child Development*, *53*, 1299–1311. doi:10.1111/1467-8624.ep8588132

Gates, G. S. (1927). The role of the auditory element in the interpretation of emotions. *Psychological Bulletin*, *24*, 175.

Goldenfeld, N., Baron-Cohen, S., & Wheelwright, S. (2006). Empathizing and systemizing in males, females and autism. *Clinical Neuropsychiatry*, *2*, 338–345.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 369–379. doi:10.1111/1469-7610.00621

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, *25*, 909–919. doi:10.1111/j.1460-9568.2007.05379.x

Harrigan, J. (1984). The effects of task order on children's identification of facial expressions. *Motivation and Emotion*, *8*, 157–169. doi:10.1007/BF00993071

Hawk, S. T., Kleef, G. A. V., Fischer, A. H., & Schalk, J. V. D. (2009). "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*, 293–305. doi:10.1037/a0015178

Hughes, C., & Dunn, J. (1997). "Pretend you didn't know": Preschoolers' talk about mental states in pretend play. *Cognitive Development*, *12*, 381–403. doi:10.1016/S0885-2014(97)90019-8

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., . . . Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *52*, 275–285. doi:10.1111/j.1469-7610.2010.02328.x

Lindner, J. L., & Rosen, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *36*, 769–777. doi:10.1007/s10803-006-0105-2

Loveland, K., Tunali–Kotoski, B., Chen, Y., Ortegon, J., Pearson, D., Brelsford, K., & Gibbs, M. C. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and psychopathology*, *9*, 579–593.

Markham, R., & Adams, K. (1992). The effect of type of task on children's identification of facial expressions. *Journal of Nonverbal Behavior*, *16*, 21–39. doi:10.1007/BF00986877

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's syndrome and high-functioning autism: The importance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders*, *37*, 1086–1095. doi:10.1007/s10803-006-0251-6

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424–453. doi:10.1037/0033-2909.126.3.424

McClure, E. B., & Nowicki, S. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior*, *25*(1), 3–19. doi:10.1023/A:1006753006870

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

Moses, L., Baldwin, D., Rosicky, J., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*, 718–735. doi:10.1111/1467-8624.00311

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*, *67*(6), 3219–3237. doi:10.1111/j.1467-8624.1996.tb01910.x

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

Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, *18*, 9–35. doi:10.1007/BF02169077

Nowicki, S., Jr., & Mitchell, J. (1998). Accuracy in identifying affect in child and adult faces and voices and social competence in preschool children. *Genetic, Social and General Psychological Monographs*, *124*, 39–59.

O'Connor, K. (2007). Impaired identification of discrepancies between expressive faces and voices in adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *37*, 2008–2013. doi:10.1007/s10803-006-0345-1

Parise, E., Cleveland, A., Costabile, A., & Striano, T. (2007). Influence of vocal cues on learning about objects in joint attention contexts. *Infant Behavior and Development*, *30*, 380–384. doi:10.1016/j.infbeh.2006.10.006

Perlman, S. B., Camras, L. A., & Pelphrey, K. A. (2008). Physiology and functioning: Parents' vagal tone, emotion socialization, and children's emotion knowledge. *Journal of Experimental Child Psychology*, *100*, 308–315. doi:10.1016/j.jecp.2008.03.007

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, *38*, 784–791. doi:10.1037/0012-1649.38.5.784

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*, 127–152. doi:10.1080/17405620344000022

Reichenbach, L., & Masters, J. (1983). Children's use of expressive and contextual cues in judgments of emotion. *Child Development*, *54*, 993–1004. doi:10.2307/1129903

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: Johns Hopkins University Press.

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

Sauter, D. A. (2006). *An investigation into vocal expressions of emotions: The roles of valence, culture, and acoustic factors* (Unpublished PhD thesis). University College London, London.

Sauter, D. A. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science*, *19*, 36–40. doi:10.1177/0963721409359290

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in non-verbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*, 2251–2272. doi:10.1080/17470211003721642

Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*, *31*, 192–199. doi:10.1007/s11031-007-9065-x

Scott, S. K., Sauter, D. A., & McGettigan, C. (2009). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. Brudzynski (Ed.), *Handbook of mammalian vocalizations* (pp. 187–198). Oxford: Academic Press.

Smith, M., & Walden, T. (1999). Understanding feelings and coping with emotional situations: A comparison of maltreated and nonmaltreated preschoolers. *Social Development*, *8*, 93–116. doi:10.1111/1467-9507.00082

Stevens, D., Charman, T., & Blair, R. (2001). Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. *The Journal of Genetic Psychology*, *162*, 201–211. doi:10.1080/00221320109597961

Stifter, C., & Fox, N. (1987). Preschool children's ability to identify and label emotions. *Journal of Nonverbal Behavior*, *11*, 43–54. doi:10.1007/BF00999606

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, *10*(5), 547–558. doi:10.1111/j.1467-7687.2007.00614.x

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science*, *7*(3), 261–269. doi:10.1111/j.1467-7687.2004.00344.x

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*, 3–28. doi:10.1007/BF00987006

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456. doi:10.1037/0033-2909.121.3.437

Wallace, S., Coleman, M., & Bailey, A. (2008). An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition and Emotion*, *22*, 1353–1380. doi:10.1080/02699930701782153

Wellman, H. M., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development*, *71*, 895–912. doi:10.1111/1467-8624.00198

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

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development*, *23*, 291–312. doi:10.1016/j.cogdev.2008.01.002

Wiggers, M., & van Lieshout, C. F. M. (1985). Development of recognition of emotions: Children's reliance on situational and facial expressive cues. *Developmental Psychology*, *21*, 338–349. doi:10.1037/0012-1649.21.2.338

Wright, B., Clarke, N., Jordan, J., Young, A., Clarke, P., Miles, J., . . . Williams, C. (2008). Emotion recognition in faces and the use of visual context in young people with high-functioning autism spectrum disorders. *Autism*, *12*, 607–626. doi:10.1177/1362361308097118

Appendix A

Table A1. Confusion matrices (%) for younger children's categorizations of non-verbal vocalizations (A) and inflected speech stimuli (B), and older children's categorizations for non-verbal vocalizations (C) and inflected speech stimuli (D)

Response	Emotion category									
	Younger children									
	Achievement	Amusement	Contentment	Relief	Neutral	Surprise	Anger	Disgust	Fear	Sadness
(A)	Non-verbal vocalizations									
Achieve	**80.83**	1.67	3.33	0.00	0.83	0.83	1.67	0.00	10.83	0.00
Amuse	1.67	**97.50**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83
Content	1.67	0.00	**68.33**	4.17	13.33	4.17	3.33	0.00	0.83	4.17
Relief	0.00	0.00	0.83	**94.17**	2.50	2.50	0.00	0.00	0.00	0.00
Neutral	0.83	0.83	11.67	0.83	**59.17**	8.33	1.67	3.33	4.17	9.17
Surprise	7.50	0.83	1.67	3.33	10.83	**63.33**	1.67	0.00	8.33	2.50
Anger	8.33	0.00	0.00	0.00	0.00	0.00	**80.00**	8.33	0.83	2.50
Disgust	0.00	0.00	0.00	0.00	0.00	0.00	0.83	**99.17**	0.00	0.00
Fear	2.50	1.67	0.83	2.50	0.00	3.33	0.00	8.33	**77.50**	3.33
Sadness	0.83	1.67	15.00	1.67	5.00	1.67	2.50	10.00	0.83	**60.83**
(B)	Inflected speech									
Achieve	**62.50**	15.83	2.50	0.83	0.00	3.33	0.83	0.83	12.50	0.83
Amuse	7.50	**61.67**	5.83	2.50	9.17	6.67	0.00	2.50	2.50	1.67
Content	6.67	0.83	**32.50**	9.17	28.33	8.33	5.83	2.50	1.67	4.17
Relief	4.17	4.17	6.67	**54.17**	13.33	2.50	5.83	0.00	1.67	7.50
Neutral	3.33	1.67	6.67	9.17	**52.50**	6.67	8.33	5.00	2.50	4.17
Surprise	15.83	1.67	8.33	10.83	15.00	**34.17**	0.00	3.33	9.17	1.67
Anger	3.33	2.50	4.17	5.83	1.67	3.33	**57.50**	7.50	3.33	10.83
Disgust	5.83	3.33	0.00	5.00	8.33	0.00	18.33	**40.00**	1.67	17.50
Fear	5.00	8.33	2.50	0.00	0.83	1.67	0.00	5.00	**65.83**	10.83
Sadness	0.83	0.00	8.33	0.83	5.00	0.00	4.17	5.00	7.50	**68.33**
	Older children									
(C)	Non-verbal vocalizations									
Achieve	**89.17**	0.83	0.00	0.00	0.00	5.00	1.67	0.00	3.33	0.00
Amuse	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Content	0.00	0.00	**80.00**	0.83	5.83	6.67	2.50	0.00	1.67	2.50
Relief	0.00	0.00	0.83	**98.33**	0.83	0.00	0.00	0.00	0.00	0.00
Neutral	0.00	0.00	21.67	0.83	**60.83**	6.67	2.50	2.50	0.00	5.00
Surprise	2.50	0.00	0.83	0.00	8.33	**81.67**	0.83	0.83	2.50	2.50
Anger	8.33	0.00	0.00	0.00	0.83	0.00	**77.50**	9.17	2.50	1.67
Disgust	0.00	0.00	0.00	0.00	0.00	0.00	1.67	**97.50**	0.83	0.00
Fear	0.00	1.67	0.00	0.83	0.00	0.00	0.00	9.17	**85.83**	2.50
Sadness	0.00	0.00	14.17	0.83	3.33	0.00	1.67	9.17	2.50	**68.33**
(D)	Inflected speech									
Achieve	**75.00**	13.33	1.67	0.00	0.00	2.50	1.67	0.00	5.83	0.00
Amuse	1.67	**69.17**	5.83	2.50	10.83	5.00	0.00	1.67	3.33	0.00
Content	0.00	0.83	**52.50**	5.83	35.00	2.50	0.83	0.83	0.00	1.67
Relief	0.00	2.50	6.67	**75.83**	10.00	0.00	1.67	0.00	0.00	3.33
Neutral	0.83	2.50	2.50	5.83	**75.00**	5.00	5.00	0.83	2.50	0.00
Surprise	5.83	0.83	2.50	5.00	3.33	**74.17**	1.67	0.83	5.00	0.83
Anger	1.67	0.00	0.00	4.17	0.00	0.00	**83.33**	1.67	3.33	5.83
Disgust	2.50	2.50	0.00	3.33	5.00	0.00	13.33	**55.00**	0.00	18.33
Fear	0.00	3.33	1.67	1.67	0.00	1.67	0.00	2.50	**73.33**	15.83
Sadness	0.00	0.00	1.67	0.00	0.83	0.83	1.67	0.83	5.00	**89.17**

Note. All horizontal lines add to 100. Achieve, achievement; amuse, amusement; content, contentment. Correct classifications are in bold.

Appendix B

Table B1. *t*-Values for tests of performance against chance for younger and older children for each emotion category of non-verbal vocalizations (A), and inflected speech (B)

Age	Emotion category									
	Achievement	Amusement	Contentment	Relief	Neutral	Surprise	Anger	Disgust	Fear	Sadness
(A)	Non-verbal vocalizations									
Young	14.2	25.9	13.8	18.5	10.6	11.3	14.9	19.7	16.7	12.0
Old	19.1	47.7	19.3	33.2	10.6	14.9	15.6	16.0	17.8	16.2
(B)	Inflected speech									
Young	9.7	10.0	4.5	9.4	8.5	5.7	7.4	8.9	13.5	11.5
Old	14.1	15.7	14.2	12.6	15.0	9.0	14.3	12.8	15.9	24.2

Note. All results were better than chance at $p < .05$, Bonferroni corrected, with one degree of freedom.