J Autism Dev Disord (2012) 42:2546–2557
DOI 10.1007/s10803-012-1511-2

ORIGINAL PAPER

# "Who Said That?" Matching of Low- and High-Intensity Emotional Prosody to Facial Expressions by Adolescents with ASD

**Ruth B. Grossman · Helen Tager-Flusberg**

Published online: 27 March 2012


**Abstract** Data on emotion processing by individuals with ASD suggest both intact abilities and significant deficits. Signal intensity may be a contributing factor to this discrepancy. We presented low- and high-intensity emotional stimuli in a face-voice matching task to 22 adolescents with ASD and 22 typically developing (TD) peers. Participants heard semantically neutral sentences with happy, surprised, angry, and sad prosody presented at two intensity levels (low, high) and matched them to emotional faces. The facial expression choice was either across- or within-valence. Both groups were less accurate for low-intensity emotions, but the ASD participants' accuracy levels dropped off more sharply. ASD participants were significantly less accurate than their TD peers for trials involving low-intensity emotions *and* within-valence face contrasts.

**Keywords** Autism · Prosody · Facial expressions · Signal intensity · Face-voice matching

## Introduction

Individuals with autism spectrum disorders (ASD) have significant social communication deficits, particularly in the realm of non-verbal communication, such as the decoding of emotion from facial expressions (Adolphs et al. 2001; Dawson et al. 2004; de Gelder et al. 1991; Gepner et al. 1996; Pelphrey et al. 2007) and tone of voice, or prosody (Diehl et al. 2008; Paul et al. 2005; Shriberg et al. 2001). There are, however, conflicting data from facial and vocal emotion recognition tasks in this population, indicating a potentially significant effect of task demand on this skill. Some studies of facial expression recognition found deficits among participants with ASD (Celani et al. 1999; Diehl et al. 2008; Grossman and Tager-Flusberg 2008; Philip et al. 2010) while others revealed facial emotion recognition skills equal to those of their typically developing (TD) peers (Gepner et al. 2001; Rosset et al. 2008; Grossman et al. 2000). A recent review of the facial expression literature for this population by Harms et al. (2010) proposed that these differences in documented facial expression comprehension levels in this population are significantly dependent on participant demographics, task selection, and stimulus type. Harms et al. (2010) hypothesize that individuals with high-functioning autism (HFA), i.e. those with IQ scores within the normal range and typical, or near-typical language skills, are susceptible to the difficulty of a facial expression emotion recognition task, showing typical performance on canonical, obvious expressions (Baron-Cohen et al. 1997; Grossman et al. 2000), but are less accurate on tasks

Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

*Present Address:*
R. B. Grossman (✉)
Department of Communication Sciences and Disorders, Emerson College, 120 Boylston Street, Boston, MA 02116, USA
e-mail: ruth_grossman@emerson.edu

R. B. Grossman
Shriver Center, University of Massachusetts Medical School, 200 Trapelo Road, Waltham, MA 02452, USA

H. Tager-Flusberg
Department of Psychology, Boston University, 64 Cummington Street, Boston, MA 02215, USA

involving more subtle expressions, or more complex tasks (Golan et al. 2007; Lindner and Rosén 2006; Philip et al. 2010).

There are fewer studies dealing with recognition of emotion from auditory stimuli, specifically prosody, but the discrepancies in their findings mirror those of studies on facial expression recognition. Some investigations of canonical, basic emotional expressions from voices found no significant impairment in individuals with ASD (Boucher et al. 2000; Jones et al. 2011; Grossman et al. 2010; O'Connor 2007), while others revealed deficits in the ASD group for similar tasks (Philip et al. 2010; Korpilahti et al. 2007) or tasks involving more subtle or complex vocal emotional expressions (Golan et al. 2006; Philip et al. 2010; Rutherford et al. 2002).

### Cue Intensity in Facial Emotion

Recently, there have been a few studies looking specifically at the comprehension of more subtle facial expressions of emotion among individuals with ASD. Law Smith et al. (2010) presented dynamic emotion facial expressions to adolescents with HFA and determined that reduced saliency or cue intensity of emotional information reduced accuracy of emotion recognition in the ASD group for stimuli depicting disgust, anger, and surprise, but had no such effect on the TD group. Kuusikko et al. (2009) showed that younger children with ASD were significantly less accurate than TD peers at recognizing more ambiguous emotional facial expressions or those with lower cue intensity. Greimel et al. (2010) collected functional MRI as well as behavioral data on accuracy levels for high- versus low-intensity facial expressions of sad and happy emotions. They found that the ASD group was significantly less accurate on trials with low-intensity expressions, but no different than their TD peers for trials with high-intensity expressions.

### Cue Intensity in Vocal Emotion

Despite this recent increase in studies of low-intensity emotional expressions of faces, there are still very few studies investigating the recognition of low-intensity emotional expressions of prosody. Mazefsky and Oswald (2007) presented emotional facial and prosodic stimuli of varying intensity levels from the diagnostic analysis system of nonverbal accuracy scale-2 (DANVA-2, Nowicki 2003) to 8–15 year olds with Asperger Syndrome (AS) and high functioning autism (HFA). Compared to the typical normative sample of the DANVA-2 the HFA group (but not the AS group, which was defined as having higher social functioning than the HFA group) was significantly impaired on all prosodic emotional expressions. When the data were divided between trials with high- and low-intensity stimuli, it was revealed that the HFA participants were significantly less accurate than the AS group only on low-intensity stimuli. There were no significant differences between the groups on high-intensity prosodic stimuli.

This apparent relationship between social competence and the ability to discern emotion from low-intensity facial and vocal stimuli reveals an important yet understudied area of social communication in ASD. Specifically, the study of low-intensity expressions of emotion represents a significant gap in the literature so far. Although Mazefsky and Oswald (2007) did investigate emotional expressions in both the facial and vocal modalities, which is after all how emotion is transmitted during daily social interaction, they investigated these components in separate trials, requiring participants to select a verbal label for emotional faces in one task and for emotional voices in another task. Social interactions, however, require the integration of both modalities to determine speaker intent and emotional state.

### Integration of Faces and Voices

Individuals with ASD have demonstrated the ability to integrate timing data of visual and auditory speech information (Grossman et al. 2009) and to derive some benefit from visual information in a speech-in-noise paradigm (Smith and Bennetto 2007). The question of whether these individuals are also able to integrate emotional information from visual and auditory channels has not yet been studied extensively.

There are some data suggesting that individuals with ASD are less accurate than mental-age matched individuals with Down Syndrome at matching an emotional voice to one of two simultaneously presented dynamic emotional faces (Loveland et al. 1995). These data were collected with a lower-functioning group of participants with ASD and the intensity level of the emotional expressions was not specified. O'Connor (2007) also found that adults with Asperger Syndrome were less accurate at judging emotional congruency between simultaneously presented facial and vocal expressions using spoken sentences and static images. Again, the intensity of the emotional expressions in either modality was not mentioned.

Successful social communication requires the integration of emotional information from faces and voices and most daily interactions rely on subtle, low-intensity expressions, rather than high-intensity emotional displays. There are emerging data suggesting that individuals with ASD have greater difficulty interpreting low-intensity emotional *facial* expressions, but only one study to date (Mazefsky and Oswald 2007) investigated whether a similar pattern exists for low-intensity emotional *prosody*, and no published studies have investigated how individuals with ASD process low-intensity versus high-intensity facial and vocal emotional expressions in the same task.

The purpose of the study presented here was to investigate the interaction of emotion recognition across auditory (prosody) and visual (facial expression) modalities for low- versus high-intensity stimuli. We used complete sentence stimuli, presenting realistic sentences spoken with low- versus high-intensity emotional prosody. In order to create task demands similar to those of daily social interactions, participants had to match the emotion of the sentence utterance to one of two static faces. The facial expressions mirrored the prosodic stimulus in terms of the emotional intensity, but differed in emotion type (happy, angry, surprised, or sad). The task required participants to indicate which of the two people depicted in the images was more likely to have produced the preceding affective utterance. Using a combination of low- versus high-intensity prosodic and facial expression stimuli and a within- versus across- valence facial expression contrast we plan to answer the question of how adolescents with ASD recognize and integrate ecologically valid, low-intensity emotional cues from voices and faces.

We hypothesize that participants with ASD will be less accurate than their TD peers at matching prosodic and facial expression affect for stimuli with low emotional intensity, particularly in trials involving the more subtle, within-valence facial expression contrast.

## Methods

### Participants

Two groups participated in this study: children and adolescents with ASD (n = 22) and TD controls (n = 22) ranging in age from 8 to 19 years old. All participants were recruited through local schools, advertisements placed in local magazines, newspapers, the internet, advocacy groups for families of children with autism, and word of mouth.

#### *Standardized Testing*

The Kaufman Brief Intelligence Test, Second Edition (K-BIT 2; Kaufman and Kaufman 2004) was used to assess IQ in all participants and receptive vocabulary ability was measured by the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn 1997). Participants were selected on the basis of standardized scores within the normal range (plus/minus two standard deviations of the mean) to create well matched and homogenous groups.

#### *Diagnosis of ASD*

The participants in the autism group met DSM-IV criteria for autistic disorder, based on expert clinical impression and confirmed by the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994) and the autism diagnostic observation schedule (ADOS) Module 3 (Lord et al. 1999), which were administered by trained examiners. Participants with known genetic disorders were excluded. Based on their ADOS scores, 15 participants met criteria for autism and seven met criteria for ASD.

The descriptive characteristics of both groups can be found in Table 1. Using a multivariate ANOVA with group as the independent variable we verified that the ASD and TD groups did not differ significantly in age, $F$ (1, 43) = .059, $p$ = .81, verbal IQ, $F$ (1, 43) = 1.96, $p$ = .17, nonverbal IQ, $F$ (1, 43) = .22, $p$ = .64 or receptive vocabulary ability, $F$ (1, 40) = .66, $p$ = .43. A Chi-squared analysis showed that the groups also did not differ in the distribution of gender ($\chi^2$ (1, N = 44) = .17, $p$ = 1),

### Materials

#### *Face Stimuli*

We selected 64 emotional faces from the MacBrain Face Stimulus Set database of facial expressions (Tottenham et al. 2009), eight different faces for each of the eight vocal stimulus conditions (four emotions × two intensity levels). Half the faces were male and half were female, and all available races (Caucasian, African American, and Asian) were represented. The face database contains stimuli grouped by emotion, and within each emotion group faces are classified as "closed mouth" or "open mouth," with the exception of the "happy" group, with contains a third classification of "happy open mouth exuberant." We selected the high-intensity stimuli of "happy" from the "exuberant" group. For the other three emotions (angry, sad, and surprise), the high-intensity stimuli were chosen

**Table 1** Descriptive characteristics of participant groups

| | ASD (n = 22) *M*(*SD*) | TD (n = 22) *M*(*SD*) |
|---|---|---|
| Age | 13:10(2:10) | 14:0(2:5) |
| | Range: 8:10–19:9 | Range: 10:2–17:11 |
| Sex | 18 Male | 19 Male |
| | 4 Female | 3 Female |
| Full scale IQ | 106.7(10.6) | 108.9(11.3) |
| | Range: 87–123 | Range: 87–123 |
| Verbal IQ | 101.2(14.3) | 108.1(14.6) |
| | Range: 83–127 | Range: 81–127 |
| Nonverbal IQ | 109.6(19.1) | 106.7(9.8) |
| | Range: 94–127 | Range: 85–116 |
| PPVT-III | 107.0(15.4) | 111.3(15.3) |
| | Range: 79–138 | Range: 79–139 |


**Fig. 1** Sample facial expression stimuli. *Note*: Images depicted here are the faces approved for publication by the NimStim database and not accurate reflections of the images used in the study. In the task all face pairs were matched by gender and race

from the "open mouth" group. Low-intensity stimuli for all emotions were chosen from the "closed mouth" set (Fig. 1).

### *Pilot Testing of Faces*

We pilot tested a large sample of facial expressions on 15 TD adults who were naïve to the intended task. Participants were seated in front of a computer and given a button box to record responses. We first showed the label of the intended emotion (angry, sad, happy, surprise) on the screen, followed by two facial expressions presented side-by-side on the screen. Pilot participants were told to select the face that best matched the emotion label they had just seen by pressing the right or left buttons on the box to represent the images on the right or left side of the screen. Only images that were rated as matching the indicated emotion with at least 70 % accuracy were maintained. The mean pilot testing accuracy for face stimuli selected for the task was 93 %. The final selection of facial expressions contained 50 % high-intensity and 50 % low-intensity images.

**Table 2** Sentences used in study

| |
|---|
| He wants to sell his car |
| I'm going to buy a new computer |
| My friend asked me for help |
| Our neighbors moved away |
| She bought a lot of soda |
| They ate all the popcorn |
| We're visiting my cousin today |
| We watched TV all day |

### *Prosodic Stimuli*

We created eight declarative sentences (Table 2), selecting the vocabulary so sentences were unlikely to carry emotional content and verified this through pilot testing. Ten pilot study participants were given the sentences in written form and were asked to rate them on a 7-point Likert scale, with "neutral" in the middle of the range, "positive emotion" on one extreme, and "negative emotion" on the other extreme. All sentences that received an average score between 3.5 and 4.5 were deemed to carry no overt emotional content and were therefore appropriate for stimulus creation.

### *Stimulus Recording*

We auditioned several acting students to find one male and one female actor who were able to produce stimuli in a range of believable emotions, as determined by three members of the study staff. Recording sessions took place in a quiet, closed room on university campus, with separate sessions for each actor. Every recording session involved at least two members of the study staff. We used a high quality USB microphone and PC computer to record stimuli in .wav format. We explained the experimental design and purpose, including the use of high- and low-intensity emotions to the actors and showed them the selected facial expression images and printed sentences. Prior to recording an emotional utterance, we showed each participant the target face image and its corresponding printed sentence. We explained that each sentence was

supposed to be produced with declarative prosodic contours, modeling the utterance-final drop in pitch characteristic of such utterances (Merewether and Alpert 1990). In order to elicit the target emotional prosody, we used descriptive cues, such as "You are completely furious about this" for high-intensity anger, versus "you find this annoying" for low-intensity anger. We modeled the emotional prosodic contours we expected for each emotion, such as higher pitch, rapid rate, and a rise-fall utterance-final pitch pattern for happy utterances, or lower pitch, slower rate, and a lower pitch ending in sad utterances (Grossman et al. 2010; Banse and Scherer 1996; Cosmides 1983; Murray and Arnott 2008). For each sentence, we recorded the actors producing 4–10 utterances each in two positive (happy, surprise) and two negative (anger, sadness) emotions, providing additional direction and cues as necessary. For the purpose of this study, surprise was always elicited and described as positive surprise. After completion of the recording session we brought the digital recordings back to the lab for pilot testing and selection.

### *Pilot Testing of Voices*

All recordings were initially verified to have good sound quality and no excessive noise using PRAAT software (Boersma & Weenink, 2009). Three members of the study staff listened to each iteration of every sentence utterance while looking at the facial expression used to elicit it. We selected the two utterances of each sentence that best expressed the emotion and intensity level portrayed in the matching facial expression. We specifically attended to whether the prosodic expressions matched the facial expressions in terms of emotional intensity. Stimulus selections were only considered valid if all three raters agreed unanimously. If consent was not unanimous, a fourth person was asked to listen to the utterance in question. Stimuli were chosen if three raters agreed that it clearly matched the target facial expression in emotion and intensity. Using this method, we preselected two iterations of each prosodic stimulus for further pilot testing. We then created a protocol that showed the facial expression presented on a computer screen, followed by the two versions of the corresponding sentence utterance. This protocol was shown to 15 naïve TD adult pilot study participants who were asked to choose the utterance that best matched the preceding facial expression through button-press responses. Each facial expression was paired once with the two sentence utterances in the emotion and intensity level that matched the face, and a second time with two utterances portraying the target emotion in the opposite intensity level. The prosodic stimulus that achieved 70 % or greater responses as matching the facial expression was chosen to be included in the final protocol.

### *Pilot Testing of Task*

Each prosodic stimulus was connected with a facial expression matched on emotion, intensity, and sex of the speaker. The next step was to add a foil face image, in order to create pairs of facial expressions to be presented for each prosodic stimulus. We chose foils from the sample of pilot-tested facial expressions, so that individual expressions could appear more than once within the task: Either as the correct match to a prosodic stimulus, or as a foil. Facial expressions were paired to provide three different contrasts: Positive–Positive, Negative–Negative (both within valence contrasts) and Positive–Negative (across valence contrast). As an example, an utterance with high-intensity sad prosody could be paired with a high-intensity sad face and a high-intensity happy face (across-valence), or a high-intensity sad face and a high-intensity angry face (within-valence). The two facial expressions presented side-by-side within a trial differed from each other only on emotion, but never on intensity, sex, or race.

The within-valence face contrast was designed to be more difficult than the across valence face contrast, since it required the more fine-grained choice of a specific emotion, rather than the more general choice of positive or negative valence. We pilot tested this complete task on a group of 10 adult TD participants who were naïve to the task and the stimuli. Participants were told to listen to the prosodic stimulus and look at the two facial expression images presented side-by-side on the screen. Their task was to determine which of the two individuals seen on screen was most likely to have produced the preceding utterance. Pilot study participants achieved a mean accuracy of 88 %. Every stimulus combination (voice, target face, foil face) reaching at least 70 % accuracy was maintained for the final version of the task. There were three positive–positive within-valence face contrast trials that achieved only 60 % accuracy in pilot testing. Investigating the pilot data further, we noticed that there were only nine trials with accuracy rates below 80 % and that seven of those trials were positive–positive within-valence contrasts. Since each of the stimuli had been pilot tested prior to inclusion in the task and confirmed to be representative of their target emotion, we concluded that the positive–positive within-valence comparison was simply too difficult even for TD adults. We decided to keep the trials in the task for the sake of symmetry (equal number of positive and negative emotion trials), but eliminate them from the final analysis. We will discuss possible explanations for this finding in the "limitations and future directions" section of this paper.

*Task Creation*

We created two pseudorandomized and counterbalanced stimulus sequences for the final task and alternated presentation of the sequences with each participant so that about half the participants saw version one and the other half saw version two. Both versions of the task contained all eight sentences, each presented in all four emotions. Eight of these 32 unique stimuli were produced with high prosodic intensity by a female speaker, eight with low intensity by a female speaker, eight with high intensity by a male speaker, and eight with low intensity by a male speaker. These 32 stimuli were repeated twice within the study, once with a within-valence facial expression contrast and once with an across-valence facial expression contrast for a total of 64 pseudo-randomized face-voice matching trials per sequence. The location of the correct and incorrect faces (left vs. right) was counterbalanced across presentations within each sequence.

## Procedure

We provided all participants and their caregivers with the IRB approved Informed Consent form, explained the study, and answered any questions. Participants 12 years and older signed Assent forms in addition to the Consent forms signed by their caregivers. We then led participants into the testing room and familiarized them with the computer and response button box. Participants were seated a comfortable distance from the computer screen and speakers were set at an easily audible volume. During the training run for the task we could adjust the volume based on participant feedback if necessary. The task was introduced using simple, easily understood language. Participants were told they would first hear an utterance spoken by somebody who was angry, happy, sad, or surprised. Sometimes that person would be very angry and other times just a little bit angry and so on for all the other emotions. After the person was done saying the utterance, participants saw two faces on the screen side-by-side. Participants were asked to decide which of the two faces was more likely to have said the preceding utterance, indicating their choice by pressing the right or left button on the button box to represent the image on the right or left. We encouraged participants to listen and look closely and make their decisions as quickly as possible without making mistakes. No part of the instructions specifically requested that participants attend to the emotional content of the face or voice. The task was simply explained as having to determine "who said that?"

Participants first completed a training run which mimicked the task, but during which they were provided with corrective feedback after each button press. Once participants pressed a button to indicate their choice, the incorrect face disappeared from the screen and only the correct face remained. At the same time study staff provided positive reinforcement for the correct choice. All participants passed the training run, which was achieved by responding correctly three out of four times, and moved on to the experimental task, which took less than 5 min to complete.

## Results

The hypothesis of this study focuses on the relative accuracy levels of low- versus high-intensity emotional stimuli and within- versus across-valence facial expression contrasts. We therefore grouped data for analysis according to emotional intensity and face contrast of the stimuli, rather than by individual emotions. All data were normally distributed.

## Main Effects

Our first analysis was to determine whether there were main effects for any of the different stimulus conditions. We therefore conducted a 2 (group) by 2 (emotional intensity) by 2 (face contrast) repeated measures ANOVA, which revealed a main effect for intensity ($F$ (1, 42) = 100.9, $p < .001$, partial $\eta^2 = .71$). Both groups were more accurate for high-intensity emotions than low-intensity emotions. There was also a main effect for contrast ($F$ (1, 42) = 5.2, $p = .028$, partial $\eta^2 = .11$) with both groups more accurate on trials with across-valence face contrasts than those with within-valence face contrasts. As expected, accuracy levels for within-valence contrasts of positive valence emotions (happy vs. surprise) were very low for both groups and even at chance for low-prosody trials. The TD group's mean accuracy was 52 % for low-intensity and 65 % for high-intensity samples of this type. In contrast, the TD group reached mean accuracy levels of 84 % for low-intensity within-valence samples with negative emotion, and 95 % for high-intensity samples of the same type, confirming our pilot results that the differentiation of happy and positive surprise was too difficult even for the control group. Accuracy levels for each trial type can be found in Table 3. As planned, we included only trials with negative valence (angry and sad) in the within-valence category for further analysis.

The same group × emotional intensity × face contrast ANOVA also revealed a significant group by intensity interaction ($F$ (1, 42) = 13.6, $p = .001$, partial $\eta^2 = .24$) showing that the accuracy levels of the ASD group dropped off more sharply for weak emotional stimuli than those of the TD group, and an intensity by contrast interaction ($F$ (1, 42) = 14.5, $p < .001$, partial $\eta^2 = .26$) indicating a combined effect of emotional intensity and face contrast on accuracy overall (Fig. 2). There was no contrast by group

**Table 3** Accuracy (in percent correct)

| | ASD (n = 22) *M*(*SD*) | TD (n = 22) *M*(*SD*) |
|---|---|---|
| Emotional intensity strong Positive-positive face contrast | 71 (19) | 65 (17) |
| Emotional intensity weak Positive-positive face contrast | 56 (17) | 52 (20) |
| Emotional intensity strong Negative-negative face contrast | 97 (9) | 95 (8) |
| Emotional intensity weak Negative-negative face contrast | 72 (14) | 84 (14) |
| Emotional intensity strong Positive–negative face contrast | 94 (6) | 94 (7) |
| Emotional intensity weak Positive–negative face contrast | 84 (10) | 89 (8) |

interaction ($F$ (1, 42) = .7, $p$ = .403, partial $\eta^2$ = .02) or intensity by contrast by group interaction ($F$ (1, 42) = 2.8, $p$ = .101, partial $\eta^2$ = .06).

## Face-Voice Interaction

To further test the hypothesis that variations in emotional intensity as well as face contrast influence accuracy levels, we conducted a one-way ANOVA (sphericity was assumed) for all four conditions (high-intensity emotion and within-valence face contrast, high-intensity emotion and across-valence face contrast, low-intensity emotion and within-valence face contrast, low-intensity emotion and across-valence face contrast), which revealed a significant group difference for trials with low-intensity emotions *and* the more subtle within-valence face contrast ($F$ (1, 43) = 8.1, $p$ = .007), as well as a trend for group differences for samples with low-intensity emotion and across-valence face contrast ($F$ (1, 43) = 3.5, $p$ = .069). There was no significant between-group difference for trials with high-intensity emotion, regardless of whether the face contrast was across-valence ($F$ (1, 43) = .0, $p$ = .998), or within-valence ($F$ (1, 43) = .4, $p$ = .512,).

### *Typical Participants*

To investigate the differences between these four conditions within each participant group, we conducted pairwise within-group t-tests. Dividing the data first according to emotional intensity, results for the TD group indicate the expected significantly higher accuracy levels for stimuli with high-intensity emotion ($M$ = 94.3, $SD$ = 7.2) versus those with low-intensity emotion ($M$ = 89, $SD$ = 8.2) in the across-valence face conditions ($t$ (21) = 2.7, $p$ = .013), as well as significantly higher accuracy levels for stimuli with high-intensity emotion ($M$ = 94.9, $SD$ = 8.3) versus those with low-intensity emotion ($M$ = 84.1, $SD$ = 14) in the within-valence face conditions ($t$ (21) = 4.1, $p$ = .001). These data show that low emotional intensity significantly and negatively affects emotion processing accuracy in the typical group, with no additional effect on accuracy caused by the face contrast variation.

Looking at the same data, but pairing conditions across face contrast, rather than emotional intensity, we found no significant within-group differences for the TD participants in accuracy levels for stimuli with across-valence face contrast ($M$ = 94.3, $SD$ = 7.2) versus those with within-valence face contrast ($M$ = 89, $SD$ = 8.2) in the high-intensity emotion conditions ($t$ (21) = −.3, $p$ = .807), and no significant within group differences in accuracy levels for stimuli with across-valence face contrast ($M$ = 89, $SD$ = 8.2) compared to those with within-valence face contrast ($M$ = 84.1, $SD$ = 14) in the low-intensity emotion conditions ($t$ (21) = 1.4, $p$ = .190). These data suggest that within each intensity condition there is no significant effect of face contrast difficulty on emotion processing accuracy levels for TD participants.

### *ASD Participants*

The results are similar, with one crucial difference, for the ASD group. The same pair-wise comparisons across emotional intensity conditions reveals significantly higher accuracy levels for stimuli with high-intensity emotion ($M$ = 94.3, $SD$ = 6.4) versus those with low-intensity emotion ($M$ = 84, $SD$ = 10) in the across-valence face conditions ($t$ (21) = 5.4, $p$ < .001), as well as significantly higher accuracy levels for stimuli with high-intensity emotion ($M$ = 96.6, $SD$ = 8.8) versus those with low-intensity emotion ($M$ = 72.2, $SD$ = 13.9) in the within-valence face conditions ($t$ (21) = 7.3, $p$ < .001). These data show that the ASD group is as susceptible to variations in emotional cue intensity as the TD group.

Looking at the data across face contrast conditions we find no significant within-group differences for the ASD participants in accuracy levels for stimuli with across-valence face contrast ($M$ = 94.3, $SD$ = 6.4) versus those with within-valence face contrast ($M$ = 96.6, $SD$ = 8.8) in the high-intensity emotion conditions ($t$ (21) = −1.2, $p$ = .261), again mirroring the results for the TD group. In contrast to the TD group, however, the ASD group does show a significant within-group difference in accuracy levels for stimuli with across-valence face contrast ($M$ = 83.8, $SD$ = 10) versus those with within-valence face contrast ($M$ = 72.2, $SD$ = 13.9) in the low-intensity emotion conditions ($t$ (21) = 3.8, $p$ = .001). This result indicates that, although both participant groups are susceptible to changes in emotional intensity, only the ASD

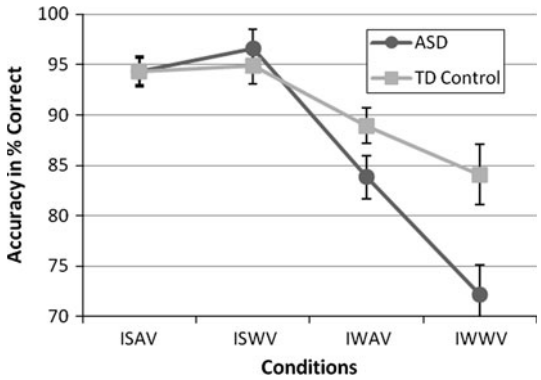


**Fig. 2** Accuracy levels for emotional intensity and face contrast combinations. *Error bars* are standard error. *ISAV* emotional intensity strong, across valence contrast; *ISWV* emotional intensity strong, within valence contrast; *IWAV* emotional intensity weak, across valence contrast; *IWWV* emotional intensity weak, within valence contrast

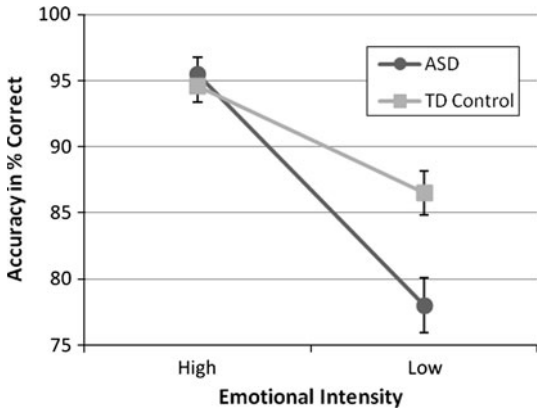


**Fig. 3** Accuracy levels for emotional intensity. *Error bars* are standard error

group shows an additional reduction in accuracy for trials involving the increased difficulty of a within-valence face contrast (Figs. 3, 4).

## Discussion

The aim of our study was to investigate whether adolescents with ASD are more susceptible than their TD peers to low- versus high-intensity emotional expressiveness in faces and voices. We hypothesized that the ASD group would show significantly lower accuracy for low-intensity versus high-intensity emotional expressions compared to their TD peers and that a within-valence facial expression matching task would lower their accuracy rates even further. Our data clearly confirm that initial hypothesis.

The results of this study show that both participant groups are susceptible to manipulation of emotional cue

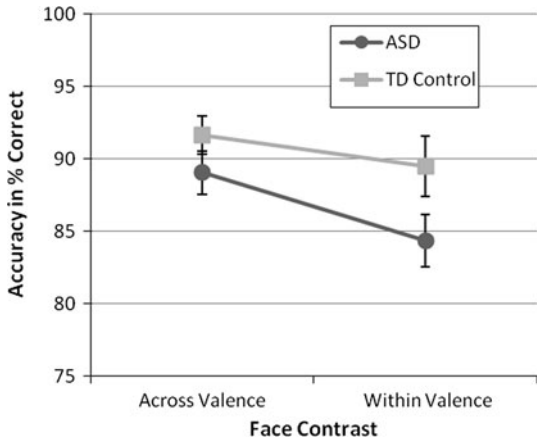


**Fig. 4** Accuracy levels for face contrast. *Error bars* are standard error

intensity in the pairing of prosody to facial expressions. There is evidence to show that decreased saliency of emotion in the voice results in reduced neuronal activity among TD participants (Leitman et al. 2010). As intensity diminishes, so does the saliency of the expressed emotion, explaining why even TD participants were less accurate for low-intensity stimuli. However, participants with ASD are clearly more vulnerable to such variations, showing accuracy levels that drop significantly more sharply for low-intensity emotional stimuli than those of their TD peers. When asked to select the most realistic facial expression from a range of mildly expressive to exaggerated, individuals with ASD were more likely than their TD peers to choose the exaggerated expression as being the most realistic (Rutherford and McIntosh 2007). These data suggest that individuals with ASD may require greater levels of expressiveness to recognize an emotional expression and are potentially less sensitive to subtle emotional expressions. The data presented here support that hypothesis, by showing that adolescents with ASD showed a significantly sharper drop in accuracy from high-intensity to low-intensity stimuli, compared to their TD peers. This finding may therefore reveal underlying deficits in adolescents with ASD that are otherwise masked through the common use of high-intensity emotional expressions in research studies.

In addition, adolescents with ASD appear to integrate and use both visual and auditory emotional information in a face-voice matching task and are highly susceptible to manipulation of saliency in both modalities. This can be seen in the fact that the ASD group was significantly less accurate for trials involving low-intensity emotional prosody *and* a more subtle, within-valence facial expression contrast for emotion matching. This effect is also demonstrated in the pairwise comparisons of the different conditions, where both participant groups showed significantly

lower accuracy for low intensity versus high intensity samples, but only the ASD group demonstrated significantly lower accuracy for the more subtle within-valence face contrast trials compared to the across valence trials for stimuli with low-intensity emotion. Overall these data strengthen and support the limited evidence available so far showing decreased accuracy for recognition of low-intensity emotional expressions. Since most of the published data relate to the responses of individuals with ASD to low-intensity *facial* expressions, our data provide an important contribution to the literature on recognition of low-intensity *prosodic* affect in this population, as well as on the *integration* of emotional cues from facial and vocal modalities within the same task.

Our data expand on the study conducted by Mazefsky and Oswald (2007) who investigated processing of high- versus low-intensity facial and prosodic expressions by individuals with ASD. There is some correspondence between our data and those reported by Mazefsky and Oswald, both indicating greater difficulty among participants with ASD in the processing of low-intensity stimuli. However, the work by Mazefsky and Oswald presented stimuli in each modality (face and voice) in separate tasks and asked participants to identify emotions using verbal written labels. In contrast, our study used a more ecologically valid task of matching an emotional sentence utterance to one of two emotional facial expressions. Our task was intended to provide stimuli and a study design that was more closely related to the requirements of social interactions, where individuals must extract speaker intent and emotional state from both auditory and visual channels and integrate those two modalities. The task presented here introduced an additional level of difficulty through the within-valence versus across-valence facial expression contrasts. This second level of difficulty revealed the important finding that individuals with ASD are significantly less skilled at choosing the correct face-voice match when the two facial expression options are more similar to each other. This result clearly indicates that task demands play an important role in determining the emotion processing abilities of individuals with ASD.

According to Harms et al. (2010) one factor contributing to the conflicting data on facial and vocal emotion recognition in the published literature may be found in the potential use of compensatory strategies by individuals with ASD. In the study presented here, the binary facial expression choice allowed for the use of such compensatory strategies in half the trials. In the across-valence samples, participants had the opportunity to verify their initial interpretation of each emotional prosodic expression, or support a guess, along the valence domain. If participants with ASD were unsure of their interpretation of the auditory emotion, they were able to use the compensatory cognitive strategy of narrowing down the possible options by valence. Without having to determine the specific emotion, participants could use a method of elimination, excluding the facial emotion clearly representing the opposite valence of the auditory stimulus. In trials where the facial expression choice was within-valence, participants were forced to commit to their choices of a specific emotion without additional cues about valence being provided. Although TD participants were less accurate for emotional expressions produced with more subtle emotion, they showed no difference in accuracy for across-valence versus within-valence face contrast trials. This indicates that they did not require secondary confirmation of their prosodic affect interpretation through the across-valence facial expression choice. Overall, the significantly lower accuracy levels of the ASD group for trials involving both the low-intensity emotion cues *and* the within-valence face contrast shows that adolescents with ASD are more susceptible to cue-intensity manipulations of both modalities than their TD peers and may use facial expression valence to confirm their interpretations of prosodic affect. The use of compensatory strategies is one possible explanation for these findings, but must be further investigated for confirmation. It is also possible that the reduced accuracy of adolescents with ASD for the low-intensity-within-valence conditions was simply caused by increased task difficulty, which may have affected this population more than their TD peers, who have less difficulty with emotion processing and face processing in general.

One important implication of our findings is that individuals with high functioning ASD, who show normal accuracy levels for high-intensity emotions, still have great difficulty interpreting low-intensity emotional expressions. During daily social interactions, adolescents with ASD are more likely to encounter subtle, low-intensity facial and vocal expressions, rather than high-intensity expressions. Our present results may begin to explain why high-functioning individuals with ASD are capable of performing emotion recognition tasks for high-intensity, canonical emotional expressions in research environments, but continue to have great difficulty interpreting more subtle emotion and speaker intent during everyday face-to-face interactions.

This interpretation is also supported by existing data on children with nonverbal learning disability (NLD). Baum and Nowicki (1998) used the diagnostic analysis of nonverbal accuracy-adult prosody (Nowicki and Duke 1994) to present high- and low-intensity prosodic stimuli to adults with NLD who were characterized using a range of standardized tests. One of their main findings was that reduced accuracy scores for low-intensity prosodic stimuli were significantly correlated with standardized scores reflecting increased social difficulty. This correlation between

recognition of low-intensity prosodic emotion and social skills may help explain the social deficits shown by individuals with high functioning autism who can recognize high-intensity prosodic emotions, but have significant deficits recognizing low-intensity prosodic emotions. Children and adolescents with ASD consistently present with deficits in social interactions as one of the central tenets of their diagnostic profile. And yet many studies document intact recognition of emotional facial and vocal expressions in this population (Baron-Cohen et al. 1997; Gepner et al. 2001; Rosset et al. 2008; Grossman et al. 2000). Based on our results, we propose that this discrepancy between intact lab-based performance and decreased ability to interpret emotional expressions in daily social interaction may at least in part be explained by the lower intensity of natural facial and vocal emotional expressions encountered during typical social engagement.

### Limitations and Future Directions

Our data present unique and novel information on the processing of high- versus low-intensity emotion in the context of an ecologically valid task using within-valence and across-valence facial expression matching. One limitation is that we were not able to collect reliable reaction time data. Using accuracy data alone, it is not possible to conclusively determine that individuals with ASD use a compensatory strategy to achieve face-voice matching of low-intensity emotional stimuli. Our accuracy data strongly indicate this to be the case, but future studies should collect reaction time data as well, to verify that theory.

In order to maintain sufficient power to conduct the targeted comparisons across all emotional intensity and face contrast conditions, we did not separate the data according to emotions. Doing so would have given us 32 conditions, rather than eight, with only two samples per condition per participant. Future investigations may want to consider using more samples in order to create a larger corpus of data and allow for analysis of each of the conditions (all variations of emotional intensity and face valence contrast) across different emotions. Another avenue for follow-up studies would be to use large enough cohorts of participants across the entire autism spectrum to allow for comparisons of participants with greater or fewer social deficits. This would enable further investigation into whether degree of social impairment is correlated with reduced ability to determine affect from low-intensity prosodic information.

Finally, the inability of either participant group to differentiate between stimuli showing happy and surprise emotions warrants further investigation. Our data show that even though each of the face and voice surprise and happy stimuli were successfully recognized in isolation by our pilot participants, the differentiation between happy and surprise proved too difficult in this face-voice matching task. There is evidence to suggest that the dynamic facial expression properties of surprise and happy expressions are very different, particularly in the speed and slope of facial feature movements (Grossman and Kegl 2006). In contrast to natural face-to-face interactions, still photographs of emotional faces don't contain these dynamic features, making the distinction between the facial expressions more difficult to determine. It is possible, that replication of this task with dynamic, as opposed to static, facial expressions would result in improved accuracy for this contrast type.

## Conclusion

High functioning adolescents with ASD are as accurate as their TD peers at matching sentence-length affective prosody to static facial expressions for high-intensity basic emotions. For low-intensity emotional expressions accuracy drops more sharply for individuals with ASD than for their TD peers. When emotional intensity is low *and* the facial expression contrast is within-valence, adolescents with ASD are significantly less accurate at matching affective voices and faces than their TD peers. These data indicate that adolescents with ASD who are capable of discerning high-intensity emotional expressions may still have significant difficulty interpreting affect from more ecologically valid, low-intensity facial and vocal expressions.

**Acknowledgments** The authors wish to thank Rhyannon Bemis, Chris Connolly, and Meaghan Kennedy, for their assistance in stimulus creation, task administration, and data analysis. We also thank the children and families who gave their time to participate in this study. Funding was provided by NAAR, NIDCD (U19 DC03610; H. Tager-Flusberg, PI) which is part of the NICHD/NIDCD Collaborative Programs of Excellence in Autism, and by grant M01-RR00533 from the General Clinical Research Ctr. program of the National Center for Research Resources, National Institutes of Health. The corresponding author is currently supported by NIDCD (R21 DC010867-01; R Grossman, PI). A version of this paper was presented as a poster at the International Meeting for Autism Research in 2009.

## References

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience, 13*(2), 232–240. doi:10.1162/089892901564289.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes?" Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition, 4*, 311–331.

Baum, K., & Nowicki, S. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior, 22*(2), 89–107. doi:10.1023/a:1022954014365.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (version 5.1.05). Retrieved May 1, 2009, from http://www.praat.org.

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry, 41*(7), 847–857.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders, 29*(1), 57–66. doi:10.1023/a:1025970600181.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 864–881.

Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science, 7*(3), 340–359.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology, 3*(1), 69–86.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language, 106*(2), 144–152. doi:10.1016/j.bandl.2008.04.002.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd edn.). Circle Pines, MN: American Guidance Service.

Gepner, B., de Gelder, B., & de Schonen, S. (1996). Face processing in autistics: Evidence for a generalized deficit? *Child Neuropsychology, 2*, 123–129.

Gepner, B., Deruelle, C., & Grynfeltt, S. (2001). Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of Autism and Developmental Disorders, 31*(1), 37–45.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without asperger syndrome. *Journal of Autism and Developmental Disorders, 36*(2), 169–183. doi:10.1007/s10803-005-0057-y.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37*(6), 1096–1106.

Greimel, E., Schulte-Rüther, M., Kircher, T., Kamp-Becker, I., Remschmidt, H., Fink, G. R., et al. (2010). Neural mechanisms of empathy in adolescents with autism spectrum disorder and their fathers. *Neuroimage, 49*(1), 1055–1065. doi:10.1016/j.neuroimage.2009.07.057.

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 53*(3), 778–793.

Grossman, R. B., & Kegl, J. (2006). To capture a face: A novel technique for the analysis and quantification of facial expressions in American sign language. *Sign Language Studies, 6*(3) 273–305.

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, 41*(3), 369–379.

Grossman, R. B., Schneps, M. H., & Tager-Flusberg, H. (2009). Slipped lips: Onset asynchrony detection of auditory-visual language in autism. *Journal of Child Psychology and Psychiatry, 50*(4), 491–497. doi:10.1111/j.1469-7610.2008.02002.x.

Grossman, R. B., & Tager-Flusberg, H. (2008). Reading faces for information about words and emotions in adolescents with autism. *Research in Autism Spectrum Disorders, 2*(4), 681–695. doi:10.1016/j.rasd.2008.02.004.

Harms, M., Martin, A., & Wallace, G. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322. doi:10.1007/s11065-010-9138-6.

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., et al. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 52*(3), 275–285. doi:10.1111/j.1469-7610.2010.02328.x.

Kaufman, A., & Kaufman, N. (2004). *Manual for the Kaufman brief intelligence test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytky, S., et al. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders, 37*(8), 1539–1549. doi:10.1007/s10803-006-0271-2.

Kuusikko, S., Haapsamo, H., Jansson-Verkasalo, E., Hurtig, T., Mattila, M., Ebeling, H., et al. (2009). Emotion recognition in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(6), 938–945.

Law Smith, M. J., Montagne, B., Perrett, D. I., Gill, M., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioing autism using dynamic stimuli of varying intensities. *Neuropsychologia, 48*, 2777–2781.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., et al. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody (Original Research). *Frontiers in Human Neuroscience, 4*. doi:10.3389/fnhum.2010.00019.

Lindner, J. L., & Rosén, 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*(6), 769–777. doi:10.1007/s10803-006-0105-2.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism diagnostic observation schedule-WPS (ADOS-WPS)*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659–685.

Loveland, K. A., Tunali-Kotoski, B., Chen, R., Brelsford, K. A., Ortegon, J., & Pearson, D. A. (1995). Intermodal perception of affect in persons with autism or Down syndrome. *Development and Psychopathology, 7*(03), 409–418. doi:10.1017/S095457940000660X.

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 (Article). *Journal of Autism and Developmental Disorders, 37*(6), 1086–1095. doi:10.1007/s10803-006-0251-6.

Merewether, F. C., & Alpert, M. (1990). The components and neuroanatomic bases of prosody. *Journal of Communication Disorders, 23*(4–5), 325–336.

Murray, I. R., & Arnott, J. L. (2008). Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. *Computer Speech & Language, 22*(2), 107–129. doi:10.1016/j.csl.2007.06.001.

Nowicki, S. (2003). Manual for the receptive tests of the diagnostic analysis of nonverbal accuracy 2. Unpublished manual.

Nowicki, S., & Duke, M. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18*(1), 9–35. doi:10.1007/bf02169077.

O'Connor, K. (2007). Brief report: Impaired identification of discrepancies between expressive faces and voices in adults with Asperger's syndrome (Article). *Journal of Autism and Developmental Disorders, 37*(10), 2008–2013. doi:10.1007/s10803-006-0345-1.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*(6), 861–869.

Pelphrey, K. A., Morris, J. P., McCarthy, G., & LaBar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Social Cognitive and Affective Neuroscience, 2*(2), 140–149. doi:10.1093/scan/nsm010.

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., et al. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine, 40*(11), 1919–1929. doi:10.1017/S0033291709992364.

Rosset, D., Rondan, C., Da Fonseca, D., Santos, A., Assouline, B., & Deruelle, C. (2008). Typical emotion processing for cartoon but not for real faces in children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(5), 919–925. doi:10.1007/s10803-007-0465-2.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32*(3), 189–194.

Rutherford, M., & McIntosh, D. (2007). Rules versus prototype matching: strategies of perception of emotional facial expressions in the autism spectrum. *Journal of Autism and Developmental Disorders, 37*(2), 187–196. doi:10.1007/s10803-006-0151-9.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A. M., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech and Language in Hearing Research, 44*(5), 1097–1115.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, 48*(8), 813–821.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249.