Ability of Children with Autism Spectrum Disorders to Identify Emotional Facial Expressions

Jill Lorenzi

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Psychology

Angela Scarpa-Friedman, Chair
Robin K. Panneton
Susan W. White

May 1, 2012
Blacksburg, VA

Keywords: Emotion Identification, Autism Spectrum Disorders, Eye Tracking, Audiovisual Integration, Children

Ability of Children with Autism Spectrum Disorders to Identify Emotional Facial Expressions

Jill Lorenzi

Abstract

Previous research on emotion identification in Autism Spectrum Disorders (ASD) has demonstrated inconsistent results. While some studies have cited a deficit in emotion identification for individuals with ASD compared to controls, others have failed to find a difference. Many studies have used static photographs that do not capture subtle details of dynamic, real-life facial expressions that characterize authentic social interactions, and therefore have not been able to provide complete information regarding emotion identification. The current study aimed to build upon prior research by using dynamic, talking videos where the speaker expresses emotions of happiness, sadness, fear, anger, and excitement, both with and without a voice track. Participants included 10 children with ASD between the ages of four and 12, and 10 gender- and mental age-matched children with typical development between six and 12. Overall, both ASD and typically developing groups performed similarly in their accuracy, though the group with typical development benefited more from the addition of voice. Eye tracking analyses considered the eye region and mouth as areas of interest (AOIs). Eye tracking data from accurately identified trials resulted in significant main effects for group (longer and more fixations for participants with typical development) and condition (longer and more fixations on voiced emotions), and a significant condition by AOI interaction, where participants fixated longer and more on the eye region in the voiced condition compared to the silent condition, but fixated on the mouth approximately the same in both conditions. Treatment implications and directions for future research are discussed.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

Autism Spectrum Disorders (ASD) are a category of developmental disorders which include Autistic Disorder, Asperger's Disorder, and Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), and are characterized by deficits related to social functioning, communication, and restricted and repetitive behaviors (American Psychiatric Association [APA], 2000). The current estimation is that approximately one in every 88 children in the United States has an ASD, with the incidence increasing in recent years (Centers for Disease Control and Prevention [CDC], 2012). As the disorder now affects more children and families than ever before, a thorough understanding of the underlying mechanisms of ASD becomes all the more critical. One way in which deficits related to social functioning and communication may present themselves in ASD is in difficulties with the ability to infer another person's expressed emotion, a crucial skill in forming and navigating social relationships. However, the majority of previous research on emotion identification in ASD has used static photographs that do not accurately characterize authentic social interactions (e.g., Capps, Yirmiya, & Sigman, 1992; Castelli, 2005). This study will build on prior research by using dynamic, talking videos, and by considering what additional benefit may be provided by the presence of a voice track. Additionally, implications for intervention and improving quality of life for individuals with ASD will also be discussed.

An accurate understanding of others' emotions is a key component of social interactions and enables one to make sense of the social world. Emotional expressions provide crucial information as to the intent of a speaker or any individual with whom a person might interact or even simply observe. The information gleaned from others' emotions allows for successful navigation of everyday life, and provides the basis for meaningful relationships. Without emotional reciprocity, what might be seen as a mutual exchange of affectionate love, for example, could otherwise be construed as a cold business meeting. Emotion recognition becomes important starting at a young age, as shown by the positive relationship in early childhood of emotion understanding and the development of adaptive social behavior (Izard, Fine, Schultz, Mostow, Ackerman, & Youngstrom, 2001). Typically developing infants as young as three months of age can discriminate between static images of basic emotions, including happy, sad, and surprised (Young-Browne, Rosenfeld, & Horowitz, 1977). Other research on 7-month-olds has demonstrated that these infants can discriminate between congruent and incongruent face-voice pairs, evidencing early support for audiovisual integration of emotion processing (Grossman, Striano, & Friederici, 2006). Later in life, difficulties related to emotion identification in adults are related to poor emotion regulation, as well as increased anxiety and depression (Taylor & Bagby, 2004). Further, psychopathy has been associated with deficits in affect recognition (Hastings, Tangney, & Stuewig, 2008). Clearly, emotion identification pervades much of the social experience and its relationship with psychological well-being speaks to its significance, potentially even cutting across diagnostic boundaries of various forms of psychopathology.

Emotion Identification in ASD

Of the previous work on emotion identification in individuals with ASD, some studies have cited a deficit in ASD compared to controls (Hobston, Ouston, & Lee, 1989; Macdonald et al., 1989), while others have failed to find a difference (Capps, Yirmiya, & Sigman, 1992; Ozonoff, Pennington, & Rogers, 1990). However, a wide range of paradigms have been used involving stimuli with different levels of subtlety and realism, and have included participants with varying levels of functioning, making it difficult to determine implications and generalizability.

Studies that use event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) to investigate facial emotion recognition have had the advantage of being able to include lower functioning individuals who might not be able to provide a behavioral response (i.e., a verbal response or pointing) to the emotion recognition task. As opposed to exclusively behavioral studies that have found inconsistent evidence as to whether there is a deficit in ASD (possibly due to differences in task demands, type of stimuli presented, etc., as mentioned above), nearly all eye-tracking and electrophysiological studies of facial emotion recognition have found abnormalities in both children and adults with ASD (for review, see Harms, Martin, & Wallace, 2010); however, the majority of studies have predominantly focused on adults. As such, methodological differences may explain the inconsistency in findings to date among children with ASD.

Additionally, the use of static photographs (e.g., Capps et al., 1992; Castelli, 2005) as opposed to dynamic videos has dominated much of the research in the area of emotion identification in ASD; unfortunately, static pictures do not capture the subtle details of dynamic real-life facial expressions that comprise actual social interactions. While studies involving static stimuli may provide preliminary information related to emotion processing, therefore, they are somewhat limited in their usefulness.

Golan, Baron-Cohen, and Golan (2008), for example, found that children with ASD were less able than a typically developing control group to recognize complex emotions presented in dynamic videos. The researchers considered their study to be more ecologically valid than most prior research since the dynamic scenes included visual input, auditory input, and context.

However, these dynamic videos consisted of scenes from children's feature films, whereby participants observed a socio-emotional interaction and were required to infer the actor's mental state. While indeed an improvement over the use of static photographs alone, many real-life situations in which a person is required to infer the mental state of another involve direct interactions between the people involved. In other words, rather than simply observing scenes as an outsider, facial expressions and auditory information are directed specifically to the individual making the inference. Golan et al. (2008) also included a portion that tested emotion recognition in a silent face task as well as a non-visual voice task, but did not consider a combination of how the emotional information from the voice might augment the information from the face. Therefore, it is unclear how children with ASD may perform in response to a dynamic video where the person is speaking directly to the child.

In another study, while viewing silent videos that transitioned from a neutral facial expression to an emotional one, children with ASD took longer to recognize emotions and were less proficient at detecting anger than the typically developing group (Bal, Harden, Lamb, Van Hecke, Denver, & Porges, 2010). Further, less severe ASD symptoms and greater attention to the eye region were also related to better emotion recognition, suggesting the potential of interventions to improve emotion identification. However, similar to the study described above, this research does not consider how children with ASD might benefit from the additional emotional information contained in a voice track when attempting to identify emotional expressions.

Golan, Baron-Cohen, Hill, and Rutherford (2007) developed a task that compared the ability of adults with ASD and those with typical development to recognize emotion from a speaker's voice alone. Results suggested "multi-modal empathizing deficits" (p. 1096) in ASD due to the task's sensitivity in distinguishing the group of adults with ASD from the group with typical development. This study suggests that information from the voice alone is not sufficient for adults with ASD to recognize emotion as well as adults with typical development. However, it is unclear whether emotional facial expressions in dynamic videos lead children with ASD to make comparable judgments to those of individuals with typical development, or whether dynamic emotional videos in combination with emotional information from the voice lead to similar performance when compared to the typical group.

Other recent research has failed to find emotion identification deficits in ASD. Jones et al. (2011) studied adolescents with ASD and reported similar performance compared to adolescents with typical development in terms of facial (using static photographs) and vocal recognition of basic emotions, except for the specific emotion of surprise. Additionally, Tracy and colleagues (2010) cited fast and accurate emotion recognition of static photographs in children and adolescents with ASD, even on some complex emotions, evidencing no significant differences compared to the group with typical development. Finally, although Sawyer, Williamson, and Young (2012) cited deficits in emotion identification of static photographs in adults with Asperger Syndrome compared to adults with typical development, the researchers did not find evidence of gaze avoidance. Findings, therefore, are quite variable and may be due to differences in diagnostic determinations, age, or paradigms employed, but additional research is needed to investigate the effect of dynamic, realistic-to-life stimuli both with and without the information from voice.

Neural Mechanisms and Emotional Face Processing

Neural mechanisms involved with the processing of emotional faces have been investigated extensively in cognitive research. Research suggests that the amygdala is crucially involved in the perception of both positive and negative emotions (Davis & Whalen, 2001), and the detection of information with high social salience, such as the understanding of facial expressions (Vuilleumier & Pourtois, 2007). According to the "amygdala theory" of autism, atypical social development in ASD may be caused by dysfunction in the amygdala and other structures of the "social brain" (Baron-Cohen et al., 2000; Brothers, 1996). Evidence in support of an amygdala abnormality in autism stems from multiple lines of research, including structural and functional neuroimaging (e.g., Abell et al., 1999; Gillberg, I. C., Bjure, J., Uvebrant, P., Vestergren, E., & Gillberg, C., 1993).

Autism is thought to also be associated with reduced activation of subcortical face processing regions including the amygdala, pulvinar, and superior colliculus, as shown in fMRI studies (Kleinhans et al., 2011). This suggests the presence of a core mechanism behind impaired socioemotional processing in ASD. Kleinhans and colleagues also concluded that abnormalities in this particular system may contribute to early deficits in social attention, thereby explaining later differences in social cognition. Related research indicates that it may be the connectivity between the amygdala and fusiform gyrus that is critical to face processing (Herrington, Taylor, Grupe, Curby, & Schultz, 2011), further providing neurological evidence for deficits in ASD.

Lack of Attention to the Eye Region in ASD

Another explanation for deficits in emotion identification in the ASD population might be the tendency of individuals with ASD to avoid looking at the eye region of the face. This region is important in deciphering emotion (e.g., Baron-Cohen, Wheelwright, & Jolliffe, 1997; Boucher & Ekman, 1975; Dalton et al., 2005). Diminished eye contact is a hallmark of the disorder (Jones, Carr, & Klin, 2008), so deficits that require this type of contact, such as emotion recognition, follow naturally. Communicative direct eye gaze and eye contact signals the intention of the speaker to convey information thought to be important to the perceiver. In a review by Itier and Batty (2009), the importance of the eye region was explicated by compiling evidence that all six basic emotions (i.e., joy, fear, anger, sadness, surprise, and disgust) involve a specific transformation of a particular aspect of the eye region, highlighting the significance of the region in emotion identification. However, this is not to suggest that simply looking at the eye region is sufficient for deciphering emotion, as successful processing also involves the interpretation of social meaning.

Kliemann, Dziobek, Hatri, Steimke, and Heekeren (2010) used static emotional stimuli to investigate gaze patterns and found that the ASD group showed a reduced preference for the eye region as compared to the control group, characterized by "active avoidance" of direct eye contact. Likewise, when viewing naturalistic social situations, participants with ASD engaged in visual tracking whereby the salience of the eyes for gathering important social information was reduced, while the importance of other areas such as the mouth, bodies, and objects were increased compared to participants with typical development (Klin, Jones, Schultz, Volkmar, & Cohen, 2002). However, Sawyer et al. (2012) failed to find evidence of gaze avoidance in adults with Asperger Syndrome, suggesting that evidence in support of gaze avoidance in ASD is not necessarily universal.

Whether from an aversive reaction or a failure to detect social saliency, reduced attention to the eyes results in a failure to notice important social information. Therefore, it is possible that emotional facial expressions in silent, dynamic video clips may not provide enough information for individuals with ASD to perform comparably to those with typical development on an emotion identification task. It is yet to be determined, however, whether children with ASD can compensate for this deficit by relying on the additional emotional information contained in the voice. Individuals without ASD do not demonstrate a lack of attention to the eyes, and therefore assumedly would not need to rely as much on information from alternative modalities. This suggests that differences in performance between the two groups may be present when the stimuli include no sound, but possibly would be less discrepant when the stimuli include both an audio and a video component. Likewise, the study by Golan et al. (2007) suggests that emotional information from the voice alone is thought to be insufficient for individuals with ASD to identify emotional expressions at a level comparable to individuals with typical development. As such, multi-modal information from both audio and video may be needed to aid individuals with ASD.

Alternative Explanation

Although individuals with ASD may benefit from the availability of multiple modalities, it is possible that stimulus overselectivity may mean that multiple modalities do not provide any additional information. Stimulus overselectivity describes the phenomenon whereby individuals with ASD may respond to only certain, limited cues in their environment, rather than gathering information from all available cues (Lovaas, Koegel, & Schreibman, 1979). For example, focusing only on the visual cue could mean that the addition of an auditory cue is simply ignored. However, considering the difficulties noted above with regard to differences in facial processing, it is thought that individuals with ASD will need the information from the voice to make accurate emotional judgments.

Multimodal Integration in Emotion Perception

In most social situations, information from both auditory (i.e., voice) and visual (i.e., face) systems must be somehow integrated in order to decode the meaning of another person's expression. In adults, this integration occurs in the early stages of processing and likely at a level outside of conscious awareness (Hietanen, Leppänen, Illi, & Surakka, 2004). Hietanen and colleagues (2004) demonstrated that when adults were provided with information from both the auditory and visual modalities but instructed to ignore one of the modalities, information from the modality that was supposed to have been ignored still influenced emotion recognition.

As for children, convincing evidence based on the results of a preferential looking task suggests that 2-month-old infants can successfully integrate information from both the face and voice (Patterson & Werker, 2003). Additional evidence for audiovisual integration is present in prelinguistic 4.5-month-old infants as it relates to speech perception and the McGurk effect, which is a phenomenon that is commonly employed to investigate the integration of auditory and visual information in speech perception (Burnham & Dodd, 2004). Developmental research indicates that successful integration of information from multiple sensory modalities gradually improves over the course of childhood, sometimes even extending into adolescence (Brandwein et al., 2011).

Children with ASD, however, have been found to evidence deficits in audiovisual integration compared to children with typical development as measured using the McGurk effect, although this deficit is thought to diminish with age (Taylor, Isaac, & Milne, 2010). Therefore, deficits in multimodal emotion identification in young children with ASD may partially be due to underdeveloped audiovisual integration. Research on adolescents with high-functioning autism has evidenced less benefit from the addition of visual information when comparing audiovisual speech perception to individuals with typical development (Smith & Bennetto, 2007), as has research on children with ASD between the ages of 8 and 19 (Mongillo et al., 2008). On the contrary, a similar study has evidenced opposite findings, whereby children with ASD were found to successfully integrate vocal and facial information in speech perception, advocating for language-learning environments that promote multimodal speech instruction (Massaro & Bosseler, 2003). Therefore, conflicting findings have resulted from studies on multimodal integration in children with ASD. Children with ASD may or may not experience deficits in audiovisual integration, but most of the previous research has focused on speech perception. As it relates to emotion perception in children with ASD, it is hypothesized that, given commonly found deficits in emotion identification in this population, children may be more likely to integrate information from both the face and the voice in order to successfully identify emotional expressions.

Implications for Intervention

Various treatments have been developed to help improve potential deficits in emotion identification, though this work still requires replication. Recent research, for example, suggests that oxytocin nasal spray might improve emotion understanding in ASD (Guastella et al., 2010). Oxytocin is thought to improve emotion understanding by reducing anxiety (via amygdala-dependent mechanisms) and to increase rewards associated with engaging in certain behaviors (via dopamine-dependent mesolimbic pathways). Also, a group-based cognitive behavioral intervention in adults with high-functioning autism has demonstrated promising results in improved face emotion identification (Turner-Brown, Perry, Dichter, Bodfish, & Penn, 2008). Research on programs designed to teach emotion understanding to children with autism has also shown potential (e.g., Bauminger, 2002; Hadwin, Baron-Cohen, Howlin, & Hill, 1996; Solomon, Goodlin-Jones, & Anders, 2004). Some preliminary research has also supported the use of neurofeedback treatment for improving reciprocal social interactions in children with ASD (e.g., Kouijzer, van Schie, de Moor, Gerrits, & Buitelaar, 2010).

However, before considering potential interventions that might increase aptitude for emotion identification, it is important to know, at a more basic level, if this skill depends on specific kinds of stimuli or stimuli features that are presented. In other words, are deficits in ASD present when the stimuli include dynamic emotional videos, both with and without the voice track? Are they more profound when the presentation is multimodal (audio and video), or might the information contained in the voice track lead to more accuracy in identifying emotion? Furthermore, many prior studies have used photographs to assess emotion identification, making it unclear whether similar results would be found in more realistic, dynamic contexts. This research intends to clarify the situations under which emotion recognition is impaired, as potential clinical implications may identify the ideal way to present stimuli in order to teach emotion recognition.

Statement of the Problem and Specific Aims

Children with ASD experience unique difficulties related to areas of social functioning and forming meaningful relationships. The ability of children to identify emotional facial expressions in dynamic video clips with both a voice and silent condition is largely unexplored. Specifically, it is unclear whether children with ASD demonstrate deficits in identifying dynamic emotional facial expressions without the aid of information from a speaker's voice. Furthermore, it is also unclear whether children with ASD might depend more on information from the voice than do children with typical development. The proposed study set out to clarify these uncertainties with the use of dynamic, talking video clips that have been absent from most related research.

This study explored three specific aims related to the questions above:

1. Determine whether children are more accurate at identifying emotions expressed in dynamic videos when the voice track is present as opposed to when the videos are silent.
2. Determine whether children with ASD perform worse than children with typical development at identifying emotions expressed in silent, dynamic videos.
3. Determine whether children with ASD benefit differentially from the addition of a voice track when identifying emotions expressed in dynamic talking videos as compared to children with typical development.

Primary Hypothesis

To address these aims, there was one primary interactive hypothesis. It was predicted that there would be a condition (voiced, silent) by group (ASD, typical development) interaction effect with regard to emotion identification accuracy, whereby children with ASD would benefit more from the presence of the voice than would children with typical development. Analyses of main effects for condition and group were also considered. Further, eye tracking analyses with primary areas of interest (AOIs) defined as the eye region and mouth explored possible explanations for any group or condition differences that were found (i.e., whether differences in looking between the silent/voice conditions explained differences in performance). Specifically, fixation durations and fixation count on specified AOIs were considered as they related to performance on the emotion identification task.

Method

Participants

Participants included 10 children with typical development between the ages of four and 12 years (9 males, 1 female), and 10 children with a diagnosis of Autistic Disorder or Asperger's Disorder between the ages of six and 12 years (9 males, 1 female). Participants in the ASD group had previously received a professional ASD diagnosis, which was required to be confirmed for the purpose of this study using the autism symptom measure described below. The two groups were matched on both mental age (MA) and gender at the group level. Exclusionary criteria for both groups included uncorrected vision problems, seizures that have not been managed for at least six months, receipt of benzodiazepine or anti-psychotic medication, or a non-English language of origin; children in the typically developing group were not permitted to have a sibling with ASD.

Participants for the group with ASD were recruited through the Virginia Tech Autism Clinic, e-mail listservs, and e-mails sent to autism organizations local to the New River Valley region in southwestern Virginia. Participants for the group with typical development were recruited through the Child Participant Database, a research registry maintained by the Department of Psychology at Virginia Tech. Parents of children in this database were invited via telephone to participate in the study.

The target sample size was determined by computing a power analysis. It was hoped that 26 participants in the group with typical development and 26 participants in the group with ASD would provide sufficient power to detect an effect comparable to those found in previous studies on facial emotion recognition (for review, see Harms, Martin, & Wallace, 2010). However, due to difficulties recruiting participants with ASD in a rural geographic area, the target sample size was later reduced to 10 participants per group. The reduced sample size suggests that this study was largely underpowered to detect differences in emotion identification accuracy.

However, the eye tracking portion of the study was better powered to detect differences in gaze patterns between small groups. Eye tracking provides hundreds of data points per participant, which enhances the power to detect group differences. Many recent eye tracking studies on ASD have included no more than 9 to 13 participants per group (e.g., Spezio, Adolphs, Hurley, & Piven, 2007; Wilson, Brock, & Palermo, 2010; Sasson, Elison, Turner-Brown, Dichter, & Bodfish, 2011), due to the nature of eye tracking data as a continuous metric.

Measures

Cognitive measure. The *Kaufman Brief Intelligence Test, Second Edition* (KBIT-2; Kaufman & Kaufman, 2004) was used to assess both verbal and nonverbal intelligence of all child participants. The KBIT-2 also provided a composite IQ score with a mean of 100 and standard deviation of 15 and took approximately 15-30 minutes to complete, making it a brief though valid measure of IQ. The KBIT-2 is acceptable for individuals between the ages of four and 90 and has good reliability (internal consistency = .93, test-retest reliability = .90) (Kaufman & Kaufman, 2004). This measure was used to determine participants' MA, the variable on which participants were matched. Since MA is a known mediator for differences in social cognition, it was important to match participants on this variable or otherwise consider it as a covariate in statistical analyses (Buitelaar, van der Wees, Swaab-Barneveld, & van der Gaag, 1999; Happe, 1995). MA was measured in months and calculated using the following formula: (age in months * composite IQ score) / 100.

Autism symptom measure. The *Social Responsiveness Scale* (SRS; Constantino & Gruber, 2005) is a parent-report questionnaire that assesses social impairments including social awareness, social information processing, capacity for reciprocal social communication, social anxiety/avoidance, and autistic preoccupations and traits, and can be used with children between the ages of four and 18 (Constantino & Gruber, 2005). More broadly, the scale includes 65 items which can be answered on a scale from "0" (never true) to "3" (almost always true), providing an overall rating of autism symptom severity as well as a score on five different subscales (social awareness, social cognition, social communication, social motivation, and autistic mannerisms). This questionnaire took parents approximately 15-20 minutes to complete.

The SRS is shown to have good sensitivity (0.77) and specificity (0.75) in terms of supporting an ASD diagnosis using the above-mentioned cutoff scores, as well as good inter-rater reliability between parents (0.91) and test-retest reliability with three months between administrations (0.88) (Constantino et al., 2003). For the current study, Cronbach's $\alpha = 0.98$.

The SRS was used to characterize the ASD sample and provide some evidence in support of the reported diagnosis. Inclusion in the typically developing group required a *T*-score below the cutoff of 65, while inclusion in the ASD group required a *T*-score above the clinical cutoff of 65.

Demographic survey. A demographic survey (Appendix A) gathered information including parent's education level, relation to the child, and income, and the child's age, race, education level, diagnosis, current symptoms (including seizures), sibling diagnostic status, medications taken, and primary language.

Emotion identification task. Participants observed video clips of Caucasian adult women talking while exhibiting one of five emotions (happiness, sadness, fear, anger, and excitement) that were presented on a computer monitor. The monitor on which participants viewed the video clips also included a Tobii T60 eye tracker, which afforded the ability for the participant to observe video clips while being eye tracked with non-intrusive infrared diodes that are built directly into the computer monitor. This technology does not restrict participants' comfort or head movement. The Tobii T60 eye tracker collects gaze data points 60 times per second for each eye.

The video clips that were presented were rated in advance by a group of 10 volunteers comprised of faculty and graduate students in the Department of Psychology. Raters were asked to judge child-appropriateness as well as the success of each video clip in communicating the desired emotion. Raters reported the emotion that they believed each clip expressed, as well as the intensity of the emotion expressed, on a scale from 1 to 5 with "1" being the lowest intensity and "5" being the highest intensity. Video clips were selected for inclusion into the set of stimuli according to which had the highest level of agreement as to the emotion type. All video clips that were included in the study had inter-rater agreement as to the emotion type of at least 0.8, and average emotion intensity of at least "3" on the scale from 1 to 5.

Prior to the start of the emotion identification videos, participants completed a manipulation check where they were shown pictures of two different animals, and asked to identify the color of each animal after each one. This manipulation check allowed the examiner to confirm that the participant understood that he or she was to respond out loud to the question recited by the computer, and to respond about the video that he or she had just seen. If needed, the participant was able to ask any questions of clarification about the task after the end of the manipulation check and before the start of the emotion identification videos.

Next, in each brief 2-4 second emotion identification video clip, the woman read a sentence (e.g., "I can't wait to go up there!" or "My best friend is moving away today.") while expressing the congruent corresponding emotion. After each clip was shown, the participant was asked, "What is she feeling?" for which his or her verbal response was documented before moving on to the next clip. If the participant did not answer in 15 seconds, he or she was prompted with the same question. After another 15 seconds without a response, the examiner documented the non-response (considered an incorrect response) and moved on to the next clip.

A total of 20 different clips were presented to each participant: two different clips for each of five emotions for both voice present and voice absent variations. All participants viewed both variations of the video clips and were asked to verbally identify the emotion demonstrated in both variations. The clips were ordered randomly and were interspersed with short cartoon clips in an effort to maintain the attention of each participant. Prior to each of the stimuli being presented, a target appeared in the middle of the screen. Participants were instructed before the task to look at the center of the target each time it appeared.

Procedure

After expressing their intent for their child to participate in this study, parents were administered a brief 10-minute phone interview in order to communicate information about the study, explain study procedures, and obtain verbal consent. The interviewer gathered information regarding the child's eligibility, including information about the child's diagnosis, verbal ability, and ability to sit unassisted. If eligible, the parent and child were scheduled to attend a laboratory session of approximately one hour.

At the laboratory session, the experimenter again explained the procedures of the study and answered any questions that the family had. Following both parental consent (Appendix B) and child assent (Appendix C), the parent completed the SRS and demographic survey while the child completed the KBIT-2. After a short break, the participant began watching the dynamic video clips on a computer monitor for the emotion identification task. Last, the participant was debriefed and thanked for their participation. Approval for these procedures was granted by Virginia Tech's Institutional Review Board.

Setup for the emotion identification task was standardized across participants. Participants sat in a chair that faced a 17-inch monitor which included the infrared corneal sensors and the aforementioned Tobii T60 system. Screen resolution was 1280 x 1024 pixels. The monitor, mounted on a moveable arm, was adjusted to position the screen approximately 65 centimeters from participants' eyes. Participants' view of their parent, the examiner, and equipment in the room was blocked by dark curtains once the trial was underway, although the examiner had a clear view of the participant at all times via a camera positioned at the top of the monitor. Immediately prior to testing, a 5-point calibration procedure was employed whereby the examiner sequentially flashed five captivating geometric figures on-screen in order to begin the process of data collection. If the child elected to take a break in between the sets of videos that were presented, the calibration procedure was repeated prior to the presentation of the next set of videos. Audio was presented via a set of two concealed external speakers.

Eye Tracking Data Reduction

Tobii Studio 3.0 software was used to facilitate the eye tracking analyses. AOIs defined as the eye region and mouth were identified by specifying key time points throughout the dynamic video clips, and the shape and position of those regions at each of the key time points. As such, the AOIs followed the movements of the speaker's face and head throughout the videos.

The eye region was ellipse-shaped and included both eyes and eyebrows, excluding as much of the cheeks and bridge of the nose as possible. The mouth region was also ellipse-shaped, and included the lips while excluding as much of the chin as possible. See Figure 1 for an example frame of an emotional video clip with the AOIs defined.

After defining each region, total duration and total count of eye fixations on each of these regions was computed for every video clip. Eye tracking data were constrained to only trials where the participant accurately identified the emotion, to determine whether gaze patterns explained differences in accuracy. For each participant, total fixation duration and count across all accurately identified voiced videos were summed for each AOI, as were total fixation duration and count across all accurately identified silent videos for each AOI.

Data Analytic Plan

An analysis of inter-rater agreement using the Kappa (κ) statistic was performed to determine consistency among two independent raters' judgments of accuracy of participants' emotion identification responses.

Emotion identification accuracy data for both groups were approximately normally distributed, with approximately equal variances and equal sample sizes between groups. To test the interactive hypothesis, where the group with ASD was hypothesized to benefit more from the presence of voice than was the group with typical development, ratio scores were calculated for each participant using the following formula: number of voiced videos identified correctly / (number of voiced videos identified correctly + number of silent videos identified correctly). For each group, an average ratio score was calculated, and tested separately with a one-sample *t*-test for the group with ASD and the group with typical development to determine whether the average ratio score for each group was significantly different from a test score of 0.5 (the score that indicated no difference between voiced and silent accuracy). A score significantly above 0.5 indicated greater accuracy on voiced emotions than silent emotions, while a score significantly below 0.5 indicated greater accuracy on silent emotions than voiced emotions. Differential benefit of the voice between groups was indicated by differences in the significance of the one-sample *t*-tests.

Further analyses were conducted to explore within- and between-group differences. To consider within-group differences in emotion identification accuracy (i.e., comparing across silent and voiced conditions for all participants combined), the aforementioned ratio score was tested for all participants combined with a one-sample *t*-test against a test score of 0.5 (the score that indicated no difference between voiced and silent accuracy). To consider between-group differences in emotion identification accuracy (i.e., comparing across groups for all emotions, silent emotions, and voiced emotions), independent samples *t*-tests were conducted.

With regard to the eye tracking data, a 2 (group) by [2 (condition) by 2 (AOI)] mixed factorial analysis of variance (ANOVA) was conducted separately for the fixation duration data (on accurately identified trials) and fixation count data (also on accurately identified trials). These tests afforded the consideration of whether differences in accuracy could be attributed to differences in gaze patterns.

Results

Descriptive Statistics of Demographics, Mental Age, and Autism Symptoms

Demographic information on child age, child gender, child minority status, mother's education, and family income is presented in Table 1. Of the 10 participants that comprised the group with ASD, five had been previously diagnosed with Autistic Disorder, while five had been previously diagnosed with Asperger's Disorder. Mothers of all participants reported their education as at least some college or beyond. The distribution of mother's education was approximately similar between groups, $\chi^2(2) = 4.67$, $p = .10$. Groups did not differ significantly in their reported distribution of family income, $\chi^2(5) = 10.67$, $p = .06$.

See Table 1 for information on mental age, age, and SRS scores. Since groups were intended to be matched on MA, *t*-tests were computed to be certain that groups were not significantly different in their mean MA, which confirmed a comparable match, $t(18)= -0.31$, $p = .76$. Additionally, groups did not differ significantly in their mean ages, $t(18) = 1.31$, $p = .21$.

However, groups differed significantly in their SRS Total *T*-scores, $t(18) = 11.63$, $p < .001$, which supports the initial selection of groups.

Inter-Correlations

As shown in Table 2, for the group with ASD, MA was significantly positively correlated with the accuracy of identifying negative emotions (i.e., sadness, fear, anger), voiced emotions, and silent emotions. However, for the group with typical development, MA was not significantly correlated with emotion accuracy in any of the conditions. For the group with ASD, age was significantly positively correlated with the accuracy of identifying silent emotions, but no significant relationships between age and positive, negative, or voiced emotions were found. For the group with typical development, no significant relationships between age and positive, negative, voiced, or silent emotions were found. Correlations between SRS scores and accuracy are also presented in Table 2.

Emotion Identification Accuracy (Silent versus Voice Conditions)

Inter-rater agreement regarding accuracy of participants' emotion identification responses was $\kappa = 0.995$. To test the hypothesis that the group with ASD would benefit more from the presence of voice, ratio scores were calculated for each participant and averaged across groups to determine whether the benefit received by the group with ASD differed from the benefit received by the group with typical development. Table 3 shows the ratio scores for each group separately and for the overall sample. For the group with ASD, according to the results of a one-sample *t*-test, the ratio of voiced accuracy to the combination of voiced and silent accuracy was not significantly different from a test score of 0.5, $t(9) = 2.15$, $p = .06$, meaning that participants in the ASD group did not perform significantly better in their ability to identify emotions in voiced videos as compared to silent videos. However, for the group with typical development, according to the results of a one-sample *t*-test, the ratio of voiced accuracy to the combination of voiced and silent accuracy was significantly different from a test score of 0.5, $t(9) = 3.50$, $p = .01$, meaning that participants in the group with typical development performed significantly better in their ability to identify emotions in voiced videos as compared to silent videos. Therefore, the group with typical development benefited more from the presence of voice, as indicated by a significant difference between a test score of 0.5 and the ratio score for the group with typical development but not for the group with ASD. Figure 2 presents a visual depiction of the emotion identification data.

Table 3 also includes the findings regarding a main effect for condition (within-subjects). According to the results of a one-sample *t*-test for all participants combined, a significant effect of condition was found, where the ratio of voiced accuracy to the combination of voiced and silent accuracy was significantly different from a test score of 0.5, $t(19) = 3.99$, $p < .01$, and in the expected direction. This means that for all participants combined, the ability to accurately identify emotions was more proficient in the voiced condition as compared to the silent condition.

Table 4 includes findings related to between-group comparisons of emotion identification accuracy, with independent-samples *t*-tests comparing the group with ASD and the group with typical development. Interestingly, the group with ASD and the group with typical development evidenced similar accuracy in their ability to identify all emotions combined, $t(18) = -0.88$, $p = .39$, voiced emotions, $t(18) = -1.69$, $p = .11$, and silent emotions, $t(18) = 0$, $p = 1$.

Eye Tracking Analyses

Supplemental analyses using eye tracking data afforded the capability to consider differences in gaze patterns across groups, as well as the relationship between looking at certain regions defined in the videos and emotion identification accuracy. For the eye tracking analyses, total fixation duration and count were summed across all accurately identified voiced videos, and all accurately identified silent videos, respectively, for each AOI, for each participant. For fixation duration, a sum of zero seconds for any given AOI and participant was considered missing, rather than considered as a zero in the analyses for that particular region. This was done because total fixation duration of zero seconds meant that the participant never fixated on that particular region on accurately identified videos. However, for fixation count, a sum of zero for any given AOI was included as zero in the analyses.

Quality of eye tracking data (e.g., accounting for eye blinks and time looking off screen) did not differ significantly between the group with typical development and the group with ASD, $t(18) = -1.88$, $p = .08$. However, only accurately identified trials were included in the eye tracking analyses, thereby removing trials whereby failure to look at the screen may have precluded the ability to accurately identify the emotional expression.

Table 5 presents information from a 2 by (2 by 2) mixed factorial ANOVA for fixation duration data on accurately identified trials. The between-subjects factor was group (i.e., ASD, typical development), while the within-subjects factors were condition of video (i.e., voiced, silent) and AOI (i.e., eye region, mouth). As such, this analysis explored differences in fixation duration between group, condition, and AOI, and their respective interactions. The mixed factorial ANOVA indicated that the three-way interaction between group, condition, and AOI was not significant, $F(1,13) = 0.04$, $p = .84$. However, results indicated a significant two-way interaction between condition and AOI, $F(1,13) = 5.46$, $p = .04$. Subsequent paired samples *t*-tests indicated that the difference in fixation duration between voiced and silent conditions differed across the two AOIs, with significantly greater fixation durations on the eye region in the voiced condition as compared to the silent condition, $t(14) = 2.97$, $p = .01$, but approximately similar fixation durations on the mouth in the voiced condition as compared to the silent condition, $t(14) = 0.98$, $p = .35$. See Table 6 for more information and Figure 3 for a visual depiction of the interaction between condition and AOI.

Results also indicated a significant main effect of group, $F(1,13) = 18.50$, $p < .01$, whereby the group with typical development evidenced longer mean fixation durations across all accurately identified videos than the group with ASD. Finally, results indicated a significant main effect of condition, $F(1,13) = 6.62$, $p = .02$, whereby participants evidenced longer mean fixation durations across all accurately identified videos in the voiced condition as compared to the silent condition. No other main or interaction effects were significant. Estimated marginal means for the fixation duration data, constrained to trials that were identified accurately, are presented in Table 7.

Table 8 presents information from a 2 by (2 by 2) mixed factorial ANOVA for fixation count data on accurately identified trials. The between- and within-subjects factors were the same as in the previous analysis (i.e., group, condition, AOI). The mixed factorial ANOVA indicated that the three-way interaction between group, condition, and AOI was not significant, $F(1,18) = 0.48$, $p = .50$. However, as with the fixation duration data, results indicated a significant two-way interaction between condition and AOI, $F(1,18) = 7.48$, $p = .01$. Similar to the fixation duration data, subsequent paired samples *t*-tests indicated that the difference in fixation count between voiced and silent conditions differed across the two AOIs, with significantly greater fixation counts on the eye region in the voiced condition as compared to the silent condition, $t(19) = 4.62$, $p < .01$, but approximately similar fixation counts on the mouth in the voiced condition as compared to the silent condition, $t(19) = 1.94$, $p = .07$. See Table 6 for more information and Figure 4 for a visual depiction of the interaction between condition and AOI.

Results also indicated a significant main effect of group, $F(1,18) = 5.15$, $p = .04$, whereby the group with typical development evidenced a greater mean fixation count across all accurately identified videos than the group with ASD. Finally, results indicated a significant main effect of condition, $F(1,18) = 17.58$, $p < .01$, whereby participants evidenced a greater mean fixation count across all accurately identified videos in the voiced condition as compared to the silent condition. No other main or interaction effects were significant. Estimated marginal means for the fixation count data, constrained to trials that were identified accurately, are presented in Table 9.

Supplemental Analyses

Although not a focus of the current study, preliminary analyses were conducted to explore differences in accuracy in emotion identification for the two different groups by type of emotion. Interestingly, the group with ASD and the group with typical development evidenced similar accuracy in their ability to identify both negative emotions, $t(18) = -0.21$, $p = .83$, and positive emotions, $t(18) = -1.52$, $p = .15$, as presented in Table 4. Additionally, when comparing benefit provided by the addition of voice by emotion type (positive vs. negative emotion) for the group with ASD and the group with typical development, independent samples *t*-tests indicated that both groups benefited similarly from the addition of the voice for positive emotions, $t(18) = -1.41$, $p = .18$, and for negative emotions, $t(18) = -1.13$, $p = .27$.

Discussion

Results did not support the primary hypothesis, which stated that children with ASD would benefit more from the presence of voice than would those with typical development when identifying emotional expressions. Instead, the results of ratio tests unexpectedly indicated that the group with typical development benefited more from the presence of voice than the group with ASD with regard to emotion identification accuracy. However, ratio tests also indicated an overall effect for condition, whereby the ability of the sample as a whole to accurately identify emotions was more proficient in the voiced condition as compared to the silent condition. In addition, independent samples *t*-tests did not support a main effect for group, as the group with ASD and the group with typical development performed similarly across both the voice and silent conditions of the emotion identification task.

The result that participants with typical development benefited more from the addition of voice than did the group with ASD needs to be further explored in future research to determine if the effect is robust, as the current study may not have had enough power to detect the effect in the group with ASD. If replicated, it is not completely clear whether this potential difference may be attributed to general emotion identification deficits in ASD, or to potential deficits in multimodal integration that have been previously cited in children and adolescents with ASD (e.g., Smith & Bennetto, 2007; Taylor, Isaac, & Milne, 2010). However, given that the group with ASD and the group with typical development performed similarly under both the voiced and silent conditions, but the group with typical development benefited more from the presence of voice than the group with ASD, this points to audiovisual integration deficits playing a role.

The finding that the group with ASD did not differ significantly from the group with ASD in the ability to identify either silent or voiced emotional expressions contradicts previous research that has evidenced a deficit in ASD (e.g., Lindner & Rosen, 2006). However, much of the previous research in this area has used static photographs that do not capture details of authentic, dynamic social interactions. This study includes dynamic videos in which the viewer is being spoken to directly, which are more real-to-life than static images and likely more representative of situations in which a person might want to infer the emotional state of another; therefore an improvement over prior research. Additionally, this study also compares the benefits of using the speaker's voice in the stimulus. Both of these aspects speak to the importance of the study and its contribution to existing literature.

Integrating information from the eye tracking data, there is evidence that participants with typical development demonstrated longer total fixation durations and greater fixation counts across accurately identified trials than participants with ASD. However, despite the result of a significant main effect of group with regard to the accuracy data, there were no significant interactions with the group variable, suggesting that although there were less total data for the group with ASD, the two groups tended to evidence similar gaze patterns across conditions, and by AOI, for both the fixation duration and count data. This finding may help to explain the lack of group differences with regard to emotion identification accuracy in this study, as both groups tended to process the accurately identified emotional stimuli in a similar manner according to their fixation patterns (when comparing across conditions, and across AOIs, for accurately identified trials). However, the finding of similar fixation patterns when accurately identifying emotions in dynamic, talking video clips between a group with ASD and a group with typical development should ideally be replicated in a larger study.

Additionally, the significant main effect of condition indicated that across accurately identified trials, participants evidenced longer total fixation durations and greater fixation counts on voiced as compared to silent emotions. Finally, the significant condition by AOI interaction indicated that the relationship between looking for the two different AOIs differed as a function of the condition of video (i.e., voiced or silent). Further tests revealed that for both fixation duration and fixation count across accurately identified trials, the difference in fixation data between the silent and voiced conditions was significant for the eye region, but not for the mouth region. In other words, for accurately identified trials, participants tended to look significantly more at the eye region when the video was voiced as compared to silent, but looked approximately the same at the mouth region in both the voiced and silent conditions. Therefore, this indicates that, for accurately identified trials, the presence of voice was associated with greater attention to the eye region than was the absence of voice.

Across the eye tracking analyses, the fixation duration and fixation count metrics typically revealed similar results (in the two mixed factorial ANOVAs, both metrics evidenced the same significant main effects and interaction effect). Therefore, in this study, neither metric emerged as the clear frontrunner in its capacity to characterize group differences relating to identifying emotional facial expressions.

The relationship between gaze patterns and emotion identification accuracy requires additional consideration. For example, a long duration of time spent fixated on the eye region or the face is not necessarily sufficient for an accurate emotion identification response. The ability to accurately identify another person's facial expression requires not just looking at emotional regions on his or her face, but also the capacity to process and successfully interpret the information gleaned from the face (and the voice, when present). Likewise, recent research has suggested that gaze avoidance does not explain deficits in emotion identification in adults with ASD (Sawyer et al., 2012).

In addition, fixating on one specific region does not completely inhibit perception of other areas outside that region. For example, a child who avoids looking at the face (or for a child for whom a certain region is not salient), and spends more time looking at the hair, ears, and background, may still obtain information from nearby areas on the face, despite having only fixated on regions besides the face. Although the majority of information is obtained and processed during fixations, the processed information is not specific to only the point of fixation.

The finding that all participants more accurately identified emotions in the voice condition as opposed to the silent conditions suggests that stimulus overselectivity is not impairing emotion recognition when multiple modalities (i.e., audio and video) are used. This is encouraging in terms of interventions for children with ASD, as it suggests that the additional information from the voice does not complicate or impede the ability to accurately identify basic emotions.

Importantly, results from this study regarding significantly better emotion identification accuracy for voiced emotions as compared to silent, and similar emotion identification performance across both the group with ASD and the group with typical development suggest two main findings: one with implications for intervention, and the other suggesting a potential overestimation of emotion identification deficits in past studies. First, greater accuracy for voiced emotions begs that interventions be mindful of ways in which children, particularly children with ASD, are most likely to learn. When designing interventions related to emotion identification and recognition, results from this study suggest that it is more difficult for children to identify emotions without the aid of voice. Therefore, it makes little sense to try to “simplify” emotional stimuli by removing important information, such as the voice track. Even children with ASD in this study benefited from having information from multiple modalities available to them, albeit not as much as children with typical development. It is likely the case that interventions that use realistic-to-life, dynamic stimuli may make it easier for children to learn. In terms of designing interventions, it may be sensible to initially use stimuli that contain plenty of emotional information (e.g., clear information from both face and voice), before advancing stimuli that are more challenging due to the limited amount of emotional information conveyed (e.g., information from either face or voice, or less exaggerated facial/vocal information). This would allow individuals to first master and learn from stimuli that are thought to be easier to identify, and to then attempt to master more challenging stimuli. Real-world interactions are sometimes characterized by imperfect information, so learning to make sense of limited information is an important skill to eventually acquire. Fortunately, research on interventions designed to teach emotional understanding to children with ASD has shown promise (e.g., Bauminger, 2002; Solomon, Goodlin-Jones, & Anders, 2004). These interventions may be improved by incorporating information from studies such as the current one regarding conditions under which children with ASD are most likely to be successful, and how their ability to navigate the social world might be improved.

Second, the overall similar emotion identification performance and gaze patterns of children with ASD as compared to children with typical development in this study suggest that previous studies that have used primarily static, unrealistic-to-life photographs may have overestimated the magnitude of the difference between these two groups. The use of dynamic, realistic-to-life videos produced similar performance on the emotion identification task and similar eye tracking gaze patterns across groups. This suggests that there may be ways in which deficits can be minimized, contingent upon the use of proper stimuli. Either way, findings from this study are encouraging in terms of the capability of children with ASD to benefit from the presence of voice, and for deficits in emotion identification to perhaps not be as global and impairing as previously suggested.

Limitations

This study is admittedly not without its limitations. First, the range of ages and MA was fairly wide, and different results may have been found had a narrower range of age or MA been targeted. For example, the hypotheses that were not confirmed may have in fact been found when looking at a more narrow range of participants (e.g., targeting the lower end of the MA range). However, using a more narrow range of participants would have then made findings less generalizable.

Another major limitation is that parents were not asked what types of or how much intervention their child had received, with social skills interventions being particularly applicable. The overall lack of significant group differences related to both emotion identification and gaze patterns could potentially be explained by participants in the ASD group having received interventions that provided specific instruction in emotion identification. This limitation is one that could be easily addressed by including a question in the parent forms concerning receipt of certain types of interventions.

While a manipulation check was included at the beginning of the emotion identification task to confirm that participants understood to state their response out loud, and to respond for the video that they just saw, the check did not confirm that participants understood that their task was to name the emotion that the person in the video was feeling. The question, “What is she feeling?” seemed to have been confusing to younger participants who sometimes took the question very literally, and named an object rather than an emotion. Therefore, while the manipulation check was successful in confirming that responses were stated out loud in response to the previously viewed video, it did not confirm that participants knew to or were able to respond with the name of an emotion. A more effective manipulation check might have also confirmed that participants understood their goal was to name an emotion or feeling as a response to the prompt.

Additionally, this study only portrayed basic emotions (i.e., happiness, sadness, fear, anger, excitement). Therefore, results are not generalizable to more complex emotions, such as shame, suspicion, and guilt, although studying them would be particularly worthwhile. Overall, there were limited group differences related to emotion identification accuracy, but the inclusion of more complex emotions may have produced more extensive group differences. Some previous research has found that deficits are present for complex emotions, but not for simple emotions (e.g., Baron-Cohen, Wheelwright, & Joliffe, 1997). This must be further investigated with the use of dynamic videos in order to determine the similarities and differences between emotion identification in simple and complex emotions.

Next, MA is a known mediator of social cognition (e.g., Buitelaar, van der Wees, Swaab-Barneveld, & van der Gaag, 1999; Happe, 1995). However, although there were some significant correlations in the group with ASD between MA and performance on the emotion identification task, there were no significant correlations in the group with typical development (for either overall MA or verbal MA). This is a surprising finding and suggests that while the task seemed to tap into social cognitive resources in the group with ASD, as evidenced by the significant correlations, it did not seem to do so for the group with typical development. However, it is not the first time that the relationship between MA and performance has been found in the group with ASD but not in the group with typical development (e.g., Dyck, Piek, Hay, Smith, & Hallmayer, 2006). It is possible that the task was not complex enough to require the capacity of social cognitive resources that would typically be related to MA. Additionally, it is possible that motivational issues may have interfered, although it appeared that all participants put forth good effort on the task.

The brief duration of the video clips (2-4 seconds per video) meant that the eye tracking data captured only the first few initial fixations, and did not capture further exploration of the emotional faces. This is a limitation because it cannot be determined how gaze patterns would differ had participants had more of an opportunity for visual exploration. Therefore, in the future, longer video clips should be used to capture more information about gaze patterns over a longer period of time.

Results are not generalizable beyond the particular age range and demographic qualities of the sample studied. Only participants with verbal abilities were included in the study, which means that findings do not generalize to participants without verbal abilities or at lower levels of functioning. The small sample size also requires caution in terms of making conclusions regarding the findings. Although a larger sample size was initially targeted, difficulties with recruitment necessitated the reduction to 10 participants per group. A larger number of participants would have provided more statistical power in order to detect any effects that might have been present. An initial power analysis suggested that 26 participants per group were necessary to detect group differences on emotion identification accuracy, assuming a large effect size, suggesting that this study was underpowered to detect group differences on the emotion identification task. However, even with only 10 participants per group, significant differences were found with regard to the interactive accuracy hypothesis. Further, the eye tracking portion of the study was better powered to detect group differences due to the continuous nature of the eye tracking data. A larger sample size may have detected additional group differences. Any future studies should aim to recruit a larger sample of participants so as to make more conclusive statements regarding the results.

Conclusions and Future Directions

Overall, results of ratio tests indicated that participants with typical development benefited more from the addition of voice than participants with ASD when identifying emotions in dynamic, talking videos in a gender- and MA-matched sample. Overall, participants were more accurate at identifying emotions for voiced as opposed to silent emotions. Both the group with ASD and the group with typical development performed similarly with regard to their ability to identify voiced emotions, as well as their ability to identify silent emotions. Fixation duration and fixation count data on the accurately identified trials evidenced significant main effects of group (longer and more fixations for the group with typical development) and condition (longer and more fixations on voiced emotions), and a significant condition by AOI interaction, where participants fixated longer and more on the eye region in the voiced condition compared to the silent condition, but fixated on the mouth approximately the same in both conditions.

Depending on the goals of the study, future research should consider using dynamic, talking videos in emotion identification tasks. Emotion identification accuracy on tasks involving dynamic videos and tasks involving static pictures may be directly compared. Additionally, a wider range of emotions, and more complex emotions, should also be explored. Duration of the emotional video clips should be increased so as to capture more information about gaze patterns over a period of time, rather than simply the first few initial fixations. The role of psychophysiological arousal in emotion identification (e.g., pupil size or measure of heart rate variability) should also be investigated in more depth. The use of a forced-choice paradigm for the emotion identification task would also allow for the inclusion of individuals with less advanced language abilities. Finally, studies investigating the perception of mismatched facial and vocal emotion would be a valuable contribution to the literature.

Implementing the results of studies such as this into interventions designed to improve emotion identification and social cognition is crucial. These interventions should be carefully designed and tested to ensure maximal benefits. Investigating and subsequently determining the most successful means of ameliorating social and communicative deficits in ASD is of utmost importance.

References

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., …Frith, U. (1999). The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. *Cognitive Neuroscience, 10*, 1647–1651.

American Psychiatric Association (APA). (2000). Diagnostic and statistical manual of mental disorders (Revised 4th ed.). Washington, DC.

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin, C., & Williams, S. (2000). The amygdala theory of autism. *Neuroscience and Behavioural Reviews, 24*, 355-364. doi:10.1016/S0149-7634(00)00011-7

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*(3), 311-331. doi:10.1080/713756761

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of Autism and Developmental Disorders*, *40*(3), 358-370. doi:10.1007/s10803-009-0884-3

Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of Autism and Developmental Disorders, 32*(4), 283-298. doi:10.1023/A:1016378718278

Boucher, J. D., & Ekman, P. (1975). Facial areas and emotional information. *Journal of Communication, 25*(2), 21-29.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. *Cerebral Cortex, 21*(5), 1042-1055. doi:10.1093/cercor/bhq170

Brothers, L. (1996). Brain mechanisms of social cognition. *Journal of Psychopharmacology, 10*, 2-8. doi:10.1177/026988119601000102

Buitelaar, J. K., van der Wees, M., Swaab-Barneveld, H., & van der Gaag, R. J. (1999). Verbal memory and Performance IQ predict theory of mind and emotion recognition ability in children with autistic spectrum disorders and in psychiatric control children. *Journal of Child Psychology and Psychiatry, 40*(6), 869–881. doi:10.1111/1469-7610.00505

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220. doi:10.1002/dev.20032

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 1169–1182. doi:10.1111/j.1469-7610.1992.tb00936.x

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

Centers for Disease Control and Prevention (CDC) (2012). Prevalence of autism spectrum disorders. *MMWR Surveillance Summary, 61*(SS03), 1-19.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 33*(4), 427-433. doi:10.1023/A:1025014929212

Constantino, J. N., & Gruber, C. P. (2005). *The Social Responsiveness Scale*. Los Angeles: Western Psychological Services.

Dalton, K.M., Nacewicz, B.M., Johnstone, T., Schaefer, H.S., Gernsbacher, M.A., Goldsmith, H.H., … Davidson, R.J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519-526. doi:10.1038/nn1421

Davis, M. & Whalen, P.J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*, 13–24. doi:10.1038/sj.mp.4000812

Dyck, M. J., Piek, J. P., Hay, D., Smith, L., & Hallmayer, J. (2006). Are abilities abnormally interdependent in children with Autism? *Journal of Clinical Child & Adolescent Psychology*, 35, 20–33. doi:10.1207/s15374424jccp3501_3

Gillberg, I. C., Bjure, J., Uverbrant, P., Vestergren, E., & Gillberg, C. (1993). SPECT in 31 children and adolescents with autism and autistic like syndromes. *European Child and Adolescent Psychiatry, 2*, 50-59.

Golan, O., Baron-Cohen, S., & Golan, Y. (2008). The “reading the mind in films” task (child version): Complex emotion and mental state recognition in children with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 38*, 1534–1541. doi:10.1007/s10803-007-0533-7

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. doi:10.1007/s10803-006-0252-5

Grossman, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*(3), 309-315. doi: 10.1111/j.1467-7687.2006.00494.x

Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., & Hickie, I. B. (2010). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biological Psychiatry*, *67*(7), 692-694. doi:10.1016/j.biopsych.2009.09.020

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development and Psychopathology*, *8*(2), 345-365. doi:10.1017/S0954579400007136

Happe, F. G. E. (1995). The role of age and verbal-ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843–855. doi:10.1111/j.1467-8624.1995.tb00909.x

Harms, M. B., Martin, A., & Wallace, G. L. (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

Hastings, M. E., Tangney, J. P., & Stuewig, J. (2008). Psychopathy and identification of facial expressions of emotion. *Personality and Individual Differences, 44*(7), 1474-1483. doi:10.1016/j.paid.2008.01.004.

Herrington, J. D., Taylor, J. M., Grupe, D. W., Curby, K. M., & Schultz, R. T. (2011). Bidirectional communication between amygdala and fusiform gyrus during facial recognition. *NeuroImage, 56*(4), 2348-2355. doi:10.1016/j.neuroimage.2011.03.072

Hietanen, J.K., Leppänen, J.M., Illi, M., and Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology, 16*(6), 769-790. doi:10.1080/09541440340000330

Hobson, R., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British Journal of Psychology, 79*, 441–453. doi:10.1111/j.2044-8295.1988.tb02745.x

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews, 33*, 843-863. doi:10.1016/j.neubiorev.2009.02.004

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science, 12*, 18–23. doi:10.1111/1467-9280.00304

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happe, 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*(3), 275-285. doi:10.1111/j.1469-7610.2010.02328.x

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*(8), 946-954. doi:10.1001/archpsyc.65.8.946

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test, Second Edition*. Bloomington, MN: Pearson, Inc.

Kleinhans, N. M., Richards, T., Johnson, C. L., Weaver, K. E., Greenson, J., Dawson, G., & Aylward, C. F. (2011). fMRI evidence of neural abnormalities in the subcortical face processing system in ASD. *NeuroImage, 54*(1), 697-704. doi:10.1016/j.neuroimage.2010.07.037

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H. R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *The Journal of Neuroscience*, *30*(37), 12281-12287. doi:10.1523/JNEUROSCI.0688-10.2010

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002) Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*, 809–816. doi:10.1001/archpsyc.59.9.809

Kouijzer, M. E. J., van Schie, H. T., de Moor, J. M. H., Gerrits, B. J. L., & Buitelaar, J. K. (2010). Neurofeedback treatment in autism. Preliminary findings in behavioral, cognitive, and neurophysiological functioning. *Research in Autism Spectrum Disorders, 4*(3), 386-399. doi:10.1016/j.rasd.2009.10.007

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*(6), 769-777. doi:10.1007/s10803-006-0105-2

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86*(6), 1236-1254. doi:10.1037/0033-2909.86.6.1236

Macdonald, H., Rutter, M., Howlin, P., Rios, P., LeConteur, A., Evered, C., et al. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry, 30*, 865–877. doi:10.1111/j.1469-7610.1989.tb00288.x

Massaro, D.W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal of Developmental and Learning Disorders, 7*, 111-144.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 38*, 1349-1358. doi:10.1007/s10803-007-0521-y

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1990). Are there emotion perception deficits in young autistic children? *Journal of Child Psychology and Psychiatry, 31*, 343–361. doi:10.1111/j.1469-7610.1990.tb01574.x

Patterson M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*, 191-196. doi:10.1111/1467-7687.00271

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders, 41*(2), 242-247.

Sawyer, A. C. P., Williamson, P., & Young, R. L. (2012). Can gaze avoidance explain why individuals with Asperger's Syndrome can't recognise emotions from facial expressions? *Journal of Autism and Developmental Disorders, 42*, 606-618. doi:10.1007/s10803-011-1283-0

Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A social adjustment enhancement intervention for High Functioning Autism, Asperger's Syndrome, and Pervasive Developmental Disorder NOS. *Journal of Autism and Developmental Disorders*, *34*(6), 649-668. doi:10.1007/s10803-004-5286-y

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lip reading in autism. *Journal of Child Psychology and Psychiatry, 48*(8), 813-821. doi:10.1111/j.1469-7610.2007.01766.x

Spezio, M.L., Adolphs, R., Hurley, R.S., & Piven, J. (2007) Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders, 37*(5), 929-939.

Taylor, G. J., & Bagby, R. M. (2004). New trends in alexithymia research. *Psychotherapy and Psychosomatics*, *73*(2), 68-77. doi:10.1159/000075537

Taylor, N., Isaac, C., & Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of Autism and Developmental Disorders, 40*(11), 1403-1411. doi:10.1007/s10803-010-1000-4

Tracy, J. L., Robins, R. W., Schriber, R. A., & Solomon, M. (2011). Is emotion recognition impaired in individuals with autism spectrum disorders? *Journal of Autism and Developmental Disorders, 41*, 102-109. doi:10.1007/s10803-010-1030-y

Turner-Brown, L., Perry, T., Dichter, G., Bodfish, J., & Penn, D. (2008). Brief report: Feasibility of Social Cognition and Interaction Training for adults with high functioning autism. *Journal of Autism & Developmental Disorders*, *38*(9), 1777-1784. doi:10.1007/s10803-008-0545-y.

Vuilleumier, P. & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia, 45*, 174–194. doi:10.1016/j.neuropsychologia.2006.06.003

Wilson, C. E., Brock, J., & Palermo, R. (2010). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. *Journal of Intellectual Disability Research, 54*(12), 1104-1115.

Young-Browne, G., Rosenfeld, H. M., & Horowitz, F. D. (1977). Infant discrimination of facial expressions*.* *Child Development, 48*, 555–562.

Table 1

Descriptive Demographic Statistics by Group

Variable	All Participants (n = 20)	ASD Group (n = 10)	TYP Group (n = 10)	Independent t-statistic	χ^2 statistic
Mean (SD) age in months	99.30 (26.02)	106.80 (21.06)	91.8 (29.34)	1.31	
Age range in months	54-145	78-145	54-139	-	
Mean (SD) KBIT-2 IQ Composite	105.85 (21.54)	95.90 (21.59)	115.80 (17.13)	-2.28*	
IQ Composite range	60-140	60-119	91-140	-	
Mean (SD) MA in months	104.72 (34.30)	102.32 (29.08)	107.12 (40.32)	-0.306	
MA range in months	50.40-172.80	50.40-135.66	55.62-172.80	-	
Mean (SD) SRS Total *T*-score	62.10 (19.99)	80.40 (7.26)	43.80 (6.81)	11.63**	
SRS Total *T*-score range	36-90	69-90	36-58	-	
Gender (% male)	90%	90%	90%		0
Ethnicity (% Caucasian)	95%	100%	90%		1.05
Mother's education (count)					4.67^{+}
Some college	6 (30%)	5 (50%)	1 (10%)		-
College degree	6 (30%)	3 (30%)	3 (30%)		-
Graduate degree	8 (40%)	2 (20%)	6 (60%)		-
Family income (count)					10.67^{+}
Under $20,000	2 (10%)	1 (10%)	1 (10%)		-
$20,000-39,999	3 (15%)	3 (30%)	0		-
$40,000-59,999	4 (20%)	2 (20%)	2 (20%)		-
$60,000-79,999	3 (15%)	3 (30%)	0		-
$80,000-99,999	2 (10%)	0	2 (20%)		-
$100,000 and above	6 (30%)	1 (10%)	5 (50%)		-

Note. ** $p < 0.01$(2-tailed). * $p < 0.05$ (2-tailed). $^{+}$ $p < 0.1$ (2-tailed). ASD = autism spectrum disorder; TYP = typical development; KBIT-2 = Kaufman Brief Intelligence Test, Second Edition; MA = mental age; SRS = Social Responsiveness Scale.

Table 2

Correlations Between Emotion Identification Accuracy and MA/Age/SRS Total T-score

Correlation	MA	Age	SRS Total *T*-score
ASD Group (*n* = 10)			
Negative emotion accuracy	.76*	.56	.41
Positive emotion accuracy	.62	.20	.17
Voiced emotion accuracy	.72*	.22	.28
Silent emotion accuracy	.76*	.65*	.39
TYP Group (*n* = 10)			
Negative emotion accuracy	.14	.19	.38
Positive emotion accuracy	.34	.60	.59
Voiced emotion accuracy	.37	.59	.67*
Silent emotion accuracy	.12	.22	.31
All Participants (*n* = 20)			
Negative emotion accuracy	.45*	.33	.08
Positive emotion accuracy	.45*	.28	-.20
Voiced emotion accuracy	.50*	.24	-.21
Silent emotion accuracy	.41	.39	.12

Note. * Correlation is significant at the 0.05 level (2-tailed). MA = mental age; SRS = Social Responsiveness Scale; ASD = autism spectrum disorder; TYP = typical development.

Table 3

Ratio of Voiced Accuracy to Combination of Voiced and Silent Accuracy, Tested as Significantly Different from 0.5

	Ratio	One sample *t*-statistic; test value = 0.5	*df*
ASD Group (*n* = 10)	0.57 (0.11)	2.15	9
TYP Group (*n* = 10)	0.61 (0.09)	3.50*	9
All Participants (*n* = 20)	0.59 (0.10)	3.99**	19

Note. ** $p < 0.01$(2-tailed). * $p < 0.05$ (2-tailed). Ratio = number of voiced videos identified correctly / (number of voiced videos identified correctly + number of silent videos identified correctly). Standard Deviations appear in parentheses below means. ASD = autism spectrum disorder; TYP = typical development.

Table 4

Mean Proportion of Videos Identified Correctly by Group, and T-values Between Groups

	Group		Independent	*df*
	ASD	TYP	*t*-statistic	
All emotions combined	51% (20%)	58% (13%)	-0.88	18
Negative emotions	57% (22%)	58% (12%)	-0.21	18
Positive emotions	43% (20%)	56% (20%)	-1.52	18
Voiced emotions	56% (20%)	69% (15%)	-1.69	18
Silent emotions	46% (22%)	46% (17%)	0	18

Note. No significant *t*-statistics. Standard Deviations appear in parentheses below means. ASD = autism spectrum disorder; TYP = typical development.

Table 5

Group x (Condition x AOI) Mixed Factorial Analysis of Variance for Fixation Duration of Accurate Trials

Source	*MS*	*df*	*F*	*p*
(A) Group	149.88	1	18.50**	<.01
(B) Condition	20.93	1	6.62*	.02
(C) AOI	3.36	1	0.18	.68
A x B (two-way interaction)	6.36	1	2.01	.18
A x C (two-way interaction)	0.00	1	0	.99
B x C (two-way interaction)	6.67	1	5.46*	.04
A x B x C (three-way interaction)	0.05	1	0.04	.84
Error	8.10	13		

Note. ** $p < 0.01$(2-tailed). * $p < 0.05$ (2-tailed).

Table 6

Pairwise Comparisons of Two-Way Interaction Between Condition and AOI for Fixation Duration and Fixation Count

	Mean (SD)	Paired samples *t*-statistic	*df*	*p*
Fixation duration				
Eye region, voiced condition	4.29 (3.30)	2.97*	14	.01
Eye region, silent condition	2.47 (3.21)			
Mouth region, voiced condition	4.09 (3.30)	0.98	14	.35
Mouth region, silent condition	3.62 (2.90)			
Fixation count				
Eye region, voiced condition	13.30 (10.31)	4.62**	19	<.01
Eye region, silent condition	6.65 (8.59)			
Mouth region, voiced condition	9.45 (7.93)	1.94	19	.07
Mouth region, silent condition	6.95 (5.91)			

Note. ** $p < 0.01$(2-tailed). * $p < 0.05$ (2-tailed).

Table 7

Estimated Marginal Means (SD) Comprising Group x (Condition x AOI) ANOVA for Fixation Duration Data, Constrained to Accurate Trials

	Group	
	ASD	TYP
Voiced emotions		
Eye region	2.52 (2.53)	6.30 (3.03)
Mouth	2.29 (1.90)	6.15 (3.45)
Silent emotions		
Eye region	1.27 (1.08)	3.85 (4.30)
Mouth	2.48 (2.44)	4.92 (3.00)

Note. Standard Deviations appear in parentheses below means. ASD = autism spectrum disorder; TYP = typical development.

Table 8

Group x (Condition x AOI) Mixed Factorial Analysis of Variance for Fixation Count of Accurate Trials

Source	*MS*	*df*	*F*	*p*
(A) Group	567.11	1	5.15*	.04
(B) Condition	418.61	1	17.58**	<.01
(C) AOI	63.01	1	0.58	.46
A x B (two-way interaction)	66.61	1	2.80	.11
A x C (two-way interaction)	59.51	1	0.55	.47
B x C (two-way interaction)	86.11	1	7.48*	.01
A x B x C (three-way interaction)	5.51	1	0.48	.50
Error	110.22	18		

Note. ** $p < 0.01$(2-tailed). * $p < 0.05$ (2-tailed).

Table 9

Estimated Marginal Means (SD) Comprising Group x (Condition x AOI) ANOVA for Fixation Count Data, Constrained to Accurate Trials

	Group	
	ASD	TYP
Voiced emotions		
Eye region	8.60 (7.17)	18.00 (11.14)
Mouth	7.00 (5.31)	11.90 (9.55)
Silent emotions		
Eye region	4.30 (3.59)	9.00 (11.42)
Mouth	5.80 (6.14)	8.10 (5.74)

Note. Standard Deviations appear in parentheses below means. ASD = autism spectrum disorder; TYP = typical development.

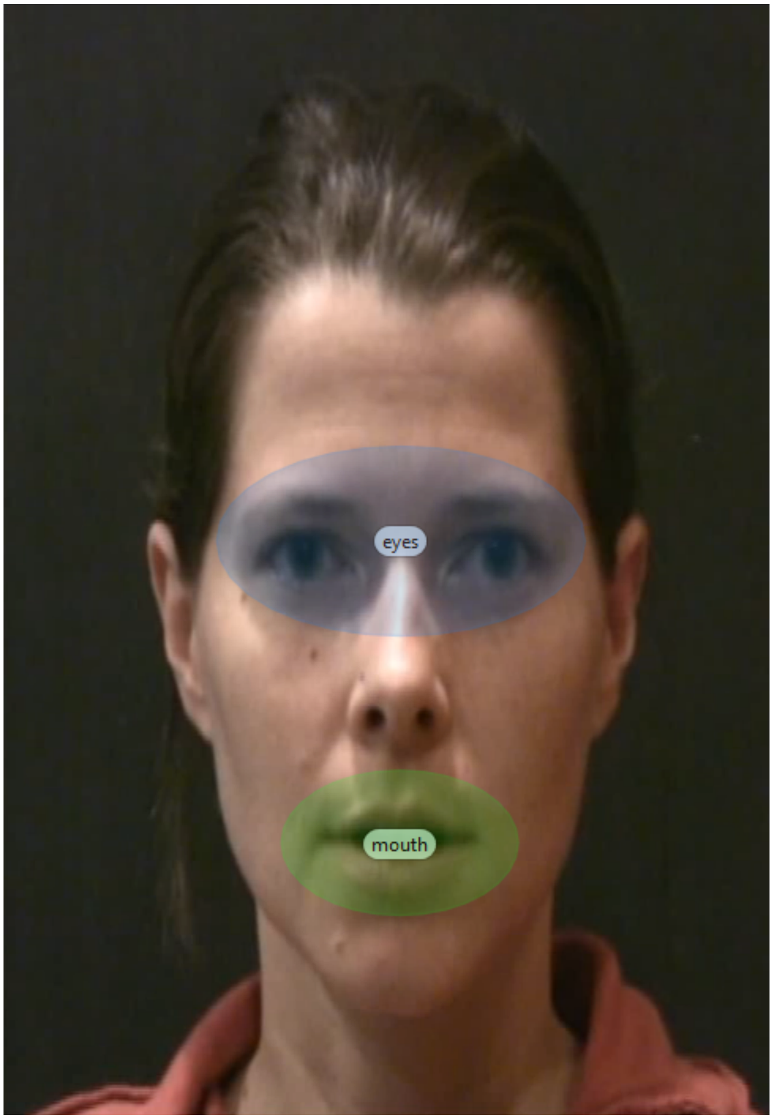

Figure 1. Example frame from video clip with eye and mouth regions identified.

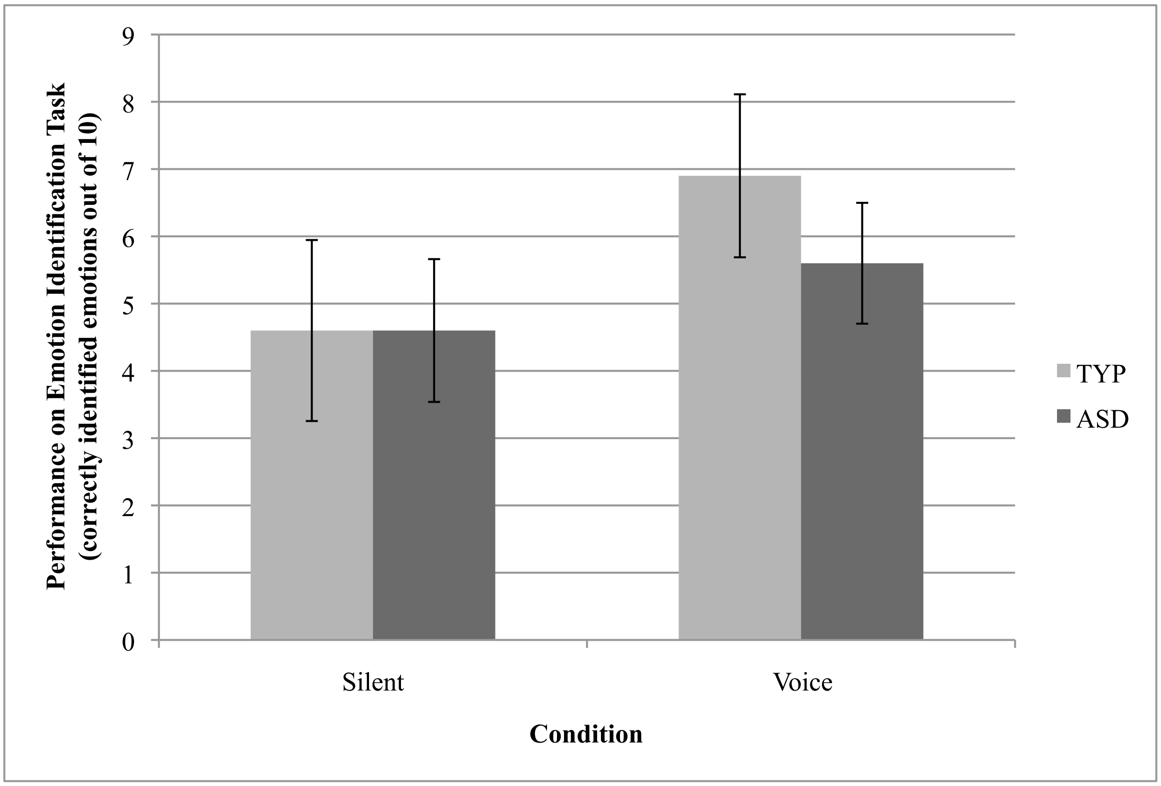

Figure 2. Emotion identification accuracy, for both silent and voice conditions, by group.

Note. Error bars represent 95% confidence interval (standard error) of the mean. TYP = typical development; ASD = autism spectrum disorder.

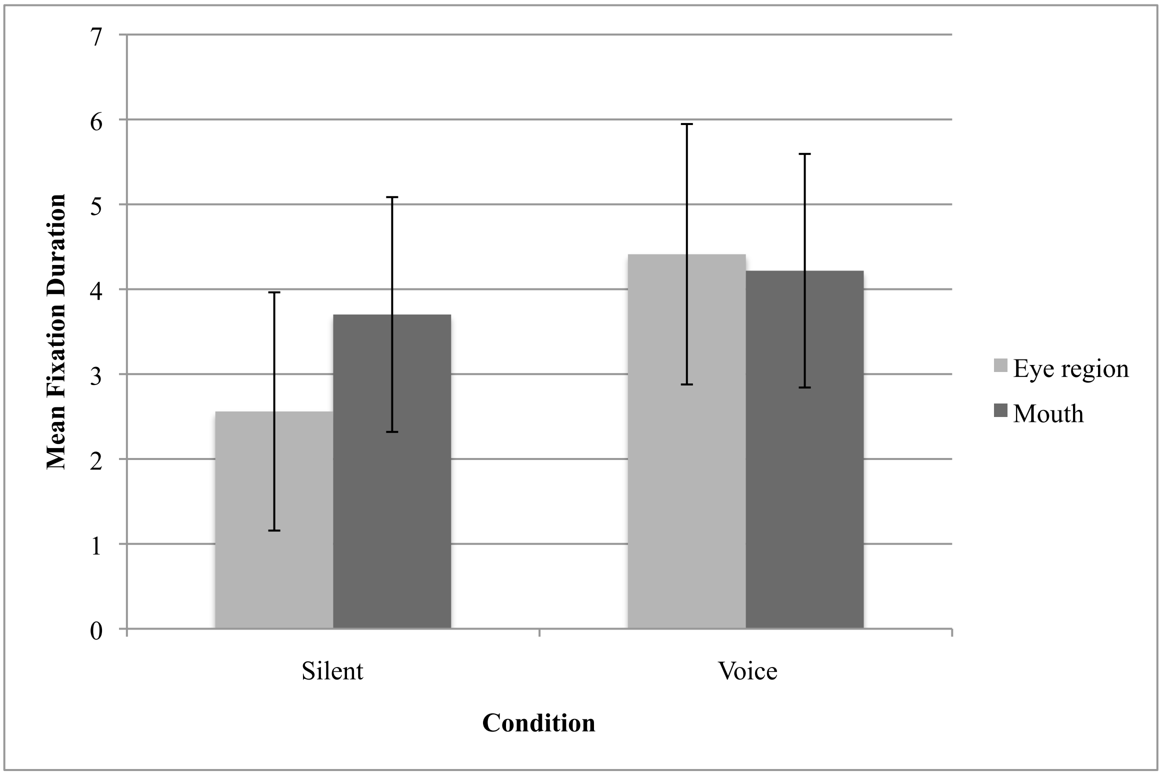

Figure 3. Two-way interaction effect (condition x AOI) of fixation duration data for accurately identified trials.

Note. Error bars represent 95% confidence interval (standard error) of the mean.

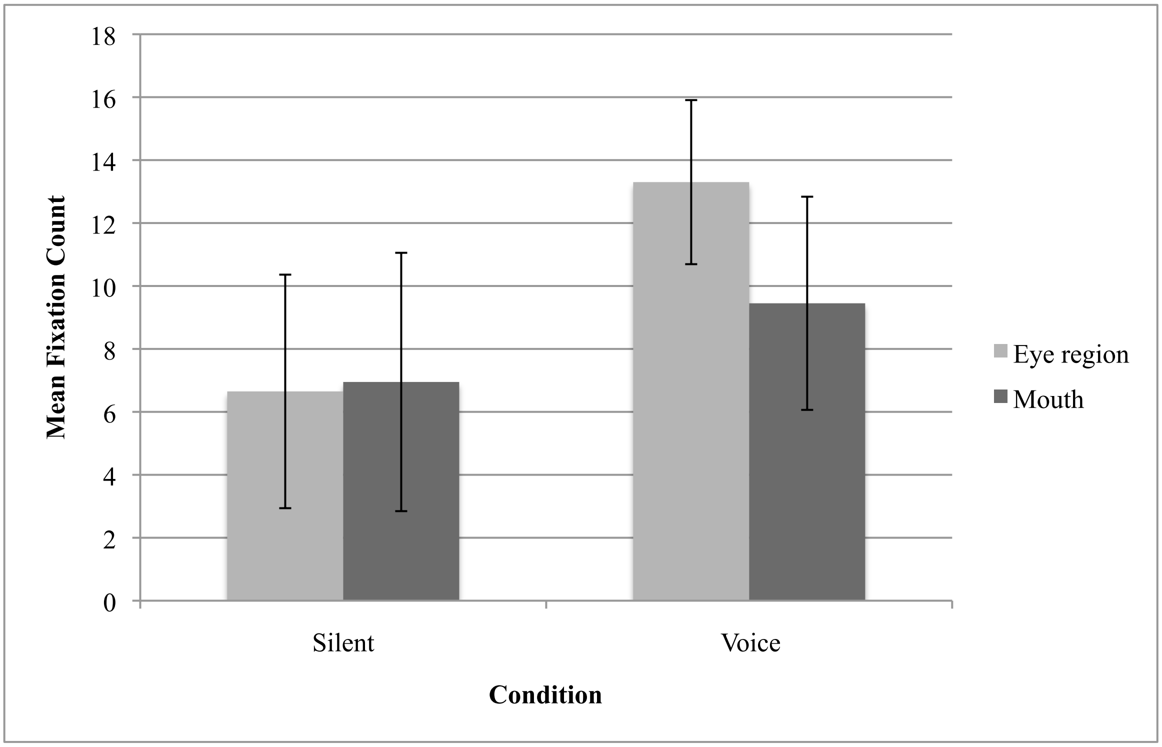

Figure 4. Two-way interaction effect (condition x AOI) of fixation count data for accurately identified trials.

Note. Error bars represent 95% confidence interval (standard error) of the mean.

Appendix A

Demographic Survey

Subject #: ________________

GENERAL INFORMATION ABOUT THE PARENT

What relation are you to the child?
____Mother
____Father
____Stepmother
____Stepfather
____Adoptive Mother
____Adoptive Father
____Legal Guardian
____Other Relative (please specify): ________________________

What is your race/ethnicity? (optional)
____African American
____Asian
____Caucasian/European American
____Native American
____Latino, Hispanic, or Chicano
____Other (please specify): ________________________

What is your highest level of completed education? (optional)
____Some High School
____High School Graduate
____Some College
____College Degree
____Some Graduate Studies
____Graduate Degree

Which of the following is closest to your annual household income? (optional)
____Under $20,000
____$20,000 - $39,999
____$40,000 - $59,999
____$60,000 - $79,999
____$80,000 - $99,999
____$100,000 and above

GENERAL INFORMATION ABOUT THE CHILD

How old is your child?
____Years ____Months

What is your child's gender?
____Male

____Female

What is your child’s race/ethnicity? (optional)
____African American
____Asian
____Caucasian/European American
____Native American
____Latino, Hispanic, or Chicano
____Other (please specify): __________________________

What level of education has your child completed?
____None
____Preschool
____Kindergarten
____Elementary School

Specify highest grade completed: __________________________

Does your child take any medications?
____Yes ____No

If yes, please specify what mediations: ____

____

GENERAL INFORMATION ON DIAGNOSIS AND PROFESSIONAL SERVICES

What is your child's current Autism Spectrum Disorder diagnosis?
____Autism or Autistic Disorder
____Asperger's Disorder
____Childhood Disintegrative Disorder
____Rett's Syndrome
____Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS)
____Other (please specify):__________________________

How old was your child when he/she got this diagnosis?
____Years ____ Months

Have any of the child's siblings been diagnosed with Autism Spectrum Disorder?
____Yes
____No

If Yes, please specify which disorder: __________________________

Does your child have any other symptoms or diagnoses? (check all that apply)
____Anxiety Disorder
____Attention Deficit Hyperactive Disorder

____Obsessive Compulsive Disorder
____Central Auditory Processing Disorder
____Depression
____Schizophrenia
____Hearing Impairment
____Vision Impairment
____'Tunnel Vision Syndrome' (peripheral vision, vision perception impairment)
____Mental Retardation
____Seizures
____Dietary Allergies
____Digestive Problems (constipation, diarrhea, bloating, or abdominal pain)
____None
____Other (please specify): _________________________

If answered yes to dietary allergies, please specify: _________________________

What type of professional diagnosed your child with Autism Spectrum Disorder?
____Developmental Pediatrician
____Psychologist
____Neurologist
____Primary Care Physician
____Psychiatrist
____Other (please specify): _________________________

Appendix B

Parental Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
Parental Permission and Consent

Title: Social Attention in Children with Autism
Principle Investigators: Dr. Angela Scarpa
Co-investigators: Dr. Robin Panneton, Katrina Ostmeyer, Jill Lorenzi, Allison Heck

I. Purpose of this Research/Project
We are studying how children with autism spectrum disorders (ASD) and typically developing children view information about people. We are looking at how this relates to their ability to identify emotions. We are also looking at how cues from people might affect learning new information.

II. Procedures
You will fill out a few forms while your child takes a short cognitive test. This will take about 15-30 minutes. Your child will then be given a short break. After the break, he/she will sit in front of a monitor where he/she will do tasks while we track his/her eyes. In the first task, your child will view short movie clips of adult females. Each movie clip will have a woman expressing one of the following emotions: happy, sad, anger, fear, surprise. All clips will show the video. Only some will have sound. Your child will be asked to name the emotion after each clip. In the second task, your child will watch short video clips of objects that are paired with words. The child's attention will be drawn to the objects by someone talking about it, looking at the object with their eyes, facing towards the object, or pointing to the object. After the eye tracking task, your child will be invited to play with some novel, exciting toys. After playing with a toy for approximately 6-10 seconds, the experimenter will abruptly ask to have the toy back. Specifically, the experimenter will say, "Oh, you weren't suppose to have that. Give it back. Give it back," while holding out his/her hand. We will code your child's behavior to this request in an attempt to examine how social processing affects affect regulation. Afterwards, we will apologize to your child and allow access to the toy telling him/her we made a mistake and he/she may have the toy after all.

III. Risks
All recorded information and data are kept confidential and kept in a secure location. All forms will be kept in locked filing cabinets in locked offices. All digital files are password protected. During the emotion naming task, some of the movie clips may be slightly aversive (e.g., anger, fear) to the children. Your child may stop participating at any time with no explanation needed and no penalty. Your child could become upset when asked to return the toy at the end of the study. To try to counteract this, we will apologize to your child, help them calm down if they become upset, and allow access to the toy after we admit our "mistake."

IV. Benefits

There are no benefits to you or your child. It is possible for you to watch the replay of your child's eye tracking after the session. While this is not a formal summary, some parents find this interesting. This also gives a chance for the parent to ask the experimenter questions about the development of children's attention to social events. Understanding how children with and without ASD understand emotion and language has potential benefits. It adds to theories on how social attention relates to developmental outcomes. It can also help guide future research and treatment for ASD.

V. Extent of Anonymity and Confidentiality

All of the data gathered in this study will be kept confidential. Data will not be identified with you in any way. You and your child's name will not be linked with this study. The results of this project may be used for scientific and/or educational purposes, discussed at scientific meetings, and/or published in scientific journals. You will be sent a summary of the general findings when this project is finished.

VI. Compensation

There is no compensation for being in this project.

VII. Freedom to Withdraw

You and your child have the right to withdraw from this project at any point and for any reason.

VIII. Participant's Responsibilities

I freely agree to allow my child to participate in this study, as well as my own participation in completing forms.

IX. Participant's Permission

I have been given an opportunity to ask more questions about this procedure. I understand that I have the right to end this session for any reason. This project has been approved by the Human Subjects Committee of the Department of Psychology and the Institutional Review Board of Virginia Tech. If I have any questions about this research, I should contact one of the persons named below. Given these procedures and conditions, I give my permission to Dr. Scarpa, Dr. Panneton, and their co-workers to test my child.

Dr. Angela Scarpa, Principle Investigator 231-2615
Dr. Robin Panneton, Co-Investigator 231-5938
Katrina Ostmeyer, MA Co-Investigator 231-6914
Jill Lorenzi, BA Co-Investigator 231-6914
Alison Heck, BA Co-Investigator 231-3972
Dr. David Harrison, Chair, Human Subjects Committee 231-4422
David M. Moore, DVM, Assistant Vice Provost for Research Compliance 231-4991

Signature of Parent: ____
Date: ____
Child's Name: ____

I would like to be contacted by phone regarding future studies: YES NO

Appendix C

Child Assent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
Informed Consent for Participants in Research Projects Involving Human Subject

Title: Social Attention in Children with Autism
Principle Investigator: Dr. Angela Scarpa
Co-investigators: Dr. Robin Panneton**,** Katrina Ostmeyer, Jill Lorenzi

Child Assent (read to child)

We are going to ask you to do some things today. First you and I are going to do an activity where I will ask you some questions about words, play with blocks, and look at pictures that go together. When we are done with that, I'll let you take a break. Next I will have you watch some videos on a television screen and I'll ask you questions about some of the movies. If you get tired, you can always take an extra break or you can stop at any time. Would you like to play with me today? Remember, you can stop at any time, you don't have to go on with what we are doing -- just let me know if you want to stop or quit.

Signature of Child: ____

Date: ____

Child's Name: ____

Appendix D

IRB Approval Letter

VirginiaTech

Office of Research Compliance
Institutional Review Board
2000 Kraft Drive, Suite 2000 (0497)
Blacksburg, Virginia 24060
540/231-4606 Fax 540/231-0959
e-mail irb@vt.edu
Website: www.irb.vt.edu

MEMORANDUM

DATE: November 18, 2011

TO: Angela Scarpa-Friedman, Katrina Ostmeyer-Kountzman, Robin Panneton, Jill Lorenzi, Alison Heck

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)

PROTOCOL TITLE: Social Attention in Children with Autism

IRB NUMBER: **10-566**

Effective November 18, 2011, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at http://www.irb.vt.edu/pages/responsibilities.htm (please review before the commencement of your research).

PROTOCOL INFORMATION:
Approved as: **Full Board Review**
Protocol Approval Date: **8/8/2011 (protocol's initial approval date: 9/13/2010)**
Protocol Expiration Date: **8/7/2012**
Continuing Review Due Date*: **6/25/2012**
*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:
Per federally regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

An equal opportunity, affirmative action institution

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

cc: File
Department Reviewer:David W. Harrison