Production of emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism

Ruth B. Grossman[1,2], **Lisa R. Edelson**[3], and **Helen Tager-Flusberg**[3]

[1]Emerson College, Department of Communication Sciences and Disorders, USA

[2]University of Massachusetts Medical School Shriver Center, USA

[3]Boston University, Department of Psychology, USA

Abstract

Background—People with high-functioning autism (HFA) have qualitative differences in facial expression and prosody production, which are rarely systematically quantified.

Purpose—Perform qualitative and quantitative analyses of prosody and facial expression productions in children and adolescents with HFA.

Method—Participants were 22 male children and adolescents with HFA and 18 typically developing (TD) controls. We used a story retelling task to elicit emotionally laden narratives, which were analyzed using acoustic measures and perceptual codes. Naïve listeners coded all productions for emotion type, degree of expressiveness, and awkwardness.

Results—The group with HFA was not significantly different in accuracy or expressiveness of facial productions, but was significantly more awkward than the TD group. Participants with HFA were significantly more expressive in their vocal productions, with a trend for greater awkwardness. Severity of social communication impairment, as captured by the ADOS, was correlated with greater vocal and facial awkwardness.

Conclusion—Facial and vocal expressions of participants with HFA were as recognizable as those of their TD peers, but qualitatively different, particularly when coding samples with intact dynamic properties. These preliminary data show qualitative differences in nonverbal communication that may have significant negative impact on the social communication success of adolescents with HFA.

Keywords

Autism; Facial Expressions; Prosody; Production; Social communication

---

Since the earliest descriptions by Kanner (1968) and Asperger (1944), one of the defining characteristics of autism spectrum disorder (ASD) is the production of unusual facial expressions and prosody. Prosody production differences have been described for measures of rhythm, speech rate, and pitch contours (McCann & Peppé, 2003; Paul, Augustyn, Klin, & Volkmar, 2005; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007; Shriberg et al., 2001; Van Santen, Prud'hommeaux, Black, & Mitchell, 2010). Young children with ASD also demonstrate differences in displays of spontaneous facial expressions during social

---

Corresponding author: Ruth B. Grossman, Emerson College, Department of Communication Sciences and Disorders, 120 Boylston Street, Boston, MA 02116, USA. ruth_grossman@emerson.edu.

interactions (Capps, Kasari, Yirmiya, & Sigman, 1993; Dawson, Hill, Spencer, Galpert, & Watson, 1991; Snow, Hertzig, & Shapiro, 1987). These differences in facial expression and prosody production can have a significant impact on how accepted individuals with ASD are and how successful they can be as social communicators (McCann & Peppé, 2003; Paul, Shriberg, et al., 2005; Shriberg et al., 2001).

Prosody

With few exceptions, studies of prosody production have focused on prosodic categories, such as lexical stress, or grammatical marking of questions vs. statements. The data indicate decreased accuracy for a range of prosodic functions (Peppé & McCann, 2003), qualitatively longer utterance durations (Diehl & Paul, 2012; Diehl & Paul, in press; Grossman, Bemis, Plesa Skwerer, & Tager-Flusberg, 2010), and greater variability in pitch ranges (Baltaxe, 1984; Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008; Fosnot & Jun, 1999). Hubbard and Trauner (2007), studying emotional prosody in sentence repetition and story completion tasks documented greater pitch ranges and lower accuracy of expressing the targeted emotion in individuals with ASD, indicating atypical or anomalous usage of prosodic contours in this population. They suggest that more work must be done to examine not just the existence of specific acoustic features, but also the qualitative usage and perception of those features.

Facial expressions

Very young children with ASD appear to show more frequent displays of negative affect and less evidence of positive facial affect compared to chronological or developmental age-matched peers (Capps et al., 1993; Snow et al., 1987). Yirmiya and colleagues (Yirmiya, Kasari, Sigman, & Mundy, 1989) showed that young children with ASD produce unique and unusual facial expressions, including blends of incompatible emotions that are not seen in TD children or children with Down syndrome. These data highlight the importance of collecting qualitative coding data on facial expressions productions to understand relevant group differences in social communication.

Faces and voices

There is only one published study of facial and vocal productions captured during the same task. Macdonald and colleagues (1989) elicited productions using pictures of emotional facial expressions and verbal situational descriptions from adults with HFA. Participants with HFA were significantly less accurate and more "odd" in both facial and vocal expressions of emotion than their TD peers. This study nevertheless has several limitations: facial expression productions were recorded via photographs, which does not allow for the analysis of dynamic changes over the course of a naturally progressing facial expression, and no acoustic measures were provided in addition to perceptual codes for prosody productions.

Purpose of the presented study

The aim of our current study is to analyze pilot data for acoustic and freeze-frame measures of prosody and facial productions in combination with perceptual coding of natural, fully dynamic facial and vocal expressions of emotion. Our hypotheses are that children and adolescents with HFA will show greater variability in acoustic measures and be rated as more awkward and less accurate than their TD peers in facial and vocal production.

Methods

Participants

Two groups participated in this study: children and adolescents with HFA (n=22, age range 8:2–19:9) and TD controls (n=18, age range 9:6–19:6). All participants were recruited through local schools, advertisements placed in local magazines, newspapers, the internet, advocacy groups for families of children with HFA, and word of mouth. Only participants who spoke English as their native and primary language were included. Descriptive characteristics for all participants can be found in Table 1.

Standardized testing—We used the Kaufman Brief Intelligence Test, Second Edition (K-BIT 2; Kaufman & Kaufman, 2004) to assess IQ, the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) to determine receptive vocabulary ability, and the Woodcock-Johnson III Diagnostic Reading Battery (WJ III DRB, Woodcock, Mather, & Schrank, 2004). All participants in both groups had standardized scores on these measures within the normal range. Despite early histories of language delay in some participants with HFA, none showed current deficits in receptive vocabulary, verbal IQ, or reading levels, consistent with a classification of high-functioning autism. Due to time constraints, we were unable to administer the PPVT to one individual with HFA and two TD participants. Since all three participants' verbal IQ scores were within normal limits, we included them in the final analysis.

Due to technical difficulties, not all analyses were performed for all participants. Recordings for the first participants were created with a small wireless camera and a separate digital voice recorder. The resulting videos were not of sufficiently high-quality to be used for video coding, but appropriate for acoustic analyses using PRAAT software (Boersma & Weenink, 2009). We then switched to recording both audio and video with a digital video camera, which produced good video quality, but some of the resulting audio files could not be read by PRAAT (2009). Acoustic analyses were therefore conducted on the prosodic productions of 18 participants with HFA and 11 TD controls, while perceptual coding of facial and vocal expressions was conducted for 14 participants with HFA and 12 TD controls. Five TD participants and ten participants with HFA received both types of analyses.

Using a multivariate ANOVA with group as the independent variable we verified that the groups of participants who received acoustic analysis (HFA, n=18 males and TD, n=11 males) did not differ significantly in age, ($F_{(1,28)}$= .882, p= .36), verbal IQ ($F_{(1,28)}$=1.68, p=.21), nonverbal IQ ($F_{(1,46)}$= .44, p=.51), full IQ ($F_{(1,28)}$= 1.62, p=.21), receptive vocabulary ($F_{(1,27)}$= .08, p=.78), or reading skills ($F_{(1,28)}$= .04, p=.84). Groups of participants whose videos were coded (HFA, n=14 males and TD, n=11 males, 1 female) also did not differ significantly in age, ($F_{(1,25)}$= .0, p= .99), verbal IQ ($F_{(1,25)}$=.38, p=.55), nonverbal IQ ($F_{(1,25)}$= 2.66, p=.12), full IQ ($F_{(1,25)}$=1.78, p=.2), receptive vocabulary ($F_{(1,22)}$= .003, p=.96), or reading skills ($F_{(1,25}$ = .02, p=.89). A chi-squared analysis showed that the groups did not differ in the distribution of gender ($\chi^2(1, N$= 26)=.112, p=.203).

Diagnosis of HFA—Trained examiners under supervision of a clinical psychologist administered the Autism Diagnostic Observation Schedule, Module 3(ADOS, Lord, Rutter, DiLavore, & Risi, 1999) to each participant, while the accompanying caregiver completed the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994). Participants in the HFA group met diagnostic criteria for autistic disorder, based on results of the ADOS, ADI-R and expert clinical impression. Participants with known genetic

disorders were excluded. Seventeen participants met ADOS criteria for autism and five met ADOS criteria for HFA.

Materials

Stimulus creation—We created four brief stories (25–32 seconds) about a protagonist, nicknamed "Safari Bob," and his adventures taking pictures of animals in the wild. Each story contained at least one phrase or sentence produced with happy, fearful, angry, and positive surprise emotion (Appendix A). We chose these four emotions because they represent an equal distribution of positive and negative affect within the set of universal emotions described by Ekman (Ekman, 1984; Ekman & Friesen, 1978). The order of emotions was counterbalanced across stories. A young adult male actor portraying Safari Bob told all stories for the video camera multiple times (Figure 1). We used the visualization scenarios inherent in the text of the stories to coach the actor on clearly differentiating each emotion within a story. After recording was completed the most fluent and naturally expressive iteration of each story was chosen by consensus of three members of the study staff.

The actor also recorded an introduction to the task, which provided participants with the opportunity to become familiar with his manner of speaking and baseline facial expressions.

Procedure

All aspects of this study were approved by the Boston University Medical School Institutional Review Board. Participants were seated comfortably in a quiet room facing a TV and digital video camera. We told participants that Bob would give them instructions on the video and they should watch and listen carefully. In the video introduction, Bob introduced himself and explained that he was the host of a television show for children where he told stories about his adventures with animals in the wild. He further stated that he was looking for somebody who could fill in for him while he took a vacation. The participant's job therefore was to listen to some of his stories and retell them"[…] in such a way that the children who'll be watching can get excited about it, just the way I told it." We chose this phrasing because it allowed us to encourage participants to be animated in their retelling, without explicitly telling them to mimic the actor's emotional portrayal of the stories. The intention was to avoid overt instructions to "be animate" for fear that it would make participants self-conscious and result in stilted and overly exaggerated productions.

After Bob's introduction, a member of the research staff answered any requests for clarification by the participant. We then played one of the stories, paused the video and asked the participant to retell the story facing the digital camera, which was placed adjacent to the TV. We provided participants with written scripts, printed in large type on a laminated card and suspended directly below the lens of the camera. This reduced cognitive load so participants could focus on telling the story in an engaging way, rather than on memory retrieval. All participants were able to complete the task, verifying that they understood the directions.

Data preparation

Since each story contained at least one sentence of angry, happy, surprised, and fearful emotion, the video and audio recordings were cut into files containing only one emotion and its corresponding complete sentence or phrase. We chose edit points based on sentence and phrase boundaries between segments with different emotions. Audio files were extracted from video recordings or captured on dedicated digital audio recorders. We then checked all audio clips using PRAAT software (Boersma & Weenink, 2009) for artifacts that might

interfere with accurate analysis of pitch or intensity (e.g., background noise, coughs, or artifacts from glottal stops) and removed them.

Coding

Clips were coded for expressed emotion, intensity of expression, and naturalness/ awkwardness of expression. All coders were blind to diagnosis and target emotion and were trained by the first author. Clips were presented to coders in randomized order to avoid bias due to position of clip within a story. A randomly selected sample of all codes was co-coded by at least one other person. Inter-coder reliability was computed using Cohen's Kappa and maintained at .85 or higher. Any discrepancies were discussed and resolved through conversation between the coders and the first author. Coders of voice files could not see the participants' faces, and coders of face videos completed their task with the sound off to eliminate cross-modal information.

Expressiveness and Naturalness codes used a 4-point Likert-type scale. Expressiveness was coded on a "flat," "mild," "moderate," "extreme" continuum and naturalness was coded on a "natural," slightly awkward," "very awkward," "unnatural" range. We used a four point scale for several reasons: 1. Using an even number of points eliminated the possibility of coders choosing the "safe" mid-point for a majority of codes, forcing coders to make thoughtful decisions. 2. The targeted distinctions were subtle to begin with and using more Likert points would have required coders to make decisions based on even more minute variations in expressiveness or naturalness. 3. Having more Likert points made inter-coder reliability harder to achieve. Future studies may want to explore more extensive scales, or even non-segmented response "meters" that allow coders to mark arbitrary points. Such scales could provide valuable fine-grained details of perceptual coding, but may only be usable for a single coder, rather than a group of coders who need to be reliable with each other. Coders in our study had no difficulty differentiating between different levels of any code and easily maintained inter-coder reliability. All codes are summarized in Table 2. More information about the coding scheme can be found in Appendix B.

Results

All data were normally distributed and can be found in Tables 3 and 4.

Acoustic analysis of voices

We conducted a multivariate ANOVA with group as the independent variable, which revealed a significant group difference for maximum F_0 ($F(1, 28) = 4.17$, $p = .05$, $\eta^2 = .134$), with the HFA group extending the upper end of the pitch range higher than the TD group (HFA mean = 274.83Hz, TD mean = 206.64Hz). There was a trend for a significant between-group difference in minimum intensity ($F(1, 28) = 3.97$, $p = .056$, $\eta^2 = .128$), with the HFA group extending the lower end of the intensity range further than their TD peers (HFA mean = 51.4dB, TD mean = 54.8dB).

There were also significant group differences in intensity range ($F(1, 28) = 4.1$, $p = .05$, $\eta^2 = .133$) and pitch range ($F(1, 28) = 4.12$, $p = .05$, $\eta^2 = .065$). In both cases the HFA group produced larger ranges than their TD peers (dynamic range: HFA mean = 27.22dB, TD mean = 23.8dB and pitch range: HFA mean = 170.0Hz, TD mean = 108.6Hz). There were no significant differences for mean pitch or intensity.

Freeze-frame coding of faces

Since the distances between anchor points on our Likert-type scales were equal we were able to average the coding scores for all video clips of a given participant to establish a

single average expressiveness score and a single average naturalness score for each individual. We then conducted a multivariate ANOVA with group as the independent variable for the resulting continuous variables (Carifio & Perla, 2007), which revealed no significant group differences in naturalness ($F(1,25) = 1.6$, $p = .21$) or expressiveness ($F(1,25) = .3$, $p = .58$).

Video-coding of faces

A multivariate ANOVA with group as the independent variable showed that the group of participants with HFA was rated as significantly more awkward for facial expressions than the TD group ($F(1, 25) = 5.25$, $p = .031$, $\eta^2 = .179$). The analysis revealed no significant group differences in accuracy ($F(1, 25) = .99$, $p = .33$) or expressiveness ($F(1,25) = 2.8$, $p = .1$).

Video-coding of voices

A multivariate ANOVA with group as the independent variable showed that the group of participants with HFA was significant more accurate ($F(1, 25) = 6.03$, $p = .022$, $\eta^2 = .201$) and more expressive ($F(1, 25) = 6.45$, $p = .018$, $\eta^2 = .213$) than the TD group. There was also a trend for the group with HFA to be more awkward ($F(1, 25) = 3.81$, $p = .063$, $\eta^2 = .136$).

Cross-modal analyses

We compared video-coding analysis of expressiveness across faces and voices and conducted a 2(group) by 2(modality, [face, voice]) repeated measures ANOVA for expressiveness, which revealed a main effect for modality ($F(1, 24) = 41.6$, $p < .001$, partial $\eta^2 = .63$), but no modality by group interaction ($F(1, 24) = .71$, $p = .41$, partial $\eta^2 = .03$). Both groups were significantly more expressive in voices (HFA mean = 1.99, TD mean = 1.70) than faces (HFA mean = 1.57, TD mean = 1.39).

We also investigated whether acoustic and perceptual coding differences in the group of participants with HFA were correlated with ADOS social communication score. There were no significant correlations to any of the acoustic measures of prosody, but analysis did reveal significant correlations between the ADOS scores and the video codes for awkwardness in faces ($r(14) = .67$, $p = .009$) and voices ($r(14) = .60$, $p = .023$). These data indicate that greater social communication impairment as documented by the ADOS is related to greater awkwardness in facial and vocal productions.

Discussion

This pilot study was designed to provide a first analysis of the qualitative differences in facial and vocal emotional expressions and lead to further, in depth studies of the relationship between objective and subjective measures of social communication in this population.

Contrary to our expectations, accuracy in the group of participants with HFA was equal to that of the TD group for facial expressions and higher than that of the TD participants for vocal expressions. Looking more closely at our voice-coding data, the most likely cause of the increased accuracy in the prosodic expressions of the group of participants with HFA was the higher expressiveness scores obtained for voice productions. Our coding scheme required that an expressiveness code of "flat" also receive the emotion code "neutral," since a lack of expressiveness cannot then be coded as a particular emotion type. As only emotional sentences were analyzed, "neutral" was never the correct target emotion. As a group, the TD participants obtained vocal expressiveness scores closer to neutral, to the point where their emotion accuracy scores were at chance level, indicating that their

emotions were not expressed strongly enough to be reliably recognized by naïve coders. Both groups exhibited very low accuracy levels overall, but the group of participants with HFA, with their significantly higher vocal expressiveness scores, had a better chance of being coded as expressing any emotion at all, including the correct one. This result may have been significantly affected by the selection of participants as falling in the high-functioning range of the autism spectrum. Future studies may consider analyzing the accuracy of emotional facial expression production in individuals with lower-functioning autism or individuals across a larger range of diagnostic severity.

As with all study designs, the task itself may have influenced our results. We asked participants to retell stories in order to control the type and number of emotions represented in their productions, which ultimately allowed us to calculate accuracy of recreating the targeted emotion. But informal observations of participants gave us the impression that older TD participants may have been embarrassed by the task and therefore less likely to be expressive. In contrast, even older adolescents with HFA engaged fully with the task and were willing to be expressive within the context of retelling these stories. Unfortunately, our sample was not large enough to create separate age groups, which might have revealed quantifiable group differences in expressiveness for older participants. Future studies should explore using natural conversation, rather than story retelling to elicit expressions, as well as use more closely constrained age groups.

The fact that the group of participants with HFA was rated as more expressive than the TD group in the prosodic codes is supported by our findings on the acoustic measures, which indicate greater ranges in pitch and intensity for the group of participants with HFA, both of which can be indicators of greater expressiveness (Banse & Scherer, 1996). This result is also in line with clinical descriptions in the literature, which often indicate greater variability in intensity and pitch, or "sing-song" prosody productions among individuals with ASD (Shriberg et al., 2001).

Our data confirm the hypothesis that the dynamic facial expressions of individuals with HFA would be rated as less natural or more awkward than those of their TD peers. Participants with HFA also showed significant correlations between increased awkwardness in facial expressions and increased severity in the ADOS social communication score. These results resonate with clinical descriptions of facial and vocal expressions in this population being "different" or socially inappropriate (Asperger, 1944; Kanner, 1968; Paul, Shriberg, et al., 2005; Shriberg et al., 2001) and provide additional evidence for the suspected connection between social communication ability and the ability to produce natural facial expressions.

An unexpected finding of our study is that the group differences in perceived awkwardness were not consistent across facial and vocal modalities. Our data indicate that individuals with HFA show more awkwardness in the facial expression component of social communication than the prosodic modality. Follow-up studies with larger participant groups are needed to determine if this result is generalizable.

Looking at the relative strengths of freeze-frame vs. video codes (in depth scrutiny vs. intact dynamic properties), it is interesting to see that group differences in facial awkwardness exist only for video-, but not freeze-frame codes. These data indicate that it is not the constellations of facial features in individual frames of a facial expression (as captured by the freeze-frame coding) that dictate the perceived awkwardness of facial expressions, but rather the dynamic transitions between those moments.

Limitations

Due to technical recording difficulties, we were not able to conduct all analyses for all participants. As a result, our participant groups are relatively small, have a wide age range, and may not be fully representative of the HFA and TD populations. There was no female with HFA included in the analyses, which is not completely unexpected in a small sample of individuals with HFA, which has a prevalence of four males to one female. For this reason, we cannot make predictions regarding the facial and vocal awkwardness of girls with HFA based on our findings. Nevertheless, these data are an important first step in presenting quantifiable differences in the facial and vocal production quality of children and adolescents with HFA and their TD peers. Follow-up studies should use larger cohorts with more tightly constrained age ranges, conduct acoustic, as well as coding analyses for all participants, and explore other methodologies for eliciting and coding emotional expressions. Future investigations should also attempt to determine the specific temporal and spatial features of facial and vocal expressions that underlie the perception of awkwardness by TD observers.

Conclusions and clinical implications

Our data reveal that participants with HFA may be significantly more awkward than their typical counterparts in facial expressions, but not prosody production during story retelling. No significant group differences were found in freeze-frame coding of facial expressions, indicating that the dynamic transitions of affective non-verbal communication may be crucial to our perception of social appropriateness.

Peppé (2009) emphasizes that it is insufficient to determine prosodic accuracy through acoustic measures and more relevant to discover how typical listeners judge prosodic productions, since it is ultimately the listener who forms an opinion of the speaker's social skills based in part on non-verbal cues, such as prosody and facial expressions. Our results indicate that individuals with HFA may produce emotional facial and vocal expressions that are categorically accurate, but qualitatively different, which creates an impression of oddness or awkwardness, contributing significantly to the overall clinical impression of social communication deficits in this population.

Acknowledgments

References

Asperger H. Die autistischen Psychopathen im Kindesalter. Archiv fur Psychiatrie und Nervenkrankheiten. 1944; 177:76–137.

Baltaxe CA. Use of contrastive stress in normal, aphasic, and autistic children. Journal of Speech Language & Hearing Research. 1984; 27(1):97–105.

Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology. 1996; 70(3):614–636. [PubMed: 8851745]

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

Capps L, Kasari C, Yirmiya N, Sigman M. Parental Perception of Emotional Expressiveness in Children With Autism. Journal of Consulting and Clinical Psychology. 1993; 61(3):475–484. [PubMed: 8326050]

Carifio J, Perla RJ. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. Journal of Social Sciences. 2007; 3(3):106–116.

Dawson G, Hill D, Spencer A, Galpert L, Watson L. Affective exchanges between young autistic children and their mothers. Journal of Abnormal Child Psychology. 1991; 19(1):115–115.10.1007/bf00910569

Diehl JJ, Bennetto L, Watson D, Gunlogson C, McDonough J. Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. Brain and Language. 2008; 106(2):144–152.10.1016/j.bandl.2008.04.002 [PubMed: 18502497]

Diehl JJ, Paul R. Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2012; 6(1):123–134.10.1016/j.rasd.2011.03.012 [PubMed: 22125576]

Diehl JJ, Paul R. Acoustic and perceptual measurements of prosody production on the PEPS-C by children with autism spectrum disorders. Applied Psycholinguistics. (in press).

Dunn, LM.; Dunn, LM. Peabody Picture Vocabulary Test. 3. Circle Pines, MN: American Guidance Service; 1997.

Ekman, P. Expression and the nature of emotion. In: Scherer, K.; Ekman, P., editors. Approaches to emotion. Hillsdale, NJ: Lawrence Erlbaum; 1984. p. 319-343.

Ekman, P.; Friesen, WV. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA: Consulting Psychologists Press; 1978.

Fosnot, SM.; Jun, S. Prosodic characteristics in children with stuttering or autism during reading and imitation. Paper presented at the 14th International Congress of Phonetic Sciences; 1999.

Grossman RB, Bemis RH, Plesa Skwerer D, Tager-Flusberg H. Lexical and affective prosody in children with high-functioning autism. Journal of Speech, Language, and Hearing Research. 2010; 53(3):778(716).

Hubbard K, Trauner DA. Intonation and emotion in autistic spectrum disorders. Journal of Psycholinguistic Research. 2007; 36(2):159–173. [PubMed: 17136465]

Kanner L. Autistic disturbances of affective contact. Acta Paedopsychiatria. 1968; 35(4):100–136.

Kaufman, A.; Kaufman, N. Manual for the Kaufman Brief Intelligence Test. 2. Circle Pines, MN: American Guidance Service; 2004.

Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule - WPS (ADOS-WPS). Los Angeles, CA: Western Psychological Services; 1999.

Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism & Developmental Disorders. 1994; 24(5):659–685. [PubMed: 7814313]

Macdonald H, Rutter M, Howlin P, Rios P, Conteur AL, Evered C, Folstein S. Recognition and Expression of Emotional Cues by Autistic and Normal Adults. Journal of Child Psychology and Psychiatry. 1989; 30(6):865–877.10.1111/j.1469-7610.1989.tb00288.x [PubMed: 2592470]

McCann J, Peppé S. Prosody in autism spectrum disorders: a critical review. International Journal of Language & Communication Disorders. 2003; 38(4):325–350. [PubMed: 14578051]

Paul R, Augustyn A, Klin A, Volkmar FR. Perception and production of prosody by speakers with autism spectrum disorders. Journal of Autism & Developmental Disorders. 2005; 35(2):205–220. [PubMed: 15909407]

Paul R, Shriberg LD, McSweeny J, Cicchetti D, Klin A, Volkmar F. Brief report: relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2005; 35(6):861–869.10.1007/s10803-005-0031-8 [PubMed: 16283080]

Peppé S, McCann J. Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. Clinical Linguistics & Phonetics. 2003; 17(4–5):345–354. [PubMed: 12945610]

Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Receptive and expressive prosodic ability in children with high-functioning autism. Journal of Speech Language & Hearing Research. 2007; 50(4):1015–1028.

Peppé SJE. Why is prosody in speech-language pathology so difficult? International Journal of Speech-Language Pathology. 2009; 11(4):258–271.

Shriberg LD, Paul R, McSweeny JL, Klin AM, Cohen DJ, Volkmar FR. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. Journal of Speech Language & Hearing Research. 2001; 44(5):1097–1115.

Snow ME, Hertzig ME, Shapiro T. Expression of Emotion in Young Autistic Children. Journal of the American Academy of Child and Adolescent Psychiatry. 1987; 26(6):836–838. [PubMed: 2448285]

Van Santen JPH, Prud'hommeaux ET, Black LM, Mitchell M. Computational prosodic markers for autism. Autism. 2010; 14(3):215–236.10.1177/1362361309363281 [PubMed: 20591942]

Woodcock, RW.; Mather, N.; Schrank, FA. Woodcock-Johnson III Diagnostic Reading Battery. Rolling Meadows, IL: Riverside Publishing; 2004.

Yirmiya N, Kasari C, Sigman M, Mundy P. Facial expressions of affect in autistic, mentally retarded and normal children. Journal of Child Psychololgy and Psychiatry. 1989; 30(5):725–735.

Figure 1.
"Safari Bob" expressing a fearful sentence

Table 1

Descriptive Characteristics of Participant groups

Acoustic Analysis	HFA (n=18) M(SD)	TD (n=11) M(SD)
Age	13:10 (3:4)	15 (3:6)
Sex	18 male	11 male
Full Scale IQ	109.17(13.61)	116.09(15.22)
Verbal IQ	108.44 (17.34)	117.09 (17.58)
Nonverbal IQ	106.94 (14.26)	110.64 (15.12)
PPVT-III	117.65 (15.00)	119.09 (9.7)
WJ III DRB	109.39 (11.71)	110.36 (14.09)
Coding Analysis	HFA (n=14)	TD (n=12)
Age	14 (2:7)	14 (2:6)
Sex	14 male	11 male 1 female
Full Scale IQ	108.57(14.52)	116.25(14.74)
Verbal IQ	109.43 (17.69)	113.67 (17.36)
Nonverbal IQ	104.86 (13.94)	114.25 (15.42)
PPVT-III	115.57 (19.53)	115.20 (8.6)
WJ III DRB	111.36 (10.75)	112.08 (15.91)

Table 2

Description of Codes

Category	Code # and name		Definition
Expressiveness	1	flat	Not necessarily "deadpan." May show non- emotional expression. This score necessitated an emotion type code of "neutral," since a lack of emotional expressiveness could not then be coded for a specific emotion type.
	2	mild	Beginning of a discernible emotion.
	3	moderate	Obvious emotional display.
	4	extreme	Highly (possibly overly) animated expression.
Naturalness/Awkwardness	1	natural	Appropriate and expected expressions.
	2	slightly awkward	Possibility that person is socially awkward.
	3	very awkward	Definite social awkwardness or difference.
	4	unnatural	Social awkwardness to the point of being outside of realm of expected expression during typical social interaction.
Emotion Type		Neutral	Used for expressions showing no discernible emotion
		Other	Emotional expression that doesn't fit any provided category
		Happy, Surprise, Fear, Anger were coded as expected	

Table 3

Acoustic Analysis

Measure	HFA (n=18) M(SD)	TD (n=11) M(SD)
Pitch range (max F_0 – min F_0) *	170.00(89.15)	108.63(56.58)
Minimum F_0	104.83(23.88)	98.00(16.17)
Maximum F_0 *	274.83(97.39)	206.64(66.45)
Mean F_0	190.89(59.54)	155.82(49.83)
Intensity range (max intensity – min intensity) *	27.22(4.19)	23.8(4.75)
Minimum intensity	51.39(3.62)	54.82(5.69)
Maximum intensity	78.61(5.1)	78.64(3.85)
Mean intensity	68.78(4.63)	69.27(3.69)

*
refers to significant difference at p=.05

Table 4

Coding Data

Analysis type	Measure	HFA (n=14) M(SD)	TD (n=12) M(SD)
Video	Face accuracy (%)	17.56(16.44)	11.63(13.49)
	Face expressiveness	1.60(.45)	1.34(.30)
	Face awkwardness *	1.44(.55)	1.07(.11)
	Voice accuracy *	38.42(22.92)	16.67(14.19)
	Voice expressiveness *	2.01(.39)	1.66(.30)
	Voice awkwardness	1.59(.7)	1.18(.26)
Freeze-Frame	Face expressiveness	1.36(.24)	1.41(.26)
	Face awkwardness	1.07(.09)	1.03(.04)

*
refers to significant difference at $p < .05$

Appendix A

Stories used in study

Story # 1	Emotion
I'm at a water hole taking pictures of elephants. I'm just getting the camera ready…	Neutral (not included in analysis)
What was that noise? A tree fell over, and all the elephants ran away!	Fear
That's it for today, I won't get any more pictures!	Anger
But look: A purple water bird landed at the water hole.	Surprise
Wow, nobody has ever seen a purple water bird before. This photo will make me famous!	Happy

Story # 2	Emotion
I'm hiding to take pictures of lemurs.	Neutral (not included in analysis)
There are some. An excellent picture!	Happy
Oh no! I'm out of film! I can't believe I could miss this shot!	Anger
I quickly reload my camera and focus …	Neutral
A tiger jumps up right in front of me! His sharp teeth are inches from my face!	Fear
But - He's not attacking! He's posing for my camera!	Surprise

Story # 3	Emotion
I'm on a safari in a place I've never been before. Suddenly my group is gone.	Neutral (not included in analysis)
Oh no! I'm lost! All by myself in the jungle! What am I going to do?	Fear
Wait, I think I hear the guys from my group!	Surprise
Yes, they're just around the next corner! I run to join them and they're – laughing? They're standing there laughing at me.	Happy
This was a prank! They left me there on purpose. This was not funny!	Anger

Story # 4	Emotion
My last journey takes me to the rainforest to take pictures of baby monkeys.	Neutral (not included in analysis)
I have to cross a rickety bridge, and I hate heights!	Anger
Oh, look! A butterfly has landed on my hat! I take some photos and move on.	Surprise
There! I see baby monkeys playing. These pictures will be terrific!	Happy
Oh no! There's the mom, coming right for me! I didn't know her teeth were going to be so long! Let's get out of here.	Fear

Appendix B

Coding Strategy

Coding Type	Data Format	Rationale	Procedure
Video Code	Video clips	To quantify the perception of real-time facial expressions	Coders viewed each 3–5second clip representing a single emotion and provided codes for overall impression of emotion type, expressiveness, and naturalness at clip conclusion (see Table 2). We did not ask coders to explain which specific features of the expressions related to their codes, so we could capture their most immediate and instinctual responses. Coders viewed clips with sound turned off.
Video code	Audio files	To quantify the perception of real-time prosody productions	Same procedure as video coding of faces. Coders listened to clips without viewing the screen.
Freeze- frame code	Video clips	To capture the fine-grain distinctions and features that represent the component parts of dynamic facial expressions	"Froze" video playback on the first frame of each clip, the frame at each subsequent second, and the final frame. Selected frames (roughly 4% of overall video frames) were coded for expressiveness, and naturalness. We did not include analysis of emotion type for this coding technique. Freeze-frames in a given clip tended to obtain different emotion type codes. Resulting data could not be suitably summarized and are therefore not included here. Coders who performed freeze-frame coding on a given participant did not also perform video coding on the same participant.
Acoustic Analysis	Audio files	To capture the fine-grain distinctions and features that represent the component parts of prosodic productions	Used PRAAT (Boersma & Weenink, 2009) to record: • minimum fundamental frequency (F_0) • maximum F_0 • mean F_0 • pitch range (max F_0 – min F_0) • minimum intensity • maximum intensity • mean intensity • intensity range (max intensity – min intensity).