

The Perception of Affective Prosody in Children with Autism Spectrum Disorders and Typical Peers

Jessica Chiew¹, Margaret Kjelgaard²

¹Therapy Hub, SPD, Singapore; ²Rocky Mountain University of Health Professions, Provo, UT, United States

Purpose: This study investigated the ability of children with ASD, including the minimally verbal subgroup, to perceive angry, neutral, and happy prosody in low-pass filtered speech when provided with a structured training paradigm.

Methods: 13 children with ASD and 21 TD children completed the experimental task and two additional measures (nonverbal cognitive abilities, social responsiveness deficits) for regression analyses.

Results: The ASD group recognized prosodic conditions significantly less accurately than the TD group, and took significantly longer times to recognize all sentences compared to the TD group. Angry prosody was consistently the most difficult to recognize across groups. Nonverbal cognitive abilities is a significant predictor variable for successful recognition of neutral and happy prosody; although low nonverbal cognitive skills do not preclude minimally verbal children with ASD from accurately perceiving affective prosody.

Conclusions: The present study shows it is possible for minimally verbal children with ASD to successfully participate in experimental research using judgment tasks when provided with appropriate training.

Keywords: Autism spectrum disorder, Affective prosody, Minimally verbal



Received: August 16, 2017
Accepted: August 28, 2017

Correspondence:
Jessica Chiew

Therapy Hub, SPD, 2 Peng Nguan Street,
SPD Ability Centre, Singapore 168955,
Singapore
Tel: +65 6268 0125
E-mail: jessica.jc.chiew@gmail.com

© 2017 The Korean Association of Speech-Language Pathologists

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Autism spectrum disorder (ASD) is characterized by impairments in two broad areas: (1) persistent difficulties in social communication and social interaction; and (2) restricted and repetitive behavior patterns [1]. As it is a spectrum disorder, there exists a high degree of heterogeneity within the ASD population. Some low-functioning individuals with ASD exhibit such severe language disabilities they fail to develop a commonly shared symbolic system of communication. In contrast, some high-functioning individuals with ASD do develop communicative competence and demonstrate only limited deficits in the pragmatics of social communication [2]. It is not known why such variability occurs. However, even for individuals with ASD who do acquire language, they often struggle with everyday communication, particularly with the prosody, manner, and content of speech [3-6].

Prosody refers to the rhythm and tune of speech [7]: it organizes speech input, influences meaning, and provides non-literal information. Prosody plays a critical role both in the interpretation of language and in language expression. Successfully grasping the

communicative intent of others relies on one's ability to accurately perceive and decode prosody. Research has found that children with ASD tend to misjudge question utterances as declarative utterances [8,9] – which has clear implications on social communication (e.g., not answering a question when asked). On the other hand, a breakdown in prosodic expression can also create communication problems like the inability to assert desired intentions, which may lead to eventual social isolation. In this respect, the speech of individuals with ASD has been described as monotonous, sing-song-like, over-exaggerated, and stilted [8,10-12], and these prosodic differences make it difficult to ascribe a clear communicative intent to their interactions.

At the most fundamental level, however, prosody can be considered an integral building block of language acquisition and social functioning. Research has shown that Infants prefer listening to infant-directed speech (also known as Motherese) from as young as one month of age [13]. Motherese is different from normal adult-directed speech in that it is linguistically simplified, has a high pitch, greater pitch excursions, and exaggerated intonation (for a discussion of “parentese”) [14]. It has been suggested that Motherese (or parentese) attracts an infant's attention, expresses emotional affect, as well as presents infants with special prosodic cues that help distinguish individual speech sounds in their native language [14,15]. However, it has been found that very young children with ASD are less responsive to the sound of their mother's voice, and human voices in general [2,16], which puts successful language acquisition at risk. There is a symbiotic relationship that develops between the voice of a mother/parent and an infant, creating a unique social bond between parent and child. It may be that a degraded ability to perceive the emotion in a parent's voice would affect later social functioning in children with ASD.

As alluded to above, there is a known subgroup of children with ASD who do not develop a robust system of communication – a subgroup variously estimated at 10-30% of the ASD population [17,18]. For this minimally verbal subgroup, verbal output is extremely limited and speech, if any, is typically not spontaneous, non-generative, and with limited functional communicative purposes (in many cases only interpretable by informed caregivers). If these individuals with ASD are unable to attend to their own mother's and human voices distinct from other sounds at an early age, then the prospect of a typical pattern of language acquisition becomes highly suspect. Could the ability to perceive affective prosody be at the

core of language and communication difficulties among the minimally verbal subgroup of children with ASD?

Studies that have investigated the specific ability of individuals with ASD to perceive affective prosody and therefore make judgments about the emotional state of others have tended to focus on the higher functioning end of the spectrum. In one study, adolescents with Asperger Syndrome (AS; diagnosed prior to DSM-V), compared to age-matched controls, were shown to have significantly poorer ability to use nonverbal cues such as facial expression, body gestures, and prosody to interpret the feelings of actors featured in videotaped scenes [19]. These adolescents with AS relied the least on prosodic information to make their interpretations and they demonstrated significant difficulties when simultaneous presentations of facial, voice, body and situational cues were presented to them, compared to when they were asked to merely label still photographs [19]. In another study, adults with AS/high-functioning autism (HFA) were presented with recorded dialogues from which they were asked to determine the mental state of the speakers. Results of this purely auditory test suggested that individuals with AS/HFA demonstrated deficits in interpreting the state of mind of others based on affective prosodic cues [20]. Some researchers have posited that in children with HFA, understanding affective prosody may be difficult because it conveys information about the speaker's perspective, and taking on the perspective of another person is an area of challenge for individuals with ASD [21].

While there has been research that indicates individuals with HFA do have intact affective prosody interpretation abilities, and they are able to identify emotions as well as their typically developing peers [22-24], more recent research suggests that this ability drops significantly when interpreting low-intensity emotional expressions [25]. The subtler the nuances of affective prosody, the more difficult it is for individuals with ASD to accurately interpret them. For example, in instances where affective prosodic cues contradict semantic information, adults tend to rely on prosodic features to interpret the utterance [26]. Various studies, however, have indicated that individuals with ASD tend to do the opposite, i.e., they focus overly on the semantic, literal, content and therefore were unable to attend to affective prosodic cues [8,27-29]. A case in point, Kjelgaard and Tager-Flusberg [28] found that children with HFA were responsive at a subconscious level to affective prosody that was congruent or incongruent with semantic information; however, they were unable to accurately label the

affective prosody [28].

In general, it appears that individuals with ASD demonstrate difficulties in interpreting emotions from affective prosody. What is unequivocal is that most of the research has focused on individuals with ASD who are high-functioning, while there seems to be little to no research examining the minimally verbal subgroup, and even fewer studies to investigate their ability to perceive and/or interpret affective prosody. The inherent heterogeneity within the ASD population adds complexity to any research on this population. Most pressing, perhaps, is the glaring lack of research in the subgroup of children with ASD who are minimally verbal. While recent scientific inquiry and technological advancements (e.g., brain imaging) have allowed us a glimpse into the world of ASDs, still very little is known about this minimally verbal subgroup, poignantly termed “the neglected end of the spectrum” [18]. For practical reasons, research has tended to focus on the high-functioning end of the spectrum because a certain level of linguistic and cognitive ability is required to comprehend expectations and perform what can be quite abstract tasks. These studies have undoubtedly added to the knowledge pool of the field as a whole, yet gaps remain.

Given the complexity of the disorder, the heterogeneity of the autism spectrum, and the seemingly contradictory results, it is difficult to make conclusions regarding the nature and the extent of prosodic deficits in individuals with ASD. The subgroups within the ASD population, as well as the different functions of prosody, need to be clearly specified experimentally before pertinent findings can be derived. Critically, studies need to begin including the subgroup of children with ASD who are minimally verbal, instead of continuing to ignore them. The present study therefore seeks to investigate the ability of children with ASD, including those who are minimally verbal, to perceive affective prosody compared to typically-developing (TD) peers. A conscious effort was made to design innovative experimental procedures that allow even minimally verbal children with ASD to participate and complete the task.

The main objective of this study is to determine the ability of children with ASD to perceive affective prosody. The prosodic qualities of happy and angry will be investigated, as recent research using near-infrared spectroscopy indicated that typically-developing infants as young as 7 months show increased response in right temporal voice-sensitive regions of the brain when hearing happy and/or angry prosody [30], suggesting that the ability to decode these salient emotions in

prosody exists in early development. Secondly, this study also seeks to determine if there is a relationship between the ability to perceive affective prosody and (1) social responsiveness and/or (2) nonverbal cognitive abilities.

The main hypothesis is that children with ASD will show poorer ability to perceive affective prosody compared to TD peers. While it remains to be seen what particular aspect(s) of affective prosody contribute to this poorer performance, it may be that some minimally verbal children never gain competence with language and with the nuances of language in conveying feelings because of deficits in perceiving affective prosody. A secondary hypothesis is that there will be a negative relationship between the ability to perceive affective prosody and the degree of deficit in social responsiveness. Finally, it is also hypothesized that the ability to perceive affective prosody is not related to nonverbal cognitive abilities.

Low-pass filtering, pretesting, and acoustic analyses

The set of stimulus sentences used in the experiment were selected from a larger group of low-pass filtered sentences based on the results of a pretest to validate the emotional manipulation of prosody and a series of acoustic analyses. The goal of low-pass filtering the sentences was to remove the spectral information associated with the perception of phonemes and therefore the semantic content of the utterance, while preserving prosodic contours. These low-pass filtered sentences were then pre-tested to identify those that were most accurately perceived as communicating the target emotions. The acoustic analyses examined the fundamental frequency (F0) contours and duration of the stimuli.

Low-pass filtering

One potentially confounding factor in the research on affective prosody is the interaction between syntax, semantics, and emotions. Humans have access to a limited set of acoustic features to express different intentions through speech. How much focus one assigns to the categories of grammatical, pragmatic, and affective prosody becomes a negotiation that speakers necessarily must engage in during the encoding of information, which in turn affects the subsequent decoding of this information. In one such study aimed at understanding the interplay between grammatical, pragmatic, and affective prosody, Pell [31] found that a speaker’s flexibility in modulating F0 parameters to communicate emotions (e.g., mean F0, F0 variation) tended to be subsumed by the prosodic demands of interrogative utterances with sentence-final stress (i.e., the

terminal rise of questions in English coupled with a “marked” content word in final position); yet this pattern was not observed in declarative utterances. The study showed that interrogative utterances exhibited less divergent F0 parameters for emotions at points where contrastive stress is obligatory [31]. This suggests that when the linguistic-communicative context warrants a question with sentence-final stress in English, grammatical and pragmatic prosody tend to override affective prosody – and listeners are left with attenuated acoustic information with which to identify emotions in these utterances.

At the same time, words tend to be nuanced with layers of meanings, and as such hold varying degrees of emotional valence that differ from person to person, even when there is a shared overall concept of a given lexical item. How one teases apart the contribution of semantics from true expressions of emotional states impacts the cogency of any research purporting to investigate affective prosody. Many researchers have used an evolutionary framework to explain the ecological significance and validity of vocal expressions of emotion, offering neurophysiological responses and social adaptation as reasons-why [31-34]. What is most striking in this tradition of research is evidence that affective prosody can be communicated independently of verbal comprehension – that listeners are able to accurately decode emotions in unfamiliar languages (e.g., pseudo-languages, foreign languages) at above chance level [32,34-36]. This suggests that there may be an underlying set of acoustic cues that listeners reliably draw upon to interpret affective prosody, a point that will be revisited later.

To circumvent the potential confound introduced by lexical and semantic contexts to the accurate interpretation of affective prosody, stimuli were low-pass filtered in this study. Low-pass filtered sentences were also designed to help to level the playing field between the ASD and TD groups, especially in accommodating the heterogeneity within the autism spectrum. Specifically, low-pass filtered sentences would (1) allow even minimally verbal ASD participants to complete the perception task despite impaired linguistic abilities; and (2) prevent TD participants from relying on lexical-semantic cues to determine affect.

We recorded the original sentences with a female native American English speaker, who is both a graduate student in communication sciences and disorders knowledgeable about prosodic theory, and an actress with very good control of her voice. We used Praat software [37], set in the mono channel at a 44,100 Hz sampling frequency, to record a total of 108 sen-

tences in a single session, in a soundproofed room. The sentences were comprised of three training items and 105 test items. These test items were made up of 35 declarative sentences each spoken with angry, neutral, and happy prosody, with an equal number of sentences matched across seven syllable lengths ranging from four to 13 syllables (i.e., 5 sentences \times 7 syllable lengths \times 3 prosodic conditions).

Following the recording session, stimuli were low-pass filtered using Praat software to remove sound wave frequencies associated with the perception of phonemes and thus eliminating the ability to perceive any lexical-semantic content, while preserving prosodic contours. Based on a review of past research that aimed to investigate the recognition of affective prosody using low-pass filtered speech, different cutoffs have been used, including: 333 Hz [38] and 150 Hz [23]. For the current study, a 0-400 Hz low-pass filter was used for sentences conveying neutral emotion, and a 0-450 Hz low-pass filter was used for sentences conveying anger or happiness. The average intensity of the sentences was scaled to 70 dB. These acoustic manipulations were determined by taking into consideration the speaker’s natural F0 and tone of voice, while ensuring that salient prosodic information was maintained at an audible level.

Pretesting

All 105 low-pass filtered sentences were pretested to determine those for which judgment regarding the target emotion communicated by the speaker was most accurate. Ten neurotypical native American English adult speakers, who were colleagues of the authors, participated in the pretest. The sound file for each sentence was embedded into a Microsoft PowerPoint document in random order, and shared with the participants. Three training items were also included in both the original and low-pass filtered versions to allow listeners to gain familiarity with the target emotion as communicated through filtered speech. Participants were instructed to listen to each stimulus sentence no more than three times, and rate each item as being angry, neutral, or happy.

Overall, the full set of low-pass filtered sentences was recognized correctly at a high mean level of 86.7%, where chance performance was 33.3%. This rate of recognition is broadly in line with existing research on the perception of vocally expressed emotions, which tend to yield accuracy scores at higher-than-chance levels [32-35]. From this set of 105 sentences, the top three sentences that best exemplified all three prosodic conditions from each syllable length (i.e., 3 sen-

Table 1. Pretest % Correct Recognition of Prosodic Condition

	Recognition (% correct)	Recognition (% correct)
	All sentences; n = 105	Experiment sentences; n = 63
Angry	80.3	87.6
Neutral	91.4	94.3
Happy	88.3	93.3

tences \times 7 syllable lengths \times 3 prosodic conditions) were selected for the experiment. The resulting 63 sentences selected were recognized correctly at a mean level of 91.7%. Table 1 shows the pretest correct recognition for each prosodic condition in the full set of all low-pass filtered sentences, as well as the smaller set selected for the actual experiment.

Acoustic analyses

Given that listeners across studies are consistently able to correctly identify emotions in unfamiliar languages at levels well above chance, researchers have posited the existence of a set of basic emotional intonations (e.g., anger, happiness, sadness, fear) that prevail cross-linguistically. Studies in the categorical perception of vocally expressed emotions support the presence of discrete emotion categories in the auditory modality [36,39]. When presented with synthesized speech for blended emotional continua (e.g., happiness-anger, anger-sadness), listeners perceived two distinct emotion categories in each continuum separated by a sudden category boundary [39] similar to the categorical perception of phonemes. As mentioned above, substantial evidence exists for a universal set of acoustic parameters that speakers manipulate and that listeners hone in on, to encode and decode affective prosody respectively. Even in the absence of lexical-semantic content, acoustic analyses have often found specific acoustic markers of prosody (e.g., F0 measures, speech rate, intensity) that differentiate the set of basic emotions [32-36].

With this in mind, acoustic analyses on the set of low-pass filtered sentences were used for two purposes:

- (1) To identify the set of acoustic parameters that differentiated the three target emotions in the current study
- (2) To validate the fidelity of the low-pass filtered sentences with reference to current knowledge regarding acoustic markers of the three target emotions in this study

Praat software was used to obtain durational measures (in seconds) and measures of pertinent F0 parameters (in Hz), including minimum F0 (MinF0), maximum F0 (MaxF0), mean F0 (MeanF0) and F0 range (F0_Range), by prosodic

Table 2. Means and Standard Deviation of Acoustic Parameters by Prosodic Condition

	Duration (s)	MaxF0 (Hz)	MinF0 (Hz)	MeanF0 (Hz)	F0_Range (Hz)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Angry	1.89 (0.61)	370 (52.55)	108 (27.04)	249 (25.75)	261 (55.54)
Neutral	1.78 (0.68)	336 (53.12)	112 (31.62)	204 (16.62)	224 (57.07)
Happy	1.79 (0.57)	484 (24.73)	152 (39.99)	321 (24.69)	331 (45.24)

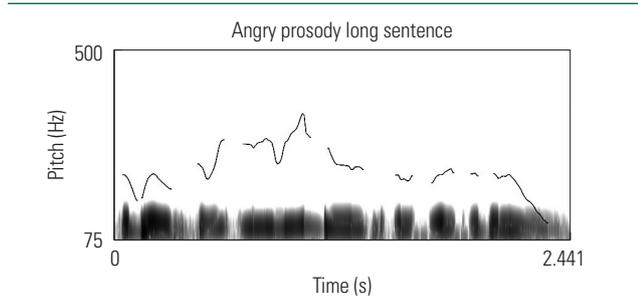
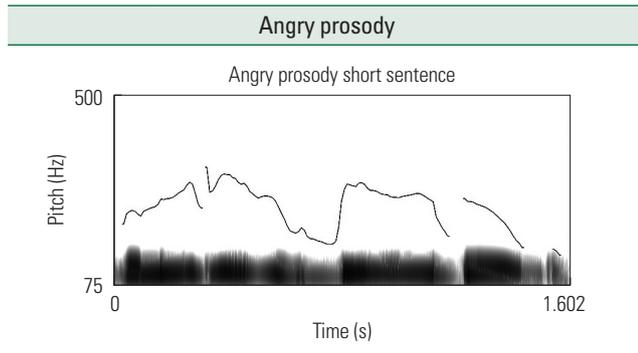
condition. The means and standard deviations of these acoustic parameters are shown in Table 2.

For each acoustic parameter, a repeated measures ANOVA was conducted across the prosodic conditions. There was a significant main effect of Prosody on all F0 acoustic parameters as well as duration: MinF0 [$F(2,19) = 10.95, p = 0.001$, partial $\eta^2 = 0.536$], MaxF0 [$F(2,19) = 105.40, p < 0.001$, partial $\eta^2 = 0.917$], MeanF0 [$F(2,19) = 335.69, p < 0.001$, partial $\eta^2 = 0.972$], F0_Range [$F(2,19) = 37.69, p < 0.001$, partial $\eta^2 = 0.799$], and Duration [$F(2,19) = 4.49, p = 0.019$, partial $\eta^2 = 0.342$].

Overall, happy prosody had the highest minimum, maximum and mean F0, and largest F0 range, closely followed by angry prosody. Neutral prosody had the lowest maximum and mean F0, and the lowest F0 range. This is generally in line with previous research on the acoustic correlates of emotional prosody – happy and angry utterances are characterized as having a higher mean F0 and greater variation in F0 [31-35]. In addition, the findings also fit well with emerging research that shows that neutral prosody has unique perceptual properties such as lower mean F0 and F0 range compared to other emotional prosody [28,35]. Thus the first goal of identifying the set of acoustic parameters – i.e., minimum, maximum, mean F0, and F0 range – that differentiated angry, neutral, and happy prosody in this current study was achieved.

To validate the fidelity of the low-pass filtered sentences, the prosodic contours of several sentences were visually examined against known acoustic markers of angry, neutral, and happy prosody. Sound waveforms were extracted and the corresponding pitch contour of several sentences in each of its angry, neutral, and happy prosodic conditions, using Praat software. Table 3 shows the prosodic contours and waveforms of one 7-syllable and one 13-syllable sentence, together with a qualitative description of key acoustic features for each prosodic condition. Overall, the set of sentences, despite being low-pass filtered, maintained prosodic contours that are qualitatively similar to findings from past research aimed at describing salient features of target emotions as conveyed through affective prosody.

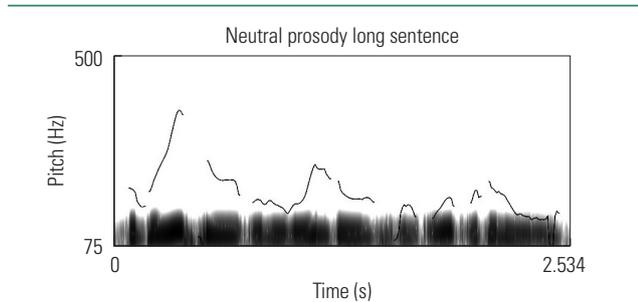
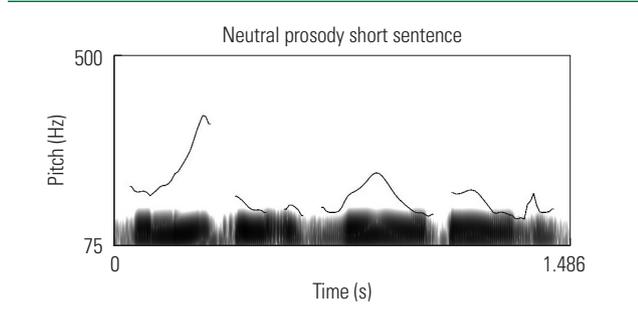
Table 3. Spectrograms and Prosodic Contours of Selected Sentences



Qualitative description:

Angry prosody is characterized by moderate-high F0, moderate-high F0 variability [33]; a greater proportion of downward directed/falling F0 contours [32,42]; and a falling contour over the last word/syllable [42].

Neutral prosody

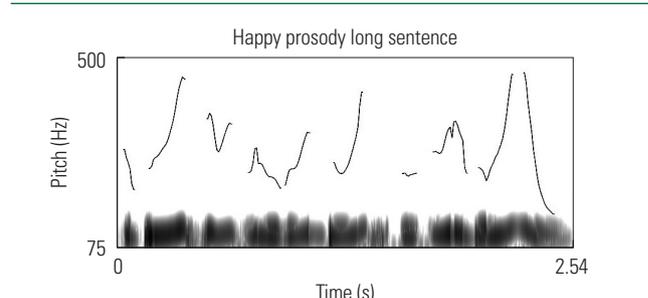
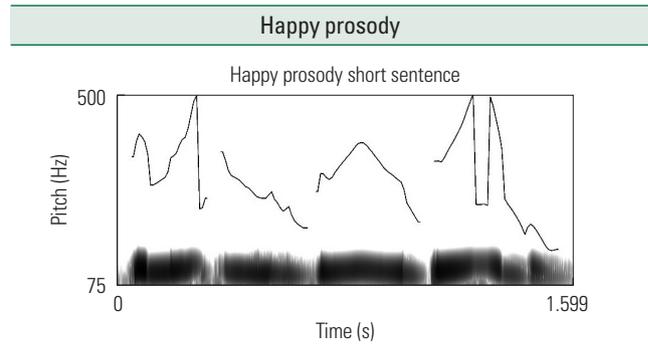


Qualitative description:

Neutral prosody is characterized by low-moderate F0, less F0 variability compared to angry prosody and happy prosody, faster speech rate [35]; and a rising contour over the last word/syllable [28,42].

(Continued to the next)

Table 3. Continued



Qualitative description:

Happy prosody is characterized by high F0, high F0 variability [33]; a greater proportion of upward directed/rising F0 contours [32,42]; and a trend towards a rising contour over the last word/syllable [28,42].

METHODS

Participants

13 children with ASD and 21 TD children participated in this study. Of the 13 children with ASD, 11 were minimally verbal (i.e., their speech was not spontaneous and non-generative, and in many cases verbal expression was extremely limited if not absent). The participants were recruited through local schools and via word of mouth through colleagues. All participants came from English-speaking homes, and were reported to have normal hearing and vision.

Autism diagnosis

ASD participants had to have met criteria for autistic disorder as stated in the DSM-IV, diagnosed by a certified clinical child psychologist/child psychiatrist/neurologist or qualified pediatrician. All the ASD participants attended schools for children with autism.

Standardized testing and caregiver survey

Participants were administered the Matrices subtest from the Kaufman Brief Intelligence Test, Second Edition (KBIT2; [40])

to assess nonverbal IQ. At the same time, parents or teachers familiar with the participants completed the Social Responsiveness Scale (SRS; [41]) to assess various dimensions of interpersonal behavior, communication, and repetitive/stereotypic behavior that are characteristic of ASD. The descriptive characteristics of both groups can be found in Table 4.

The ethics of this study with regard to human subject participants and the procedures were approved by the Spaulding Rehabilitation Hospital Institutional Review Board. The parents of all children gave written informed consent prior to their participation, and all procedures were conducted according to the approved protocol.

Materials

The training stimuli consisted of a mix of unfiltered and low-pass filtered sentences ($n = 18$), and the experimental stimuli consisted of a set of 21 low-pass filtered sentences that each conveyed happy, angry, and neutral prosody (i.e., $n = 63$ sentences). As discussed above, the sentences were selected

based on perceptual saliency given results of the pretest, as well as acoustic features highly characteristic of happy, angry, and neutral prosody. The sentences ranged in length from four to 13 syllables, with an equal number of sentences matched on syllable length (i.e., 3 sentences \times 7 syllable lengths \times 3 prosodic conditions). Matching an equal number of sentences on syllable length allowed us to investigate the effect of sentence length on various aspects of performance.

Procedure

Participants were tested in a quiet room. The experiment was run using FLXLab software [43] on a laptop computer. The stimuli were presented through stereo headphones with the volume adjusted to a comfortable listening level. Participants were told that they would hear a sentence through the headphones, and they should determine how the speaker was feeling (i.e., “happy”, “angry/mad”, or “in the middle”), as quickly and as accurately as possible. Responses were collected using three keys on the keyboard that had *Mayer-Johnson Board-maker* face symbols to represent the three emotions. Accuracy and reaction times were recorded with the software. Visual supports in the form of *First/Then* boards were used with ASD participants to provide directions and to scaffold comprehension of task expectations (e.g., “First listen to sentence, then pick one face”).

Training

The training component was specifically designed to allow even ASD children with very low language levels to be able to successfully participate in the experiment. The levels of progression of the training block are shown in Table 5. All participants in the study underwent the same training protocol.

Experiment

After successful completion of the training block, participants

Table 4. Descriptive Characteristics of Participants by Group

		ASD	TD
Gender (#)	M	12	8
	F	1	13
Age (yr;:mth)	Mean (SD)	11;7 (1;0)	8;0 (0;5)
	Min/Max	9;10/12;10	6;11/8;8
SRS total T-score ^a	Mean (SD)	76.46 (13.43)	48.62 (9.48)
	Min/Max	61/108	40/76
KBIT2	Mean (SD)	52.38 (18.94)	106.43 (10.66)
	Min/Max	40/107	79/122
Nonverbal intelligence age-equivalent (yr;:mth)	Mean (SD)	5;4 (2;11)	8;11 (1;10)
	Min/Max	3;11/14;8	5;2/12;8

^aSRS total T-score ≤ 59 T indicates normal range; ≥ 76 T indicates severe range strongly associated with autistic disorder.

Table 5. Training Block – Levels of Progression

Training level	Task description
Level 1 (Maximum scaffolding)	Unfiltered sentences ($n = 3$) with matching semantic content and affective prosody presented, direct teaching by the tester with verbal explanation and gestural modeling (e.g., “I am sitting” and pointing to neutral face).
Level 2a (Moderate scaffolding)	Filtered Level 1 sentences ($n = 3$) maintaining prosodic contour without semantic content presented, direct teaching by the tester with verbal explanation and gestural modeling (e.g., “she feels in the middle” and pointing to neutral face).
Level 2b (Minimal scaffolding)	Filtered Level 1 sentences ($n = 3$) maintaining prosodic contour without semantic content presented, independent response by the participant with verbal correction and gestural modeling if an error was made.
Level 3 (No scaffolding)	Filtered sentences ($n = 9$) from the original stimulus set used during the pretest presented, independent response by the participant with no feedback.

completed the experiment. There were 63 sentences in the experiment phase, divided into three blocks. All trials were initiated with the presentation of a photograph of a boy wearing headphones to alert the participants to listen, followed by a 1,500 ms inter-stimulus interval. Next, the stimulus sentence was presented in its entirety. Reaction time was measured from the offset of the stimulus to the time of the response key press. After participants rendered an emotion decision, a reinforcing image was shown indicating success. The examiner then initiated the next trial. The order of presentation of experiment blocks was systematically rotated across participants, and the order of presentation of stimulus sentences within each block was automatically randomized within the FLXLab program.

RESULTS

Pearson correlation coefficients across groups were computed between the stimuli lengths and reaction time (RT). The relationships among the RTs for the different lengths of stimuli were tested for the neutral prosodic condition to determine stimulus length categories with the greatest discriminate validity. Results of the correlational analyses produced the following two categories:

- Short sentences: 4-, 5-, and 7-syllable sentences
- Long sentences: 8-, 10-, 11-, and 13-syllable sentences

All correlation coefficients were statistically significant within the two categories except for one in the long sentence category (10-syllable and 13-syllable sentences with a coefficient of $r = 0.33$, $p = 0.063$).

Accuracy

The mean accuracy for short sentences was 44.1% (SD = 14.1)

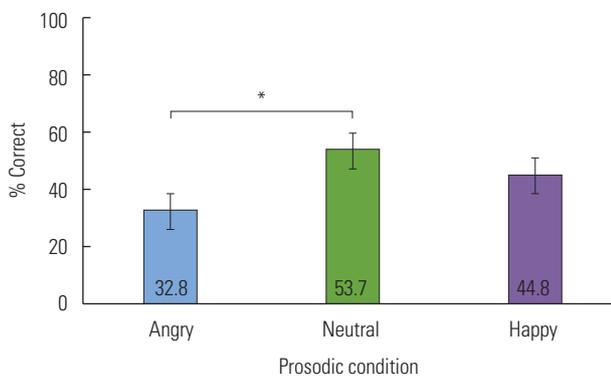


Figure 1. Mean accuracy of all sentences by prosodic condition. * $p < 0.05$.

while the mean accuracy for long sentences was not significantly different at 43.4% (SD = 15.4), $t(33) = 0.313$, $p = 0.756$. To evaluate whether the prosodic conditions differed in how accurately they were recognized, a 3 (Prosody) \times 2 (Length) \times 2 (Group) repeated measures ANOVA was conducted on the mean correct recognition (% correct). Results indicated a significant main effect of Prosody, $F(2,31) = 4.297$, $p = 0.023$, partial $\eta^2 = 0.217$; but no significant interactions were reported between Prosody \times Group, or Prosody \times Length. Angry prosody was consistently the most difficult for participants to identify compared to neutral and happy prosody (see Figure 1). Further analysis also revealed a significant Group effect as a between-subjects factor, $F(1,32) = 26.224$, $p < 0.001$, partial $\eta^2 = 0.450$. The ASD group recognized prosodic conditions less accurately than the TD group (see Figure 2). Specifically, they were significantly less accurate in recognizing neutral and happy prosody compared to the TD group.

Reaction time

The mean RT across groups for short sentences was 1,518.4 ms (SD = 522.4), and the mean RT across groups for long sentences was statistically significantly less, 1,333.6 ms (SD = 582.0), $t(33) = 4.936$, $p < 0.001$. To determine the effects of prosody on reaction time, a 3 (Prosody) \times 2 (Length) \times 2 (Group) repeated measures ANOVA was performed. Results indicated a significant interaction for Length \times Group, $F(1,32) = 4.745$, $p = 0.037$, partial $\eta^2 = 0.129$. The TD group required significantly longer time to recognize shorter sentences compared to longer sentences, while the ASD group showed no differences (see Figure 3).

There was also a main effect of Length, $F(1,32) = 20.702$, $p < 0.001$, partial $\eta^2 = 0.393$. Recognition time of shorter sentences was consistently prolonged compared to longer sen-

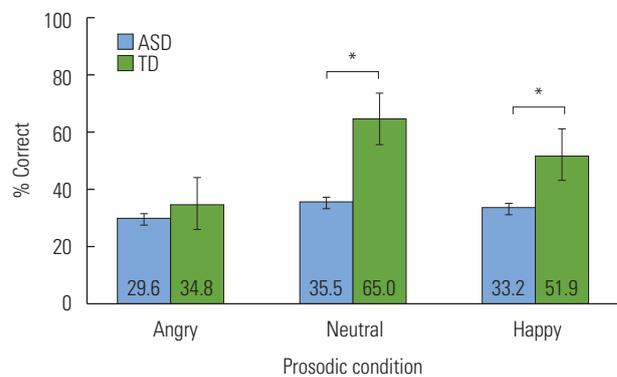


Figure 2. Mean accuracy of all sentences by group. * $p \leq 0.05$.

tences (see Figure 4). Angry prosody tended to take the longest to be recognized compared to neutral and happy prosody. There was also a significant Group effect, $F(1,32) = 10.304$, $p = 0.003$, partial $\eta^2 = 0.244$. The ASD group demonstrated consistently longer times to recognize all sentences compared to the TD group (see Figure 5).

Error analysis

An error matrix was generated to further analyze the patterns of responses made by the participants. Table 6 presents the correct recognition for each prosodic condition between the

two groups. A 3 (Prosody) × 3 (Emotion response) crosstab was performed to determine if the target emotions were assigned differently across the prosodic conditions. There were significant differences for the TD group, Cramer's $V = 0.315$, $p < 0.001$; however, no significant differences were detected for the ASD group, Cramer's $V = 0.069$, $p = 0.118$. The TD group correctly recognized neutral and happy prosody at significantly higher rates than chance, whereas angry prosody was recognized at approximately chance level. However, the ASD group recognized all the prosodic conditions (angry, neutral, and happy) at chance levels. Both the ASD and TD groups were most likely to erroneously recognize angry prosody as neutral-sounding.

To further analyze the apparent difficulty that TD participants have in recognition of angry prosody, data collected across all TD participants in both the pretest and experimental study were compared. A small pilot study was also conducted with TD children from 9-15 years of age. Table 7 summarizes the correct recognition of prosodic conditions by TD participants.

Accuracy across prosodic conditions generally shows a developmental trend, the most significant of which appears to be for angry prosody.

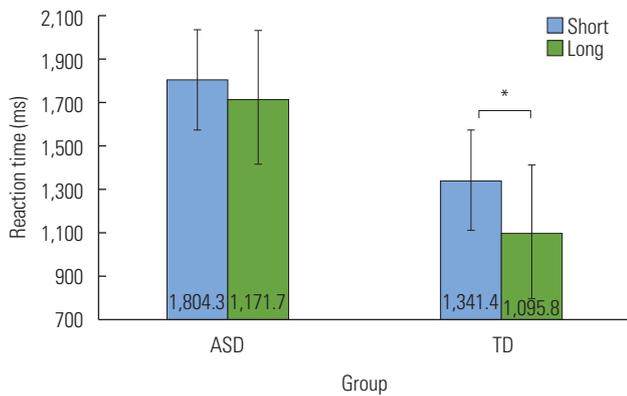


Figure 3. RT for short vs. Long sentences by group. * $p < 0.05$.

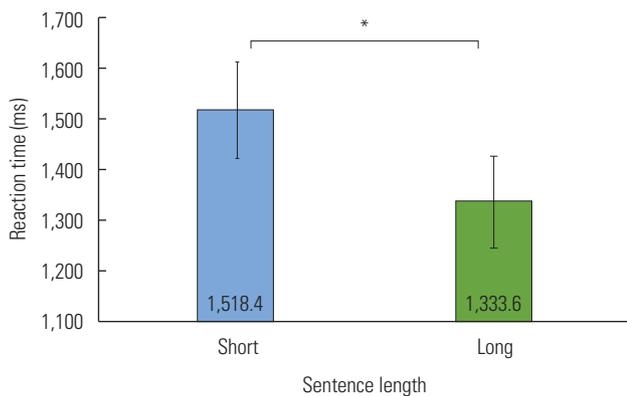


Figure 4. RT for All sentences by length. * $p < 0.05$.

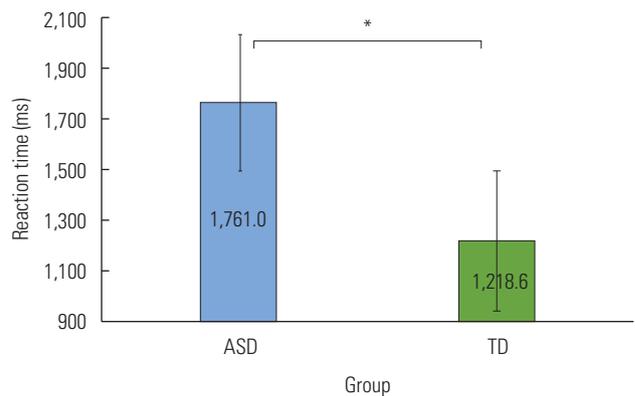


Figure 5. RT for All sentences by group. * $p < 0.05$.

Table 6. % Correct Recognition across Prosodic Conditions by Group

	ASD recognition (% correct)			TD recognition (% correct)		
	Angry	Neutral	Happy	Angry	Neutral	Happy
Angry prosody	30.2	41.9	27.9	34.2	47.8	17.9
Neutral prosody	32.3	36.6	31.1	25.6	65.1	9.4
Happy prosody	34.4	30.1	34.4	21.9	25.3	52.7

Table 7. % Correct Recognition of Prosodic Conditions by TD Participants across Age Clusters

	Recognition (% Correct) in Experiment	Recognition (% Correct) in Pilot		Recognition (% Correct) in Pretest
	6-8 y.o.; (n = 21)	9-12 y.o.; (n = 3)	13-15 y.o.; (n = 1)	Adults; (n = 10)
Angry	34.2	34.9	61.9	87.6
Neutral	65.1	85.7	81.0	94.3
Happy	52.7	79.4	61.9	93.3

Table 8. Chronological Age, Nonverbal Intelligence Age-Equivalence, and % Correct Recognition by Individual ASD Participant

	CA (yr;mth)	NVIQ AE (yr;mth)	Recognition (% Correct)			
			Overall	Angry	Neutral	Happy
ASD1	12;2	6;3	38.1	42.9	33.3	38.1
ASD2	12;10	4;0	39.7	33.3	42.9	42.9
ASD3	10;7	4;0	36.5	23.8	47.6	38.1
ASD4	12;6	6;0	31.7	38.1	38.1	19.0
ASD5	12;1	14;8	46.0	14.3	61.9	61.9
ASD6	11;3	4;10	31.7	38.1	28.6	28.6
ASD7	11;5	4;8	19.0	14.3	28.6	14.3
ASD8	11;9	3;11	30.2	19.0	38.1	33.3
ASD9	10;0	3;11	34.9	42.9	23.8	38.1
ASD10	10;11	4;10	41.3	52.4	38.1	33.3
ASD11	12;1	3;11	17.5	23.8	19.0	9.5
ASD12	9;10	3;11	34.9	23.8	38.1	42.9
ASD13	12;10	3;11	25.4	14.3	28.6	33.3

CA=chronological age; NVIA AE=nonverbal intelligence age-equivalence.

Nonverbal cognitive abilities and correct recognition of affective prosody

Each ASD participant's chronological age and nonverbal age-equivalence on the KBIT2 were examined together with their correct recognition of affective prosody to explore the relationship between these few factors (see Table 8).

At first glance, the data appears to trend towards a positive correlation between nonverbal intelligence age-equivalence on the KBIT2 and correct recognition of affective prosody. However, three participants (Participants 2, 3, and 12) had low nonverbal intelligence age-equivalence on the KBIT2, yet demonstrated the ability to consistently recognize neutral and/or happy prosody at levels that were above chance, suggesting that perhaps nonverbal cognitive abilities may not always predict poor ability to recognize affective prosody.

Regression analysis

A series of all-possible-subsets regression analyses were con-

Table 9. Significant Regression Models for Accuracy

Model 1
Dependent variable: Mean accuracy for neutral sentences Adjusted $R^2 = 0.64$, $F(2,10) = 11.626$, $p = 0.002$ Predictor Variables: <ul style="list-style-type: none"> ■ KBIT2, $B = 0.328$, $p = 0.013$ ■ SRS T-score, $B = 0.382$, $p = 0.033$
Model 2
Dependent variable: Mean accuracy for happy sentences Adjusted $R^2 = 0.22$, $F(1,11) = 4.336$, $p = 0.061^*$ Predictor Variables: <ul style="list-style-type: none"> ■ KBIT2, $B = 0.374$, $p = 0.061^*$

* p -value approaching significance.

ducted to identify the significant effects of behavioral score profiles on the dependent experimental variables of accuracy and reaction time. Specifically, we were interested in the KBIT2 and SRS total T-score as predictor variables. A decision was made to perform regressions only for the ASD group,

given the clinical relevance of this population. In addition, only the statistically significant variables from the repeated measures ANOVA were entered into the regression models as criterion variables (i.e., mean accuracy by prosodic condition and mean reaction time by length). One significant model and one model approaching significance emerged with respect to the accuracy data (see Table 9) while no significant models emerged with respect to the RT data.

DISCUSSION

The aim of the study was to investigate the ability of children with ASD, including those who are minimally verbal, to perceive affective prosody when presented with low-pass filtered sentences as stimuli. The inclusion of this subgroup of children with ASD who are minimally verbal represents a departure from a long tradition of research that has tended to focus on high-functioning individuals on the autism spectrum. As discussed, there is a dearth of research into the minimally verbal “neglected end of the spectrum” [18]. The present study incorporated multiple strategies aimed at supporting minimally verbal children with ASD to task completion. Low-pass filtered sentences were used as stimuli in order to eliminate the confounding factors in linguistic stimuli that would be associated with language abilities (e.g., knowledge of vocabulary, syntax) and to accommodate the heterogeneity within the autism spectrum. A structured training component was specifically designed to allow ASD participants to progress towards levels of greater independence through the task. Visual supports in the form of First/Then boards were also used with the ASD group to provide directions and scaffold comprehension of task expectations. The fact that there were some minimally verbal children with ASD who were able to comprehend testing instructions in order to complete a fairly abstract judgment task is exceedingly encouraging for translational purposes.

The goal of the present study to investigate the perception of affective prosody in children with ASD and TD peers is driven by the fundamental role that prosody plays in language acquisition. Given findings that there are some very young children with ASD who are less responsive to their mother’s voice [2], and show less motivation to attend to social interactions [44], the resulting impact on language acquisition is of grave concern. It may be that minimally verbal children with ASD possess deficits in perceiving affective prosody that hinders their ability to become competent users of spoken lan-

guage systems.

The main hypothesis that children with ASD will show a poorer ability to perceive affective prosody compared to TD peers was borne out in this study. The ASD group recognized prosodic conditions significantly less accurately than the TD group. This is perhaps not surprising, and is in line with the majority of past research that has shown that individuals with HFA/AS typically demonstrate deficits in recognition of affective prosody [8,19,20,27]. While some studies have found intact receptive abilities for affective prosody in individuals with HFA [22], including one that also used low-pass filtered speech with the HFA population [23], the inclusion of minimally verbal children with ASD in the present study yields different and interesting contrasts to the prior work on the higher end of the spectrum.

The ASD group also evidenced a different pattern in time to recognize affective prosody. The TD group required a significantly longer time to recognize the affective prosody in shorter sentences compared to longer sentences. However, the ASD group took more time on all sentences in general but did not differ significantly between short and long sentences. This finding suggests that TD children may need extra processing time to judge emotions in prosody when less verbal information is available; although there is also the possibility that the longer sentences allowed the TD participants to form a judgment while the sentence was still playing resulting in a shorter reaction time. In contrast, individuals with ASD spent more time overall processing all stimuli. While sentence length has not been a focus of research on affective prosody in the past, findings from the present study may motivate further investigation into the role of stimulus length in supporting successful decoding of emotions from the prosody of speech (e.g., as an instructional strategy).

In any perception study, it is often informative to analyze what and where errors were made that influenced accuracy. In the present study, both the ASD and TD groups struggled with recognizing angry prosody. Not only was angry prosody recognized at approximately chance level across both groups, it was also most often mistakenly categorized as neutral-sounding. While there may have been an inherent bias in the selection of neutral as a response, this finding is important because it shows that both ASD and TD participants show similar error patterns when it comes to angry prosody. It is an interesting contrast to the research cited above in which infants from around 5 months are able to discriminate between happy, angry, and sad emotional prosody in familiar contexts

[30]. It raises the question: why are the children in the present study unable to consistently recognize angry prosody? Some researchers have suggested that the ability to decode basic emotions from filtered speech appears to improve with age probably into early adulthood [33]. Preliminary data from the present study support the hypothesis that while affective prosody may be differentially processed in critical brain regions a few months after birth, the ability to perceive and decode emotions from (filtered) speech develops over time. It is likely that as a child gains more exposure to and experience with social interactions, he or she will grow to become more adept at identifying emotions from the prosody of speech.

A secondary hypothesis was that a negative relationship would exist between the ability to perceive affective prosody and the degree of social responsiveness deficits. One finding here suggests that the more deficits one has in social responsiveness, the better one will be at recognizing neutral prosody; contrary to the hypothesis. One reason for this counterintuitive finding may be an artifact of the SRS questionnaire (used to measure social responsiveness deficits) as administered in the present study. That is, one subset of the ASD group had questionnaires completed by parents while the other subset of the ASD group had questionnaires completed by teachers. Inherent differences in the relationship to the child, as well as the depth and quality of interaction spent with the child may have introduced a confound to the SRS scores.

A final hypothesis was that the ability to perceive affective prosody was not related to nonverbal cognitive abilities. The significant models generated from the series of all-possible-subsets regression analyses rejected this hypothesis. Nonverbal cognitive abilities was a significant predictor variable for mean accuracy scores for neutral and happy prosody – a higher nonverbal cognitive ability (as measured on KBIT2) was a reliable predictor that a child with ASD was better able to correctly recognize neutral and happy prosody. While it seems logical that nonverbal cognitive abilities will have a direct impact on one's ability to perceive and accurately recognize affective prosody, it is also the contention of these authors that low nonverbal cognitive skills do not preclude one from being able to accurately perceive affective prosody. At the very least, the current study provides some initial finding that suggests low nonverbal cognitive abilities may not necessarily hinder one's capacity to correctly recognize neutral and/or happy prosody.

One limitation of the present study is the small sample size, especially in the ASD group. There was also lack of experi-

mental control in terms of including both high- and low-functioning individuals with ASD as well as controlling for nonverbal cognitive abilities in order to better tease out differences in the perception of affective prosody. The two different administrations of the SRS questionnaire may also have skewed findings somewhat, especially with regards to properly identifying predictor variables for accuracy rate and reaction time. At the same time, a better definition of the minimally verbal child with ASD may also be helpful to properly identify and include this subgroup in future research. The inclusion of different prosodic conditions in future studies with minimally verbal children with ASD will shed further light on what is being differentially processed by this population of children.

In conclusion, the driving force behind the present study is to extend research into the subgroup of children with ASD who are minimally verbal. Through conscious decision and careful deliberation, an innovative and structured experimental design came to fruition in the present study. While the mean accuracy rates for the ASD group tended to approximate chance levels, individual differences were found, a necessary consequence of the inherent heterogeneity within the population. Most importantly, however, this is one of a few studies that have included minimally verbal children with ASD in the sample. The present study has shown that it is possible for minimally verbal children with ASD to successfully participate in experimental research using judgment tasks when provided with appropriate training and scaffolding to task expectations. To this end, it remains our hope that more can and will be done to further our understanding of this long neglected other end of the autism spectrum.

REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author. 2013.
2. Tager-Flusberg H, Paul R, Lord C. Language and communication in autism. In Volkmar F, Paul R, Klin A, Cohen D. (Eds), Handbook of autism and pervasive developmental disorders. Hoboken, NJ: John Wiley & Sons, Inc. 2005. p. 335-364
3. Frith U. The neurocognitive basis of autism. *Trends in Cognitive Sciences*. 1997;1:73-77.
4. Tager-Flusberg H. Understanding the language and communicative impairments in autism. *International Review of Research in Mental Retardation*. 2001;23:185-205.
5. Young EC, Diehl JJ, Morris D, Hyman SL, Bennetto L. The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, Speech, and Hearing Services in Schools*. 2005;36:62-72.
6. Diehl JJ, Friedberg C, Paul R, Snedeker J. The use of prosody dur-

- ing syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*. 2015; 27(3):867-884.
7. Pierrehumbert J, Hirschberg J. The meaning of intonational contours in the interpretation of discourse. In Cohen P, Morgan J, Pollack M. (Eds). *Intentions in Communication*. Cambridge, MA: MIT Press. 1990. pp. 271-311
 8. Järvinen-Pasley A, Peppé S, King-Smith G, Heaton P. The relationship between form and function level receptive prosodic abilities in autism. *Journal of Autism and Developmental Disorders*. 2008; 38:1328-1340.
 9. 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, and Hearing Research*. 2007; 50:1015-1028.
 10. Lord C, Rutter M, DiLavore PC, Risi S. *Autism Diagnostic Observation Schedule*. Los Angeles: Western Psychological Services. 1999.
 11. Paul R, Shriberg LD, McSweeney 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;35:861-869.
 12. DePape AMR, Chen A, Hall GB, Trainor LJ. Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology*. 2012;3:72.
 13. Cooper R, Aslin R. Preference for infant-directed speech in the first month after birth. *Child Development*. 1990;61:1584-1595.
 14. Cohen D, Cassel RS, Saint-Georges C, Mahdhaoui A, Laznik MC, Apicella F, et al. Do parentese prosody and fathers' involvement in interacting facilitate social interaction in infants who later develop autism? *PLoS ONE*. 2013;8(5):e61402.
 15. Kuhl P. Early language acquisition: Cracking the speech code. *Neuroscience*. 2004;5:831-843.
 16. Klin A. Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*. 1991;21:29-42.
 17. Koegel R, Shirotova L, Koegel L. Brief report: Using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism & Developmental Disorders*. 2009;39:1587-1592.
 18. Tager-Flusberg H, Kasari C. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*. 2013;6:468-478.
 19. Koning C, Magill-Evans J. Social and language skills in adolescent boys with Asperger syndrome. *Autism: Journal of Research and Practice*. 2001;5:23-36.
 20. Rutherford M, Baron-Cohen S, Wheelwright S. Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*. 2002;32:189-194.
 21. McCann J, Peppé S, Gibbon F, O'Hare A, Rutherford M. Prosody and its relationship to language in school-age children with high-functioning autism. *International Journal of Language and Communication Disorders*. 2007;42:682-702.
 22. Loveland K, Tunali-Kotoski B, Chen Y, Ortegon J, Pearson D, Brelsford K, Gibbs M. Emotion recognition in autism: Verbal and nonverbal information. *Development and Psychopathology*. 1997;9:579-593.
 23. Grossman R, Bemis R, Skwerer D, Tager-Flusberg H. Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*. 2010;53:778-793.
 24. Chevallier C, Noveck I, Happé F, Wilson D. What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*. 2011;49:507-517.
 25. Grossman RB, Tager-Flusberg H. "Who Said That?" Matching of low-and high-intensity emotional prosody to facial expressions by adolescents with ASD. *Journal of Autism and Developmental Disorders*. 2012;42:2546-2557.
 26. Fujiki, M., Spackman, M.P., Brinton, B., & Illig, T. Ability of children with language impairment to understand emotion conveyed by prosody in a narrative passage. *International Journal of Language & Communication Disorders*. 2008;43:330-345.
 27. Lindner J, Rosén L. Decoding of emotion through facial expression, prosody, and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*. 2006;36:769-777.
 28. Kjelgaard MM, Tager-Flusberg H. The perception of the relationship between affective prosody and the emotional content in utterances in children with autism spectrum disorders. *Perspectives on Language Learning and Education*. 2013;20(1):20-32.
 29. Stewart ME, McAdam C, Ota M, Peppé S, Cleland J. Emotional recognition in autism spectrum conditions from voices and faces. *Autism*. 2013;17(1):6-14.
 30. Grossmann T, Oberecker R, Koch S, Friederici A. The developmental origins of voice processing in the human brain. *Neuron*. 2010;65:852-858.
 31. Pell M. Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America*. 2001;109:1668-1680.
 32. Banse R, Scherer K. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*. 1996;70:614-636.
 33. Juslin P, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*. 2003;129:770-814.
 34. Thompson W, Balkwill L. Decoding speech prosody in five languages. *Semiotica*. 2006;158:407-424.
 35. Pell M, Paulmann S, Dara C, Alasser A, Kotz S. Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*. 2009;37:417-435.
 36. Jaywant A, Pell M. Categorical processing of negative emotions from speech prosody. *Speech Communication*. 2011;54:1-10.
 37. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. 2011. Version 5.2.26. Available from <http://www.praat.org/>
 38. Mitchell R, Elliott R, Barry M, Cruttenden A, Woodruff P. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*. 2003;41:1410-1421.
 39. Laukka P. Categorical perception of vocal emotion expressions. *Emotion*. 2005;5:277-295.
 40. Kaufman A, Kaufman N. *Kaufman Brief Intelligence Test, Second Edition*. Bloomington, MN: Pearson, Inc. 2004.
 41. Constantino J, Gruber C. *Social Responsiveness Scale (SRS)*. Los Angeles: Western Psychological Services. 2005.
 42. Oudeyer PY. The production and recognition of emotions in

- speech: Features and algorithms. *International Journal of Human-Computer Studies*. 2003;59:157-183.
43. Haskell T. FLXLab [Computer program]. 2010. Version 2.5alpha. Available from <http://xlab.sourceforge.net>
 44. Carter A, Davis N, Klin A, Volkmar F. Social development in autism. In Volkmar F, Paul R, Klin A, Cohen D. (Eds), *Handbook of autism and pervasive developmental disorders*. 2005. p. 312-334. Hoboken, NJ: John Wiley & Sons, Inc.