Article

FIRST LANGUAGE

First Language
2020, Vol. 40(1) 41–63

DOI: 10.1177/0142723719885270
journals.sagepub.com/home/fla
SAGE

A deficit in using prosodic cues to understand communicative intentions by children with autism spectrum disorders: An eye-tracking study

Peng Zhou
Tsinghua University, China

Weiyi Ma
University of Arkansas, USA

Likan Zhan
Beijing Language and Culture University, China

Abstract

The present study investigated whether Mandarin-speaking preschool children with autism spectrum disorders (ASD) were able to use prosodic cues to understand others' communicative intentions. Using the visual world eye-tracking paradigm, the study found that unlike typically developing (TD) 4-year-olds, both 4-year-olds with ASD and 5-year-olds with ASD exhibited an eye gaze pattern that reflected their inability to use prosodic cues to infer the intended meaning of the speaker. Their performance was relatively independent of their verbal IQ and mean length of utterance. In addition, the findings also show that there was no development in this ability from 4 years of age to 5 years of age. The findings indicate that Mandarin-speaking preschool children with ASD exhibit a deficit in using prosodic cues to understand the communicative intentions of the speaker, and this ability might be inherently impaired in ASD.

Keywords

Autism spectrum disorder, communicative intentions, eye movements, Mandarin Chinese, preschool children, prosody

Corresponding author:
Peng Zhou, Department of Foreign Languages and Literatures, Child Cognition Lab, Tsinghua University, Beijing 100084, China.
Email: zhoupeng1892@mail.tsinghua.edu.cn

According to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5; American Psychiatric Association, 2013), autism spectrum disorder (ASD) is diagnosed on the basis of two symptom clusters: (a) persistent deficits in social communication and interaction, and (b) restricted and repetitive patterns of behaviour, interests, or activities. Compared to DSM-IV-TR (American Psychiatric Association, 2000), DSM-5 has introduced significant changes in the diagnostic criteria of ASD, among which one is that impairments in language abilities that are not employed in social communication have been removed as a core symptom. However, this does not preclude the centrality of language deficits within the ASD phenotype, because a delay in language onset or an atypical language profile is one of the primary early triggers for parents seeking referrals for their children who are eventually diagnosed with ASD (Robins et al., 2014). In addition, some components of language abilities are closely related to the social and communicative domain. Thus, investigations of these components in ASD will help us better understand the sources of their deficits in social communication and interaction. In the present study, we focus on a particular domain of language that is intimately connected with the communicative functions and intentions of the speaker, namely, prosody; and we investigate whether preschool children[1] with ASD use prosodic information to understand the communicative intentions of the speaker, which we discuss below.

Prosody refers to the suprasegmental characteristics of speech deriving from variations in duration, amplitude and fundamental frequency of speech sounds. The prosodic patterns of spoken sentences often signal a speaker's emotional and mental states and instigate certain kinds of pragmatic functions (Ladd, 1996; Pierrehumbert & Hirschberg, 1990; Prieto, 2015; Wilson & Wharton, 2006). The present study examines one of the core pragmatic functions of prosody in which prosody signals the communicative intentions of the speaker. Understanding a speaker's prosody is directly relevant to the ability to infer the speaker's communicative intentions (Verschueren, 1999). In addition, the present study focuses on a less studied language, Mandarin Chinese. Mandarin is ideally suited for investigating this issue, because the same sequences of words can be used to express either a question or a statement depending on prosody. Consider sentence (1), for example. Negative sentences containing a *wh*-phrase like (1) are ambiguous in Mandarin. Sentence (1) can be used to ask a question, as in (1a): 'What animal did Xiaoming not pat?' Alternatively, it can be used to make a statement, as in (1b): 'Xiaoming didn't pat any animal'. Xiaoming is a typical boy character in Mandarin. Note that the two readings correspond to two speech acts, i.e. asking a question versus making a statement, which are the two most basic speech acts and are used frequently in daily conversation.

(1) Xiaoming meiyou mo shenme dongwu

Xiaoming not pat what animal

(1a) Question reading: What animal did Xiaoming not pat?

(1b) Statement reading: Xiaoming didn't pat any animal.

What is more interesting about Mandarin is that prosodic cues can be used to distinguish between the two speech acts. More specifically, a rising intonation on the *wh*-phrase *shenme dongwu* 'what animal' indicates a question, whereas a level intonation (lack of rising intonation) on the same *wh*-phrase indicates a statement. Note that in English prosody can also be used to distinguish a question from a statement. For instance, English-speakers can ask a yes-no question by applying the question intonation to a statement: *Mary bought an apple?* However, English and Mandarin differ in that an English sentence like *Mary bought an apple* by itself is not ambiguous but changing the prosody can convert it into a question. By contrast, a Mandarin sentence like (1) by itself is ambiguous (i.e. the question and statement readings are equally accessible when the sentence is presented visually), and prosody can be used reliably to disambiguate between the two readings. Mandarin-speakers frequently use sentences like (1) either to ask a *wh*- question or to make a statement, by employing different intonation patterns. The present study takes advantage of this special property of Mandarin to investigate whether preschool children with ASD are able to use prosodic cues to infer the speaker's intended meaning. Before presenting the study, we briefly review previous research on prosody in ASD.

There is a growing body of research exploring whether children with ASD have deficits in the use of prosody during language comprehension (see Diehl, Friedberg, Paul, & Snedeker, 2015 for review). Deficits in both perceiving and producing prosody have been observed in children with ASD (Arciuli & Bailey, 2019; DeMyer et al., 1973; Diehl & Paul, 2013; Kanner, 1943, 1971; Lord & Paul, 1997; McAlpine, Plexico, Plumb, & Cleary, 2014; McCann & Peppé, 2003; Paul, Augustyn, Klin, & Volkmar, 2005; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2006, 2007; Rutter & Lockyer, 1967; Shriberg et al., 2001; Su, Jin, Wan, Zhang, & Su, 2014; Tager-Flusberg, 1999; Terzi, Marinis, & Francis, 2016). For example, Peppé, McCann, Gibbon, O'Hare, and Rutherford (2007) found that high-functioning English-speaking children with ASD (aged between 6;1 and 13;6) were not able to use prosody to distinguish questions from statements. As discussed, English-speakers can ask a yes-no question by applying the question intonation to a statement. Peppé et al. found that when a sentence was presented with a question intonation (rising intonation), the participants tended to wrongly judge it as a statement. In addition, prior studies have shown that even high-functioning individuals with ASD exhibit deficits in inferring the speaker's emotion using vocal cues (Golan, Baron-Cohen, Hill, & Rutherford, 2007; Järvinen-Pasley, Peppé, King-Smith, & Heaton, 2008; Kleinman, Marciano, & Ault, 2001; Peppé et al., 2006, 2007; Rutherford, Baron-Cohen, & Wheelwright, 2002).

However, challenges in prosody have not been universally reported. For instance, Baltaxe and Simmons (1985) found that only four out of the seven adolescents with ASD (between 9 and 17 years of age) in their study had significant problems in their suprasegmental characteristics of speech. Paul, Shriberg, et al. (2005) reported that atypical prosodic patterns were observed in only 14 out of the 30 adolescents and adults with ASD (between 10 and 49 years of age) in their study. In addition, in contrast to Peppé et al. (2007), Paul, Augustyn, et al. (2005) found that high-functioning adolescents with ASD (between 14 and 21 years of age) can use prosody to distinguish questions from statements, e.g. this sample showed an understanding that the speaker is asking a yes-no question when he/she says the sentence *John ate an apple* with a rising intonation and the speaker is making a statement when the same sentence is uttered using a level intonation (lack of rising intonation).

To take stock, extant research that investigated the ability of children with ASD to comprehend the communicative intentions using prosodic cues has yielded mixed results. This inconsistency of results might be partly due to the different methodologies used in these studies. In addition, the age variations of the participants in different studies make cross-study comparisons difficult (Chevallier, Noveck, Happé, & Wilson, 2009; McCann & Peppé, 2003). Some studies tested young children and some tested adolescents and adults. It is, therefore, possible that the differences between children and adolescents observed in different studies might reflect the development of their ability to infer the speaker's communicative intentions using prosodic information. The ability to comprehend the communicative intentions of others is crucial to being able to function effectively in the social world (Akhtar & Martinez-Sussmann, 2007; Hala, 1997; Moore & Corkum, 1994). It is, therefore, important to tap into the development of this ability in children with ASD. Investigations of the developmental trajectory of this ability will inform us about the nature of the observed impairment in previous research, e.g. whether this ability is inherently impaired, or it is simply a delayed scenario relative to typical development.

In fact, the study by Diehl et al. (2015) seems to suggest a developmental trend in the use of prosodic cues by children with ASD. Diehl et al. used eye tracking as a tool to examine the role of prosody during online sentence comprehension in English-speaking children with ASD, which is directly relevant to the experimental method used in our study. So, we briefly review the study before introducing the advantages of using eye tracking to investigate the use of prosody by children with ASD. In their study, two groups of ASD participants (i.e. children between 8 and 12 years of age vs adolescents between 12 and 18 years of age) were investigated to see whether they were able to use prosodic cues to constrain their interpretation of sentences like *You can feel the frog with the feather*. The sentence is ambiguous between an instrument reading (i.e. you can feel the frog using the feather) and a modifier reading (you can feel the frog that has a feather), but a prosodic boundary/break (indicated by *[]*) can be used to disambiguate the two interpretations (*you can feel the frog [] with the feather* vs *you can feel [] the frog with the feather*). The eye gaze patterns reflected that both ASD groups were sensitive to the initial prosodic cues that were presented. However, when a different cue was presented in subsequent trials, unlike the older group (between 12 and 18 years of age), the younger group (between 8 and 12 years of age) were more likely to respond in a manner consistent with the initial prosodic cue rather than the new one, indicating that they had trouble shifting their interpretation as the prosodic cue changed. The contrast between the younger group and the older group seems to suggest a potential developmental trend in the use of prosodic cues in the ASD population, though this development seems to occur at a rather late developmental stage, i.e. at around adolescence. The study also revealed that eye tracking is a sensitive measure of how sentence comprehension proceeds in real time by children with ASD.

In the present study, we examined much younger children than in previous research. More specifically, the present study aimed to investigate the ability of preschool children with ASD to comprehend the communicative intentions using prosodic cues. To chart the developmental trajectory of this ability, we compared the performance of children with ASD in two different age groups (i.e. 4 and 5 years of age) relative to TD children. To our knowledge, very limited work has examined prosody in children with ASD as young as 4 years of age.

Like Diehl et al. (2015), the present study used an on-line measure, eye tracking, to investigate the use of prosody by children with ASD. More specifically, we used the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In the visual world paradigm participants' visual exploration of a scene is monitored while they hear spoken language. Under most natural circumstances, one can assume that where the participant is looking reflects what he or she views as relevant to the task and to the ongoing comprehension process (Trueswell, 2008; Trueswell & Gleitman, 2007). The visual world paradigm is based on a linking assumption that when participants are simultaneously presented with spoken language while viewing a visual scene, their eye movements are very closely synchronized to the referential processing of the concurrent linguistic input (Tanenhaus et al., 1995).

Note that much prior research that assessed the comprehension abilities of children with ASD used off-line tasks. These off-line tasks often require high response demands or interactions with the experimenters, which might pose particular difficulties for children with ASD, because children with ASD often exhibit various kinds of challenging behaviours or symptoms which might interact with the high task and communication demands, and thus frequently mask the comprehension abilities of these children (Kasari, Brady, Lord, & Tager-Flusberg, 2013; Plesa-Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016; Zhou, Zhan, & Ma, 2019). Compared with off-line tasks, the visual world paradigm has several advantages. First, it is sensitive to the time course of sentence comprehension. Second, it minimizes the task and communication demands involved, thus is a better-suited method that can be used with children exhibiting challenging behavioural features. Third, it simply records eye movements as automatic responses to linguistic input without asking participants to provide conscious judgements about the input, significantly reducing the computational burden of the participants. These features of the paradigm make it ideal for studying younger children and minimally verbal children with ASD who often exhibit various kinds of challenging behaviours and symptoms.

Recent studies have begun to explore the potential of using eye tracking to study language comprehension in children and adolescents with ASD, and have proven that it is a sensitive testing paradigm to demonstrate the comprehension abilities of children with ASD (Bavin, Kidd, Prendergast, & Baker, 2016; Brock, Norbury, Einav, & Nation, 2008; Diehl et al., 2015; Hahn, Snedeker, & Rabagliati, 2015; Kelly, Walker, & Norbury, 2013; Norbury, 2017; Zhou, Ma, Zhan, & Ma, 2018). The eye gaze patterns children exhibit in a visual world eye-tracking study should directly reflect their (in)abilities to use prosodic cues during online sentence comprehension.

The present study used the visual world paradigm to investigate the comprehension of communicative intentions using prosodic cues by 4- and 5-year-old Mandarin-speaking children with ASD. In the task, they were presented with spoken sentences as in (1) while viewing the corresponding visual images, and their eye movements were recorded. If 4- and 5-year-old Mandarin-speaking children with ASD exhibited a deficit in using prosodic cues to understand the communicative intentions, then they might fail to exhibit an eye gaze pattern that is associated with the successful use of prosody to infer the speaker's intended meaning.

Methods

Participants

ASD group. Forty-six high-functioning Mandarin-speaking children with ASD[2] were recruited from the Rehabilitation and Education Centre for Children with ASD affiliated with the Peking University Sixth Hospital, Beijing, China. Their diagnoses were made by paediatric neurologists at the Peking University Sixth Hospital using both DSM-IV-TR (American Psychiatric Association, 2000) and DSM-5 (American Psychiatric Association, 2013). To further confirm the participants' diagnoses, each participant was independently evaluated by our research team using the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). Ten of the 46 participants were not included in the final analyses, because two did not meet the autism cut-off on the ADOS, four did not complete the study and four dropped out because we were unable to calibrate them on the eye-tracker. The other 36 participants successfully completed the task and were included in the final analyses, and they were divided into two age groups: 18 4-year-olds (mean age 4;4, range 4;1–4;6) and 18 5-year-olds (mean age 5;6, range 5;1–5;8). A sample size of 18 participants per group is fairly typical in language development studies using continuous data (see e.g. Bergmann et al., 2018, for a review).

Participants' verbal IQ was assessed using the Chinese-Wechsler Young Children Scale of Intelligence (C-WYCSI) – a standardized IQ test designed for Mandarin-speaking children between the ages of 4 and 6.5 (Gong & Dai, 1992). The test showed that all the participants had a verbal IQ score above 80. In addition, we measured the mean length of utterance (MLU) of each participant.[3] We recorded 100 utterances for each participant. The participants' utterances were recorded from their interactions with their teachers either in the classroom or individual training sessions. In Mandarin Chinese, MLU was often calculated by dividing the total number of words by the number of utterances in each speech sample. In the present study, we followed the common practice of calculating Mandarin MLU by dividing the total number of words in each participant's utterances by the total 100 utterances that were recorded for each participant. MLU is often used as an indicator of sentence complexity levels.

TD comparison group. Twenty-seven TD 4-year-olds (mean age 4;4, range 4;1–4;6) were recruited from the Beijing Taolifangyuan Kindergarten, Beijing, China. Previous research has shown that TD Mandarin-speaking 4-year-olds are already able to use intonational cues to distinguish questions from statements (Zhou, Crain, & Zhan, 2012). According to the parental reports, none of the participants had a history of clinical diagnosis, language disorder, or special educational services. Participants were also screened for potential undiagnosed ASD by an expert clinician in our research team using DSM-5, and none of them met the DSM-5 criteria for ASD. Participants' verbal IQ was also assessed using the C-WYCSI (Gong & Dai, 1992). All the participants had a verbal IQ score above 80. Two of the 27 participants were not included in the final analyses, because we were unable to calibrate them on the eye-tracker. The other 25 participants all successfully completed the task and were included in the final analyses. Again, their MLU was calculated by dividing the total

Table 1. Verbal IQ scores and MLU of each participant group (SD in parentheses).

Group	Number	Verbal IQ	MLU
ASD group			
4-year-olds	18	95.56 (9.46)	4.77 (1.25)
5-year-olds	18	101.17 (9.53)	5.92 (1.26)
TD group			
4-year-olds	25	101.28 (9.02)	5.89 (1.22)

number of words in each participant's utterances by the total 100 utterances that were recorded for each participant.

Groups and matching. Table 1 shows the descriptive characteristics of the two ASD groups and the TD group divided by age. The 5-year-olds with ASD were matched with the TD 4-year-olds on both verbal IQ ($t(41) = 0.54$, $p = 0.26$, Cohen's $d = 0.01$) and MLU ($t(41) = 0.85$, $p = 0.65$, Cohen's $d = 0.02$). Compared with their age-matched TD peers, the 4-year-olds with ASD had significantly lower verbal IQ score ($t(41) = 4.86$, $p < .01$, Cohen's $d = 0.62$) and MLU level ($t(41) = 13.76$, $p < .001$, Cohen's $d = 0.91$). Participants in both the ASD and TD groups were native speakers of Mandarin, and they had no reported history of hearing or visual impairment.

The study was approved by the Ethics Committee of the School of Medicine, Tsinghua University, 20170018. Written informed consent has been obtained from each participant's parents.

Materials and design

The present study was modelled closely on the task by Zhou et al. (2012) and used similar visual stimuli as in Zhou et al. (2012). A total of 18 target items were constructed, each comprising a visual stimulus and two spoken sentences, one with a rising intonation on the *wh*-phrase (hereafter Question Prosody) and one with a level intonation on the same *wh*-phrase (hereafter Statement Prosody). The picture stimuli were about two characters (either the boy character Xiaoming or the girl character Xiaohong). The two characters had stereotypical boy and girl appearances. In each picture, there were five objects with three objects belonging to one category and two belonging to a different category, and the character always chose one object from the category that had three objects. To control for potential preferences for looking at a particular displayed object, the gender and the position of the character were counterbalanced across trials. All the target sentences had the same structure: Subject noun phrase (Subject NP) + Negation + Verb + *Wh*-word + Object noun phrase (Object NP). The subject NP was either the boy character's name *Xiaoming* or the girl character's name *Xiaohong*. The negation marker was always *meiyou* 'not'. The verb was always monosyllabic and indicated an action. The *wh*-word was always *shenme* 'what'. The object NP was always disyllabic.

The spoken sentences were produced by a female native speaker of Beijing Mandarin. When constructing the target sentences, the elements prior to the *wh*-word (i.e. the

Table 2. Duration and intensity of each element of the target sentences in the question prosody and in the statement prosody.

Element	Duration	Intensity
	Question/Statement	Question/Statement
Subject NP	820 ms /820 ms	65.50 dB/65.50 dB
Negation	500 ms/500 ms	66.20 dB/66.20 dB
Verb	400 ms/400 ms	66.50 dB/66.50 dB
Wh-word	750 ms/650 ms	65.50 dB/70.50 dB
Object NP	920 ms/920 ms	64.20 dB/64.20 dB

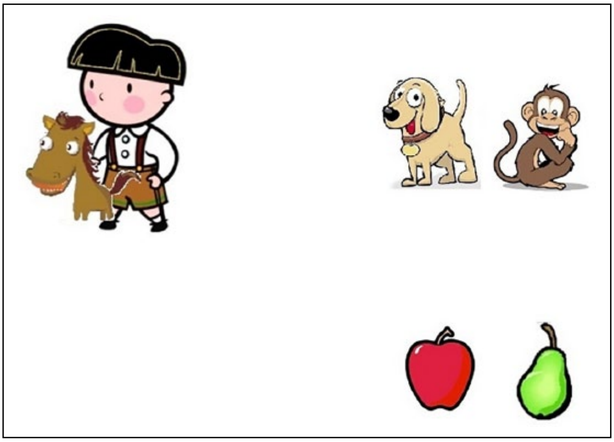

Figure 1. An example test image in the study; the image has been adapted from Zhou et al. (2012).

subject NP, the negation marker and the verb) and the object NP following the *wh*-word were made of equal duration and intensity (see Table 2). A distinction in intonation (Rising versus Level) leads to a difference in duration and intensity between the *wh*-words in the question prosody versus in the statement prosody (see Table 2). The duration and intensity analyses in Table 2 confirm that the difference between the two prosodic conditions lies in the different acoustic features of the *wh*-words. All the sentences with question prosody were 3390 ms long, and those with statement prosody were 3290 ms long.

A typical visual stimulus is given in Figure 1. As indicated in the figure, there were three animals (a dog, a monkey and a horse) and two fruits (an apple and a pear), and the boy character Xiaoming patted one of three animals – the horse. For this visual image, two prosodic versions of the target sentence were recorded, one with question prosody and one with statement prosody. The question prosody indicates a question, ‘what animal did Xiaoming not pat?’ as in (1a), and the statement prosody indicates a statement, ‘Xiaoming didn’t pat any animal’ as in (1b).

The 18 target items were divided into two experimental lists with each participant seeing each visual stimulus but listening to only one of the two target sentences recorded for that visual image. Target sentences with question prosody and those with statement prosody were counterbalanced across the two lists with nine in each prosodic condition for each list. In addition, 16 filler items were added to each experimental list. Each filler item consisted of a visual image and a spoken sentence. The visual images were similar to those used in the target items. The spoken sentence in each item was either an unambiguous *wh*-question (8 questions) or a statement with negation (8 statements). Two examples are given in (2) and (3), where (2) is a filler *wh*-question, and (3) is a filler negative statement. Half of the statements were true descriptions of the corresponding visual images, and half were false descriptions of the corresponding images. The filler items were included to verify whether the participants could understand simple *wh*-questions and negative sentences, since the target sentences involved both negation and *wh*-words.

(2) Xiaohong mai-le shenme wanju?
Xiaohong buy-ASP what toy
'What toy did Xiaohong buy?'

(3) Xiaoming mei chi shuiguo.
Xiaoming not eat fruit
'Xiaoming didn't eat fruit.'

In each experimental list, the 18 target and 16 filler items were arranged in random order. Participants in each group were randomly assigned to one of the two lists: nine 5-year-olds with ASD, nine 4-year-olds with ASD and 12 TD 4-year-olds were assigned to List 1; nine 5-year-olds with ASD, nine 4-year-olds with ASD and 13 TD 4-year-olds were assigned to List 2.

Procedure

Both the children with ASD and the TD children were tested using the visual world paradigm (Tanenhaus et al., 1995). The children with ASD were tested individually in a quiet room at the Rehabilitation and Education Centre for Children with ASD affiliated with the Peking University Sixth Hospital. The TD children were tested individually in a quiet room at the Beijing Taolifangyuan Kindergarten. In the task, they were presented with a spoken sentence as in (1) while viewing a visual scene as in Figure 1. Their eye movements were recorded using an EyeLink 1000 plus eye-tracker. The EyeLink 1000 plus allows remote eye tracking, without a head support. It provides information about the participant's point of gaze at a sampling rate of 500 Hz, and it has accuracy of 0.5 degrees of visual angle. The visual stimuli were displayed on the monitor. Spoken sentences were presented to the participants through the PC computer connected to two external speakers. The distance between the participants' eyes and the monitor was about 60 cm. The spoken sentence started 500 ms after the appearance of the visual stimulus. The participants were simply told to listen to the sentences while looking at the visual images, and they were not asked to make any conscious judgements about the spoken sentences. We measured their eye movements that arose as automatic responses to the linguistic input. Participants' eye movements were recorded for 4000 ms from the onset of the spoken sentence.

Data processing

In the final analysis, we only included the eye movement data of the participants whose eye gaze patterns reflected their correct responses to at least 14 out of the 16 filler trials (87.50%). All the participants correctly responded to the filler trials above 87.50% of the time, and thus their data were all included in the final analysis. To analyse the eye movement data, we first categorically partitioned the data from the onset of the *wh*-word into nine temporal bins, each with a duration of 300 ms. Among the nine bins, two were located prior to the onset of the *wh*-word, thus providing a baseline for the comparison, and the remaining seven bins were located after the onset of the *wh*-word. We then divided each visual image into three areas of interest: the question-compatible area, the statement-compatible area and the irrelevant area. On the example trial (Figure 1), the question-compatible area contained Xiaoming and the horse, the statement-compatible area consisted of the dog and the monkey, and the irrelevant area comprised the apple and the pear. The proportion of fixations on a particular area in a specific temporal bin was treated as the dependent variable. For example, if we recorded five fixation points in a temporal bin, with one fixation point located in a specific area, then the proportion of fixations on that area was 1/5. The two critical areas were the question-compatible and the statement-compatible areas in the two prosodic conditions.

Predictions

If the participants were sensitive to intonational cues and were able to use the cues to distinguish between the two speech acts, they would be expected to look more at the question-compatible area in the visual image when listening to the target sentences with question prosody than when listening to those with statement prosody. By contrast, they should fixate more on the statement-compatible area when hearing sentences with statement prosody than when hearing those with question prosody. These differences between the two prosodic conditions should occur after the onset of the *wh*-word, because the different prosodic features (duration and intensity) of the two conditions occurred from the onset of the *wh*-word, as indicated in Table 2. On the example trial, hearing sentence (1) with question prosody should trigger more fixations on the area containing the dog and the monkey in Figure 1 (question-compatible area), because the sentence asked a question, 'what animal did Xiaoming not pat?' In contrast, hearing the same sentence with statement prosody should trigger more fixations on the area containing Xiaoming and the horse (statement-compatible area), because it made a statement 'Xiaoming didn't pat any animal', which was a false description of Figure 1. The effects should occur after hearing the *wh*-word *shenme* 'what'.

Results

To provide an overview of the eye movement data, the results are first presented in the form of descriptive graphs followed by more detailed statistical analyses. Figures 2–4 summarize the average fixation proportions in the two critical areas across the two prosodic conditions by the TD 4-year-olds, the 4-year-olds with ASD and the 5-year-olds with ASD, respectively. '0' indicates the onset of the critical word – the *wh*-word. As shown in Figure 2, the TD 4-year-olds showed increased looks to the question-compatible area after hearing sentences with question prosody and they started to launch fewer fixations to this area after hearing sentences with statement prosody. By contrast, an opposite pattern was observed in the statement-compatible area. There was a higher proportion of fixations in this area for sentences with statement prosody than for those with question prosody. The observed effects occurred after the onset of the *wh*-word. The two ASD groups, however, exhibited a different eye gaze pattern than the TD group, as

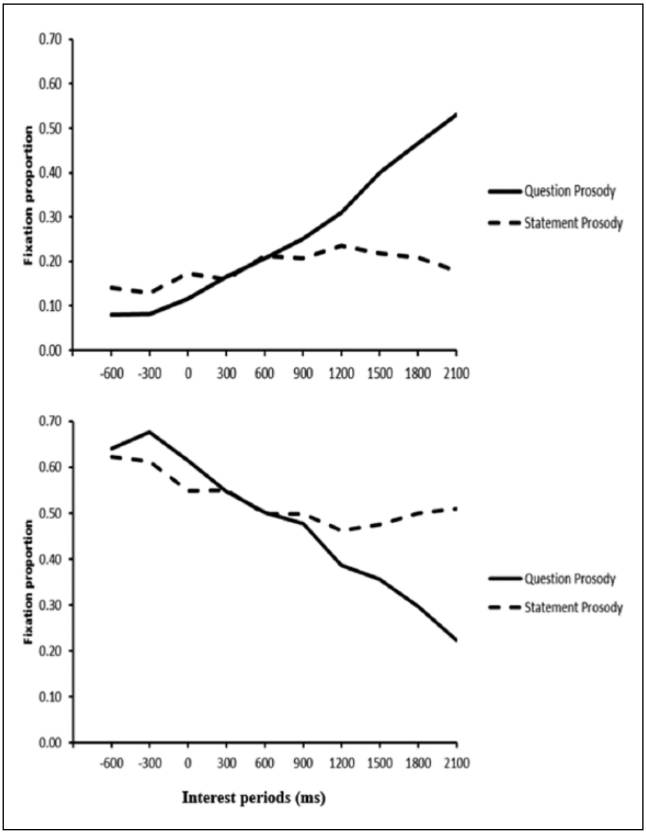

Figure 2. Average fixation proportions in the question-compatible area (upper panel) and in the statement-compatible area (lower panel) across the two prosodic conditions by the TD 4-year-olds; '0' indicates the onset of the *wh*-word.

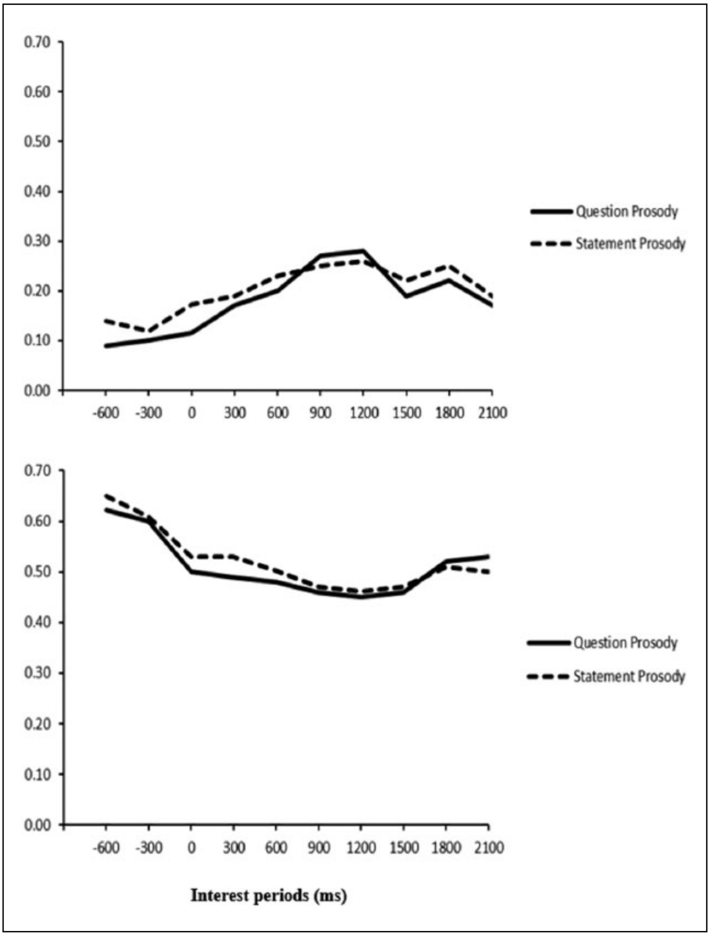

Figure 3. Average fixation proportions in the question-compatible area (upper panel) and in the statement-compatible area (lower panel) across the two prosodic conditions by the ASD 4-year-olds; '0' indicates the onset of the *wh*-word.

illustrated in Figures 3 and 4.[4] For both the 4-year-olds with ASD and the 5-year-olds with ASD, hearing sentences with question prosody did not trigger more fixations to the question-compatible area than hearing those with statement prosody. Both groups exhibited more looks at the statement-compatible area regardless of which prosodic version of the target sentence was presented to them.

To assess the fixation patterns statistically, we transformed the fixation proportions using the empirical logit formula (Barr, 2008): probability = $\ln((y+0.5)/(n-y+0.5))$, where y is the number of fixations on the areas of interest during a particular temporal bin; n is the total number of fixations in that temporal bin. We then fitted a series of linear mixed-effects models to the transformed data for each participant group. In the full model, the fixed effects included the temporal bin, the prosodic condition and

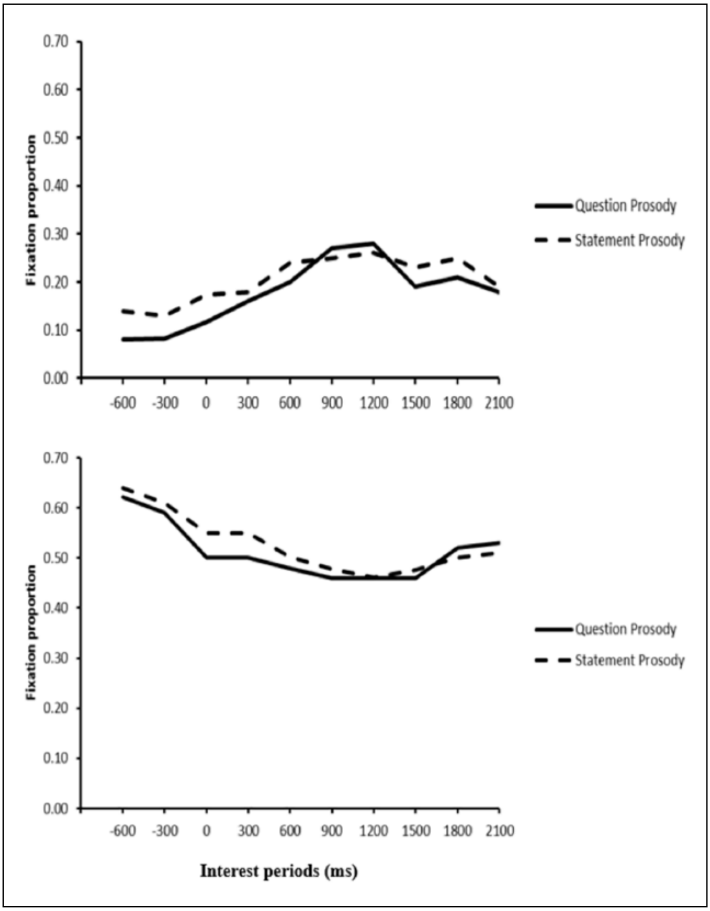

Figure 4. Average fixation proportions in the question-compatible area (upper panel) and in the statement-compatible area (lower panel) across the two prosodic conditions by the ASD 5-year-olds; '0' indicates the onset of the *wh*-word.

their interactions; the random effects included items and participants, where both their intercepts and slopes were allowed to vary among all the fixed effects (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). The full model's complexity was then reduced to see whether the reduced model could explain the same variance as the full model (Bates, Kliegl, Vasishth, & Baayen, 2015). If it could, we then accepted the simplified model. The final model we used for each group can be found in the footnote of Tables 3–5 respectively. Analyses were conducted on the raw data with no aggregation. When conducting the analyses, the temporal bins were rescaled and grand-mean centred, to avoid issues involving collinearity. We conducted the fitting process via functions lmer from package lme4 (v1.1-12) (Bates, Maechler, & Bolker, 2013) of the R (v3.2.5) software environment (R Development Core Team, 2017). We then used the Wald test to compute *p*-values for each fixed effect.

Table 3. Fixed effects from the best-fitting model, TD 4-year-olds.

Interest areas	Fixed effects	Estimate	SE	*t*-value
Q-compatible				
	(Intercept)	−0.28	0.08	−2.87*
	Condition (S-prosody)	−0.48	0.09	−4.33***
	Bin	0.21	0.02	15.58***
	Condition (S-prosody) × Bin	−0.27	0.02	−14.57***
S-compatible				
	(Intercept)	−0.24	0.08	−2.67*
	Condition (S-prosody)	0.37	0.09	3.82***
	Bin	−0.17	0.02	11.06***
	Condition (S-prosody) × Bin	0.21	0.02	12.01***

Model used in R: logit proportion ~ Condition * Bin + (1 + Condition + Bin | Participant) + (1 + Condition + Bin | Item).
*$p < .05$, *** $p < .001$.

Table 4. Fixed effects from the best-fitting model, 4-year-olds with ASD.

Interest areas	Fixed effects	Estimate	SE	*t*-value
Q-compatible				
	(Intercept)	−0.24	0.09	–2.39*
	Condition (S-prosody)	0.04	0.09	1.19
	Bin	0.03	0.04	0.79
	Condition (S-prosody) × Bin	0.04	0.03	1.26
S-compatible				
	(Intercept)	−0.27	0.08	−2.88*
	Condition (S-prosody)	0.04	0.08	1.01
	Bin	−0.03	0.03	0.98
	Condition (S-prosody) × Bin	−0.04	0.03	−1.32

Model used in R: logit proportion ~ Condition * Bin + (1 + Condition + Bin | Participant) + (1 + Condition + Bin | Item).
*$p < .05$.

In the models, the fixed effect *condition* has two levels: 'question prosody' and 'statement prosody', where the 'statement prosody' was treated as the baseline. Table 3 shows the model results for the TD 4-year-olds in the two critical areas: the question-compatible (Q-compatible) and the statement-compatible (S-compatible) areas. There was a significant main effect of prosodic condition (= 'Condition (S-prosody)'). The negative coefficient indicates that the TD 4-year-olds looked significantly more at the question-compatible area when hearing question prosody than when hearing statement prosody. The temporal bin (= 'Bin') is also a reliable predictor and it interacts with prosodic condition (= 'Condition (S-prosody) × Bin') such that the probability of looking at the question-compatible area decreased over time after hearing sentences with statement

Table 5. Fixed effects from the best-fitting model, 5-year-olds with ASD.

Interest areas	Fixed effects	Estimate	SE	*t*-value
Q-compatible				
	(Intercept)	−0.26	0.09	−2.46*
	Condition (S-prosody)	0.04	0.08	1.02
	Bin	0.03	0.03	0.87
	Condition (S-prosody) × Bin	0.05	0.03	1.21
S-compatible				
	(Intercept)	−0.25	0.08	−2.67*
	Condition (S-prosody)	0.04	0.05	1.04
	Bin	−0.03	0.04	−0.77
	Condition (S-prosody) × Bin	−0.05	0.02	−1.24

Model used in R: logit proportion ~ Condition * Bin + (1 + Condition + Bin | Participant) + (1 + Condition + Bin | Item).
$^{*}p < .05$.

prosody. By contrast, an opposite pattern was observed in the statement-compatible area. The positive coefficient for the main effect of prosodic condition reflected that hearing sentences with statement prosody triggered more fixations on the statement-compatible area. The positive coefficient for the interaction between temporal bin and prosodic condition confirms that the probability of looking at the statement-compatible area increased over time after hearing sentences with statement prosody. To explore when the observed effects started to occur, we applied an LMM model (excluding the temporal bin from both the fixed effects and the random effects) to each temporal bin. The results indicated that the difference between the two prosodic conditions reached significance 1500 ms after the onset of the *wh*-word ($b = 0.45$, $t = 3.12$, $p < .01$).

Table 4 summarizes the model results for the 4-year-olds with ASD, and Table 5 gives the model results for the 5-year-olds with ASD. As indicated in the tables, the two ASD groups exhibited similar eye gaze patterns in the two critical areas. But their eye gaze patterns were different from the TD 4-year-olds. Unlike the TD 4-year-olds, in both the question-compatible and the statement-compatible areas the ASD groups did not exhibit a main effect of prosodic condition. In addition, there was no significant main effect of temporal bin, nor a significant interaction between temporal bin and prosodic condition.

In order to see statistically whether there exist differences between the three groups in the time course of the eye gaze patterns, new models were developed in which group was included as an experimental factor. More specifically, models were fitted to the entire data set (the TD 4-year-olds, the ASD 4-year-olds and the ASD 5-year-olds) in the two critical areas, treating prosodic condition, temporal bin and group as fixed effects, with random intercepts and slopes for both participants and items. The model results are summarized in Table 6.

As indicated in Table 6, in both the question-compatible (Q-compatible) and the statement-compatible (S-compatible) areas, significant effects of group were observed between the TD 4-year-olds and the 4-year-olds with ASD (= ‘ASD 4-yr-olds (TD 4-yr-olds)’), and between the TD 4-year-olds and the 5-year-olds with ASD (= ‘ASD

Table 6. Fixed effects from the best-fitting model, three groups combined.

Interest areas	Fixed effects	Estimate	SE	*t*-value
Q-compatible				
	(Intercept)	−0.21	0.04	−5.44**
	Condition (S-prosody)	−0.59	0.06	−2.57*
	Bin	0.13	0.04	2.49*
	ASD 4-yr-olds (TD 4-yr-olds)	0.05	0.03	5.25**
	ASD 5-yr-olds (TD 4-yr-olds)	0.04	0.03	5.19**
	Condition (S-prosody) × Bin	−0.24	0.04	−2.65*
S-compatible				
	(Intercept)	0.19	0.05	5.22**
	Condition (S-prosody)	0.53	0.04	2.38*
	Bin	−0.15	0.06	−2.54*
	ASD 4-yr-olds (TD 4-yr-olds)	−0.05	0.04	−5.12**
	ASD 5-yr-olds (TD 4-yr-olds)	−0.05	0.04	−4.89**
	Condition (S-prosody) × Bin	0.26	0.03	2.48*

Model used in R: logit proportion ~ Condition*Bin*Group + (1 + Condition + Bin | Participant) +(1 + Condition + Group+ Bin | Item).
*$p < .05$, ** $p < .01$.

5-yr-olds (TD 4-yr-olds)'), confirming the significant differences between the TD group and the two ASD groups in the time course of the eye gaze patterns. The eye gaze patterns displayed in Figures 2–4 were supported by the statistical modelling.

Discussion

The present study sought to investigate whether high-functioning Mandarin-speaking preschool children with ASD are able to use prosodic cues to understand the communicative intentions of the speaker. Using the visual world paradigm, we found that unlike their TD peers, high-functioning children with ASD failed to exhibit an eye gaze pattern that is associated with the successful use of prosody to infer the speaker's intended meaning. More specifically, the eye gaze patterns showed that the TD 4-year-olds effectively used prosodic cues to arrive at the intended meaning; the two ASD groups, however, failed to do so. Both the 4-year-olds with ASD and the 5-year-olds with ASD preferred the statement reading over the question reading regardless of the prosodic cues, indicating a tendency to interpret questions as statements. The findings are consistent with the study by Peppé et al. (2007) that high-functioning English-speaking children with ASD tended to wrongly judge questions as statements, suggesting that the statement reading seems to be the default reading for children with ASD. The findings indicate that high-functioning Mandarin-speaking preschool children with ASD are not able to use prosodic cues to distinguish between the two speech acts (asking a question versus making a statement).

To see whether the ability to understand communicative intentions using prosodic cues develops in preschool children with ASD, the present study included two age groups: the 4-year-olds with ASD and the 5-year-olds with ASD. Their eye gaze patterns were then compared to those of the TD 4-year-olds with whom the 5-year-olds with ASD were matched both in terms of verbal IQ and MLU. We found that the 4-year-olds with ASD and the 5-year-olds with ASD exhibited similar eye gaze patterns that reflected their inability to use prosodic cues. The results indicate that the ability to comprehend the speaker's communicative intention using prosodic cues is relatively independent of their verbal IQ and sentence complexity levels, or structural language in general (MLU). Furthermore, the fact that the 5-year-olds with ASD were still impaired relative to their verbal IQ- and MLU-matched TD peers also seems to suggest that inferring intentionality from prosody relies on social processes, and the impaired prosody in the ASD group is not simply due to impaired language abilities. Our findings are in line with previous research showing that the use of prosody is atypical even in individuals with no structural language impairments (see Tager-Flusberg, Paul, & Lord, 2005). In addition, the results show that there is no development from 4 years of age to 5 years of age, indicating that this ability is perhaps inherently impaired. However, we also wish to point out that the lack of development from 4 to 5 years of age does not necessarily exclude the possibility that there is a late development of this ability in the ASD population, because development of the ability may proceed beyond 4 and 5 years of age in ASD children. In fact, the study by Diehl et al. (2015) as discussed in the introduction seems to suggest a late development of prosodic ability in ASD (i.e. at around adolescence). Further research is required in order to see whether there is late development of this ability in ASD.

Overall, the results of the present study show that preschool children with ASD exhibit a deficit in using prosodic cues to understand the communicative intentions of the speaker. Unravelling the kinds of cues preschool children with ASD can and cannot utilize when making inferences about others' communicative intentions is crucial to understanding the nature of their deficits in social communication and interaction. More specifically, identifying both types of cues will help us better understand the impaired and the preserved components associated with their social interaction and communication, which in turn will enable us to specify which components (i.e. impaired components) contribute to their deficits in social communication and interaction. Identifying both the impaired and the preserved components also has important clinical implications. Once the two types of components are identified, we can take advantage of the preserved components and use them to compensate for the impaired components when designing treatment plans for their abilities in the social and communicative domain.

It remains an open question, though, regarding the source of the deficit in using prosodic cues to understand others' communicative intentions. This deficit might reflect a general pragmatic difficulty in understanding the communicative function of prosody, which might ultimately be connected with their core deficits in the social and communicative domain. Further research is required to systematically investigate their ability to use various types of linguistic cues that are closely related to the social and communicative domain.

We also wish to acknowledge a few limitations of the current study. First, our sample size is relatively small with 18 participants in each ASD group. Therefore, the eye gaze patterns obtained from the current study need to be further validated using a much larger sample. Second, like most previous work, our sample of children is highly verbal and does not represent a broader spectrum sample (e.g. our sample does not include minimally verbal children with ASD), so inferences about the overall spectrum should be made with caution. Further research is required to examine the ability to infer intentionality from prosody in minimally verbal children with ASD. In addition, the current study mainly focused on the comprehension abilities of children with ASD. Further research is required to examine both comprehension and production in ASD children, so that a more comprehensive picture can be obtained. Finally, we wish to point out that our findings allow us to generalize about prosody interpretation for one type of construction in one language. However, since some other studies have found no differences between highly verbal children with ASD and verbal IQ-matched TD peers on prosody interpretation for other languages and other types of communicative intent (see e.g. A. T. Wang, Lee, Sigman, & Dapretto, 2006 for ironic vs literal language), future research is required to identify which specific types of communicative intent in which specific language are impaired in children with ASD.

Conclusions

The present study shows that 4- and 5-year-old Mandarin-speaking children with ASD exhibit a deficit in using prosodic cues to understand the communicative intentions of the speaker. The ability to infer intentionality from prosody might rely on social processes, and the impaired prosody in the ASD population might not be simply due to impaired language abilities.

Although no development of this ability was observed from 4 to 5 years of age, the possibility remains that development of the ability may proceed beyond 4 and 5 years of age and might occur at a rather late stage. Further studies are required in order to explore whether there is indeed a late development of this ability. Longitudinal studies would be particularly ideal to observe the developmental changes.

Acknowledgements

The authors would like to thank the children, the parents and the teachers at the Rehabilitation and Education Centre for Children with ASD and at the Taolifangyuan Kindergarten, Beijing, China, for their assistance and support in running the study.

Author contributions

PZ, WM and LZ contributed equally to this work.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by the National Social Science Foundation of China (16BYY076) to Peng Zhou.

ORCID iDs

Peng Zhou https://orcid.org/0000-0002-0818-2545
Weiyi Ma https://orcid.org/0000-0002-2299-9834
Likan Zhan https://orcid.org/0000-0002-9275-3557

Notes

1. In China, 0–6 years of age are often defined as preschool age, because Chinese children typically go to primary school at the age of 7, which marks the beginning of the nine-year formal compulsory education.
2. ASD was not recognized in China until 1982 when the first cases of ASD were reported by Tao (1982). According to the National Bureau of Statistics of the People's Republic of China (2016), ASD has become the most prevalent mental disability among Chinese children. While no nationwide prevalence survey of ASD in China has been published, some epidemiological studies based on meta-analyses suggest that prevalence is likely to be between 10.18 and 11.80 per 10,000 individuals (Sun & Allison, 2010; Sun et al., 2013; F. Wang et al., 2018). In China, the diagnoses of ASD are typically made by paediatric neurologists at hospitals using DSM-5, supplemented with screening instruments, such as the Chinese versions of the Autism Behaviour Checklist, the Childhood Autism Rating Scale and the Clancy Autism Behaviour Scale (see Su, 2014, for a review of the instruments). In recent years, the 'gold standard' diagnostic instrument – ADOS – has been adopted by some hospitals in Beijing to confirm the diagnoses. Chinese children typically receive confirmed diagnoses at 3.5 to 4 years of age.
3. One anonymous reviewer asked whether the speech samples used to calculate MLU show any evidence for atypical prosodic patterns in ASD children. We wish to note that the current study mainly focused on children's comprehension of prosody, and MLU was mainly used to match ASD children and TD children on sentence complexity levels. So, we did not measure the prosodic features of these speech samples. In addition, in the speech samples only 24 questions were observed in the total of 3600 utterances recorded for the two ASD groups, and all 24 questions were short questions (see Appendix). Based on such limited question samples, it would be hard to make any meaningful generalizations about the prosodic use of prosody in ASD children.
4. One anonymous reviewer asked whether there were any differences across the ASD children. We wish to note that the statistical results were based on individual data. Also, we examined the eye gaze pattern of each individual participant before running the statistical models, and their eye gaze patterns showed that they consistently exhibited an impairment in inferring intentionality from prosody. No significant variability was observed with regard to their prosodic ability. We believe that this is partly due to the relative homogeneity of our participants. As discussed, we only looked at highly verbal children with ASD. In addition, we wish to point out that the current study mainly focused on children's comprehension, without looking at their production. There is a possibility that a higher variability might be observed in production than in comprehension. The asymmetry between comprehension and production is not unusual in child language research, in particular in the area of prosody and interpretation (see Speer & Ito, 2009; Zhou, Su, Crain, Gao, & Zhan, 2012 for a comprehensive review).

References

Akhtar, N., & Martinez-Sussmann, C. (2007). Intentional communication. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 201–220). New York, NY: Guilford.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, text revision (DSM-IV-TR)*. Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, 5th edition (DSM-5)*. Washington, DC: Author.

Arciuli, J., & Bailey, B. (2019). An acoustic study of lexical stress contrastivity in children with and without autism spectrum disorders. *Journal of Child Language*, *46*, 142–152.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baltaxe, C., & Simmons, J. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. Mesibov (Eds.), *Communication problems in autism* (pp. 95–125). New York, NY: Plenum Press.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models*. Retrieved from https://arxiv.org/abs/1506.04967

Bates, D. M., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes*. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bavin, E. L., Kidd, E., Prendergast, L. A., & Baker, E. K. (2016). Young children with ASD use lexical and referential information during on-line sentence processing. *Frontiers in Psychology*, *7*, 171.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M. B., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, *89*, 1996–2009.

Brock, J., Norbury, C. F., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*, 896–904.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger syndrome. *Research in Autism Spectrum Disorders*, *3*, 502–516.

DeMyer, M. K., Barton, S., DeMyer, W. E., Norton, J. A., Allen, J., & Steele, R. (1973). Prognosis in autism: A follow-up study. *Journal of Autism and Childhood Schizophrenia*, *3*, 199–246.

Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, *27*, 867–884.

Diehl, J. J., & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics*, *34*, 135–161.

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

Gong, Y. X., & Dai, X. Y. (1992). *Chinese-Wechsler Young Children Scale of Intelligence (C-WYCSI)*. Changsha, China: Map Press.

Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research*, *8*, 717–726.

S., Hala (Ed.) (1997). *The development of social cognition*. Hove, UK: Psychology Press.

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism. *Journal of Autism and Developmental Disorders*, *38*, 1328–1340.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kanner, L. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, *1*, 119–145.

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, *6*, 479–493.

Kelly, D. J., Walker, R., & Norbury, C. F. (2013). Deficits in volitional oculomotor control align with language status in autism spectrum disorders. *Developmental Science*, *16*, 56–66.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*, 29–36.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

Lord, C., & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed., pp. 460–483). New York, NY: Wiley.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.

McAlpine, A., Plexico, L. W., Plumb, A. M., & Cleary, J. (2014). Prosody in young verbal children with autism spectrum disorder. *Contemporary Issues in Communication Science & Disorders*, *41*, 120–132.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*, 325–350.

Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review*, *14*, 349–372.

National Bureau of Statistics of the People's Republic of China. (2016). Communiqué on major statistics of the second China national sample survey on disability. *Chinese Journal of Rehabilitation Theory and Practice*, *12*, 1013 (in Chinese).

Norbury, C. F. (2017). Eye-tracking as a window on language processing in autism spectrum disorder. In L. Naigles (Ed.), *Innovative investigations of language in Autism* (pp. 13–33). New York, NY: APA Books.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*, 205–220.

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

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*, 1776–1791.

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

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Plesa-Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with Autism Spectrum Disorders. *Autism*, *20*, 591–604.

Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*, 371–381.

R Development Core Team. (2017). *R: A language environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.r-project.org/

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, *133*, 37–45.

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

Rutter, M., & Lockyer, L. (1967). A five to fifteen year follow-up study of infantile psychosis: I. Description of sample. *The British Journal of Psychiatry*, *113*, 1169–1182.

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

Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition – Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass*, *3*, 90–110.

Su, L.-Y. (2014). *Child psychiatry*. Changsha, China: Hunan Science and Technology Press (in Chinese).

Su, Y. E., Jin, Y., Wan, G. B., Zhang, J. S., & Su, L. Y. (2014). Interpretation of wh-words in Mandarin-speaking high-functioning children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *8*, 1364–1372.

Sun, X., & Allison, C. (2010). A review of the prevalence of autism spectrum disorder in Asia. *Research in Autism Spectrum Disorders*, *4*, 156–167.

Sun, X., Allison, C., Matthews, F. E., Sharp, S. J., Auyeung, B., Baron-Cohen, S., & Brayne, C. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: A systematic review and meta-analysis. *Molecular Autism*, *4*, 7.

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry*, *11*, 325–334.

Tager-Flusberg, H., Paul, P., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, A. Klin, R. Paul, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 335–364). Hoboken, NJ: John Wiley.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tao, K. (1982). The problems of diagnosis and classification of infantile autism. *Journal of Chinese Neuropsychiatry*, *2*, 104 (in Chinese).

Terzi, A., Marinis, T., & Francis, K. (2016). The interface of syntax with pragmatics and prosody in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *46*, 2692–2706.

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. A. Sekerina, E.M. Fernández, & H. Clahsen (eds) *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 73–96). Amsterdam, The Netherlands: John Benjamins.

Trueswell, J. C., & Gleitman, L. R. (2007). Learning to parse and its implications for language acquisition. In G. Gaskell (ed.), *Oxford handbook of psycholinguistics* (pp. 635–656). Oxford, UK: Oxford University Press.

Verschueren, J. (1999). *Understanding pragmatics*. London, England: Edward Arnold.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*, 932–943.

Wang, F., Lu, L., Wang, S. B., Zhang, L., Ng, C. H., Ungvari, G. S., & Xiang, Y. T. (2018). The prevalence of autism spectrum disorders in China: A comprehensive meta-analysis. *International Journal of Biological Sciences*, *14*, 717–725.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*, 1559–1579.

Zhou, P., Crain, S., & Zhan, L. (2012). Sometimes children are as good as adults: The pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language*, *67*, 149–164.
Zhou, P., Ma, W., Zhan, L., & Ma, H. (2018). Using the visual world paradigm to study sentence comprehension in Mandarin-speaking children with autism. *Journal of Visualized Experiments*, *140*, e58452.
Zhou, P., Su, Y., Crain, S., Gao, L., & Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: A view from Mandarin Chinese. *Journal of Child Language*, *39*, 687–730.
Zhou, P., Zhan, L., & Ma, H. (2019). Predictive language processing in preschool children with autism spectrum disorder: An eye-tracking study. *Journal of Psycholinguistic Research*, *48*, 431–452.

Appendix

Questions in the speech samples of the children with ASD: 24 questions were observed in the total of 3600 utterances recorded for the two ASD groups. In the 24 questions, there were four instances of the *shei* 'who' question in (1), four instances of the *shenme* 'what' question in (2), three instances of the *shenme* 'what' question in (3), three instances of the *shenme* 'what' question in (4), three instances of the *shenme* 'what' question in (5), two instances of the *weishenme* 'why' question in (6), two instances of the yes-no question in (7), two instances of the yes-no question in (8) and one instance of the yes-no question in (9). In (7), (8) and (9), *ma* is the yes-no question marker.

(1) Shei?
who

(2) Shenme?
what

(3) Shei lai?
who come

(4) Shenme ren?
what person

(5) Shenme shuiguo?
what fruit

(6) Weishenme?
why

(7) Dui ma?
right Q-marker

(8) Daxiang ma?
elephant Q-marker

(9) Chi dangao ma?
eat cake Q-marker