RESPONSES TO PROSODIC PROMINENCE IN CHILDREN WITH AUTISM SPECTRUM DISORDER

Kiwako Ito[1] and Elizabeth Kryszak[2]

[1]University of Newcastle, Australia and [2] Nationwide Children's Hospital
kiwako.ito@newcastle.edu.au, elizabeth.kryszak@nationwidechildrens.org

ABSTRACT

Prosodic emphasis facilitates referential resolution [1, 5, 6, 16] and recall of discourse contents [3, 4], yet these effects have not been fully confirmed in young children with developmental disorders. The present eye-tracking study tested how prosodic prominence affects object search and attention maintenance in young children with Autism Spectrum Disorder (ASD) with interactive videos. Typically-developing children swiftly switched their gazes from the actor's face to the object with or without prominence on X in "Where is X?" and maintained longer gazes on the object with than without prominence. Children with ASD failed to switch gazes and were overall slower in detecting objects, although the prominence led to relatively faster eye-movements to the object. The data demonstrate the sensitivity to prosodic prominence in children with ASD, yet suggest that prominence cannot trigger robust attention shift and maintenance.

Keywords: Prosodic prominence, Autism Spectrum Disorder, eye-tracking, object search, attention.

1. INTRODUCTION

Prosodic prominence accompanied by a dynamic movement of fundamental frequency (F0), longer segmental duration, higher intensity, and distinctive voice quality, etc. affects how we process spoken messages. Past eye-tracking studies have demonstrated that listeners respond to prosodic prominence for referential expressions immediately, with swift eye-movements for object search. Adults are generally very efficient in allocating attention according to the presence of a pitch accent in a referential expression, which signals novelty or contrastive status of the referent in a given discourse [1, 16, 7, 8]. Children are not as proficient as adults, yet they exhibit the sensitivity to prosodic prominence and start making use of it predictively for visual search during the preschool age [5, 6, 9].

Another line of research on prosodic prominence and memory have shown that people better recall the contents of short passages when a particular member of a referential set is narrated with than without a prominent pitch accent [3,4].

Taken together, prosodic prominence has a direct impact on how we process speech input and interact with referential context, and affects how we store and retrieve memory of discourse. While these functions of prosodic prominence may play critical roles in the development of communication skills, research on prosody in younger children, especially with developmental disorders, have not progressed largely due to methodological difficulties [10-12].

The present study investigates how young children with Autism Spectrum Disorder respond to a prosodic prominence for visual search, and whether the prominence has any effect on how they maintain attention to a particular referent after a referential expression marks its relevance to the communicative goal. To overcome the methodological pitfalls of the past eye-tracking studies with passive scene observations, which did not clarify the communicative goals of the experimental tasks [2, 13, 14], we tested young children with ASD with interactive video stimuli that elicited responses to both linguistic and gestural cues. While young children with autism may show general sensitivity to acoustic prominence in speech, the presence of facial movements accompanying speech signals may distract these participants with limited cognitive resources (reflected in their general delay in development of memory and executive function) [15] and thus make them less efficient in utilizing prosodic prominence for object search.

2. EXPERIMENT

2.1. Participants

This paper presents data from 17 children diagnosed with ASD at [institution name] and 17 typically-developing children who were individually age-matched to the clinical group. Table 1 summarizes the two groups' performances in the Peabody Picture Vocabulary Test-IV (PPVT-IV).

Table 1: Participants' age and PPVT scores.

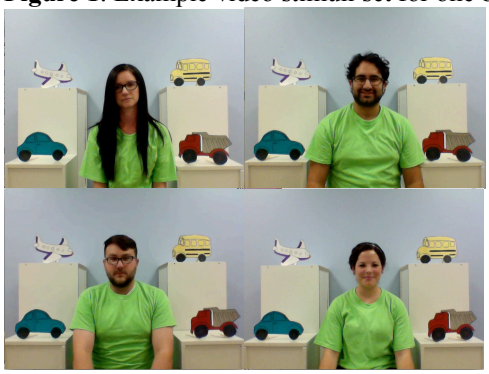
Group	age	PPVT (ss)
ASD (n=17)	3;01 (1;07-5;03)	53.18 (21.27) Range: 20-93
TD (n=17)	3;09 (1;10-4;09)	99.65 (22.97) 51-133

2.2. Materials and Design

A total of 32 videos were presented across 4 lists. Four sets of objects (farm animals, kitchen items, fruits, and vehicles) were prepared for object search. Each session consisted of four blocks, each of which showed 4 video clips with the same set of objects. Within each video clip, an actor appeared with a set of four objects in the background (e.g., car, plane, bus and truck: Figure 1), and asked for the participant's help in finding an object, using the plot such as:

Hi, my name is Erin. I'm looking for my favourite vehicle. Can you help me find it? (pause)
Where is the BUS/bus? (pause)
Over where? (pause)
Oh, I see. [2s later, she turns her head to the object, holds the gaze for 2 seconds, points to the object, and grabs it.]
Thank you! (total duration approx. 30 sec)

Figure 1: Example video stimuli set for one block



For each block, four actors asked for four objects in a random order. After the 4th video, a static image of one of the actors reappeared with the same set of objects (of which the locations are shuffled) and participants heard a question:
Which one did he/she find?

2.1.1. Acoustic analyses of Where is X?

Each of 8 actors (4 males and 4 females) produced two versions of the same video, altering only the tune for the critical question *Where is X?* - one with a prominence on the noun X and one without. See Figure 2 for example two tunes.

Figure 2: Example stimuli pair "*Where is the BUS* (left) / *bus* (right)"?

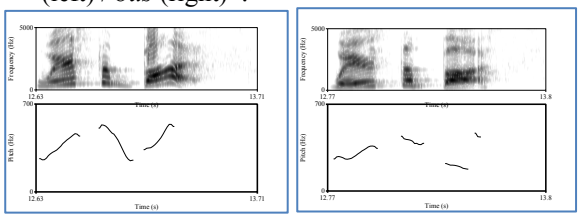


Table 2: Where is X? acoustic analyses
Mean (std) and results of paired t-tests of duration and F0

Condition	Duration (ms)	Ave. F0 (Hz)	X dur (ms)	X F0 (Hz)
Emphatic	1074 (171)	260.3 (73.3)	537 115	277.2 (43.6)
Non-Emphatic	1028 (137)	256.6 (66.1)	468 (111)	200.9 (79)
t (df=15)	1.317	0.705	4.186	4.203
p-val	0.208	0.491	<0.001	<0.001

Acoustic analyses revealed that the utterance duration and F0 did not differ between the two conditions, but the noun was significantly longer and had higher F0 in Emphatic than in Non-Emphatic condition (Table 2). Further analysis also confirmed significantly higher relative amplitude of the object (log of the ratio between object RMS and utterance RMS) for the Emphatic as compared to Non-Emphatic condition (t=9.82, $p<.0001$).

2.3. Procedure

Each child participant was seated within 50-60cm from the Tobii 60XL monitor in a dimmed room. Participant's eye-movements were quickly calibrated using Clearview software before each block. Unless the child was distracted, researchers remain silent during each video and coded the child's verbal and gestural responses to the video. Short cartoon videos were played between blocks to keep the child entertained. After the final block, the child faced the researcher and PPVT-IV was administered.

Each child viewed a total of 16 video clips (4 blocks of 4 videos): 2 blocks presented Emphatic versions, and the other 2 blocks presented Non-Emphatic versions. The order of the blocks and the conditions were counterbalanced across the four presentation lists.

2.4. Results

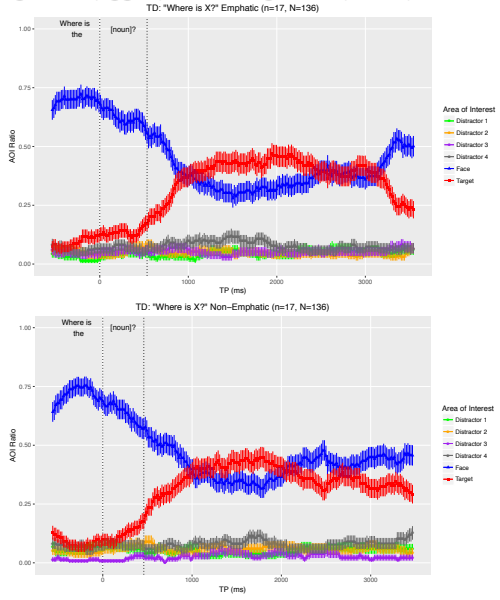
Analyses of oral and gestural responses to the video stimuli confirmed the general lack of responsiveness to speech input in ASD [15]. While TD children made gestural responses to "*Where is X?"* nearly 50% of the time with or without emphasis, children with ASD did so less than 20% of the time (Table 3).

The emphasis on the critical object impacted the patterns of visual search differently across the groups. As shown in Figure 3, TD children started launching the eye-movements to the target object toward the end of the critical noun, and shifted their gazes from the actor's face (blue) to the mentioned object (red) in a few hundred milliseconds after the offset of the noun. After hearing an emphasis, however, they maintained the gaze on the object longer, delaying the shift back to the actor by about 1sec as compared to when they heard a non-emphatic object (prosody effect was tested with linear mixed effects regression models on ln(targ/face) for 2100-2400ms after noun offset: t=1.98, p=0.05).

Table 3: Mean (std) Oral/Gestural responses to *Where is X?*

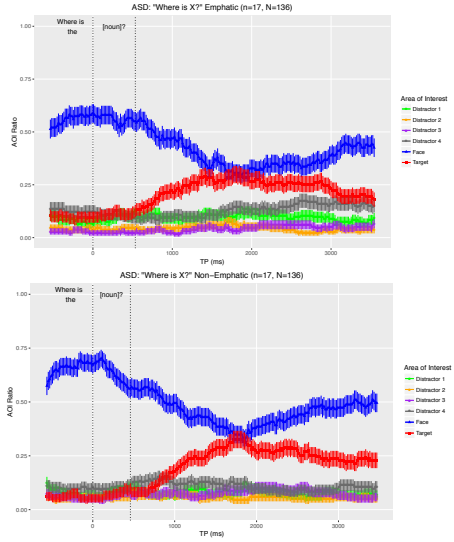
Condition	Emphatic: Where is X?		Nonemphatic: Where is x?	
oral/gesture	Oral %	Gesture %	Oral %	Gesture %
ASD (n=17)	15.63 (29.21)	15.63 (27.77)	15.13 (24.14)	17.11 (29.23)
TD (n=17)	25.00 (36.17)	49.26 (37.88)	21.32 (33.88)	48.53 (43.50)

Figure 3: TD gazes during *Where is X?*: Emphatic (upper) Non-Emphatic (lower)



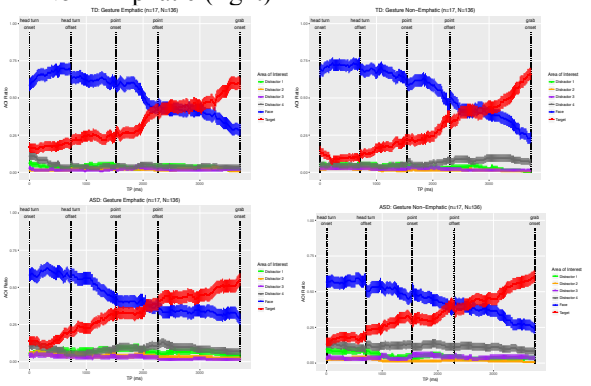
Children with ASD were overall slower to detect the object as compared to their TD peers (group effect on ln(targ/all) for 600-900ms after noun offset: t=-3.11, p<0.01) and did not maintain the attention as long as their TD peers (group effect on ln(targ/all) for 900-2100ms after noun offset: t=-2.09~-3.12, p<0.01). Although the gazes to the target object increased after *Where is X?*, they never exceeded the looks to the actors face in both conditions (Figure 4). With the emphasis, however, their looks to the target object increased relatively faster (prosody effect on ln(targ/face) for 600-900ms after noun offset: t=1.88, p=0.07). In addition, the emphasis on the object noun seemed to have led relatively more looks to other objects/distractors in the scene, although this did not bear statistical reliability.

Figure 4: ASD gazes during *Where is X?*: Emphatic (upper) Non-Emphatic (lower)



Comparisons of the eye-movements during the head-turn, pointing and grabbing (that was accompanied by no speech) revealed interestingly similar patterns between the two groups (Figure 5). The looks to the mentioned object increased steadily after the head-turn, and exceeded the looks to the actor's face after the pointing to the object. Importantly, children with ASD did not show any delay in this gaze switch as compared to their TD peers (no group effect on ln(targ/all) and ln(face/all) from the beginning of head-turn and the end of grabbing). This provides counterevidence to a traditional view that children with ASD lack the ability to process cues to joint-attention or that they are not sensitive to social cues [13-15].

Figure 5: eye-movements during gesture sequence TD (upper) vs. ASD (lower), Emphatic (left) vs. Non-Emphatic (right)



3. DISCUSSION & CONCLUSION

Using interactive videos, this study elicited spontaneous responses to prosodic prominence in young children with ASD and their age-matched TD peers. As shown in the difference in the PPVT scores, the many children of the clinical group were minimally verbal at the time of experiment. Group differences in the gestural and oral responses to the videos also confirmed the general assumption that children with ASD are less responsive to speech input.

However, participants' spontaneous eye-movements during the videos revealed unpredicted similarities more than predicted differences across the groups. First, both TD children and children with ASD mostly looked at the actor's face during speech, unlike a previous study [15]. Second, the looks to the target object increased after *Where is X,* and then gazes were shifted back to the actor in both groups. This happened regardless of the prosody: that is, young children with ASD generally processed spoken messages in a similar manner as in their TD peers, even though they failed to respond with words and gestures. Finally, the timings of the gaze shift during the sequence of gesture (head-turn →pointing → grabbing) were strikingly similar between the groups, despite that the deficit in processing joint-attention cues has been repeatedly reported for children with ASD [2, 14, 15]. Children with ASD tested in the present study rather swiftly shifted their gazes from the actor's face to the object according to the actor's gaze direction and pointing gesture. The timing of such gaze shift during the non-verbal gestural sequence show that the delay in responses to *Where is X* cannot be attributed to the general delay in oculomotor control or the lack of interest in communicative cues.

We argue that the observed differences in the timing of the effect of prosodic prominence across groups are due to the developmental delay specific to speech processing in ASD. TD children swiftly processed segmental cues for word recognition with or without prosodic prominence, and the effect of prosody had no room to show during the initial gaze shift with its ceiling speed. The effect of prominence instead appeared in the degree of attention maintenance to the object in a later window. Children with ASD, in contrast, were generally slower to process speech input for word recognition and therefore the facilitative effect of prominence appeared in the earlier window where the looks to the target increased gradually. Importantly, however, neither group showed an effect of prosodic prominence on the gaze shift during the later non-verbal gesture sequence. Thus, prosodic prominence does not seem to have long-term effect to maintain discourse salience in young children, with or without a developmental disorder.

7. REFERENCES

[1] Dahan, D., Tanenhaus, M. K., Chambers, C. G. 2002. Accent and reference resolution in spoken-language comprehension. *J Mem and Lang* 47, 292–314.

[2] Falck-Ytter, T., Bölte, S., Gredebäck, G. 2013. Eye tracking in early autism research. *J of Neurodev Disord,* 5(1): 28

[3] Fraundorf, S. H., Watson, D. G., Benjamin, A. S. 2010. Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J Mem Lang*, 63, 367–386.

[4] Fraundorf, S. H., Watson, D. G., Benjamin, A. S. 2012. The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, 27(1), 88–98.

[5] Ito, K., Bibyk, S., Wagner, L., Speer, S. R. 2014. Interpretation of contrastive pitch accent in 6- to 11-year-old English speaking children and adults. *J Child Lang*, 41(1), 84–110.

[6] Ito, K., Jincho, N., Minai, U., Yamane, N., Mazuka, R. 2012. Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds. *J Mem Lang*, 66(1), 265–284.

[7] Ito, K., Turnbull, R., Speer, S. R. 2017. Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology*, 8(1): 6, 1–29.

[8] Ito, K., Speer, S. R. 2008. Anticipatory effect of intonation: eye movements during instructed visual search. *J of Mem Lang* 58, 541–73.

[9] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., Tanenhaus, M. K. 2014. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. Cognition, 133, 335–342.

[10] Peppé, S. 2009. Aspects of identifying prosodic impairment. International *J of Sp-Lang Path*, 11(4), 332–338.

[11] Peppé, S. Cleland, J., Gibbon, F., O'Hare, A., Martínez Castilla, P. 2011. Expressive prosody in children with autism spectrum conditions. *J Neuroling, 24, 41-53.*

[12] Peppé, S., & McCann, J. 2003. Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, 17, 345–354.

[13] Pierce, K., Marinero, S., Hazin, R., McKenna, B., Carter Barnes, C., Malige, A. 2016. Eye-tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated with Increased Symptom Severity. *Biol Psychiatry 79(8): 657-666.*

[14] Shic, F., Bradshaw, J., Klin, A., Scassellati, B., Chawarsca, K. 2011. Limited activity monitoring in toddlers with autism spectrum disorder. *Brain Research 1380, 246-254.*

[15] Shic, F., Macari, S., Chawarska, K. 2014. Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biol Psychiatry 75(3): 231-237.*

[16] Weber, A., Braun, B. & Crocker, M. W. 2006. Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49(3), 367–92.