

Slipped lips: onset asynchrony detection of auditory-visual language in autism

Ruth B. Grossman,¹ Matthew H. Schneps,² and Helen Tager-Flusberg¹

¹Lab of Developmental Cognitive Neuroscience, Boston University School of Medicine, USA;

²Harvard-Smithsonian Center for Astrophysics, USA

Background: It has frequently been suggested that individuals with autism spectrum disorder (ASD) have deficits in auditory-visual (AV) sensory integration. Studies of language integration have mostly used non-word syllables presented in congruent and incongruent AV combinations and demonstrated reduced influence of visual speech in individuals with ASD. The aim of our study was to test whether adolescents with high-functioning autism are able to integrate AV information of meaningful, phrase-length language in a task of onset asynchrony detection. **Methods:** Participants were 25 adolescents with ASD and 25 typically developing (TD) controls. The stimuli were video clips of complete phrases using simple, commonly occurring words. The clips were digitally manipulated to have the video precede the corresponding audio by 0, 4, 6, 8, 10, 12, or 14 video frames, a range of 0–500ms. Participants were shown the video clips in random order and asked to indicate whether each clip was in-synch or not. **Results:** There were no differences between adolescents with ASD and their TD peers in accuracy of onset asynchrony detection at any slip rate. **Conclusion:** These data indicate that adolescents with ASD are able to integrate auditory and visual components in a task of onset asynchrony detection using natural, phrase-length language stimuli. We propose that the meaningful nature of the language stimuli in combination with presentation in a non-distracting environment allowed adolescents with autism spectrum disorder to demonstrate preserved accuracy for bi-modal AV integration. **Keywords:** Autism, AV-integration, language, autistic disorder, communication, language, face, voice. **Abbreviations:** ASD: autism spectrum disorders; TD: typically developing; AV: auditory-visual; ADOS: Autism Diagnostic Observation Schedule; ADI-R: Autism Diagnostic Interview-Revised; K-BIT 2: Kaufman Brief Intelligence Test, Second Edition; PPVT-III: Peabody Picture Vocabulary Test, Third Edition.

Individuals with autism spectrum disorders (ASD) have a number of core deficits, including repetitive or stereotyped behaviors, social-emotional disturbances, and verbal and non-verbal language difficulties, which significantly disrupt their ability to maintain successful face-to-face communication. Everyday conversation requires rapid integration of verbal and nonverbal sensory information from face and voice to interpret a speaker's expressed thoughts and emotional intent. This type of multi-sensory processing may pose specific difficulties for individuals with ASD.

There are data suggesting atypical sensory profiles among individuals with ASD (e.g., Watling, Deitz, & White, 2001) as well as atypical reactions to sounds and sights, which may be among the earliest signs of autism in very young children (Osterling, Dawson, & Munson, 2002). Furthermore, there is a body of literature suggesting that individuals with ASD have difficulty integrating stimuli from multiple sensory modalities, such as vision and audition, which is a crucial skill for the processing of everyday conversation. It is as yet unclear whether these reported sensory integration deficits are based on their overall cognitive style (Frith, 1989; Frith & Happé, 1994; Happé, 2005), or neuro-anatomical/neuro-physio-

logical differences (Townsend, Harris, & Courchesne, 1996; Just, Cherkassy, Keller, & Minshew, 2004; Russo et al., 2007; Castelli, Frith, & Happé, 2002; see Iarocci & McDonald, 2006 for review). There is clear evidence that typically developing (TD) individuals integrate facial speech information, i.e. lipreading, with auditory speech input to enhance language comprehension (Sumby & Pollack, 1954) and that lipreading improves recognition of speech in noisy environments (Calvert, Brammer, & Iversen, 1998; Massaro, 1984; Summerfield & McGrath, 1984). The question remains whether individuals with ASD are also able to integrate visual speech information to enhance their comprehension of everyday spoken language.

There have been only a few studies on the ability of individuals with ASD to integrate visual lipreading information with auditory speech perception and the evidence so far is contradictory. Some studies have found decreased influence of visual speech on the perception of syllables using McGurk-type paradigms (McGurk & MacDonald, 1976). In these studies, participants are simultaneously presented with auditory information for one syllable, e.g., /ga/, and mouth movements for a different syllable, e.g., /ba/. In typical individuals, these two incongruent presentations are perceptually integrated and merged into a novel syllable, /da/. Individuals with

Conflict of interest statement: No conflicts declared.

© 2008 The Authors

Journal compilation © 2008 Association for Child and Adolescent Mental Health.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

ASD, however, have demonstrated a less consistent or weaker AV integration effect for these incongruent syllables (Irwin, 2006; Condouris et al., 2004; Magneé, de Gelder, van Engeland, & Kemner, in press). There is also evidence that individuals with ASD derive less benefit from lipreading for speech-in-noise paradigms. When asked to repeat sentences presented in auditory background noise, typical individuals' accuracy rates significantly improve with the addition of relevant visual speech (lipreading) information, but individuals with ASD show less such improvement (Smith & Bennetto, 2007). Another study found that 4–6-year-old children with ASD show no preference for looking at the face of a speaker whose lips are moving in synchrony with the simultaneous auditory presentation of a sentence, vs. a speaker who is three seconds out of synch (Bebko, Weiss, Demark, & Gomez, 2006). However, several studies seem to indicate that individuals with ASD have at least preserved lower-level, or sensory, integration for AV information. When asked to count visual flashes, the simultaneous presentation of simple auditory beeps produces illusory additional flashes that are reported equally by participants with ASD and their TD peers (van der Smagt, van Engeland, & Kemner, 2007). ERP data also suggest that low-level, sensory integration for auditory and visual speech sounds is intact in adolescents with ASD (Magneé et al., in press) and there is some evidence of normal AV integration abilities even beyond basic sensory information, for the repetition of single AV syllables from a naturalistic computer-generated speaker (Massaro & Bosseler, 2003; Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004).

In light of these inconclusive prior findings, the aim of our current experiment was to investigate whether individuals with ASD can integrate visual and auditory speech information that is presented in a meaningful language context at normal speaking rate, as it would be during natural face-to-face communication. Our task involved onset asynchrony detection with full, phrasal-length language stimuli representing basic, conversational speech. Onset asynchrony detection has been shown to be a reliable task to test AV integration in typical populations (e.g., van Wassenhove, Grant, & Poeppel, 2007), but has not been used in ASD populations. We chose to use meaningful phrase-length language, rather than single syllables, because of prior evidence showing that the cognitive processing of meaningful language is not necessarily related to the processing of non-word syllables (Grant & Seitz, 1998). In order to maximize the ecological validity of this task, our aim was to use natural language stimuli that could easily occur in everyday conversation. Our prior research on lipreading of complete words (Grossman & Tager-Flusberg, in press) has led us to hypothesize that individuals with ASD are better at recognizing meaningful language stimuli than non-word syllables, a difference also found in typically developing

individuals (Grant & Seitz, 1998). Building on their preserved lower-level sensory AV integration abilities, we hypothesize that participants with ASD will be able to use the meaningful language context of our stimuli to enhance their performance and show higher-order AV integration skills for phrasal speech that are equal to those of their TD peers.

Method

Participants

Two groups participated in this study: 25 adolescents with autism spectrum disorder (20 males, 5 females, mean age 14;5) and 25 typically developing (TD) controls (21 males, 4 females, mean age 13;7). The participants were recruited through advocacy groups for parents of children with autism, local schools, and advertisements placed in local magazines, newspapers, and on the internet. Only participants who spoke English as their native and primary language were included. The Kaufman Brief Intelligence Test, Second Edition (K-BIT 2; Kaufman & Kaufman, 2004) was used to assess IQ and the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) was administered to assess receptive vocabulary. We did not administer the PPVT-III to one participant with ASD due to time constraints, but since his verbal IQ data fell within the matched range (117), he was included in the final sample. Using a multivariate ANOVA with group as the independent variable we confirmed that the groups were matched on age ($F(1,49) = 1.02, p = .32$), sex ($\chi^2(1, N = 50) = .14, p = .5$), verbal IQ ($F(1,49) = 2.82, p = .1$), nonverbal IQ ($F(1,49) = .02, p = .89$), and receptive vocabulary ($F(1,48) = 2.27, p = .14$) (Table 1).

Diagnosis of autism

Eighteen of the participants in the autism group met DSM-IV criteria for full autism and 7 met criteria for ASD, based on expert clinical impression and confirmed by the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter & Le Couteur, 1994) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, and Risi, 1999), which were administered by trained examiners. Participants with known genetic disorders were excluded.

Stimuli

We digitally video-recorded a woman's detailed description of baking various dessert items. We chose this topic because of its easy accessibility to all

Table 1 Descriptive characteristics of participants

	ASD ($N = 25$) Mean (St. Dev.)	Typical controls ($N = 25$) Mean (St. Dev.)
Age	14;5 (2;11)	13;7 (2;7)
Verbal IQ	109.24 (17.85)	116.96 (14.49)
Nonverbal IQ	113.20 (10.36)	113.60 (10.21)
PPVT	113.08 (20.02)	120.08 (11.50)

adolescent participants and its use of brief, simple sentences of commonly used words. The camera captured her entire face and most of her neck, excluding the rest of her body. This tight shot was chosen to maximize visual presentation of the face and especially the mouth, while eliminating other sources of communicative information, such as body lean or gestures. The eye gaze of the speaker was directed towards the camera at all times and her face was captured facing fully front for the duration of the narrative. The acoustic information was recorded by a microphone integrated with the camera to ensure synchronicity of the audio and visual channels in the captured video.

The resulting digital video file was then cut into individual clips and edited so that each clip began at the start of a complete phrase to ensure a natural and meaningful language presentation. The clips were viewed by two coders, who selected the 12 clips deemed most natural in tone, facial expression, and verbal content. Each clip was approximately 3 seconds long. We then digitally separated the audio from the video track so they could be slipped out of synch, with the audio preceding or following the video. We conducted pilot testing to determine the optimal slip direction (audio preceding or following video) and slip rate (number of frames difference between audio and video onset) for our adolescent target population. We tested slip rates of 4, 6, 8, 10, 12, and 14 frames in both slip directions. Assuming the standard video rate of 30 frames/second, a 4-frame asynchrony translates into a 120-millisecond (ms) audio delay, while a 14-frame shift represents approximately a 500ms delay. Van Wassenhove et al. (2007) established that fusion perception indicating AV integration occurred up to 267ms of audio delay. Our chosen slip rates were therefore both inside and outside the window of audio delays typically perceived as simultaneous, providing our participants with an appropriate range to demonstrate either decreased or heightened AV integration skills. Our pilot data showed that typical adolescents were able to reliably detect even an onset asynchrony of only four frames, or 120 ms when the audio preceded the video, but required at least a 10-frame or 300ms slip-rate to detect asynchrony above chance when the video preceded the audio (Figure 1). This result is confirmed

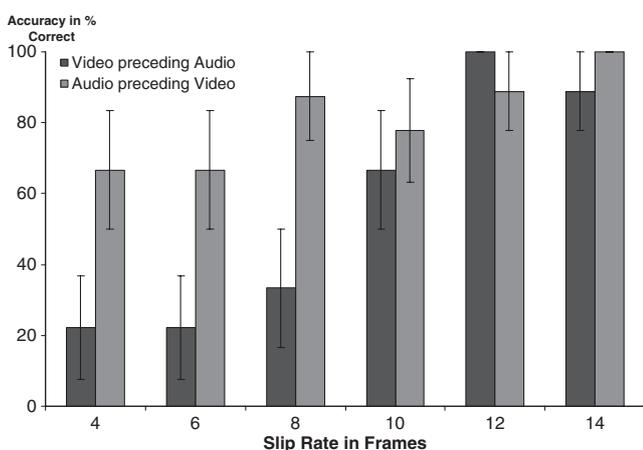


Figure 1 Pilot study accuracy in percent correct (error bars are standard error of mean)

by prior data showing a shorter window for the perception of AV synchronicity for audio preceding video (video delay), than video preceding audio (audio delay) (see Conrey & Pisoni, 2006 for review). We decided to use stimuli with audio delay, since the higher average threshold of asynchrony detection in this slip direction would enable us to capture greater individual variations in onset asynchrony accuracy.

The typical adolescents in our pilot study were able to reach above 50% accuracy for 10 frames of shift, but not 8. We therefore decided to use the midpoint between 8 and 10 frames of shift as the detection threshold and present three versions with higher slip rates (10, 12, 14) and three with lower slip rates (4, 6, 8), as well as one version of each clip in-synch, ultimately creating 7 versions of each of the 12 original clips, for a total of 84 presented stimuli. We ensured that the offset of all clips occurred simultaneously for audio and video, so that there was no cue given regarding synchronicity based on whether the sound continued after the conclusion of the video presentation. All clips were randomized and counterbalanced to ensure that two clips with the same slip rate or same verbal content did not follow each other. We created two different counterbalanced sequences and alternated their presentations between participants.

Procedure

Participants arrived at the testing site with a primary caregiver who gave informed consent. We also obtained assent of any minor participant over the age of 12. Participants were seated in front of a screen where the stimuli were presented. We explained the task to each participant, showing them several sample stimuli of in-synch and out-of-synch trials. 'In-synch' was defined as being the way people normally speak, without any artificial delays between the speech sounds and their corresponding mouth movements as one might see on television when the soundtrack slipped from the video. Participants saw the face of the speaker on the screen of a Dell laptop computer, while hearing her speak through attached loudspeakers. We refrained from using headphones, because some participants with ASD had previously been easily distracted by them. The instructions were to state whether each trial had been in-synch or not and participants' verbal responses of 'yes' or 'no' were recorded on a score sheet for each presented stimulus. Their responses were also digitally audio-recorded and the score sheet was double-checked for errors using that recording after each participant's visit had concluded.

Results

Mean accuracy scores and confidence intervals for each group's performance on all slip rates can be found in Table 2. Accuracy for in-synch trials was 97% for the ASD group and 93% for the TD group with no significant difference between the two groups ($t(48) = 1.86, p = .07$, two-tailed). These results confirm that both cohorts were able to follow the task directions and accurately identify in-synch trials.

Table 2 Means and standard deviations of accuracy levels

Slip-rate	ASD (<i>N</i> = 25)			Typical controls (<i>N</i> = 25)		
	Mean (St. Dev.)	95% Confidence		Mean (St. Dev.)	95% Confidence	
		Lower bound	Upper bound		Lower bound	Upper bound
4 frames	3.7 (5.8)	1.5	6.0	4.0 (5.4)	1.8	6.3
6 frames	13.9 (18.4)	7.8	20.1	10.1 (11.6)	4.0	16.3
8 frames	39.4 (24.4)	30.1	48.7	33.1 (21.9)	23.7	42.4
10 frames	63.8 (20.9)	55.1	72.5	64.3 (22.3)	55.6	73.0
12 frames	82.9 (13.5)	77.0	88.8	84.1 (15.7)	78.2	90.0
14 frames	89.5 (15.4)	84.4	94.6	92.0 (9.3)	86.9	97.1

After excluding the in-synch trials, we computed a 2 (group) by 6 (slip rate) repeated measures ANOVA for accuracy of onset asynchrony detection and found an expected main effect for slip rate ($F(5, 240) = 385.9, p < .0001$, partial $\eta^2 = .89$), showing that higher slip rates resulted in greater accuracy of asynchrony detection. We found no effect for group ($F(1, 48) = .09, p = .77$, partial $\eta^2 = .002$) and no group by slip-rate interaction ($F(5, 240) = .82, p = .54$, partial $\eta^2 = .017$), (Figure 2). The sphericity assumption was met. Pearson correlations revealed no significant correlations between accuracy scores and diagnostic and standardized test data in the ASD group (ADI Social: $r^2 = .002$, ADI Verbal Communication: $r^2 = .08$, ADOS Communication: $r^2 = .09$, ADOS Social: $r^2 = .005$, ADOS Combined: $r^2 = .03$, VIQ: $r^2 = .1$, NVIQ: $r^2 = .11$, Full IQ: $r^2 = .14$, PPVT: $r^2 = .14$, age: $r^2 = .003$), or the TD group (VIQ: $r^2 = .05$, NVIQ: $r^2 = .02$, Full IQ: $r^2 = .007$, PPVT: $r^2 = .02$, age: $r^2 = .13$).

To determine whether our results were skewed by a large number of participants with milder autism symptoms, we repeated the analysis using only the 18 adolescents who met criteria for full autism on the ADOS and the original 25 TD control subjects. Our results showed the same main effect for slip rate ($F(5, 205) = 300.95, p < .0001$, partial $\eta^2 = .88$) and again revealed no main effect for group ($F(1, 41) = .071, p = .79$, partial $\eta^2 = .002$) or group by

slip-rate interaction ($F(5, 205) = 1.26, p = .28$, partial $\eta^2 = .03$).

Discussion

The aim of our study was to determine whether adolescents with ASD are able to integrate auditory and visual information of meaningful language stimuli presented in a task of onset asynchrony detection. The results clearly show that, as we predicted, the participants with ASD were as accurate as their TD peers at all slip rates. The small between-group effect sizes and overlapping confidence intervals for the accuracy means strengthen the argument that the good performance of the ASD group is based on a true lack of group difference, rather than a lack of statistical power.

Studies of AV integration in ASD using McGurk-type paradigms of congruent and incongruent non-word syllables have revealed decreased influence of the visual speech component on perception in individuals with ASD compared to their TD peers (Irwin, 2006; Condouris et al., 2004; Magneé et al., in press; deGelder, Vroomen, & van der Heide, 1991), seemingly contradicting our results. However, it has been established that the temporal component of language, which is not measured by a McGurk paradigm, has a significant effect on AV integration (Shams, Kamitani, & Shimojo, 2004), even in participants who do not demonstrate a McGurk blending effect (Brancazio & Miller, 2005), indicating that AV integration goes beyond the type of perceptual blending tested in a McGurk paradigm.

The use of full, meaningful phrase-length speech, rather than single syllables may also have affected our results. Grant and Seitz (1998) used four tasks involving the recognition and AV integration of syllables and sentences in noise, quiet presentations of congruent and incongruent syllables (McGurk), and onset asynchrony of sentences in noise with a group of hard-of-hearing adults. Their aim was to determine whether the ability to integrate AV information of non-word syllables was correlated to the AV integration skills of meaningful sentences. Their results showed that AV benefit for meaningless speech sounds was not correlated with AV benefit for

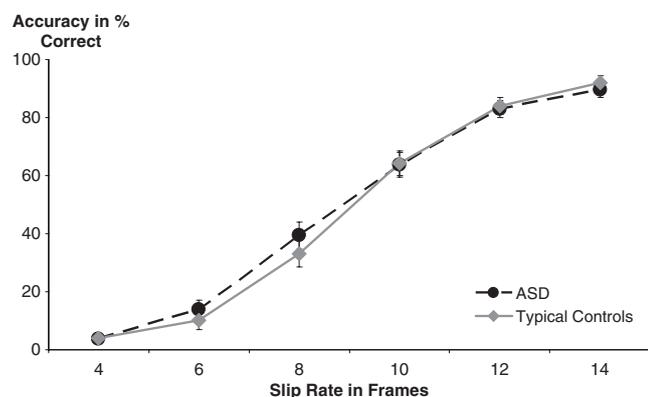


Figure 2 Accuracy by slip rate in percent correct (error bars are standard error of mean)

meaningful sentences. This distinction between the AV integration of non-word syllables, as used by McGurk-type paradigms, in contrast to meaningful phrases, may explain the differences between the impaired AV performance in participants with ASD found by studies using syllables and the preserved AV integration accuracy for meaningful, phrasal language we describe here. In a separate study, Grant and Seitz (2000) also demonstrated that language context is relevant for the accurate processing of language information. They found that a group of hard-of-hearing adults was significantly more accurate at recognizing meaningful sentences than isolated words at three different levels of background noise. These data provide strong evidence that the verbal meaningfulness of a linguistic stimulus is crucial to its accurate perception and support our result of preserved AV integration skills for meaningful sentences in adolescents with ASD.

Several studies have documented intact low-level integration of auditory and visual stimuli, using beeps and flashes (van der Smagt et al., 2007) or non-word syllables (Magnée et al., in press). This raises the question of whether the adolescents with ASD in the present study were able to detect onset asynchrony simply using those low-level, pre-phonetic integration skills. Magnée et al. (in press) discuss two levels of AV integration, an early, low-level and probably pre-phonetic effect, as well as a later ERP measure indicative of processing complex phonological stimuli, in their case a two-syllable non-word. We propose that the phrase-length stimuli used in the present study fall within the category of complex phonological stimuli and could not be adequately dealt with using only low-level, pre-phonetic processing, especially because participants could not respond to the stimuli until the entire phrase had been presented. Since there are a multitude of acoustic and visual transitions taking place over the course of these stimuli, it would be extremely difficult for participants to ignore all but the very first and very brief auditory and visual events without being influenced by the richness of the subsequent visual and auditory speech information or the overall meaning of the presented phrase, thereby requiring participants to use higher-level processing for these complex language stimuli.

There have been only a few studies of AV integration in ASD using meaningful language stimuli. Bebko et al. (2006) found decreased AV integration for linguistic stimuli in very young children (aged 4–6 years) with ASD, using a preferential looking paradigm. The children were asked to attend to two screens, both showing the same speaker's face while listening to an auditory sentence that was in-synch with only one of the faces. The researchers recorded the children's looking patterns and found no preference in the ASD group to visually attend to the face of the in-synch speaker. It is possible that the significantly larger onset asynchrony (3 seconds) between visual and auditory speech provided too much

of a sensory challenge for the young children with ASD to deal with complex language information. It is also possible that previously documented deficits in divided attention among individuals with ASD (Courchesne, Townsend, Akshoomoff, & Saitoh, 1994; Rinehart, Bradshaw, Moss, Brereton, & Walker, 2001), negatively affected the target group in Bebko et al.'s (2006) study due to the difficulty of attending to two screens that were relatively far apart (60cm). Furthermore, the age difference between these young children and the adolescents studied here makes it difficult to compare the results across these two studies.

Smith and Bennetto (2007) used complete sentence stimuli in a language-in-noise paradigm with a group of high-functioning adolescents with ASD and concluded that participants with ASD showed less improvement with bi-modal over auditory-only stimuli compared to their TD peers. They also demonstrated that individuals with ASD were significantly worse than their TD peers in the uni-modal condition of auditory-only language-in-noise recognition, which contributed significantly to their bi-modal accuracy scores. It has been established that individuals with ASD are often overly sensitive to auditory stimulation (Grandin & Scariano, 1986; Boatman, Alidoost, Gordon, Lipsky, & Zimmerman, 2001) and are worse than their TD peers in detecting language in noise (Alcántara, Weisblatt, Moore, & Bolton, 2004). When comparing the accuracy rates for the auditory-only and AV conditions, Smith and Bennetto (2007) found that both participant groups significantly improved with the addition of visual speech information. This result thus supports our claim that adolescents with ASD are able to integrate AV information for stimuli containing meaningful language information.

The aim of our experiment was to test the ability of adolescents with ASD to integrate AV speech components in a task that would allow them to derive the maximum benefit from meaningful language context, while at the same time minimizing distractions that may be related to their putative deficits in filtering out auditory background noise (Alcántara et al., 2004). We propose that presentation of meaningful language stimuli, rather than non-word syllables, in a quiet environment may have enabled our adolescent participants with ASD to accurately integrate the temporal aspects of auditory and visual language information. Our data cannot determine the underlying processing strategies of the two participant groups, but rather show that there was no difference in the behavioral accuracy for this AV integration task between adolescents with ASD and their typically developing peers. Further study of sensory processing and integration in individuals with ASD is required to determine whether the deficits and strengths for sensory integration found in different experiments are based on underlying cognitive profiles, differences in neurological processing, or artifacts of methodology, specifically the different types of AV integration skills

tapped by studies of non-words syllables vs. meaningful language, and language-in-noise vs. onset asynchrony detection.

Conclusion

Adolescents with ASD can detect onset asynchrony in phrase-length language stimuli. We propose that meaningful linguistic context presented without background noise provides a more successful basis for AV integration in this population than single non-word syllables, or competitive auditory stimulation. These results could influence our approaches to intervention by focusing on presentation of sentence-length AV language in non-competitive sensory backgrounds.

Acknowledgements

The authors wish to thank Chris Connolly, Sarah Delahanty, Danielle Delosh, Bob Duggan, Alex B.

Fine, Keri Green, Alex Griswold, Robert M. Joseph, Meaghan Kennedy, Maria Kobrina, Janice Lomibao, and Toby McElheny for their assistance in stimulus creation, task administration, and data analysis. We also thank the children and families who gave their time to participate in this study.

Funding was provided by NAAR, NIDCD (U19 DC03610; H. Tager-Flusberg, PI) which is part of the NICHD/NIDCD Collaborative Programs of Excellence in Autism, and by grant M01-RR00533 from the General Clinical Research Ctr. program of the National Center for Research Resources, National Institutes of Health.

Correspondence to

Ruth B. Grossman, Lab of Developmental Cognitive Neuroscience, Boston University School of Medicine, 715 Albany Street, L-814, Boston, Massachusetts, 02118, USA; Tel: (6 17) 414 1300; Fax: (617) 414 1301; Email: ruthberg@bu.edu

- Individuals with autism spectrum disorders (ASD) have been reported to have difficulty integrating multi-modal sensory information.
- Many studies of auditory-visual (AV) integration in ASD use non-word syllables as stimuli, rather than meaningful language.
- Our study investigated AV integration of meaningful full-sentence stimuli using onset-asynchrony detection.
- Adolescents with ASD were as accurate as their typically developing peers in completing this task.
- Clinical interventions for individuals with ASD should not shy away from using AV integrated language and sentence-length tokens.

References

- Alcántara, J.I., Weisblatt, E.J.L., Moore, B.C.G., & Bolton, P.F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*, 1107–1114.
- Bebko, J.M., Weiss, J.A., Demark, J.L., & Gomez, P. (2006). Discrimination of temporal synchrony in inter-modal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, *47*, 88–98.
- Boatman, D., Alidoost, M., Gordon, B., Lipsky, F., & Zimmerman, W. (2001). Tests of auditory processing differentiate Asperger's syndrome from high-functioning autism. *Annals of Neurology*, *50*, S95.
- Brancazio, L., & Miller, J.L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception and Psychophysics*, *67*, 759–769.
- Calvert, G.A., Brammer, M.J., & Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Science*, *2*, 247–253.
- Castelli, F., Frith, C., & Happé, F. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*, 1839–1849.
- Condouris, K., Joseph, R.M., Ehrman, K., Connolly, C., & Tager-Flusberg, H. (2004). *Facial speech processing in children with autism*. Third International Meeting for Autism Research, Sacramento, CA.
- Conrey, D., & Pisoni, D.B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America*, *119*, 4065–4073.
- Courchesne, E., Townsend, J., Akshoomoff, N.A., & Saitoh, O. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience*, *108*, 848–865.
- deGelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*, 69–86.
- Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test* (3rd edn). Circle Pines, MN: American Guidance Service.
- Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Basil Blackwell.
- Frith, U., & Happé, F. (1994). Autism: Beyond 'theory of mind.' *Cognition*, *50*, 115–132.
- Grandin, T., & Scariano, M.M. (1986). *Emergence: Labeled autistic*. New York: Arena Press.
- Grant, K.W., & Seitz, P.F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America*, *104*, 2438–2450.

- Grant, K.W., & Seitz, P.F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *Journal of the Acoustical Society of America*, 107, 1000–1011.
- Grossman, R.B., & Tager-Flusberg, H. (in press). Reading faces for information about words and emotions in adolescents with autism. *Research in Autism Spectrum Disorders*.
- Happé, F. (2005). The weak central coherence account of autism. In F.R. Volkmar, A. Klin, R. Paul, & D.J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders, Vol. 1: Diagnosis, development, neurobiology, and behavior* (3rd edn, pp. 640–649). New York: John Wiley & Sons.
- Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, 36, 77–90.
- Irwin, J.R. (2006). Audiovisual speech integration in children with autism spectrum disorders. *Journal of the Acoustic Society of America*, 120(5, Pt 2), 3348.
- Just, M.A., Cherkassy, V.L., Keller, T.A., & Minshew, N.J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain*, 127, 1811–1821.
- Kaufman, A., & Kaufman, N. (2004). *Manual for the Kaufman Brief Intelligence Test Second Edition*. Circle Pines, MN: American Guidance Service.
- Lord, C., Rutter, M., DiLavore, P.C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Magnée, M.J.C.M., de Gelder, B., van Engeland, H., & Kemner, C. (in press). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry*.
- Massaro, D.W. (1984). Children's perception of visual and auditory speech. *Child Development*, 55, 1777–1788.
- Massaro, D.W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal on Development and Learning Disorders*, 7, 111–144.
- McGurk, H., & MacDonald, J.W. (1976). Hearing lips and seeing voices. *Nature*, 2, 746–748.
- Osterling, J.A., Dawson, G., & Munson, J.A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, 14, 239–251.
- Rinehart, N.J., Bradshaw, J.L., Moss, S.A., Brereton, A.V., & Walker, P. (2001). A deficit in shifting attention present in high-functioning autism but not Asperger's disorder. *Autism*, 5, 67–80.
- Russo, N., Flanagan, T., Iarocci, G., Berringer, D., Zelazo, P.D., & Burack, J.A. (2007). Deconstructing executive deficits among persons with autism: Implications for cognitive neuroscience. *Brain and Cognition*, 65, 77–86.
- Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulation of visual perception by sound. In G. Calvert, C. Spence, & B.E. Stein (Eds.), *The handbook of multisensory processes* (pp. 26–33). Cambridge, MA: MIT Press.
- Smith, E.G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 48, 813–821.
- Sunby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.
- Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology. A*, 36, 51–74.
- Townsend, J., Harris, N.S., & Courchesne, E. (1996). Visual attention abnormalities in autism: Delayed orienting to location. *Journal of the International Neuropsychological Society*, 2, 541–550.
- van der Smagt, M.J., van Engeland, H., & Kemner, C. (2007). Can you see what is not there? Low-level auditory-visual integration in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 37, 2014–2019.
- van Wassenhove, V., Grant, K.W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45, 598–607.
- Watling, R.L., Deitz, J., & White, O. (2001). Comparison of sensory profile scores of young children with and without autism spectrum disorders. *American Journal of Occupational Therapy*, 55, 416–423.
- Williams, J.H.G., Massaro, D.W., Peel, N.J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities*, 25, 559–575.

Manuscript accepted 25 July 2008