Journal of Child Psychology and Psychiatry 48:8 (2007), pp 813–821 doi:10.1111/j.1469-7610.2007.01766.x

Audiovisual speech integration and lipreading in autism

Elizabeth G. Smith and Loisa Bennetto

Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY, USA

Background: During speech perception, the ability to integrate auditory and visual information causes speech to sound louder and be more intelligible, and leads to quicker processing. This integration is important in early language development, and also continues to affect speech comprehension throughout the lifespan. Previous research shows that individuals with autism have difficulty integrating information, especially across multiple sensory domains. **Methods:** In the present study, audiovisual speech integration was investigated in 18 adolescents with high-functioning autism and 19 well-matched adolescents with typical development using a speech in noise paradigm. Speech reception thresholds were calculated for auditory only and audiovisual matched speech, and lipreading ability was measured. **Results:** Compared to individuals with typical development, individuals with autism showed less benefit from the addition of visual information in audiovisual speech perception. We also found that individuals with autism were significantly worse than those in the comparison group at lipreading. Hierarchical regression demonstrated that group differences in the audiovisual condition, while influenced by auditory perception and especially by lipreading, were also attributable to a unique factor, which may reflect a specific deficit in audiovisual integration. **Conclusions:** Combined deficits in audiovisual speech integration and lipreading in individuals with autism are likely to contribute to ongoing difficulties in speech comprehension, and may also be related to delays in early language development. **Keywords:** Speech reception threshold, speech in noise, audiovisual speech integration, autism. **Abbreviations:** SNR: speech to noise ratio; SRT: speech reception threshold.

One of the hallmarks of autism is an impairment in communication, which can range from severe delays in language development to relatively intact language accompanied by problems with functional communication (Tager-Flusberg, Paul, & Lord, 2005). Perception of speech is a particular aspect of communication that may be altered in autism, and further investigation of this domain may shed light on the development of communication deficits. For example, understanding a person's speech often requires that listeners integrate information from the speaker's voice, lips, face, and body. This audiovisual speech integration increases identification and comprehension of the information being communicated (Calvert, Brammer, & Iverson, 1998). However, individuals with autism often show deficits in crossmodal integration (Iarocci & McDonald, 2006), which might put them at a disadvantage during speech perception.

Audiovisual speech integration in typical development

Audiovisual speech perception has primarily been investigated in typical development using the 'McGurk effect' paradigm (McGurk & Macdonald, 1976). In this paradigm, unisyllabic or disyllabic, nonword utterances are presented either visually (i.e., individual sees model's lips move without sound), auditorily (i.e., individual hears utterance without visual information), or audiovisually (i.e., individual hears utterance and sees model's lips move). Results from McGurk's initial work showed an interesting effect when mismatching auditory and visual stimuli were presented together: the reported percept sometimes represented a fusion between the auditory and visual modes (e.g., auditory/ba/and visual/ga/are perceived as/da/). Studies using the McGurk effect have shown that multisensory speech integration is mandatory and unmodulated by attention (Soto-Faraco, Navarra, & Alsius, 2004). In addition, Driver (1996) used the ventriloquist effect to show that audiovisual information guides attention and also that it is processed prior to attentional modulation. Studies using the McGurk effect have also shown that audiovisual speech perception is present in very young infants and plays an important role in speech production (Desjardins, Rogers, & Werker, 1997; Patterson & Werker, 1999). Audiovisual speech continues to assist older child and adult listeners in comprehension of speech in daily social situations.

Audiovisual integration in autism

There is evidence that individuals with autism have difficulty integrating information across auditory and visual modes (see Iarocci & McDonald, 2006 for

Conflict of interest statement: No conflicts declared.

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

a review), including matching voices to faces (Boucher, Lewis, & Collis, 1998; Loveland et al., 1995), forming crossmodal associations between sound beeps and light flashes (Martineau et al., 1992), discriminating temporal synchrony of audiovisual speech (Bebko, Weiss, Demark, & Gomez, 2006), and blending auditory and visual speech (Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004). However, deficits in either auditory or visual perception alone might account for differences in multimodal integration in autism (Čeponienė et al., 2003; Williams et al., 2004).

Until recently, relatively little research has explored audiovisual integration of speech in autism. Investigation of the McGurk effect has shown that children and adolescents with autism report fewer fusions than typical children, reflecting that children with autism are less likely to take the non-matching, visual syllable into account during speech perception (de Gelder, Vroomen, & van der Heide, 1991). Williams and colleagues (2004) replicated this finding, and examined the contributions of unisensory components to this deficit. When visual accuracy (i.e., lipreading) was controlled for, individuals with autism were no longer significantly worse than those in the comparison group in audiovisual integration. Thus, the essential step to be taken in this field requires characterization of both unisensory and multisensory speech perception in autism using ecologically valid stimuli.

Speech in noise paradigm

Ecological validity can be improved by employing a paradigm that is correlated with everyday perceptual challenges, such as the speech in noise paradigm. Speech is often heard in varying levels of background noise (e.g., at a loud party), requiring listeners to filter the background noise out of the speech. Individuals with autism appear to have a relative weakness in understanding speech presented in background noise compared to individuals with typical development (Alcantara, Weisblatt, Moore, & Bolton, 2004).

While typical individuals are better at understanding speech in noise, they are also able to use visual information to enhance the auditory signal of speech presented in noise (Schwartz, Berthommier, & Savariaux, 2004). The addition of visual information does not simply add unimodal information; the visual information actually enhances the individual's ability to perceive the auditory information. Thus, the ability to use visual information when listening to speech (i.e., audiovisual speech perception) is linked to the ability to perceive and comprehend speech in noise (Rudmann, McCarley, & Kramer, 2003). If, indeed, individuals with autism experience deficits in both audiovisual speech perception and speech in noise perception, these deficits could produce an additive, deleterious effect on comprehension in everyday situations.

In the current study, we investigated whether individuals with autism were capable of using visual information to enhance an auditory signal embedded in background noise. Based on previous research, we predicted that individuals with autism would be worse at processing audiovisual speech in background noise, and that these deficits in integration would not be explained by auditory or visual processing deficits alone.

Methods

Participants

Participants were 18 adolescents with autism and 19 adolescents with typical development matched by group on chronological age, gender, Full Scale IQ, and the Receptive Language Index (RLI) from the Clinical Evaluation of Language Essentials, 4th Ed. (CELF-4; see Table 1). Since there is a small verbal memory compon ent in the speech in noise paradigm, we also administered the Recalling Sentences subtest from the CELF-4 to ensure that all individuals were able to recall sentences at least as long as those presented in the stimuli.

Diagnoses of autism were confirmed in the autism group and ruled out in the comparison group with a combination of the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999) and the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003). Individuals with autism were excluded if they had a diagnosis of a genetic syndrome (e.g., fragile X syndrome) or a definable postnatal etiology for their developmental symptoms

Table 1 Descriptive characteristics of the autism and comparison groups

	Autism Group *M (SD)* [range]	Comparison Group *M (SD)* [range]	*F* or χ^2	*p*
n	18	19		
Age	15.84 (2.17) [12.42–19.50]	16.08 (2.04) [12.00–19.17]	.11	.73
CELF-4 RLI[a]	104.05 (8.72) [88–119]	104.26 (5.48) [96–117]	.01	.93
FSIQ[b]	108.1 (14.23) [77–129]	112.37 (8.81) [94–124]	1.24	.27
Gender (M:F)	13:5	15:4	.23	.63

[a]Clinical Evaluation of Language Fundamentals, 4th Ed. Receptive Language Index.
[b]FSIQ was measured with the WISC-IV or WAIS-III.

(e.g., head trauma). Individuals in the comparison group were excluded if there were concerns about learning disabilities, mental retardation, language delays, head trauma, or other psychiatric conditions, or if there were concerns about autism spectrum disorders in their first- or second-degree relatives.

All participants were native English speakers, and had normal or corrected vision and hearing acuity within the range required for speech perception. Visual acuity was evaluated for each eye using the Snellen chart. Auditory acuity was measured separately for each ear with an audiometer (Maico MA32 Advanced Diagnostic Audiometer). All participants were capable of hearing tones represented at frequencies of 1000–4000 hertz and loudness of 25 decibels.

This research was approved by the University of Rochester's Research Subjects Review Board. Prior to testing, written informed consent was obtained from parents and from individuals aged 18 or 19; written assent was also obtained from younger adolescents.

Measures

Stimuli were short sentences (5–7 words long, 4 seconds in duration) containing three key words (e.g., The **cat jumped** over the **fence**). A person's response was considered correct when he or she was able to correctly report all three key words for a trial. Sentences were presented in Auditory Only and Audiovisual conditions, and were presented in a speech in noise paradigm. Manipulating the loudness of speech relative to background noise in a systematic way reveals a particular speech to noise ratio (SNR) for each condition. Differences in SNR across conditions reveal the benefit provided by the addition of visual information. In addition, a lipreading condition evaluated perception of visual information presented alone.

We used 48 sentences from Rosenblum, Johnson, and Saldana's (1996) list, which was composed for American English listeners and designed to provide a balance of lexical and semantic clarity across the sentences. Similar sentence lists have been used to measure speech in noise perception in children, including those with specific language impairments (Stollman, van Velzen, Simkens, Snik, & van den Broek, 2003, 2004).

The speakers for the target sentences included five females between the ages of 23 and 28 years. Previous research has shown that younger women are the easiest to lipread and understand auditorily (Bench, Daly, Doyle, & Lind, 1995). We used several speakers to minimize learning across the session. Each of the speakers was recorded speaking 9 or 10 sentences in a sound-attenuated chamber with a professional-quality digital video camera equipped with a unidirectional microphone. Movements of the neck and throat can aid audiovisual speech perception (Thomas & Jordan, 2004), so our speakers wore a black turtleneck to cover these areas.

Background noise. Four additional females were recorded reading excerpts from children's books. All articulatory sounds were low-pass filtered out from the speech by removing frequencies containing segmental content with Praat, a computer phonetics program (Boersma & Weenink, 2006). The filtered streams were then overlapped in Final Cut Pro and divided into 48, 4-second blocks. Since the type of information typically available in background noise (e.g., temporal and spectral dips) differentially affects speech intelligibility for individuals with autism (Alcantara et al., 2004), combining multiple speech streams and removing segmental content from articulatory sounds yielded background noise that was equally difficult for both groups. Background noise was presented at 70 decibels across all stimuli and conditions.

Pilot study to determine final sentence lists. We piloted all 48 sentences with 16 young adults to determine the average auditory SNR for each sentence when presented without visual information. Each sentence was presented first at the quietest, or most difficult level (i.e., 40 dB), resulting in an SNR of −30. The speech stream volume was then raised by 2 dB steps until the individual could accurately report all three key words. This allowed us to determine the average SNR at which each of the 48 sentences was intelligible across subjects.

To determine lipreadability, the 48 sentences were shown to a different group of 15 young adults, who were asked to report (or guess) any words they could lipread. The average number of correctly identified key words yielded a mean lipreading accuracy score for each sentence.

We used our piloting data to balance our five speakers across conditions and also to ensure homogeneity of our lists (see Table 2). We first constructed the list for the Lipreading condition from the 12 sentences with the highest average lipreadability scores. This resulted in a list with a high probability of evoking lipreading abilities for all individuals and a low probability of a floor effect. The remaining three lists of 12 sentences each were used in the Auditory Only and Audiovisual conditions, and were matched in terms of average SNR and

Table 2 Characteristics of sentence stimuli

	List 1 *M* (*SD*) [range]	List 2 *M* (*SD*) [range]	List 3 *M* (*SD*) [range]	List 4 *M* (*SD*) [range]	*F*	*p*
Speech to noise ratio (SNR)	19.32 (3.43) [14.85–25.18]	19.10 (2.47) [13.35–22.68]	18.08 (3.86) [10.16–24.02]	20.55 (3.34) [14.46–25.23]	1.12	.35
Lipreading	17.67 (5.45) [11.20–26.40]	5.20 (4.32) [0–14.4]	5.13 (4.9) [0–15.2]	4.87 (4.14) [1.60–15.20]	.02	.98

Note: Data were obtained from two separate groups of young adults in the stimulus development pilot study ($n = 16$ for SNR; $n = 15$ for lipreading). Each list included 12 unique sentences. List 1 was used for the Lipreading condition; Lists 2–4 were used for the Auditory Only and Audiovisual conditions. The SNR is the level at which key words in sentences were intelligible, averaged across sentences. The Lipreading score was the average number of key words lipread across the 12 sentences (of 36 maximum). The ANOVA for Lipreading did not include List 1, since this list was designed to be easier to lipread.

lipreading accuracy scores. Three presentation sets were constructed, distributing the three balanced lists across the two conditions, so that each person received a different list for each condition. The presentation sets were distributed equally across groups; later analyses revealed no effects of presentation set on performance.

Procedure

Participants were seated 30 inches in front of a 21-inch LCD monitor and professional-quality speakers in a quiet room. Sentences were presented with DirectRT software in three conditions in the following fixed order: Auditory Only, Audiovisual, Lipreading. For the Auditory Only and Audiovisual conditions, the MacLeod and Summerfield (1990) method was used to obtain a speech reception threshold (SRT). The first sentence in each list was presented beginning at the hardest level (SNR = −30, i.e., background noise presented at 70 dB and speech presented at 40 dB). Participants were asked to report any words they had heard. The loudness of the speech stream was increased in 2 dB steps until the person's report included all three key words. The remaining 11 sentences were presented only once via an up–down adaptive staircase procedure to converge on the SRT. For example, the SNR at which the second sentence was presented was decreased by 2dB from the loudness required to hear the first criterion sentence. If the participant could identify all three key words, the next sentence was presented 2 dB lower; if not, the next sentence was presented 2 dB higher. The SRT for each condition was estimated as the average of the SNRs at which sentences 4–13 were presented (the SNR of the 13th sentence can be deduced from performance on the 12th sentence). The SRT revealed by this method provides an efficient and reliable threshold estimation of the SNR for each condition at which an individual is able to report all three key words 50% of the time (e.g., an SRT of −20 indicates that the individual could report the three key words correctly 50% of the time when speech was 20 dB quieter than noise). In order to counteract learning effects across each condition, the 12 sentences in the Auditory Only and Audiovisual conditions were ordered by SNRs from the pilot data in ascending order.

For the Lipreading condition, 12 sentences were shown and the number of key words the individual correctly lipread for each sentence was recorded. This lipreading measure has been shown to have a test–retest reliability greater than .90 (Summerfield, 1992).

Results

All dependent variables were examined for deviations from the required assumptions of normality and sphericity. Effect sizes were calculated with partial eta squared (η^2_{partial}). Values between .01 and .06 indicate a small effect, between .06 and .14 a medium effect, and above .14 a large effect.

Auditory Only and Audiovisual conditions

Our primary hypothesis was that individuals with autism would not benefit from the addition of visual information to auditory speech. We analyzed this initially by examining SRTs in a mixed-model multivariate analysis of variance (MANOVA), with group as the between-subjects factor and condition (Auditory Only, Audiovisual) as the within-subjects factor (see Figure 1). As expected, we found a strong main effect of condition, $F(1,35) = 141.54$, $p < .001$, $\eta^2_{\text{partial}} = .80$, with the Auditory Only condition associated with higher SRTs than the Audiovisual condition. This very large effect size strongly supports the validity of our experimental design, and indicates that speech had to be louder (relative to background noise) in the Auditory Only condition than in the Audiovisual condition for individuals to correctly report the sentence. We also found a main effect of group, $F(1,35) = 16.02$, $p < .001$, $\eta^2_{\text{partial}} = .31$, with individuals with autism requiring higher SRTs across conditions. Consistent with our hypothesis, we found a significant group × condition interaction, $F(1,35) = 11.94$, $p = .001$, $\eta^2_{\text{partial}} = .25$.

Follow-up analyses of variance (ANOVAs) on both conditions indicated that there was no difference in the groups' SRTs in the Auditory Only condition, $F(1,35) = 1.06$, $p = .31$, $\eta^2_{\text{partial}} = .03$. However, there was a significant group difference for the Audiovisual condition, $F(1,35) = 30.46$, $p < .001$, $\eta^2_{\text{partial}} = .46$. The group with autism had significantly higher SRTs in this condition (mean = −22.35 dB, $SD = 2.36$) compared to the comparison group (mean = −26.05 dB, $SD = 1.68$), indicating that the comparison group could understand quieter speech in the multimodal condition.

To test whether the group with autism showed any improvement in performance with the addition of visual information, we performed paired *t*-tests on the two conditions for each group. These analyses

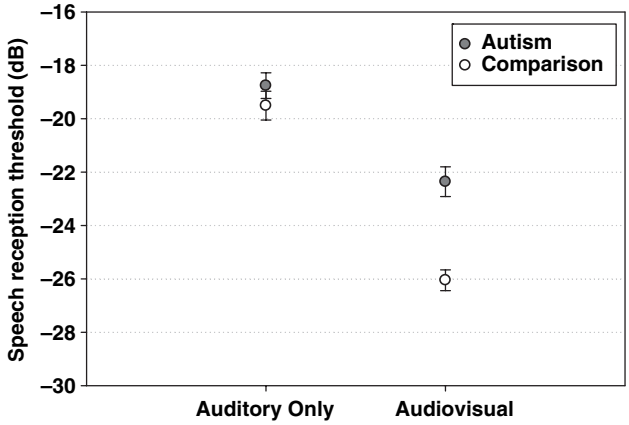

Figure 1 Mean speech reception thresholds for auditory only and audiovisual speech. Speech reception thresholds represent the speech to noise ratio at which individuals correctly reported the 3 key words 50% of the time, with background noise presented at a constant 70 dB. More negative values indicate better performance (i.e., individual accurately reported the speech signal when it was presented at quieter levels). Error bars represent standard error of the mean

demonstrated that both groups performed better in the Audiovisual condition relative to the Auditory Only condition: Autism group, t (17) = 6.53, p < .001; Comparison group, t (18) = 10.15, p < .001. Thus, although individuals with autism were differentially worse on the Audiovisual condition relative to individuals with typical development, they were able to accurately identify the key words at lower thresholds when they also saw the model's lips move, compared to when they only heard the sentences.

Lipreading

A one-way ANOVA showed that Lipreading scores for individuals with autism were significantly lower than for individuals with typical development, F (1,35) = 21.68, p < .001, η^2_{partial} = .38. On average, individuals with autism were capable of lipreading 14% of the key words presented, while those with typical development lipread an average of 39% (see Figure 2). As predicted by our pilot data, there were no ceiling effects for this condition in either group; no individuals lipread more than 23 of the 36 words (64%). However, one individual with autism could not lipread any words.

To examine the independent influences of unisensory abilities on Audiovisual SRT, we used a hierarchical multiple regression procedure. Auditory Only SRT and Lipreading scores were entered at step 1, and group membership was entered at step 2. The model was evaluated for multivariate outliers using Cook's distance (Cook & Weisberg, 1982); no cases were removed. The results from this analysis are presented in Table 3. The final model, which included all three predictors, accounted for 59% of the variance in audiovisual speech perception, F (3,33) = 15.62, p < .0001. Lipreading was the strongest predictor, but auditory perception also contributed a significant amount of unique variance. Group was entered at the final step, and accounted for significant, 11% of unique variance beyond what was predicted by the unisensory factors. Together, these data suggest that, while lipreading deficits are clearly implicated in group differences, an additional, unique component, such as a deficit in preattentive integration, contributes to audiovisual speech perception difficulties in autism.

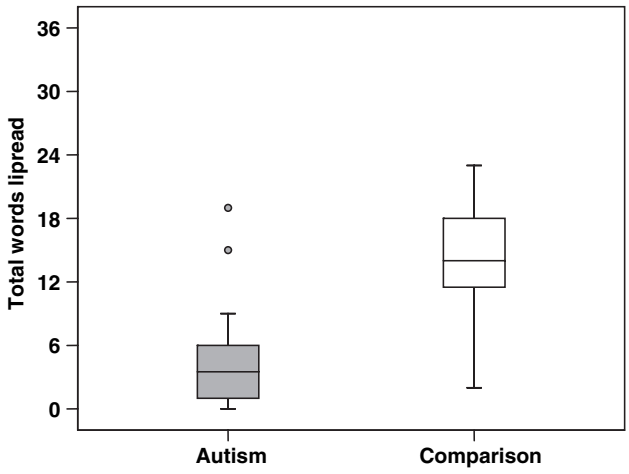

Figure 2 Mean lipreading scores. Boxplots show the group medians and interquartile ranges for the total number of key words reported in the visual only condition (36 possible words)

Table 3 Regression analyses predicting audiovisual speech reception threshold (SRT) from unisensory factors

Variable	β	p	R^2	ΔR^2	$F\Delta$	p
Step 1						
Lipreading	−.65	<.0001	.42	.42	25.08	<.0001
Step 2						
Lipreading	−.63	<.0001				
Auditory Only SRT	.26	.04	.48	.07	4.24	.04
Step 3						
Lipreading	−.37	.01				
Auditory Only SRT	.20	.08				
Group	.42	.007	.59	.11	8.37	.007

Note: Variables at steps 1 and 2 were entered in the order of highest statistical significance.

Discussion

The results of this study provide evidence of an audiovisual integration impairment in autism. While the comprehension of speech in noise of both groups improved with the addition of visual information, this improvement was markedly stronger for individuals with typical development compared to those with autism. We also found that individuals with autism were significantly less skilled on a lipreading task that was closely matched to the audiovisual paradigm. Regression analyses showed that even after accounting for the variance due to unisensory factors, the group differences in audiovisual speech remained.

Unisensory auditory and visual processing in autism

As predicted, we found no group differences in the SRTs for the Auditory Only condition. This result was expected given the way the stimuli were engineered; neither spectral nor temporal dips were routinely present, making the stimuli equally difficult for both groups (Alcantara et al., 2004).

Our finding of impaired lipreading in autism is consistent with some previous evidence of lipreading deficits in autism. For example, Williams and colleagues (2004) found that children and adolescents with autism were worse at lipreading consonant–vowel syllables than an age-matched group with typical development. However, de Gelder and colleagues (1991) did not find differences in lipreading vowel–consonant–vowel syllables in children

with autism compared to children with Down syndrome. The increased difficulty of our task, which required lipreading whole words, may have made it more sensitive to lipreading differences, and thus may explain the discrepant findings across these studies.

Audiovisual integration in autism

In the present study, individuals with typical development showed a significantly greater benefit in the Audiovisual condition compared to those with autism. The very large effect size ($\eta^2_{\text{partial}} = .46$) suggests that this group difference is particularly robust, and is especially notable when compared to the small effect found in the Auditory Only condition ($\eta^2_{\text{partial}} = .03$). This group difference in audiovisual processing may reflect an impairment in the super-additive process of multisensory integration in individuals with autism. Previous research indicates that such a deficit would affect the non-conscious enhancement of the auditory signal (Soto-Faraco et al., 2004). This would likely occur via the temporal cues provided by lip movements (Fingelkurts, Fingelkurts, Krause, Mottonen, & Sams, 2003), and is consistent with recent evidence of poor discrimination of temporal synchrony for linguistic stimuli in autism (Bebko et al., 2006). While it is possible that this difference is related to other, linguistic strategies (e.g., deriving words from semantic context), such an effect would have led to group differences in the Auditory Only condition and also would not explain previous findings of audiovisual speech impairments in autism (de Gelder et al., 1991; Williams et al., 2004).

Implications for the neurobiology of autism

Our finding of an audiovisual integration deficit in autism may illuminate research on the neurobiology of autism. Studies of typical adults have suggested that super-additive audiovisual speech integration activates the superior temporal sulcus (STS; Calvert, 2001). In individuals with autism, the STS is activated differently during several social cognition tasks, including perception of biological motion (Pelphrey et al., 2003), attribution of mental states (Castelli, Frith, Happé, & Frith, 2002), perception of faces (McCarthy, Puce, Belger, & Allison, 1999), and speech-sound detection (Boddaert et al., 2004). Electrophysiological studies of typical audiovisual integration have also shown that the information gained from lips is largely temporal (Kim & Davis, 2004), and that typical perception of audiovisual speech is associated with cortical temporal synchrony and thus depends on the integrity of connections between brain regions (Fingelkurts et al., 2003). Imaging studies in persons with autism using diffusion tensor imaging to study white matter tracts have found decreased connectivity between the STS and other integrative centers (Brambilla et al., 2004), indicating that the temporal aspects of audiovisual speech integration may be affected in this way. Finally, lipreading and audiovisual speech have been associated with activation of the mirror neuron system, a network that responds not only during self-initiated action but also when observing another person performing the action (Watkins, Strafella, & Paus, 2003). Atypical activation of mirror neurons has also been proposed as a primary neuropathological feature of autism (Dapretto et al., 2006; Williams et al., 2006).

A critical next step to test these associations is to investigate neural correlates of audiovisual speech integration in autism. Functional MRI studies of related processes (e.g., action observation) have found decreased activity within the mirror neuron system (Dapretto et al., 2006); an important question is whether dysfunction in this or related networks may also underlie autism-specific impairments in multimodal speech processing. Similarly, fMRI investigations of neural activity in autism during auditory, visual, and audiovisual speech perception, as well as temporal and spatial coordination during such activities, will further elucidate the possible role of decreased connectivity in an audiovisual integration impairment.

Clinical implications

An impairment in audiovisual integration would have a significant impact on daily functioning for individuals with autism, beyond the challenges inherent to the unisensory auditory processing impairments shown by others (Alcantara et al., 2004; Boddaert et al., 2004; Čeponienė et al., 2003). For example, in a classroom setting, children with autism are likely to perceive and understand less of what is spoken by their teachers compared to their typical peers, even if their attention and verbal conceptual skills are intact. Further research might explore ways in which teachers and parents can modify their speech in order to improve comprehension of children with autism, as has been done for children with specific language impairments (Bradlow, Kraus, & Hayes, 2003). Future research might also investigate whether lipreading training can help individuals with audiovisual speech perception. For example, individuals with autism trained to recognize lip movements for consonant–vowel syllables improved their audiovisual speech perception for the same syllables (Williams et al., 2004). However, teaching lipreading at the level of whole words and sentences is considerably more challenging (Summerfield, 1992).

Previous research on typical development has shown that audiovisual speech integration is present in very young infants (e.g., as early as 2 months; Patterson & Werker, 2003), and is robust in 4.5-month-old infants (Patterson & Werker, 1999). The ability to attend to matching audiovisual speech may

become especially important by the end of the first year of life, when babies become experts at selectively attending to important speech information, despite background noise (Newman, 2005). In addition to directing attention, there is evidence that audiovisual speech perception is related to spontaneous babbling in infants and speech production in preschoolers (Desjardins et al., 1997; Patterson & Werker, 1999). Latent deficits in audiovisual speech perception in autism might therefore be involved in atypical or delayed language development. Preliminary data from the present study suggest this may be a promising avenue for further investigation: a correlational analysis between Audiovisual SRT and age of first word in the autism group yielded a marginal positive relationship, $r\,(16) = .40$, $p < .10$. In order to examine the role that an audiovisual integration deficit might play in the pathogenesis of communication impairments in autism, the ability of very young children with autism to perceive audiovisual speech must be investigated. In addition, studies of high-risk infant siblings of children with autism could examine audiovisual speech integration at a critical, earlier time in development (e.g., Patterson & Werker, 1999).

Limitations

The present findings are based on the performance of verbal, high-functioning individuals with autism. The role of audiovisual speech integration in atypical language development could also be addressed in lower-functioning individuals with autism, particularly those with limited verbal skills. In fact, comparing these individuals' performance with the performance of higher-functioning individuals with autism might elucidate any relationships between language development and audiovisual speech integration, and may help to further define endophenotypes within the autism spectrum (Kjelgaard & Tager-Flusberg, 2001). The present findings were also based on a typically-developing comparison group. To determine whether audiovisual speech integration deficits are specific to autism, the abilities of individuals with other developmental delays, and particularly those with other language impairments, should be investigated.

Conclusion

The results of this study suggest that along with a lipreading deficit, individuals with autism have an impairment in audiovisual speech integration. This impairment results in speech that sounds less intelligible to individuals with autism than their typically-developing peers. When combined with the differences in auditory speech perception in noise shown by others, these difficulties may impact early language development, and likely contribute to the challenge of everyday speech comprehension for individuals with autism spectrum disorders.

Acknowledgements

We are very grateful to the children and families who participated in this study. This research was supported by an NIMH Center for Studies to Advance Autism Research and Treatment (U54 MH066397) and an NIH General Clinical Research Center Grant (5 M01 RR00044).

Correspondence to

Loisa Bennetto, Department of Clinical & Social Sciences in Psychology, University of Rochester, Box 270266, Rochester, NY 14627, USA; Tel: 585-275-8712; Fax: 585-273-1100; Email: bennetto@psych.rochester.edu

References

Alcantara, J.I., Weisblatt, E.J.L., Moore, B.C.J., & 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 intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, *47*, 88–98.

Bench, J., Daly, N., Doyle, J., & Lind, C. (1995). Choosing talkers for the BKB/A Speechreading test: A procedure with observations on talker age and gender. *British Journal of Audiology*, *29*, 172–187.

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., Mouren-Simeoni, M.-C., Philippe, A., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Perception of complex sounds in autism: Abnormal auditory cortical processing in children. *American Journal of Psychiatry*, *161*, 2117–2120.

Boersma, P., & Weenink, D. (2006). *Praat: Doing phonetics by computer* (Version 4.4.30).

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry*, *39*, 171–181.

Bradlow, A.R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*, 80–97.

Brambilla, P., Hardan, A.Y., di Nemi, S.U., Caverzasi, E., Soares, J.C., Perez, J., & Barale, F. (2004). The functional neuroanatomy of autism. *Functional Neurology*, *19*, 9–17.

Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.

Calvert, G.A., Brammer, M.J., & Iverson, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, *2*, 247–253.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*, 1839–1849.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Science*, *100*, 5567–5572.

Cook, R.D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28–30.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*, 69–86.

Desjardins, R.N., Rogers, J., & Werker, J.F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*, 85–110.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lipreading. *Nature*, *381*, 66–68.

Fingelkurts, A.A., Fingelkurts, A.A., Krause, C.M., Mottonen, R., & Sams, M. (2003). Cortical operational synchrony during audio-visual speech integration. *Brain and Language*, *85*, 297–312.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, *36*, 77–90.

Kim, J., & Davis, C. (2004). Investigating the audio-visual speech detection advantage. *Speech Communication*, *44*, 19–30.

Kjelgaard, M.M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*, 287–308.

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

Loveland, K.A., Tunali-Kotoski, B., Chen, R., Brelsford, K.A., Ortegon, J., & Pearson, D.H. (1995). Intermodal perception of affect in persons with autism or Down syndrome. *Development and Psychopathology*, *7*, 409–418.

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*, 29–43.

Martineau, J., Roux, S., Adrien, J., Garreau, B., Barthelemy, C., & Lelord, G. (1992). Electrophysiological evidence of different abilities to form cross-modal associations in children with autistic behavior. *Electroencephalography and Clinical Neurophysiology*, *82*, 60–66.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, *9*, 431–444.

McGurk, H., & Macdonald, J.W. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Newman, R.S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, *41*, 352–362.

Patterson, M.L., & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*, 237–247.

Patterson, M.L., & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*, 191–196.

Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, *23*, 6819–6825.

Rosenblum, L.D., Johnson, J.A., & Saldana, H.M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research*, *39*, 1159–1170.

Rudmann, D.S., McCarley, J.S., & Kramer, A.F. (2003). Bimodal displays improve speech comprehension in environments with multiple speakers. *Human Factors*, *45*, 329–336.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles: Western Psychological Services.

Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*, B69–B78.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92*, B13–B23.

Stollman, M.H.P., van Velzen, E.C.W., Simkens, H.M.F., Snik, A.F., & van den Broek, P. (2003). Assessment of auditory processing in 6-year-old language-impaired children. *International Journal of Audiology*, *42*, 303–311.

Stollman, M.H.P., van Velzen, E.C.W., Simkens, H.M.F., Snik, A.F.M., & van den Broek, P. (2004). Development of auditory processing in 6–12-year-old children: A longitudinal study. *International Journal of Audiology*, *43*, 34–44.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, *335*, 71–78.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F.R. Volkmar, R. Paul, K. A.J. & D.J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders*, Vol. 1 (3rd edn, pp. 335–364). Hoboken: Wiley.

Thomas, S.M., & Jordan, T.R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 873–888.

Watkins, K.E., Strafella, A.P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994.

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.

Williams, J.H.G., Waiter, G.D., Gilchrist, A., Perrett, D.I., Murray, A.D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia*, *44*, 610–621.

Manuscript accepted 6 February 2007