autism © 2010
SAGE Publications
and The National
Autistic Society
Vol 14(4) 305–320; 353615
1362-3613(2010)

Visual influences on speech perception in children with autism

GRACE IAROCCI *Simon Fraser University, Burnaby, British Columbia, Canada*

ADRIENNE ROMBOUGH *Simon Fraser University, Burnaby, British Columbia, Canada*

JODI YAGER *Simon Fraser University, Burnaby, British Columbia, Canada*

DANIEL J. WEEKS *Simon Fraser University, Burnaby, British Columbia, Canada*

ROMEO CHUA *University of British Columbia, Vancouver, British Columbia, Canada*

ABSTRACT The bimodal perception of speech sounds was examined in children with autism as compared to mental age–matched typically developing (TD) children. A computer task was employed wherein only the mouth region of the face was displayed and children reported what they heard or saw when presented with consonant-vowel sounds in unimodal auditory condition, unimodal visual condition, and a bimodal condition. Children with autism showed less visual influence and more auditory influence on their bimodal speech perception as compared to their TD peers, largely due to significantly worse performance in the unimodal visual condition (lip reading). Children with autism may not benefit to the same extent as TD children from visual cues such as lip reading that typically support the processing of speech sounds. The disadvantage in lip reading may be detrimental when auditory input is degraded, for example in school settings, whereby speakers are communicating in frequently noisy environments.

KEYWORDS *autism; visual and auditory processing; bimodal speech perception*

ADDRESS *Correspondence should be addressed to* GRACE IAROCCI, *Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada. e-mail: giarocci@sfu.ca*

Numerous anecdotal, clinical, and empirical reports of cross-modal integration problems among individuals with autism exist (see Iarocci and McDonald, 2006, for a review). Atypical sensory-perceptual behaviours are commonly observed in persons with autism, are not solely attributable to

DOI: 10.1177/1362361309353615

sensory organ deficits (e.g., Scharre and Creedon, 1992), and appear to persist throughout their development (Greenspan and Wieder, 1997; O'Neill and Jones, 1997). The seamless integration of input from multiple senses is integral to the development of perception in a variety of domains. In particular, speech comprehension, a process that relies upon the accurate integration of information from the visual and auditory modalities, is an area of difficulty for many children on the autism spectrum. In this paper, we focus on whether the perception of speech sounds and lip movements are overly salient, diminished and/or not well integrated across the visual and auditory modalities.

Typical audiovisual speech perception

There is an extensive literature on the multiple sources of input that are needed to support the perception of speech and the development of language (Massaro, 1998). In addition to using audition to hear another's speech, visual cues such as lip, face, and body movements contribute to the audibility of speech and are integral to speech perception (Calvert et al., 1998; Marschark et al., 1998; Campbell, 1989). In noisy environments when speech signals are poor or distorted, visual information supplements the auditory signal and improves perceptual accuracy (Massaro, 1984; Middelweerd and Plomp, 1987). However, using visual cues from the speaker's face to improve speech perception occurs automatically and implicitly, even when the auditory input is not impoverished. The McGurk effect is the most empirically robust example of the influence of visual cues on perception of speech sounds (McGurk and MacDonald, 1976). A phenomenon known as the McGurk effect occurs in response to incongruous auditory and visual speech stimuli whereby most participants report hearing a sound that has been visually influenced or represents a fusion of the two syllables presented. For example, upon hearing 'aba' presented simultaneously with the visual stimuli 'aga', a fused perception may consist of 'ada'. The McGurk effect demonstrates that speech perception is bimodal. Moreover, it provides evidence that the common assumption that speech perception is primarily an auditory task is false, instead it suggests that the processing of visual input plays a significant role in speech perception, when occurring in the context of face-to-face communication (McGurk and MacDonald, 1976). However, the extent of influence that visual cues have on speech perception depends on a person's age. Children under 8 years of age do not consistently display audiovisual integration in the form of fusions or combinations (i.e., they fail to demonstrate the typical McGurk response consistently) (e.g., Massaro, 1984). Thus, the use of visual cues in speech perception is a developmental process that is acquired over time and is fine-tuned in middle childhood.

Bimodal speech perception in autism

Children with autism may have sensory integration difficulties that impact their speech perception and language development. Difficulties with the processing of one type of sensory input or a decreased benefit from concurrent visual and auditory input may lead to variability in children's speech perception and development over time (de Gelder and Vroomen, 1998; de Gelder et al., 1998). De Gelder and colleagues (de Gelder et al., 1991) used the McGurk task (McGurk and MacDonald, 1976) to assess audiovisual integration among children with autism. They presented children with autism (mean chronological age [CA] 10.9 years) three audiovisual conditions: discrepant auditory and visual lip movements of common syllables, the visual lip movements alone, and the auditory sound alone. In the discrepant audiovisual condition, auditory syllables such as /aba/ or /ana/ were synchronously dubbed with visual lip movements that did not match the sound that was uttered (e.g., /ada/, /apa/). De Gelder and colleagues (1991) found adequate lip reading and auditory speech abilities among children with autism but relatively little influence of the visual input on auditory speech as compared to their typically developing (TD) peers matched on receptive language ability. Further, TD participants who demonstrated better lip reading ability were more influenced by visual speech but this was not the case in participants with autism. The authors also found no correlation between face recognition and lip reading performance in children with autism, whereas these two variables were positively correlated in the TD group.

Williams and colleagues (2004) replicated and extended de Gelder et al.'s (1991) findings by examining the independent contributions of lip reading and audition among children within the autism spectrum (mean CA 9 years) within the context of a computer-animated version of the McGurk task. The findings indicated that when lip reading abilities were controlled, the group of children with autism spectrum disorder (ASD) no longer showed diminished integration of auditory and visual input. Williams et al. (2004) concluded that children with ASD may use visual information less consistently than TD children (mean CA 9.5 years); however, their ability to integrate the audio and visual input was not impaired.

Smith and Bennetto (2007) examined both lip reading and perception of whole words in noisy conditions among high functioning children and adolescents with autism (mean CA 15.8 years). Contrary to Williams et al. (2004), they found that unlike their IQ-matched TD peers, the participants with autism did not improve when their performance was compared across audio only and audiovisual conditions. Furthermore, this lack of improvement could not be accounted for by their poorer lip reading abilities. Smith and Bennetto (2007) concluded that individuals with autism may have an audiovisual speech integration difficulty in addition to poor lip reading abilities.

Discrepant findings across studies may be due to methodological issues such as differences in the homogeneity of diagnostic groups (i.e., ASD versus autism) or differences in the task stimuli (consonant-vowel syllables versus whole words; computer animated image versus digital video images). One possible explanation for the discrepant findings is that children with autism may have difficulty interpreting information from realistic whole face images and processing rich facial information during a speech perception task. Boucher, Lewis and Collis (1998) asked children with autism and a comparison group of children with specific language impairment and a moderate learning disability to match familiar faces with their respective voices. The groups were matched on chronological age (mean CA 7.9 years) and verbal mental age (mean VMA 4.4 years). The children with autism performed significantly more poorly than the comparison group on familiar face recognition, familiar voice recognition and familiar voice-face identity matching. Loveland and colleagues (1995) investigated the intermodal perception of affect among children and adolescents with autism and a comparison group of those with Down syndrome, and found that, after VMA and IQ differences were considered, the individuals with autism displayed greater difficulty matching facial and vocal/linguistic affect. These findings indicate that children with autism have difficulty reading facial cues in a bimodal context.

In the present study, we explored whether children with autism would show unusual patterns of visual or auditory influence during the audiovisual perception of simple consonant-vowel syllables as compared to mental age–matched TD children. Whereas previous studies presented auditory syllables within the context of a realistic face image, the current study used a task in which only the mouth region of the face was presented. In light of recent evidence suggesting abnormal face processing among individuals with autism (see Dawson et al., 2005, for a review), it was thought that presenting the entire face might place children with autism at a disadvantage compared to their same-age peers. In addition, evidence from eye-tracking studies suggests that persons with autism may be more tuned into information from the mouth region of the face (Klin et al., 2002). Thus, isolating the mouth region served to decrease the number of facial and affective cues in the visual display and eliminated the contextual cues that typically support audiovisual integration such as emotion or facial movement cues (Thomas and Jordan, 2004; Scheinberg, 1980; Lansing and McConkie, 1999; Preminger et al., 1998) as well as highlight a region of the face that is more salient to persons with autism.

Visual cues (mouth movements) were manipulated to modify the perception of speech sounds (consonant-vowel syllables) during a computer task. Participants were presented with discrepant auditory and visual sounds. An auditory syllable, /ba/, was synchronously dubbed with discrepant visual syllables: /tha/, /va/, or /da/. Based on previous findings, we expected that when presented with discrepant audio-visual stimuli, the children with autism would be less influenced by the concurrent visual and auditory input than the TD children matched on verbal mental age. Further, we hypothesized that unimodal difficulties would not account for the lack of bimodal speech sound integration in autism.

Method

Participants

Twelve children and adolescents with autism were matched individually to 12 TD children on the basis of verbal mental age (within 12 months). The autism group did not differ significantly from the TD group on mean chronological age ($t = .22$, $p = .83$), verbal mental age ($t = .13$, $p = .45$), or non-verbal mental age ($t = .77$, $p = .45$). There was only one female in the autism group whereas there were an equal number of male and female participants in the TD group. This difference in the number of male and female participants is typical for studies of autism; however, there is no evidence of sex differences in lip reading ability or speech perception in the literature. Parents of the participants were informally queried and confirmed that their child did not show any frank deficits in vision or hearing. Verbal mental age was assessed using the Peabody Picture Vocabulary Test (PPVT-III) (Dunn and Dunn, 1997) and was selected as the matching criterion due to the verbal nature of the experimental task. However, non-verbal mental age was assessed using the Raven's Coloured Progressive Matrices (Raven, 1962) and was not found to be significantly different between the groups. The descriptive characteristics of the autism and TD groups are shown in Table 1.

The children and adolescents with autism were recruited from the central and southwestern regions of British Columbia. The clinical diagnoses of autism were confirmed with the Autism Diagnostic Interview–Revised (ADI-R; Rutter et al., 2003), which was administered by a clinical psychology graduate student who was trained by the developers of the tool. The TD children were recruited from communities surrounding Brandon, Manitoba, and in Qualicum, British Columbia.

Table 1 **Mean CA, verbal and nonverbal mental age (all in months), and gender ratios for participants in the autism and typically developing (TD) groups**

	Autism *M (SD)*	*TD* *M (SD)*	*t*	*p*
N	12	12		
Age	127.25(33.63)	123.80(39.64)	.22	.83
PPVT-III[a] – Verbal Mental Age	150.75(63.64)	147.58(57.24)	.13	.90
Raven's[b] – Nonverbal Mental Age	133.58(34.39)	123.25(31.30)	.77	.45
Gender (M:F)	11:1	6:6	–2.42	.02

[a] Peabody Picture Vocabulary Test, 3rd Edition.
[b] Raven's Coloured Progressive Matrices.

Stimuli and apparatus

A photorealistic, dynamic visual image of the mouth and nose region of an adult male face was created using a digital video camera and displayed on a 17-inch monitor. The stimuli included one-syllable, consonant-vowel sounds and/or images that were presented on a portable PC system with a 17-inch monitor using Windows Media Player software. The auditory stimuli were presented via computer speakers that were located equidistant from the monitor on the left and right sides. Three task conditions were presented that differed in the format of stimuli presentation. In the unimodal auditory condition, a blank, blue screen was presented while a recorded male voice delivered one of the following consonant-vowel sounds: /ba/, /tha/, /va/, or /da/. In the unimodal visual condition, a photorealistic, dynamic visual image of the mouth and nose region of an adult male face was displayed on the screen. The mouth region moved as if articulating one of the consonant-vowel sounds (i.e., /ba/, /tha/, /va/, or /da/) although no sound was presented during this condition. During the bimodal (audiovisual) condition, the articulating mouth and consonant-vowel sounds were presented simultaneously. The auditory stimuli remained consistent throughout the bimodal condition (i.e., /ba/) whereas the visual stimuli varied randomly (i.e., /tha/, /ba/, /va/, or /da/).

Procedure

This research was approved by Simon Fraser University's Office of Research Ethics. Written informed consent was obtained from the parents of all the child participants prior to scheduling the testing session.

During the administration of the audiovisual task, participants were seated 50 cm in front of the computer monitor. Experimental conditions were administered in the following fixed order: unimodal visual, bimodal audio-visual condition, unimodal auditory. Participants were observed by the experimenter during each trial in order to ensure adequate attention and focus on the stimuli. All verbal responses were recorded by the experimenter on paper following each trial.

The unimodal visual condition consisted of 12 trials of visual stimuli in which participants were required to lip-read the consonant-vowel sounds that the mouth appeared to be articulating. Prior to the administration of the unimodal visual condition, participants were told that they were going to see the mouth speak but with no sound and that the mouth would be saying consonant/vowel sounds like: /pa/, /tha/, /ba/, /ta/, /da/, /la/, /va/, /fa/, or /ka/. They were instructed to watch the image closely and tell the experimenter what they saw the mouth say after each trial. Participants then gave free verbal responses after each visual trial, which were recorded on paper by the experimenter. They were encouraged to try their best and guess if necessary.

The bimodal audiovisual condition consisted of 20 trials in which both visual and auditory stimuli were presented simultaneously. On five trials, the auditory and visual stimuli were congruent (i.e., an auditory /ba/ sound was accompanied by a visual /ba/ articulation). These trials served as catch trials with correct answers (/ba/) indicating that participants adequately understood task demands. Only participants correctly responding to at least four of the five catch trials (80% accuracy) were included in the analyses. On the remaining 15 trials, the stimuli were incongruent (e.g., an auditory /ba/ sound accompanied by a visual /va/ articulation). Participants were informed that they were going to see and hear the mouth speaking and were instructed to watch carefully and to tell the experimenter what they *heard* the mouth say after each trial.

The unimodal auditory condition consisted of 12 trials in which the participants were instructed to look at a blue screen (no mouth image was presented) and report what sound they heard. Auditory stimuli consisted of /ba/, /tha/, /da/, and /va/ sounds spoken by a male voice.

Scoring

Scoring techniques based upon criteria established by Campbell and colleagues (1990) were employed. In the unimodal visual condition, responses were scored as correct if they were deemed visually compatible with the stimuli. For instance, bilabial responses (i.e., /ba/, /ma/, or /pa/) were considered consistent with a visually presented /ba/, and thus were scored as correct. Velar or alveolar responses (i.e., /ga/, /da/, /ka/, /ta/, or /na/) were scored as compatible with a visually presented /da/. Labiodental responses (i.e., /va/ or /fa/) were scored as correct for a visually presented /va/. Only the interdental response of /tha/ was scored as correct for a visually presented /tha/. All other responses to the visual stimuli were coded as incorrect.

In the unimodal auditory condition, responses that matched the actual auditory stimuli (i.e., /ba/, /tha/, /va/, or /da/) were scored as correct.

In the bimodal audiovisual condition, responses on both congruent and incongruent trials were coded as either visually compatible (VC), auditory compatible (AC), or other. Visually compatible responses were scored using the same criteria that were applied in the unimodal visual condition. Responses were scored as auditory compatible if they matched the auditory stimuli (i.e., /ba/). Responses were coded as /other/ if they were deemed not compatible with either the visual or auditory stimulus (e.g., /da/ in response to an auditory /ba/ and a visual /tha/).

Results

Scores consisted of proportions (percentages) and thus data was not normally distributed. Therefore, data was transformed using an arcsine square root transformation. Transforming the data was chosen over nonparametric analyses because we wished to conduct more sophisticated analyses than would be permitted with non-parametric tests.

In the unimodal auditory condition, the accuracy for the autism group ($M = 78.47\%$, $SD = 17.64\%$) was not significantly different from the accuracy of the TD group ($M = 75.75\%$, $SD = 13.98\%$), $t(22) = .11$, $p = .92$ (two-tailed). This finding indicates that children with autism were able to identify auditory stimuli as accurately as TD comparisons. For the unimodal visual condition (lip reading), the accuracy of the autism group ($M = 37.50\%$, $SD = 25.75\%$) was significantly different from the accuracy of the TD group ($M = 59.75\%$, $SD = 23.79\%$), $t(22) = -2.32$, $p < .05$ (two-tailed). Children with autism demonstrated difficulties with lip reading as compared to TD children. Mean rates of percentage accuracy for each group in the unimodal conditions are presented in Figure 1.

Within the bimodal condition on catch trials where an auditory /ba/ was matched with a visual /ba/, the mean accuracy ($M = 91.67\%$, $SD = 10.29\%$) for the group of children with autism did not differ significantly from catch trial accuracy ($M = 93.33\%$, $SD = 9.85\%$), $t(22) = -.405$, $p = .69$ (two-tailed) of the TD children.

Our primary hypothesis was that individuals with autism would not integrate visual and auditory information as effectively as TD children when perceiving bimodal speech sounds. Specifically, we predicted that visual influences on auditory speech perception during the bimodal condition would be decreased in children with autism relative to that of the TD comparison group. Within the bimodal condition, the mean number of audio compatible responses made by children with autism (M = 38.33%, SD = 33.40%) was significantly greater than the mean number of audio compatible responses given by TD children (M = 11.17%, SD = 13.34%), $t(22) = 2.75$, $p < .05$ (two-tailed). Children with autism were significantly more likely than their TD peers to make auditory compatible responses. The mean number of visual compatible responses (M = 33.89%, SD = 21.74%) in the autism group differed significantly from the mean (M = 59.41%, SD = 26.67%), $t(22) = -2.58$, $p < .05$ (two-tailed) in the TD group. When visual and auditory stimuli were incongruent, children with autism were less likely than their TD peers to make visually compatible responses. No significant differences were found in the mean number of 'other' responses given by the autism group (M = 27.77%, SD = 17.94%) and the TD group (M = 29.44%, SD = 19.12%), $t(22) = -.15$, $p = .88$ (two-tailed).

In order to explore the finding that the children with autism made fewer visually compatible responses, a measure of 'visual effect' was derived. Visual effect was calculated by subtracting mean accuracy on incompatible trials from mean accuracy on compatible trials in the bimodal condition. Since participants were instructed to report what they heard, accuracy was defined as auditory compatible responses (i.e., /ba/). The resulting scores

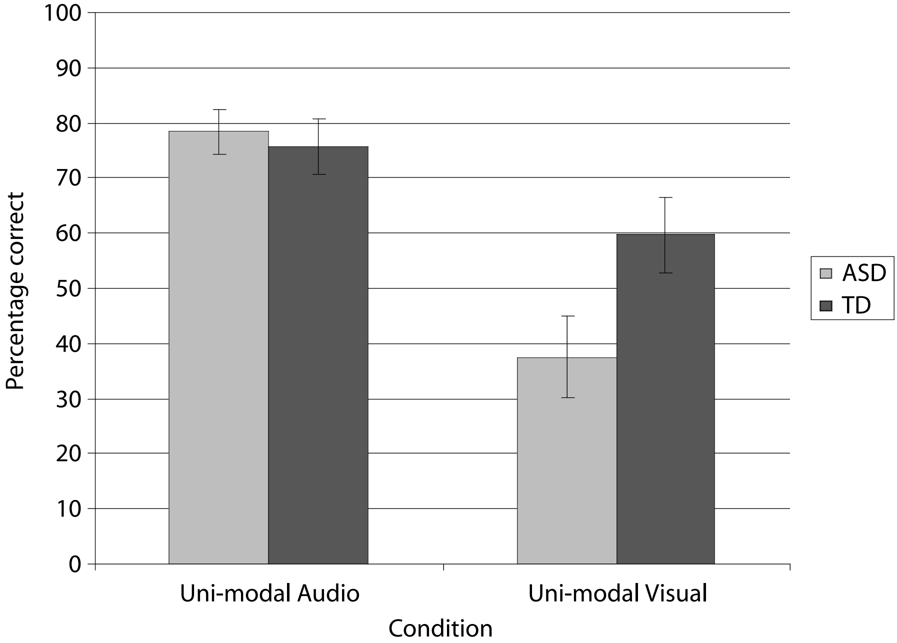

Figure 1 **Mean percentage of accurate responses in the unimodal visual and unimodal auditory conditions for participants with autism and TD comparisons**

reflect the extent to which incompatible visual stimuli influenced perception of speech sounds. A significant mean difference was found between the mean visual effect scores of the autism group (M = 53.33%, SD = 30.22%) and the TD group (M = 82.17%, SD = 14.52%), $t(22) = -3.06$, $p < .05$ (two-tailed). Children with autism did not respond in a manner that suggested their perception of bimodal speech was influenced by incompatible visual information to the same extent as their TD peers. Thus, the children with autism appeared to be less influenced by visual speech information during bimodal speech perception.

A medium to large significant correlation was detected between the performance of the children with autism in the unimodal visual condition (lip reading) and the number of visual compatible responses they gave in the bimodal condition ($r = .72$, $p < .05$). A similar positive correlation (between lip reading and visually compatible responses) was found for the TD children ($r = .68$, $p < .05$). Specifically, within both groups, poorer lip reading ability was associated with lower visual influence on speech perception. Conversely good lip reading ability was associated with more visual influence during bimodal speech perception. In order to determine whether group differences in visual effect were driven by poor lip reading ability among the children with autism, visual effect differences were

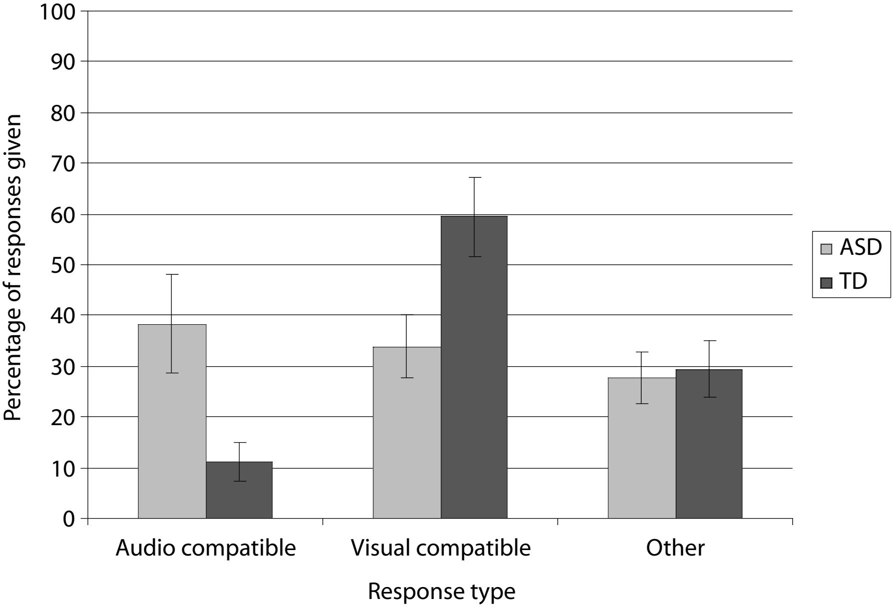

Figure 2 **Mean percentage of response types for participants in the autism and TD groups in the bimodal condition**

examined while controlling for lip reading abilities. An Analysis of Covariance (ANCOVA) revealed that the significant main effect for group (autism vs. TD) on visual effect scores disappeared once lip reading ability was controlled for $F(1, 21) = 3.11$, $p = .09$. Similarly, a between-group ANCOVA with bimodal visual compatible responses as a dependent variable and mean unimodal visual performance accuracy as a covariate found no significant between-group difference in visually compatible responses once lip reading accuracy was accounted for $F(1, 2) = 1.39$, $p = .25$. Thus, it appears that difficulty in lip reading accounted for less visual influence on speech perception in both groups.

In order to explore the possibility that lip reading may be associated more generally with early language development, we examined correlations between lip reading accuracy and caregiver reports of the acquisition of words and phrases (using the ADI-R age at 1st words and age at phrase speech data). Lip reading was not significantly correlated with age of 1st words ($r = -.44$, $p = .15$) or age at phrase speech ($r = -.22$, $p = .50$). Further, we did not find evidence that lip reading was associated with receptive vocabulary in the autism group. Lip reading accuracy and PPVT were not significantly correlated, $r = -.04$, $p = .51$.

Discussion

In this study, a computer task was employed wherein the mouth region of the face was isolated in order to decrease the number of facial and affective cues within the context of a realistic face image display and to highlight a region of the face that has been found to be salient to persons with autism. The task required children with autism to report what they heard or saw when presented with consonant-vowel sounds in one of three task conditions: unimodal auditory condition, unimodal visual condition, and the bimodal (audiovisual) condition. Children with autism showed less visual influence and more auditory influence on their bimodal speech perception when compared to TD children. Groups did not differ in accurately perceiving auditory input alone. However, the autism group performed significantly worse in the unimodal visual condition (lip reading). Further analysis revealed that group differences in the audiovisual condition were largely attributable to lip reading difficulties that were more pronounced in the autism group.

Our results are consistent with previous studies (Smith and Bennetto, 2007; Williams et al., 2004) in identifying poor lip reading as a significant source of the difference in bimodal speech perception among children with and without autism. They extend previous findings by showing that the lip reading difference is evident even when children with autism are presented with a less perceptually and contextually rich image of the mouth, as opposed to the whole face, that typically supports bimodal speech perception (Thomas and Jordan, 2004; Scheinberg, 1980; Lansing and McConkie, 1999; Preminger et al., 1998). Moreover, highlighting the mouth region of the face that was found to be more salient to persons with autism (Klin et al., 2002) did not improve their lip reading ability and bimodal speech perception as compared to their TD-matched peers who typically use the whole-face affective and motion cues to support bimodal speech perception (Davis and Kim, 2006; Lansing and McConkie, 2003).

Differences in the abilities to integrate audio and visual features of speech driven by lip reading difficulties likely impact daily speech comprehension and language performance in children with autism and may be associated with atypical or delayed language acquisition (Smith and Bennetto, 2007). Smith and Bennetto found a marginally positive relation between the ability of children with autism to perceive bimodal speech embedded noise and age of first word. In typical development, lip reading ability is closely associated with language processing skills (Campbell, 1989; Massaro, 1987). Potentially, in early development, lip reading may facilitate the auditory tracking of speech within noisy environments. Typically developing infants selectively attend to speech information (Newman, 2005) whereas children with autism demonstrate a lack of orienting to speech sounds that is not due to auditory deficits alone (Čeponienė et al., 2003). The findings in this study suggest that children with autism may not benefit to the same extent as TD children from visual cues such as lip reading that typically support the processing of speech sounds. Moreover, the disadvantage in lip reading may be especially detrimental when auditory input is degraded as in school settings wherein multiple speakers are communicating in frequently noisy environments. For example, Plaisted and colleagues found that individuals with ASD have difficulty understanding speech embedded in background noise, due to a wide auditory filter that creates a greater susceptibility to interfering sounds during speech processing (Plaisted et al., 2003). Thus a subgroup of persons with autism may have difficulty with both the auditory processing of speech signals as well as the visual lip reading of speech in background noise. Alternatively, difficulties dividing attention between speech and non-speech noise may be integral to our understanding of communicative development in persons with ASD (Kenworthy et al., 2009). Future studies could explore how critical lip reading is to the perception of speech in children with autism by comparing the extent to which visual cues impact language acquisition as development unfolds by examining very young children with autism, or children with varying degrees of language impairment.

The current findings indicate that at approximately 10 years of age, TD children were quite adept at using visual cues such as lip reading when perceiving bimodal speech sounds and may have developed a general tendency to process visual cues implicitly while engaged in a speech perception task. However, the children with autism were significantly less skilled at lip reading and may also have been less inclined to attend to visual cues when engaged in processing speech sounds. This begs the question of whether a lack of attending to and thus, experience over time, would contribute to lip reading difficulties in at least a subgroup of children with autism. Bimodal speech perception, like many other perceptual abilities that we study in typical and atypical development, represents the outcome of perceptual developmental processes that have occurred over time. The intermediary processes are often elusive in typical development because multisensory integration occurs in a seamless manner to produce a coherent percept. However, children with autism may show atypical development of bimodal speech perception and thus provide a unique window into the elusive intermediary processes that may be necessary for typical perceptual learning to occur.

In future studies, the role of instructions and task demands needs to be considered in bimodal speech perception. For example, in previous studies, bimodal speech perception in TD individuals was found to be susceptible to cognitive manipulations (Summerfield and McGrath, 1984; Colin et al., 2005). Summerfield and McGrath (1984) highlighted the impact of instructions on audiovisual processing by demonstrating that visual influence during a speech perception task was less prominent when participants were instructed to repeat what they *heard* versus what the speaker *uttered*. The authors proposed that varying attentional mechanisms were implicated, with the latter instructions drawing less attention to the auditory modality. Similarly, Colin et al. (2005) examined the McGurk effect under two different instruction conditions. They found that when participants were provided with a list of possible syllables that they might have heard (multiple choice), more fusions were reported than when they were instructed to write down the syllables they heard (free response). The authors suggested that cognitive factors, including task instructions, influence speech perception at a later processing stage (i.e., response selection). They further proposed that the degree of automaticity of audiovisual integration processes underlying speech perception likely depends on the context – with integration occurring at an earlier perceptual stage under optimal conditions and later cognitive processes enhancing modality-specific processing under less favorable conditions (e.g., instructions favoring the auditory modality).

Acknowledgements

This project was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) grant and a Michael Smith Foundation for Health Research (MSFHR) scholar award to Grace Iarocci. Jodi Yager and Adrienne Rombough's contributions were supported by a fellowship from the Autism Research Training Program and a CIHR Doctoral Award. Jodi Yager's work was also supported by scholarships from the SSHRC and MSFHR. Daniel Weeks and Romeo Chua were supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Human Early Learning Partnership. We express our deep appreciation to the children and families that participated in this study.

References

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.

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

Campbell, R. (1989) 'The Sensory Imperative: Modularity and the Development of Face Processing in the Neonate', *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 9: 55–59.

Campbell, R., Garwood, J., Franklin, S., & Howard, D. (1990) 'Neuropsychological Studies of Auditory-Visual Fusion Illusions: Four Case Studies and Their Implications', *Neuropsychologia* 28: 787–802.

Čeponienė, R., Lepistö, T., Alku, P., Aro, H., & Näätänen, R. (2003) 'Event-Related Potential Indices of Auditory Vowel Processing in 3-Year-Old Children', *Clinical Neurophysiology* 114: 652–661.

Colin, C., Radeau, M., & Deltenre, P. (2005) 'Top-Down and Bottom-Up Modulation of Audiovisual Integration in Speech', *European Journal of Cognitive Psychology* 17: 541–560.

Davis, C. & Kim, J. (2006) 'Audio-Visual Speech Perception off the Top of the Head', *Cognition* 100(3): B21–B31.

Dawson, G., Webb, S.J., & McPartland, J. (2005) 'Understanding the Nature of Face Processing Impairment in Autism: Insights from Behavioral and Electrophysiological Studies', *Developmental Neuroscience* 27(3): 403–424.

de Gelder, B., & Vroomen, J. (1998) 'Impaired Speech Perception in Poor Readers: Evidence from Hearing and Speech Reading', *Brain and Language* 64: 269–281.

de Gelder, B., Vroomen, J., & Bachoud-Levi, A. (1998) 'Impaired Speech Reading and Audio-Visual Speech Integration in Prosopagnosia', in R. Campbell, B. Dodd & D. Burnham (eds.) *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech* (pp. 195–207). Hove, UK: Psychology Press/Erlbaum (UK) Taylor & Francis.

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.

Dunn, L.M., & Dunn, L.M. (1997) *Peabody Picture Vocabulary Test – Third Edition*. Minnesota: American Guidance Service.

Greenspan, S.I., & Wieder, S. (1997) 'An Integrated Developmental Approach to

Interventions for Young Children with Severe Difficulties in Relating and Communicating', *Zero to Three* 17: 5–18.

Iarocci, G., & McDonald, J. (2006) 'Sensory Integration and the Perceptual Experience of Persons with Autism', *Journal of Autism and Developmental Disorders* 36: 77–90.

Kenworthy, L., Black, D.O., Harrison, B., della Rosa, A., della Rosa, W., & Gregory, L. (2009) 'Are Executive Control Functions Related to Autism Symptoms in High-Functioning Children?' *Child Neuropsychology* 9(1): 1–6.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002) 'Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals with Autism', *Archives of General Psychiatry* 59(9): 809–816.

Lansing, C.R., & McConkie, G.W. (1999) 'Attention to Facial Regions in Segmental and Prosodic Visual Speech Perception Tasks', *Journal of Speech, Language, and Hearing Research* 42: 526–539.

Lansing, C.R., & McConkie, G.W. (2003) 'Word Identification and Eye Fixation Locations in Visual and Visual-Plus-Auditory Presentations of Spoken Sentences', *Perception & Psychophysics* 65(4): 536–552.

Loveland, K.A., Tunali-Kotoski, B., Chen, R., & Brelsford, K.A. (1995) 'Intermodal Perception of Affect in Persons with Autism or Down syndrome', *Development and Psychopathology* 7: 409–418.

Marschark, M., LePoutre, D., & Bement, L. (1998) 'Mouth Movement and Signed Communication', in R. Campbell, B. Dodd & D. Burnham (eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech* (pp. 245–266). Hove, UK: Psychology Press/Erlbaum (UK) Taylor & Francis.

Massaro, D.W. (1984) 'Children's Perception of Visual and Auditory Speech', *Child Development* 55: 1777–1788.

Massaro, D. (1987) 'Speech Perception by Ear and Eye', in B. Dodd & R. Campbell (eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 53–83). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Massaro, D. (1998) *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle.* Cambridge Massachusetts: MIT Press.

McGurk, H., & MacDonald, J. (1976) 'Hearing Lips and Seeing Voices', *Nature* 264: 746–748.

Middelweerd, M., & Plomp, R. (1987) 'The Effect of Speechreading on the Speech-Reception Threshold of Sentences in Noise', *Journal of the Acoustical Society of America* 82(6): 2145–2147.

Newman, R.S. (2005) 'The Cocktail Party Effect in Infants Revisited: Listening to One's Name in Noise', *Developmental Psychology* 41: 352–262.

O'Neill, M., & Jones, R.S.P. (1997) 'Sensory-Perceptual Abnormalities in Autism: A Case for More Research?' *Journal of Autism and Development Disorders* 27: 283–293.

Plaisted, K., Saksida, L., Alcántara, J., & Weisblatt, E. (2003) 'Towards an Understanding of the Mechanisms of Weak Central Coherence Effects: Experiments in Visual Configural Learning and Auditory Perception', *Philosophical Transactions of the Royal Society: Part B* 358: 375–386.

Preminger, J.E., Lin, H., Payen, M., & Levitt, H. (1998) 'Selective Visual Masking in Speechreading', *Journal of Speech, Language, and Hearing Research* 41: 564–575.

Raven, J.C. (1962) *Raven's Progressive Coloured Matrices*. London: E.T. Heron & Co., Ltd.

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

Scharre, J.E., & Creedon, M.P. (1992) 'Assessment of Visual Function in Autistic Children', *Optometry and Vision Science: Official Publication of the American Academy of Optometry* 69: 433–439.

Scheinberg, J.C. (1980) 'Analysis of Speechreading Cues Using an Interleaved Technique', *Journal of Communication Disorders* 13: 489–492.

Smith, E., & Bennetto, L. (2007) 'Audiovisual Speech Integration and Lipreading in Autism', *Journal of Child Psychology and Psychiatry* 48(8): 813–821.

Summerfield, Q., & McGrath, M. (1984) 'Detection and Resolution of Audio-Visual Incompatibility in the Perception of Vowels', *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 36: 51–74.

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.

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: 569–575.