J Autism Dev Disord (2014) 44:1470–1477
DOI 10.1007/s10803-013-1992-7

BRIEF REPORT

Brief Report: Arrested Development of Audiovisual Speech Perception in Autism Spectrum Disorders

Ryan A. Stevenson · Justin K. Siemann · Tiffany G. Woynaroski · Brittany C. Schneider · Haley E. Eberly · Stephen M. Camarata · Mark T. Wallace

Published online: 12 November 2013

Abstract Atypical communicative abilities are a core marker of Autism Spectrum Disorders (ASD). A number of studies have shown that, in addition to auditory comprehension differences, individuals with autism frequently show atypical responses to audiovisual speech, suggesting a multisensory contribution to these communicative differences from their typically developing peers. To shed light on possible differences in the maturation of audiovisual speech integration, we tested younger (ages 6–12) and older (ages 13–18) children with and without ASD on a task indexing such multisensory integration. To do this, we used the McGurk effect, in which the pairing of incongruent auditory and visual speech tokens typically results in the perception of a fused percept distinct from the auditory and visual signals, indicative of active integration of the two channels conveying speech information. Whereas little difference was seen in audiovisual speech processing (i.e., reports of McGurk fusion) between the younger ASD and TD groups, there was a significant difference at the older ages. While TD controls exhibited an increased rate of fusion (i.e., integration) with age, children with ASD failed to show this increase. These data suggest arrested development of audiovisual speech integration in ASD. The results are discussed in light of the extant literature and necessary next steps in research.

Keywords Autism · Sensory processing · Multisensory integration · Speech perception · Audiovisual · McGurk effect

R. A. Stevenson (✉)
Department of Hearing and Speech Sciences, Vanderbilt Kennedy Center, Vanderbilt Brain Institute, Vanderbilt University Medical Center, 7110 MRB III BioSci Bldg, 465 21st Ave South, Nashville, TN 37232, USA
e-mail: ryan.andrew.stevenson@gmail.com

R. A. Stevenson
Department of Psychology, University of Toronto, 100 Saint George St., Toronto, ON, Canada

J. K. Siemann
Neuroscience Graduate Program, Vanderbilt University, Nashville, TN 37232, USA

T. G. Woynaroski · S. M. Camarata
Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA

B. C. Schneider · H. E. Eberly
Program in Neuroscience, Vanderbilt University, Nashville, TN 37232, USA

M. T. Wallace
Department of Hearing and Speech Sciences, Psychology and Psychiatry, Vanderbilt Kennedy Center, Vanderbilt Brain Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA

Introduction

Autism Spectrum Disorders (ASD) represent a class of highly prevalent developmental disabilities that have been characterized by deficits in communication and social reciprocity, and by the presence of restricted and/or repetitive behaviors (APA 2000; APA DSM-5Task Force 2013). Individuals with ASD also exhibit alterations in sensory processing, a finding first noted in Kanner's original description of the disorder (Kanner 1943). Since that time, unusual responses to sensory stimuli have been reported across multiple sensory domains in individuals with autism (for review, see Iarocci and McDonald 2006). The prevalence and persistence of sensory symptoms led to the proposal that sensory processing alterations be considered a core deficit in ASD (Billstedt et al. 2007). Indeed, the recently revised diagnostic criteria (APA 2000; APA DSM-5Task Force 2013) recognize that children with ASD often present with atypical sensory processing, including marked difficulty integrating sensory information (Bebko et al. 2006; Mongillo et al. 2008a, b; Foss-Feig et al. 2010; Kwakye et al. 2011).

One key skill that impacts our day-to-day social interaction and communication and that necessitates successful integration of sensory information from multiple modalities is speech perception (Massaro and Cohen 1995). Speech perception in everyday settings involves integration of both acoustic cues and visual information from the moving face and neck (Massaro 1998). The benefit of multisensory cues in speech perception is evidenced by a boost in perceptual accuracy and an increase in processing speed across a variety of listening conditions (Calvert et al. 1998; van Wassenhove et al. 2005; Sekiyama and Burnham 2008). Visual cues especially enhance perception when the auditory signal is degraded (Sumby and Pollack 1954; Middelweerd and Plomp 1987) and when the message is particularly difficult to comprehend (Arnold and Hill 2001).

However, visual cues also impact speech perception under controlled listening conditions when the input is limited to meaningless syllables. The influence of visual information on perception of clearly audible nonsense syllables is illustrated by the McGurk effect (McGurk and MacDonald 1976; Macdonald and McGurk 1978), a psychophysical illusion in which the presentation of discrepant visual and auditory cues (e.g., visual "ga" and auditory "ba") often induces an illusory percept (e.g., "da" or "tha"). The illusion appears to be relatively automatic and robust (Soto-Faraco et al. 2004), persisting in typically developing individuals even under conditions of rather large spatial discrepancy (i.e., up to at least 90 degrees in either direction; Jones and Munhall 1997) and temporal asynchrony (i.e., up to approximately 180–250 ms; Massaro et al. 1996; Munhall et al. 1996).

The McGurk effect is early-emerging, seen even in infants and young children (Massaro 1984; Rosenblum et al. 1997). However, studies in typically developing individuals have found that preschoolers and young children tend to experience the McGurk effect to a lesser degree than adults (McGurk and MacDonald 1976; Massaro 1984). Cross-sectional studies indicate that the tendency to perceive the McGurk effect increases steadily throughout early childhood, with most typically developing, English-speaking children reaching adult levels of audiovisual integration by late childhood (i.e., by approximately 12 years old; Hockley 1994; Sekiyama and Burnham 2008).

Several studies to date have employed the McGurk illusion to investigate audiovisual speech perception in ASD. A number of these investigations have found that children with ASD show reduced levels of integration in response to incongruent McGurk stimuli (e.g., de Gelder et al. 1991; Williams et al. 2004a, b; Mongillo et al. 2008a, b; Irwin et al. 2011a, b). However, at least two studies have failed to find significant differences in the mean magnitude of fusion reported by children with ASD and TD controls (Iarocci et al. 2010; Woynaroski et al. 2013). Inconsistency in findings across studies suggests variability in multisensory speech perception across children with ASD.

Previous studies suggest that variability across children with ASD may be explained in part by differences in child profile, such as degree of social deficit or severity of sensory symptoms (Mongillo et al. 2008a, b; Woynaroski et al. 2013). Another possible difference may originate from differences in attention, as multisensory integration is known to be impacted by attention (Talsma et al. 2010) and individuals with ASD show differences in attending social stimuli, such as faces (Dawson et al. 1998; Dawson et al. 2004). Variability in audiovisual integration in children with ASD may also be explained in part by chronological age. One recent study reported that children with ASD show a protracted or "delayed" developmental trajectory of audiovisual speech integration, whereby they tend to report fused percepts to a lesser extent than their typically developing peers at younger ages (i.e., approximately 7–8 years), but to report similar levels of fusion in comparison to controls at older ages (i.e., in adolescence Taylor et al. 2010). These results seem to suggest that children with ASD may "catch up" in audiovisual integration processing after their typically developing peers have reached a developmental plateau (Taylor et al. 2010). If this is the case, then we should be more likely to find reduced reports of fusion relative to age-matched controls in younger children with ASD, than in older children with ASD.

The present study tests this hypothesis by measuring the rate of McGurk perceptions in younger (ages 6–12) and older (ages 13–18) children with ASD and TD controls matched on chronological age. In accordance with the bulk of the prior literature (e.g., de Gelder et al. 1991; Williams et al. 2004a, b; Mongillo et al. 2008a, b; Irwin et al. 2011a, b), we hypothesize that individuals with ASD will likely show reduced perception of the McGurk effect on average relative to TD controls. However, based on the findings of Taylor and colleagues (Taylor et al. 2010), we anticipate that these between-group differences will be more pronounced in younger children than in older children, as evidenced by an age-by-diagnosis interaction.

Methods

Participants

Participants included 30, 6–18 year old children with ASD and 31 TD controls matched on chronological age, nonverbal general intellectual ability as indexed by the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-2; Wechsler 1999), and full-scale intellectual ability. The ASD group had significantly lower verbal IQ scores on average relative to the TD control group (see results for analysis of this difference). See Table 1 for subgroup statistics.

Participants in the ASD group were diagnosed with either Autistic Disorder (25 %), Asperger's Syndrome (66 %), or Pervasive Developmental Disorder-Not Otherwise Specified (9 %) based on the Autism Diagnostic Observation Schedule (Lord et al. 2000) and/or Autism Diagnostic Interview, Revised (Lord et al. 1994) and clinical diagnosis by a practitioner familiar with ASD. Diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders IV-TR (APA 2000). Individuals in the TD group had no diagnoses of ASD or any other developmental disorder or related medical diagnosis, including, but not limited to, Fragile X, tuberous sclerosis, or seizure disorders. Individuals in both groups were screened for normal or corrected-to-normal vision using a tumbling E chart and were reported to have normal hearing. All experimental protocols were approved by Vanderbilt University Medical Center's Institutional Review Board.

Stimuli

All stimuli were presented using MATLAB (MATHWORKS Inc., Natick, MA) software with the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997). Visual, Auditory, congruent Audiovisual, and "McGurk" stimuli were derived from digital video clips of a female speaker uttering the syllables "ga" and "ba". These stimuli have been used successfully in previous studies of multisensory integration (Quinto et al. 2010; Stevenson et al. 2012). Visual stimuli were presented on a NEC MultiSync FE992 monitor positioned approximately 60 cm from the participants. Auditory stimuli were presented from centrally-aligned speakers. Syllable presentations were 2 s in duration, with each presentation comprising the entire production of the syllable, including pre-articulatory gestures.

Procedure

Participants were seated inside a light and sound attenuating WhisperRoom™ (Model SE 2000; Whisper Room Inc, Morristown, TN) and were asked to fixate on a central cross. Participants were monitored by closed circuit infrared cameras throughout the experiment to ensure fixation to the monitor during stimulus presentation. The task began with an instruction screen prompting participants to

Table 1 Group demographics

	All subjects					Younger (6–12 years old)					Older (13–18 years old)				
	ASD	TD	*t*	*p*	*d*	ASD	TD	*T*	*p*	*d*	ASD	TD	*t*	*p*	*d*
ASD Severity Score	7.9 ± 1.2	N/A				7.9 ± 1.2	N/A				8.0 ± 1.4	N/A			
Sex—% male (n)	80 % (24)	58 % (18)	2.47^{χ}	0.12	0.28^{ϕ}	82 % (14)	56 % (10)	1.28^{χ}	0.26	0.20^{ϕ}	77 % (10)	62 % (8)	0.18^{χ}	0.67	0.08^{ϕ}
Chronological Age (years)	12.1 ± 3.0	11.0 ± 3.1	0.65	0.52	0.36	7.9 ± 1.2	7.9 ± 1.2	0.31	0.76	0.00	14.9 ± 1.8	14.7 ± 1.4	0.37	0.71	0.12
Full Scale IQ	111.1 ± 16.1	115.7 ± 12.1	079	0.27	0.32	106.3 ± 15.7*	117.3 ± 10.4*	0.96	<0.05	0.82	116.9 ± 15.3	113.8 ± 14.1	0.64	0.53	0.21
Verbal IQ (T-score)	53.6 ± 10.6*	61.9 ± 8.7*	1.00	<0.05	0.86	49.2 ± 7.3*	59.7 ± 8.7*	1.00	<0.05	1.31	59.0 ± 11.8	64.5 ± 8.3	0.83	0.22	0.54
Performance IQ (T-score)	57.5 ± 8.4	53.7 ± 8.0	0.89	0.14	0.46	57.1 ± 10.2	57.6 ± 5.3	0.39	0.88	0.06	57.9 ± 15.3*	49.3 ± 8.6	0.98	<0.05	0.69

ASD Autism spectrum disorders. *TD* Typically developing control group

* Significant between group difference ($p < 0.05$)

$^{\chi}$ A χ^2 value with Yates correction

$^{\phi}$ A ϕ value for the effect size calculation of the respective χ^2 test indicate *what syllable the speaker said* by pressing the letter on a keyboard corresponding to the first letter of the syllable they perceived (i.e., "b" for "ba", "g" for "ga", "d" for "da", and "t" for "tha"). For simplicity, we will henceforth refer to illusory perceptions (i.e., reports of "da" or "tha") only as "da". Button presses were employed to reduce verbal response demands on participants. Each participant verbally confirmed that he or she understood the instructions and completed practice trials.

Each experimental trial consisted of: (a) a fixation screen presented for 500 ms plus a random jitter ranging from 1 to 1000 ms; (b) a stimulus presentation; (c) a 250 ms fixation screen; and (d) a response screen asking the participant, "What did she say?" below the fixation cross. Following the participant's response, a fixation screen reappeared signaling the start of the subsequent trial. Participants were presented with auditory only (with the fixation cross remaining on the screen), visual only, and congruent audiovisual versions of the "ba" and "ga" stimuli. Additionally, they were presented with incongruent audiovisual McGurk stimuli in which the visual "ga" was presented with the auditory "ba". Thus, a total of 7 stimulus conditions were presented, with 20 trials in each condition. The order of trial types was randomly generated for each participant for each experiment.

Analysis

Participants' data were split into groups of younger (i.e., 6–12 years; ASD n = 17, TD n = 18) and older (i.e., 13–18 years; ASD n = 13, TD n = 13) year olds in accordance with known developmental trajectories for audiovisual speech integration and perception (Hockley 1994; Taylor et al. 2010; Ross et al. 2011). Mean responses were then calculated in response to each of the 7 conditions. Response rates to the McGurk stimuli were then evaluated via a planned 3-way ANOVA (diagnosis × age group × condition). When a significant interaction effect was found, follow-up t-tests were performed to clarify the nature of the interaction. Independent samples t-tests were performed to determine if children with ASD showed reduced identification accuracy for unisensory and congruent audiovisual control trials in comparison to TD controls at either the younger or older age ranges. These tests were collapsed across syllable type to limit the number of significance tests conducted and to improve our estimation of the true score for the constructs of interest (Baggaley 1988).

Results

McGurk (Incongruent) Audiovisual Presentations

While our primary analysis of interest here is the comparison of responses to the incongruent McGurk stimuli, and initial 3-way ANOVA was run across diagnostic group, age group, and condition, with accuracies collapsed across control conditions. This revealed a 3-way interaction between these variables ($F_{(1,57)} = 4.16$, $p < 0.05$, partial-$\eta^2 = 0.068$), allowing for our planned analysis of the McGurk effect. The planned 2-way ANOVA revealed a significant diagnosis × age interaction effect ($F_{(1,57)} = 4.16$, $p < 0.05$, partial-$\eta^2 = 0.07$) wherein a larger difference in proportion of McGurk percepts reported was observed between the older diagnostic subgroups relative to the younger diagnostic subgroups (Fig. 1). Also, a main effect of diagnosis was found ($F_{(1,57)} = 5.32$, $p < 0.03$, partial-$\eta^2 = 0.08$), but no main effect of age ($F_{(1,57)} = 2.16$, $p = 0.15$, partial-$\eta^2 = 0.02$). Individuals with ASD on average reported the illusory McGurk percept at a reduced rate relative to TD controls, but younger

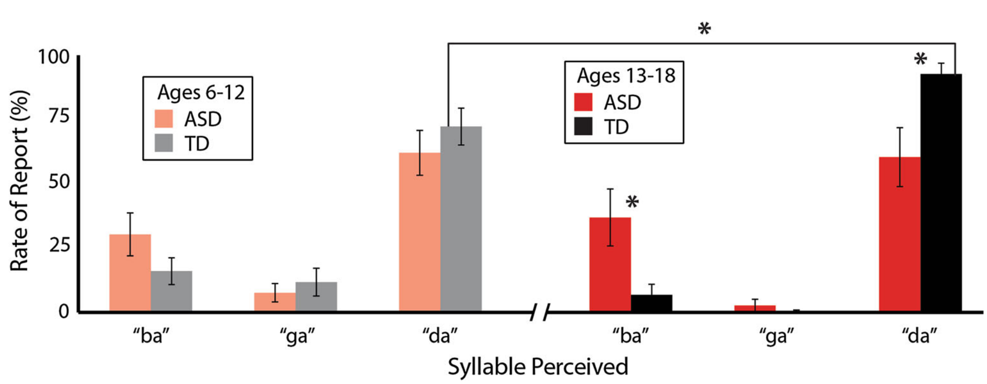

Fig. 1 Changes in the perception of the McGurk Effect with age. ASD = Autism spectrum disorders. TD = Typically developing control group. In the younger age groups (Panel A), no differences were found between individuals with ASD (*red*) and age-matched TD controls (*black*). In TD controls, the rate of perception of the McGurk effect is increased in the older age group relative to the younger age group. As a result, significant differences are observed between individuals with ASD and TD controls in the older groups in magnitude of the McGurk effect experienced (Color figure online)

children on average did not differ from older children in the rate at which they experienced the McGurk Effect (Table 2).

Follow-up independent samples t-tests revealed that there was not a statistically significant difference in the proportion of fused percepts reported by younger children with ASD and TD controls ($t_{(33)} = 0.94$, $p = 0.36$, $d = 0.33$), but that there was a significant difference in reports of fusion for the older children with ASD and TD controls ($t_{(24)} = 2.53$, $p = 0.02$, $d = 0.99$). As seen in Fig. 1, older children with ASD showed reduced reports of the McGurk effect relative to their TD peers. As seen in previous reports (Mongillo et al. 2008a, b; Iarocci et al. 2010; Irwin et al. 2011a, b) this older subgroup of children with ASD showed an increased rate of reporting the auditory token "ba" in response to incongruent multisensory trials in comparison to controls ($t_{(24)} = 2.42$, $p = 0.02$, $d = 0.95$) whereas the younger group did not ($t_{(24)} = 0.50$, $p = 0.13$, $d = 0.54$).

Additional analyses confirm that the older TD subgroup showed a higher rate of illusory McGurk perception relative to the younger TD subgroup ($t_{(29)} = 2.06$, $p = 0.04$, $d = 0.85$), as expected. In contrast, there was not a statistically significant difference in the rate of fusion reported by younger and older children with ASD.

Congruent Audiovisual and Unisensory Control Conditions

An initial repeated-measures ANOVA was run with control condition as a within-subjects factor and both age group and diagnosis as between subject factors. No age group × diagnosis interaction was found ($F_{(1,1)} = 0.99$, p = 0.33, partial-$\eta^2 = 0.02$), and no main effect of diagnosis was found ($F_{(1,57)} = 0.05$, p = 0.82, partial-$\eta^2 < 0.01$). There was, however, a main effect of age group ($F_{(1,57)} = 4.62$, $p = 0.04$, partial-$\eta^2 = 0.08$), where the older group was more accurate collapsed across conditions and diagnosis. See Fig. 2 for a synopsis of data from control conditions.

Accounting for Verbal IQ Differences

It should be noted here that when collapsed across age groups, show a significant difference in verbal IQ (see Table 1 for detailed statistics) that may impact these findings, given that the McGurk effect is verbal by nature. To ensure verbal IQ was not significantly contributing to our analysis of interest, analyses of verbal scores within each sub-group was examined; differences across diagnostic groups was observed in the younger cohort that exhibited no group differences in the McGurk effect, while the older cohort did not show significant differences in verbal IQ, suggesting that the group difference in McGurk effect in the older cohort was not driven by differences in verbal IQ. While there were no significant differences in between verbal IQ scores in the older group, the lack of significant differences does not necessarily imply that these two groups are perfectly matched according to more stringent criteria (Frick 1995; Mervis and Klein-Tasman 2004), and as such, the possibility of verbal IQ having an influence on McGurk effect in adolescents should be explored in depth in future experiments.

Discussion

Communication impairments are a core deficit in ASD (APA 2000; American Psychiatric Association and American Psychiatric Association. DSM-5 Task Force 2013), and are likely to play a central role in a wide range of the behavioral and cognitive characteristics that accompany ASD (Kuhl 2004; Tsao et al. 2004; Kuhl et al. 2005; Kuhl 2007; Pierce and Redcay 2008). Successful spoken communication skills are strongly reliant upon an individual's ability to process not only the information contained within the auditory stream, but also on an ability to effectively process the visual cues that accompany and complement the auditory content. Indeed, such visual cues not only contain information about the content of the communication signal, but also convey meaningful information concerning the affect and other attributes of the speaker, and thus could extend the significance of the current finding into the social domain as well. Collectively, the current results suggest that decreases in the ability to properly synthesize audiovisual speech signals may be the result of a delay in the maturation of this integrative capacity in individuals with ASD.

Table 2 A priori statistical comparisons across diagnostic group

Modality	Syllable	Perception	All subjects			Younger (6–12 years old)			Older (13–18 years old)		
			p	*t*	*d*	*p*	*t*	*d*	*p*	*t*	*d*
Audiovisual	McGurk "da"	ba	*0.01**	*2.80**	*0.72**	0.13	0.50	0.54	*0.02**	*2.42**	*0.95**
		ga	0.59	0.54	0.14	0.53	1.10	0.22	0.13	0.87	0.34
		da	*0.03**	*2.20**	*0.56**	0.36	0.12	0.33	*0.05**	*2.53**	*0.99**

* Significant between group difference ($p < 0.05$)

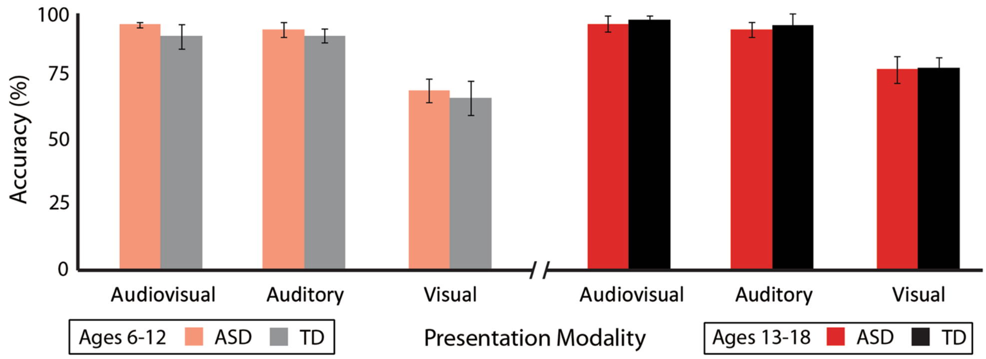

Fig. 2 Perception of multisensory and unisensory controls. ASD = Autism spectrum disorders. TD = Typically developing control group. Both younger and older groups showed high accuracies of congruent audiovisual, audio-only, and visual-only speech perception. No significant differences between diagnostic groups were observed in either the younger or the older age groups

The data in the current study, separated into age intervals, suggest that the maturational process for the ability to integrate audiovisual speech cues as measured by the McGurk effect diverges for ASD and TD groups, with the ASD population failing to show changes at the later time intervals when compared to younger ages. Intriguingly, the current data, which suggests a stalling of the normal sequence of developmental events associated with audiovisual speech stimuli, seems to run counter to prior work that has reported that children with ASD "catch up" to their peers with age in terms of their rates of McGurk perceptions (Taylor et al. 2010). There are a number of differences between these two studies that may explain these discrepant results. The most salient of these include the use of different stimuli, the employment of different and more varied phonemes, and the mode of response.

With that said, contradictory findings in McGurk perceptions in ASD is not limited to the differences between the current study and Taylor et al.'s (2010) study, but are in fact far from uncommon. While many studies show decreased McGurk fusions in ASD relative to TD (e.g., de Gelder et al. 1991—age range = 6–16 years; Williams et al. 2004a, b—mean age = 8.8 years, range = 5–13 years; Mongillo et al. 2008a, b—mean age = 13.7 years, range = 8–17 years; Irwin et al. 2011a, b—mean age = 9 years, range = 5–15 years), others find no difference (Iarocci et al. 2010—mean age = 10.6 years, range not reported; Woynaroski et al. 2013—mean age = 12.3 years, range = 8–17 years). Despite the number of studies investigating the McGurk effect in ASD, only this report and Taylor et al.'s (2010) factor in differences across chronological ages, and both find significant differences. Given the reliable finding of age-related difference coupled with the conflicting findings of the direction of that difference, future longitudinal studies of sensory integration across development should be implemented, avoiding both experimental design differences as well as between-individual differences inherently present in cross-sectional studies.

In the current study, the significant difference in the McGurk effect seen with older children with ASD relative to TD controls was driven by a higher frequency of reporting the auditory token ("ba"). The tendency to report the auditory component of the multisensory stimulus at a higher rate than TD controls is consistent with previous work (Mongillo et al. 2008a, b; Iarocci et al. 2010; Irwin et al. 2011a, b).

This over-reliance upon the auditory channel is not a result of an increase in reported "ba" percepts in the ASD group across age groups, but instead is a result of a failure of growth in the reports of the multisensory perceptions. A straightforward explanation of this could be a deficit in visual processing of more complex social stimuli such as faces (Blake et al. 2003; Dawson et al. 2005). However, our data, though not specifically designed to test this hypothesis, do not support it. First, there were no significant differences in low-level unisensory visual performance between any of the groups. Additionally, the only significant difference in the unisensory auditory performance was an increased reporting of auditory-only "ba" presentations as "da". This observed unisensory difference would predict an *increase* in McGurk perceptions in the ASD group (i.e. a tendency to report the fused "da" more frequently), the exact opposite of what was found. With that said, the majority of results associated with unisensory presentations was at or near ceiling, which may have masked possible unisensory differences. Taken in their entirety, these data suggest that this failure of the ASD group to display the developmental growth in audiovisual integration seen in their TD counterparts is likely not the result of a unisensory processing deficit, but rather is specific to the integration of information across the auditory and visual modalities.

The current study also suggests that multisensory processing may be plastic, changing throughout development. Recent work focused on the plasticity of multisensory processes may hold promise as a tool in the remediation of speech and language deficits in ASD. This work has utilized perceptual learning paradigms (Powers et al. 2009; Stevenson et al. 2013) and speech-reading training (Williams et al. 2004a, b) are capable of increasing the precision of multisensory integration. Providing further support for these prospects is the finding that such plastic changes in *behavioral* measures of multisensory processing are also accompanied by *neural* changes (Powers et al. 2012) that are concordant with known neurological differences between groups with ASD and typical development (Boddaert and Zilbovicius 2002; Boddaert et al. 2004; Pelphrey and Carter 2008a, b). Perceptual training may also be utilized in conjunction with other more conventional speech-language interventions (SLIs) that may impact speech processing in individuals with ASD (Yoder and Stone 2006). Further work investigating the ability of multisensory perceptual training to improve speech communication in ASD is warranted both relative to and in conjunction with SLIs.

Acknowledgments Funding for this work was provided by National Institutes of Health F32 DC011993, Multisensory Integration and Temporal Processing in ASD, National Institutes of Health R34 DC010927, Evaluation of Sensory Integration Treatment in ASD, a Simons Foundation research grant, Exploring Links Between Multisensory and Cognitive Function in Autism, a Vanderbilt Institute for Clinical and Translational Research grant VR7263, Neuroplasticity of Sensory Processing in Autism Spectrum Disorders, a Vanderbilt Kennedy Center MARI/Hobbs Award, the Vanderbilt Brain Institute, and the Vanderbilt University Kennedy Center. We also acknowledge the help of Zachary Barnett for technical assistance and Lena Quinto for the stimuli, and thank G.O.B. for his assistance with the multisensory illusions.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders-IV-TR*. Washington, D.C.: American Psychological Association.

American Psychiatric Association, & American Psychiatric Association. DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, VA: American Psychiatric Association.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speech reading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339–355.

Baggaley, A. R. (1988). The relative usefulness of multiple measures for differentiating among groups of people. *Educational and Psychological Measurement, 48*, 89–92.

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*(1), 88–98.

Billstedt, E., Gillberg, I. C., & Gillberg, C. (2007). Autism in adults: Symptom patterns and early childhood predictors. Use of the DISCO in a community sample followed from childhood. *Journal of Child Psychology and Psychiatry, 48*(11), 1102–1110.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science, 14*(2), 151–157.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., et al. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: A voxel-based morphometry MRI study. *Neuroimage, 23*(1), 364–369.

Boddaert, N., & Zilbovicius, M. (2002). Functional neuroimaging and childhood autism. *Pediatric Radiology, 32*(1), 1–7.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in cognitive sciences, 2*(7), 247–253.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders, 28*(6), 479–485.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*(2), 271–283.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27*(3), 403–424.

de Gelder, B., Vroomen, J., & Van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology, 3*(1), 69–86.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., et al. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research, 203*(2), 381–389.

Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition, 23*(1), 132–138.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22*(6), 713–722.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders, 36*(1), 77–90.

Iarocci, G., Rombough, A., Yager, J., Weeks, D. J., & Chua, R. (2010). Visual influences on speech perception in children with autism. *Autism, 14*(4), 305–320.

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011a). Can children with autism spectrum disorders "hear" a speaking face? *Child Development, 82*(5), 1397–1403.

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011b). Can children with autism spectrum disorders "hear" a speaking face? *Child Development, 82*(5), 1397–1403.

Jones, J. A., & Munhall, K. G. (1997). Effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics, 25*(4), 13–19.

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

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843.

Kuhl, P. (2007). Is speech learning 'gated' by the social brain? *Developmental Science, 10*(1), 110–120.

Kuhl, P., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science, 8*(1), F1–F12.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience, 4*, 129.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H, Jr, Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

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*(5), 659–685.

Macdonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Attention, Perception, & Psychophysics, 24*(3), 253–257.

Massaro, D. (1984). Children's perception of visual and auditory speech. *Child Development, 55*(5), 1777–1788.

Massaro, D. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1995). Perceiving talking faces. *Current Directions in Psychological Science, 4*(4), 104–109.

Massaro, D., Cohen, M., & Smeele, P. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America, 100*(3), 1777–1786.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

Mervis, C. B., & Klein-Tasman, B. P. (2004). Methodological issues in group-matching designs: Alpha levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders, 34*(1), 7–17.

Middelweerd, M. J., & Plomp, R. (1987). The effect of speech reading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America, 82*(6), 2145–2147.

Mongillo, E., Irwin, J., Whalen, D., Klaiman, C., Carter, A., & Schultz, R. (2008a). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(7), 1349–1358.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008b). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(7), 1349–1358.

Munhall, K., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Attention, Perception, & Psychophysics, 58*(3), 351–362.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pelphrey, K. A., & Carter, E. J. (2008a). Brain mechanisms for social perception: Lessons from autism and typical development. *Annals of the New York Academy of Sciences, 1145*, 283–299.

Pelphrey, K. A., & Carter, E. J. (2008b). Charting the typical and atypical development of the social brain. *Development and Psychopathology, 20*(4), 1081–1102.

Pierce, K., & Redcay, E. (2008). Fusiform function in children with an autism spectrum disorder is a matter of "who". *Biological Psychiatry, 64*(7), 552–560.

Powers, A. R, 3rd, Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience, 32*(18), 6263–6274.

Powers, A. R, 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience, 29*(39), 12265–12274.

Quinto, L., Thompson, W. F., Russo, F. A., & Trehub, S. E. (2010). A comparison of the McGurk effect for spoken and sung syllables. *Attention, Perception, & Psychophysics, 72*(6), 1450–1454.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Attention, Perception, & Psychophysics, 59*(3), 347–357.

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience, 33*(12), 2329–2337.

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science, 11*(2), 306–320.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition, 92*(3), B13–B23.

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research, 225*(4), 479–89.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517–29.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*(9), 400–410.

Taylor, N., Isaac, C., & Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of Autism and Developmental Disorders, 40*(11), 1403–1411.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75*(4), 1067–1084.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181–1186.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio: Harcourt Brace.

Williams, J., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004a). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*(6), 559–575.

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004b). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*(6), 559–575.

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech rerception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1836-5.

Yoder, P., & Stone, W. L. (2006). A randomized comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Journal of Speech, Language and Hearing Research, 49*(4), 698–711.