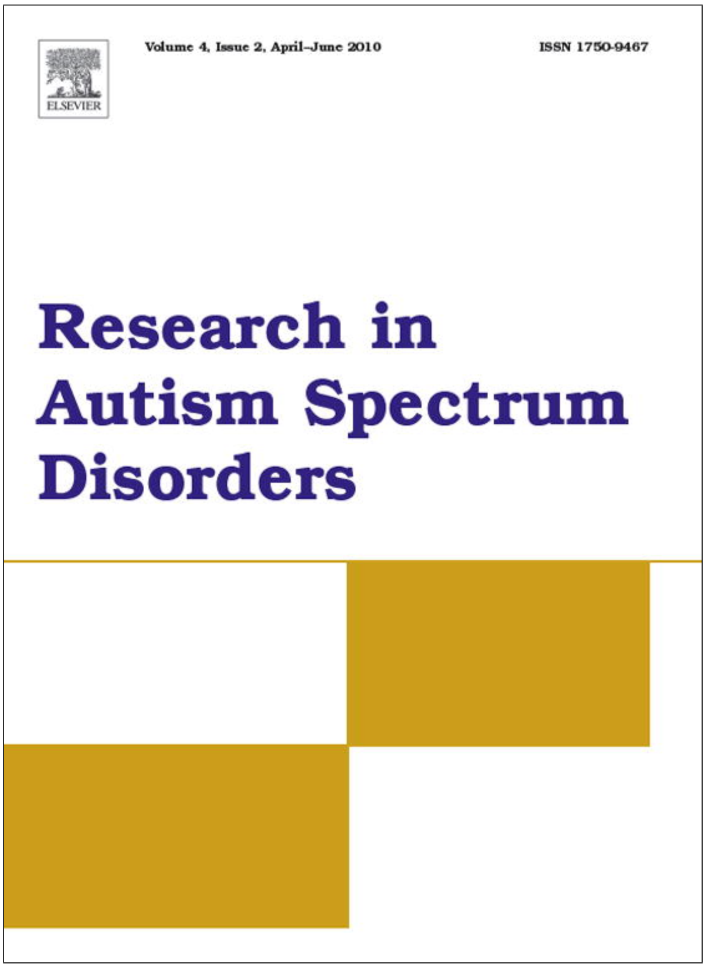

Volume 4, Issue 2, April–June 2010
ISSN 1750-9467

Research in Autism Spectrum Disorders

Research in Autism Spectrum Disorders 4 (2010) 276–289

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

Audiovisual integration in high functioning adults with autism

Brian P. Keane [a], Orna Rosenthal [b], Nicole H. Chun [a], Ladan Shams [a,*]

[a] Department of Psychology, University of California, Los Angeles, United States
[b] School of Psychology, University of Birmingham, UK

ARTICLE INFO

Article history:
Received 16 September 2009
Accepted 21 September 2009

Keywords:
Audiovisual integration
Multisensory integration
High-functioning autism
Speech integration
Crossmodal dynamic capture
Sound-induced flash illusion

ABSTRACT

Autism involves various perceptual benefits and deficits, but it is unclear if the disorder also involves anomalous audiovisual integration. To address this issue, we compared the performance of high-functioning adults with autism and matched controls on experiments investigating the audiovisual integration of speech, spatiotemporal relations, and temporal numerosity. In each experiment, performance for both groups was faster and more accurate when audiovisual information was congruent rather than incongruent. Importantly, audiovisual congruency did not affect the control group more than the autism group. These results suggest that the ability to integrate between the auditory and visual sense modalities is unimpaired among high-functioning adults with autism.

1. Introduction

Persons with autism typically show a unique set of perceptual strengths and weaknesses. On the one hand, they show memory for exact pitches (Heaton, Hermelin, & Pring, 1998), enhanced local processing of music (Heaton, 2003), superior visual search (O'Riordan, Plaisted, Driver & Baron-Cohen, 2001), and a superior ability to differentiate confusable visual patterns (Plaisted, O'Riordan, & Baron-Cohen, 1998). On the other hand, people with autism are typically impaired at detecting coherent motion (Bertone, Mottron, Jelenic, & Fauber, 2003; Milne et al., 2002; Tsermentseli, O'Brien, & Spencer, 2008), biological motion (Blake, Turner, Smoski, Pozdol, & Stone, 2003) and global/configural stimuli (Brosnan, Scott, Fox, & Pye, 2004; see Happé & Frith, 2006, for a review). The purpose of the present paper was to examine psychophysically whether autism also involves abnormalities in integrating information across the visual and auditory sensory modalities. This inquiry, aside from elucidating autism's broad phenotype, will at least indirectly bear on neurophysiological and functional theories of the disorder.

1.1. Multisensory integration in autism: an unsettled matter

Although multisensory integration is central to constructing a coherent representation of the world, the process is not well understood in the context of autism. Some studies suggest that an endophenotype of autism includes a failure to fully integrate. In a study by Smith and Bennetto (2007), a group of adolescents with autism and a typically developing (TD) group of adolescents heard spoken sentences that were embedded in noise. The sentences were presented either with their corresponding faces or without any visual information. It was found that visual information decreased volume thresholds (the amount of volume required to identify key words of each sentence) more for the TD group. This effect persisted even

---

[*] Corresponding author at: University of California, Los Angeles, Department of Psychology, 1285 Franz Hall, Los Angeles, CA 90095-1563, United States. Tel.: +1 310 206 3630; fax: +1 310 267 2141.

E-mail addresses: Brian.Keane@gmail.com (B.P. Keane), ladan@psych.ucla.edu (L. Shams).

1750-9467/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rasd.2009.09.015

after lip-reading ability was controlled (apparently compromised in autism, see Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004). In a similar study, when children with autism were instructed to repeat what a woman had said, and the heard and viewed syllables (as observed at the mouth) were incongruent (de Gelder, Vroomen, & Van Heide, 1991), members of the autism group combined/fused the spoken and observed syllables less than the TD group (see also, Mongillo et al., 2008). Some have tried to link these kinds of results to functional and/or neuroanatomical under-connectivity between brain areas (e.g., Brock, Brown, Boucher, & Rippon, 2002; Courchesne & Pierce, 2005).

Not all studies are in agreement with the foregoing. Williams et al. (2004) controlled for lipreading ability and showed that children with autism normally utilized visual information in identifying auditory speech. In a related ERP study of speech integration, early (presumably sensory) processing of congruent and incongruent audiovisual speech was comparable to controls (Magnée, de Gelder, van Engeland, & Kemner, 2008). When children with autism determined whether the specific sound of a bouncing ball matched its physical appearance, the performance of clinical group was again not compromised (Mongillo et al., 2008; see also, Morsanyi & Holyoak, in press). The ability to judge the number of pulsations of a visual flash (temporal numerosity integration task) during concurrent presentations of sound bursts also is preserved in autism (van der Smagt, van Engeland, & Kemner, 2007). Finally, and perhaps most surprisingly, when an electrical stimulation was applied to the median nerve of the wrist, loudness perception of a simultaneous tone was more altered for persons with autism than TD persons (Møller, Kern, & Grannemann, 2005; see also, Cesaroni & Garber, 1991, p. 305, and Kern et al., 2007). These results, taken together, do not yet yield a consensus as to what extent, if any, multisensory integration is impaired in autism.

1.2. *Motivation and approach*

We approached this issue by comparing the performance of high-functioning adults and typically developed matched controls on three audiovisual perceptual tasks. In Experiment 1, participants attempted to detect an auditory syllable when observing a mouth speaking the same or different syllable. Because audiovisual incongruence produces a non-veridical perception of either fused or combined syllables (Kaiser, Hertrich, Ackermann, Mathiak, & Lutzenberger, 2005; McGurk & MacDonald, 1976) and because audiovisual congruence causes bisensory facilitation (Kaiser et al., 2005; Sumby & Pollack, 1954; see Campbell, 2008 for a review), performance was expected to be faster and more accurate on congruent than on incongruent trials. In Experiment 2, participants attempted to discriminate the order of a sequence of two spatially displaced beeps while simultaneously observing a sequence of two spatially displaced flashes. When the modalities were incongruent, participants were expected to rely more on visual information through an effect sometimes referred to as *crossmodal dynamic capture* (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Soto-Faraco, Spence, & Kingstone, 2004). Performance was consequently expected to be faster and more accurate when the visual and auditory directions were congruent rather than incongruent. In Experiment 3, participants attempted to detect the number of times that a disk flashed, while two speakers synchronously emitted one or two beeps. Because incongruence between the number of beeps and flashes causes illusions (Shams, Kamitani, & Shimojo, 2000; Shams, Kamitani, & Shimojo, 2002; Wozny, Beierholm, & Shams, 2008), performance in these incongruent trials was expected to be less accurate or slower than when the number of beeps and flashes matched.

In all experiments, audiovisual integration was assessed by comparing reaction time (RT) and accuracy differences between congruent and incongruent trials. Larger or smaller differences for the autism group relative to the TD group would suggest that autism involves increased or reduced crossmodal integration, respectively.[1] Our two-pronged measure of performance furnished a unique advantage over the assessment methods of existing studies. Crossmodal differences have been explored almost exclusively with respect to accuracy or bias (de Gelder et al., 1991; Møller, Kern, & Granneman 2005; Mongillo et al., 2008; Smith & Bennetto, 2007; van der Smagt et al., 2007; Williams et al., 2004), but unless RT is also measured, there remains the possibility of speed-accuracy trade-offs. By analyzing both accuracy and RT, we were able to obtain a more complete picture of audiovisual integration in autism.

The paradigms in our study were selected for a number of reasons. First, all three yield robust effects in TD individuals. Crossmodal speech interactions can occur even when the audiovisual stimuli are out of synch by 250 ms (Grant, Greenberg, Poeppel, & Wassenhove, 2004), when faces are blurred (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004), or when faces are presented as point light displays (Rosenblum, Johnson, & Saldaña, 1996). Crossmodal dynamic capture, although more recently discovered, has been repeatedly shown to yield nearly perfect and below-chance performance on congruent and incongruent trials, respectively (Sanabria, Soto-Faraco, & Spence, 2004; Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Soto-Faraco, Spence, Lloyd, & Kingstone, 2004). Finally, the sound-induced flash illusion can occur with a variety of flash and beep durations and interstimulus intervals (ISIs), with various kinds of auditory and visual stimuli, and even at large spatial discrepancies between visual and auditory stimuli (Mishra, Martinez, Sejnowski, & Hillyard, 2007; Shams et al., 2000, 2002; van der Smagt et al., 2007; Watkins, Shams, Josephs, Rees, 2007; Watkins, Shams, Tanaka, Haynes, Rees, 2006).

---

[1] It should be emphasized that, for present purposes, the term 'audiovisual integration' is defined rather broadly to signify any effect of one modality on the other. We consequently do not distinguish between when the irrelevant modality degrades, suppresses, or clearly transforms the percept of a presented stimulus in the target modality.

Our tasks were selected also because they likely involve different brain areas and processes, and thus can provide a more general perspective of audiovisual integration in autism. For example, congruent and incongruent audiovisual speech produces super- and sub-additive responses, respectively, in the posterior superior temporal sulcus (pSTS) compared with visual/auditory unisensory baselines (Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003; see also, Campbell, 2008). Judgments about the relation between auditory and visual motion direction—which is akin to judgment of sequence order used in our study—involve enhanced response in the intraparietal sulcus, anterior midline, and anterior insular cortex (Lewis, Beauchamp, & DeYoe, 2000; see also, Baumann & Greenlee, 2007). Finally, the sound-induced flash illusion incorporates some of the earliest multisensory sites, including the visual and auditory cortices (Mishra et al., 2007; Shams, Iwaki, Chawla, & Bhattacharya, 2005). Since audiovisual integration of speech, motion and temporal numerosity signals involve neural networks and processes that are largely distinct, our results will apply to the processing that occurs at a variety of levels.

1.3. General methods

1.3.1. Participants

Ten highly functioning adults with autism took part in the study. They were recruited through the UCLA Center for Autism Research and Treatment (CART), and local clinics (see Table 1). These participants met criteria for autistic disorder according to DSM-IV (American Psychiatric Association, 2000) as evidenced by ADOS assessments (see Section 1.3.2). All participants with autism had ~80 or higher Full-Scale IQ, and 90 or higher performance IQ (see Section 1.3.2).

Potential TD participants were recruited in the West Los Angeles area via posted fliers and Internet/radio advertisements. In order to participate, candidates could not have a history/presence of developmental, neurological, psychiatric or learning disorders. Nine of the ten persons with autism were individually matched on the basis of sex, age (±2 years) and performance IQ (within the range of 95% confidence intervals). Data from the unmatched person with autism were not included in our analyses.

One person with autism and two TD participants were either unwilling or unable to complete Experiment 3. Thus data from only 6 of the 9 pairs of participants were used to assess potential differences in temporal numerosity integration. Members of both groups were screened for sensory and cognitive disabilities. All participants were confirmed to have normal or corrected-to-normal visual acuity (20/25 or better), except for one participant with autism (20/50). All participants were informally verified to have the abilities to identify heard syllables and understand spoken speech. Each participant had to possess at least an 8th grade verbal IQ and thus each demonstrated competence in reading and understanding the informed consent form. Finally, all were naive to the purpose of the experiments.

1.3.2. Methods of assessment

Eligibility for the autism group was decided on the basis of the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). For six of the matched participants, the Autism Diagnostic Interview—Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) was also considered in the assessment. When possible, diagnosis was based on previous authenticated clinical records, provided by the participant. In all other cases, the assessment was carried-out by—or under the direct supervision of—a senior CART clinician.

Verbal, Performance and Full-Scale IQ scores of 6 autistic participants, and 8 TD participants were assessed with WASI (Wechsler Abbreviated Scale of Intelligence; Wechsler, 1999). IQ assessments of the remaining participants were based on previous official records on WAIS-III (Wechsler Adult Intelligence Scale; Wechsler, 1997), and were furnished by CART or by the participants themselves.

Table 1

Pair	Sex		Age		PIQ		VIQ		FSIQ	
	A	TD	A	TD	A	TD	A	TD	A	TD
1	M	M	23	25	90	94	58	101	79	98
2	F	F	29	28	122	121	135	133	133	131
3	M	M	49	47	105	100	105	98	105	100
4	M	M	27	25	107	111	104	117	106	116
5	M	M	32	33	137	140	90	135	112	142
6	M	M	18	19	112	117	127	98	123	108
7	M	M	31	32	99	103	80	94	88	99
8	M	M	34	32	90	91	75	99	80	95
9	M	M	25	25	129	124	139	135	139	133
Mean			30	30	110	111	101	112	107	114
SD			9	8	17	16	28	18	22	18
Range			18–49	19–47	90–137	91–140	58–139	94–135	79–139	95–142

Note: Individual pairings (above) and group statistics (below) are shown for the autism group (A) and typically developed group (TD), where PIQ, VIQ, and FSIQ denote Performance, Verbal, and Full-Scale IQ, respectively. Each of the nine participants with autism was matched individually on the dimensions of sex, age, and PIQ.

Table 2
Performance data for unisensory and catch trials.

	Auditory-Only				Visual-Only				Catch			
	% Correct		RT (s)		% Correct		RT (s)		False Alarms (%)		Misses (%)	
	A	TD	A	TD	A	TD	A	TD	A	TD	A	TD
Experiment 1												
M	92	94	0.80	0.50	82	90	0.98	0.73	<1	<1	2	1
SD	9	5	0.39	0.26	12	8	0.53	0.45	<1	<1	3	2
Experiment 2												
M	89	89	1.82	1.39	99	99	1.26	0.96	<1	<1	2	2
SD	9	6	0.47	0.45	2	1	0.35	0.22	<1	<1	2	2
Experiment 3												
M	98	100	0.89	0.70	85	86	0.99	0.93	<1	<1	3	2
SD	4	0	0.25	0.19	12	11	0.25	0.23	<1	<1	2	3

Note: Means and standard deviations are shown for the autism (A) and typically developed (TD) groups for each experiment; RT = reaction time. Experiment 1–3 involved audiovisual speech, motion, and temporal numerosity integration, respectively. The target-modality was auditory for Experiments 1 and 2, and visual for Experiment 3.

1.3.3. General design

The three experiments shared a similar design. In each case, there were four unique blocks, and each block involved a yes–no task. The first block was always practice, involved only the target modality (e.g., in Experiment 3, identification of a single flash), and served to ensure that each participant was able to engage in the task. During practice trials, but not other trials, feedback was provided after each response in written form (e.g., ''Wrong'' or ''Right'' in red and green fonts, respectively). The practice block terminated when participant's responses responded correctly in 9 of 10 consecutive trials. Tasks were designed so that they could be easily explained and so that criterion performance could be reached with minimal practice. In the auditory-only block, participants attempted to detect an auditory stimulus without any visual stimulus. In the audiovisual block, participants attempted to detect the target stimulus in the presence of the irrelevant modality. In the visual-only block, participants attempted to detect a visual stimulus in the absence of any auditory input. Participants were asked to respond accurately, and were not given any instructions on how fast to respond. In all experiments, the last block was always the non-target modality to ensure that participants would not have to switch back-and-forth between auditory- and visual-detection tasks within an experiment.

In each experiment, participants were instructed to keep visually focused on the fixation point region throughout a trial. To further promote vigilance to the visual display, we added a secondary task that required participants to detect—via a button-press—a briefly appearing white stimulus. These ''catch'' stimuli had the same dimensions and location as the fixation cross, except in Experiment 3, where the catch stimulus was a white dot that appeared within the cross. Catch trials occurred randomly and non-consecutively on about 20% of the trials of each block. When a catch stimulus was missed or falsely identified (which happened rarely, see Table 2), a screen text message informed the participant to contact the experimenter to enable continuation of the experiment.

1.3.4. Apparatus

Experiments were run on a Macintosh computer (OS 9.2.2), and were programmed with Matlab (5.2.1) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Visual stimuli were displayed on a monitor that had: a resolution of 1024 × 768 pixels, a refresh rate of 100 Hz (Experiments 1 and 2) or 75 Hz (Experiment 3), and a viewable screen of 14 in. × 10.5 in. Participants were seated with a chinrest that was 22.5 in. (57 cm) from the monitor, so that the viewable screen subtended an area of 35 × 26° of visual angle. Auditory stimuli were presented by two speakers—one on each side of the monitor. The speakers were separated from one another by 22 in., and were located 26 in. from the participant at approximately ear-level. The room was dimly lit and had background noise of 39 db.

1.3.5. Analyses

Data were analyzed in the same way for all experiments. Catch trials or trials falsely identified as having the catch stimulus were excluded from data analyses, unless stated. RT outliers were defined as points that fell beyond ±2 SD from a participant's mean for a trial-type and were also removed. Outliers were calculated regardless of accuracy, and separately for auditory-only, visual-only, congruent, and incongruent trials.

Performance in the unisensory trials of the target modality served as a baseline. Therefore, prior to conducting ANOVAs, average performance in the target-modality, unisensory trials was subtracted from the average performance of each trial-type in the bisensory block. This procedure was performed on both accuracy and reaction time data, and was done separately for each participant. Trial-types of the bisensory block were categorized according to whether they were congruent or incongruent, and also according to whether the target stimulus item (e.g., /pa) was or was not presented. We took into account this last variable in order to explore whether potential integration differences between the groups depended on attention to the specified target. Accuracy and RT data sets were each submitted to a 2 (congruency)[2] by 2 (participant group) by 2 (target stimulus presence) within-subjects, repeated-measures analysis of variance (ANOVA).[3] All follow-up *t*-tests were paired and two-tailed.

2. Experiment 1: audiovisual speech integration (McGurk illusion)

The primary aim of Experiment 1 was to determine whether autism is associated with anomalous audiovisual speech integration. In some trials the auditory and visual stimuli were congruent to produce bisensory facilitation; in other trials, the mouthed and spoken syllables were incongruent, so that illusory syllables would be heard. If RT and accuracy differences between congruent and incongruent trials in the autism group were larger or smaller than the differences found in the TD group, then that would suggest increased or reduced integration in autism, respectively.

2.1. Method

2.1.1. Stimuli

In all trials, there was a fixation cross (a black plus sign, Helvetica 18 pt) that appeared between the trials, though not within a trial itself. The location of the fixation cross was the same location as the upper chin of the visually displayed mouths.

The auditory stimulus in the test blocks consisted of one of two consonant–vowel syllables: /pa or /ta.[4] Syllables derived from two adult females. When presented with auditory /pa and visual 'ta', observers tend to 'fuse' the information from the two modalities and perceive a /ka syllable. In order to preserve this illusory syllable percept in our experiment, and not to bias the participant into expecting exactly two syllable types for the upcoming bisensory block, a third /ka syllable was added to the syllable repertoire in the practice block. The /ka syllable was produced by one of the same female speakers that produced the /pa or /ta syllable. The sound level was low but still detectable (reaching an amplitude of about 14 db above the background noise). This sound level was selected on the basis of pilot studies and assured strong crossmodal effects. The visual stimulus consisted of a close-up video of the mouth and chin (but not the eyes) of one of the two female speakers against a black background. During trials that involved auditory and visual information, voice timing was matched as closely as possible to the movement of the mouth in each video clip. The effectiveness of the dubbed stimuli in inducing fusion and combination illusions was confirmed in a pilot experiment. The clips were centered on a black screen, lasted between 2.1 and 2.4 s, and spatially spanned a rectangle that measured 12.5° horizontally and 8.4° vertically.

2.1.2. Procedure and design

Prior to the beginning of the practice block, the participants were presented with a written "pa", and were asked to read it and say it out-loud. All participants correctly identified and produced the syllable, and were then told that this syllable would be the target of the experiment. Next, participants were seated in the experiment room, and received instructions for the practice trials. They were asked to keep fixated on the screen throughout each trial, and were told to press one key if they believed the speaker said the "pa", and to respond with the alternative key, otherwise. A third key was to be pressed, if they witnessed the catch stimulus. RT was measured from the offset of the movie clips.

In each trial, after the video clip was played, the clip display area became black (if it was not black already). The fixation cross then appeared at the center of the display. Participants entered a response, and feedback was provided (if the trial was practice). Participants could proceed to the next trial by pressing a key, after which point there was a 250 ms pause.

After the practice block, each participant received the three test blocks in the following order: audiovisual (128 non-catch trials), auditory-only (32 non-catch trials), and visual-only (32 non-catch trials). For the audiovisual block, a video showed one of two females speaking either 'pa' or 'ta' with the voice of the same speaker saying either the same (congruent) or different (incongruent) syllable. There were four audiovisual pairings of the syllables: /pa-'pa', /pa-'ta', /ta-'pa', /ta-'ta'; each occurred with equal frequency and pseudo-randomly within the bisensory block. Trials in the practice and auditory-only test blocks were identical to the audiovisual block, with the exception that the movie area appeared black (unless there was a catch stimulus). The visual-only block contained trials that were silent versions of the audiovisual trials. On catch trials, the catch stimulus occurred 200 ms into a movie clip, and remained visible for 20 ms.

The instructions in all but the last block were the same as the practice, and were repeated at the beginning of each block. Instructions in the last (visual-only) block were to identify whether or not the mouthed syllable was 'pa'. Participants were further told to respond 'pa' if they thought the syllable was 'ba', since mouths speaking either syllable have an almost

---

[2] In the upcoming discussion, "congruency" is an independent variable, and "congruent" and incongruent" are the levels of that variable.

[3] We treated participant group as a within-subject variable, since participants were matched individually (rather than as a group) on the dimensions of age, sex, and performance IQ.

[4] A slash denotes an auditory syllable (/pa); quotations denote a visual syllable, e.g., "pa".

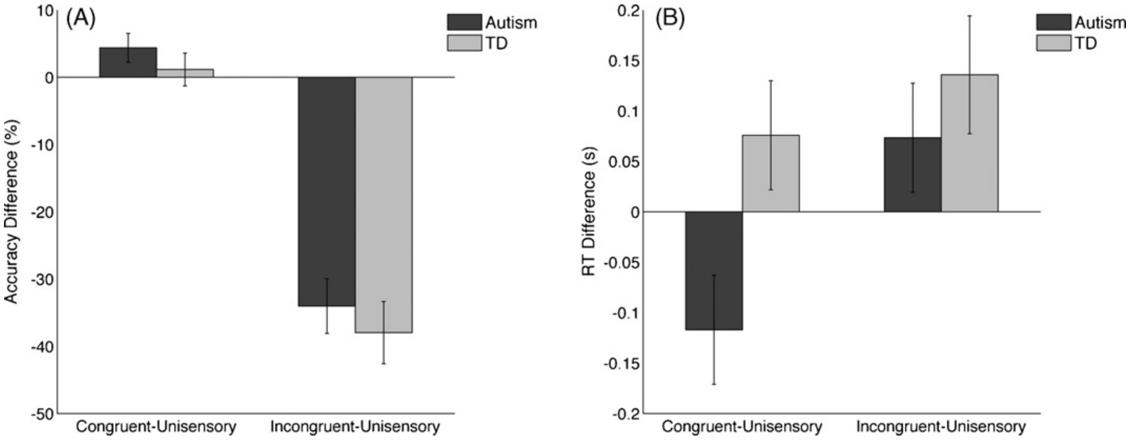

Fig. 1. Results from the audiovisual speech integration task in Experiment 1. (a) For the autism and TD groups, average accuracy is shown for the congruent and incongruent conditions, after subtracting the average accuracy in the unisensory (auditory-only) condition. (b) The same differences are shown for each group for the reaction time data. Error bars in both graphs depict ±1 SEM.

indistinguishable appearance (based on pilot studies). All other procedural aspects of the trials in the test blocks were identical.

2.2. Results and discussion

The two groups performed comparably on a host of measures, including: accuracy during auditory-only trials, RT and accuracy in the visual-only trials, rate of falsely identified catch trials, and rate of missed catch trials (see Table 2). The two groups also required a comparable number of practice trials to reach the performance criterion. The groups differed in RT for the auditory-only trials, with the autism group taking longer to respond, $t(8) = 2.6$, $p = 0.03$. However, this effect would not be statistically significant with a Bonferroni-correction ($p_{crit} = 0.007$).

Accuracy results from Experiment 1 are shown in Fig. 1a. There was a main effect of congruency, such that congruent trials overall produced more accurate responses than incongruent trials, $F(1,8) = 63.0$; $p < 0.001$, $\eta_p^2 = 0.89$. This indicates that, across both groups, crossmodal speech integration was strong. There was a significant main effect of target presence, where trials involving the target (/pa) yielded worse performance than /ta, $F(1,8) = 39.6$, $p < 0.001$, $\eta_p^2 = 0.83$. This outcome probably owed—at least, in part—to the illusion being stronger, and thus performance being lower, with a /pa-“ta” combination than with a /ta-“pa” combination (Kaiser et al., 2005). The outcome also may have occurred as a result of heightened attention to the target syllable (Alsius, Navarra, Campbell, & Soto-Faraco, 2005). The effect of congruency was more pronounced when the auditory stimulus was /pa, $F(1,8) = 102$, $p < 0.001$, $\eta_p^2 = 0.93$. This result too likely owed to either stronger illusions during /pa-“ta” trials, heightened attention towards the target syllable, or both. Most importantly, the congruency variable did not affect performance differently in the two groups, $F(1,8) = 0.02$, $p = 0.90$. No other effects in the three-way ANOVA were significant ($p > 0.20$).

RT results are shown in Fig. 1b. An ANOVA applied to the RT data revealed an interaction between the congruency and target presence variables such that the effect of congruency was stronger when the target was presented, $F(1,8) = 5.8$; $p < 0.05$, $\eta_p^2 = 0.42$. As noted above, this outcome probably arose from an asymmetry in illusion strength or to heightened attention to the target syllable. More importantly, there was a significant interaction between participant group and congruency ($F(1,8) = 7.1$; $p < 0.05$, $\eta_p^2 = 0.47$) with the autism group *more* affected by congruency. (NB: This interaction was robust: removing data outside ±1.5 or ±2.5 standard deviations or outside of the 0.1 and 2.5 s range did not change the results.) To determine whether the interaction was due to enhanced bisensory facilitation or a stronger illusion in the autism group, we performed Bonferroni-corrected *t*-tests ($p_{crit} = 0.03$) on the congruent and incongruent conditions. Prior to performing the tests, data were collapsed across the /pa and /ta trials, since the three-way interaction was not significant, $F(1,8) = 0.69$; $p > 0.43$. The tests showed that while the TD and autism participants responded similarly to incongruent trials relative to their respective unisensory baselines ($t(8) = 0.68$, $p = 0.52$), the autism group exhibited more benefit in the congruent trials, $t(8) = 2.72$, $p < 0.03$.

The TD group was not faster in the congruent than the unisensory trials, $t(8) = 1.4$, $p = 0.20$. This may reflect a floor effect: the relatively easy unisensory task may have produced speeded responses that could not be further improved by adding audiovisual congruence.

To our knowledge, the present experiment is the first to utilize both RT and accuracy measures to assess audiovisual speech integration in autism. Our results suggest that people with autism integrate at least to the same degree as TD individuals. In particular, while the autism group did not exhibit a different congruency effect with respect to accuracy, RT differences were greater among those with autism. The accuracy results fit well with the findings of Williams et al. (2004) and Massaro and Bosseler (2003) but are inconsistent with certain other findings of reduced audiovisual speech fusion in autism (de Gelder et al., 1991; Smith & Bennetto, 2007). The RT results are novel, and suggest that people with autism may benefit from audiovisual speech congruence more than previously supposed. Both sets of results will be further considered (and qualified) in Section 5.

3. Experiment 2: crossmodal dynamic capture

The first experiment examined whether people with and without autism differ in the tendency to integrate auditory and visual speech information. In Experiment 2, we examined whether the two groups differed in crossmodal dynamic capture. Observers were presented with a left–right or right–left auditory (beep) sequence, and—at the same time—a left–right or right–left visual (flash) sequence. The primary task was to indicate whether or not the beep sequence began on the left. Incongruent visual and auditory sequence orders are known to induce crossmodal dynamic capture, where the perceived order of auditory sequence is dominated by the perceived order of the visual sequence (e.g., Soto-Faraco et al., 2002). The effect has been suggested to reflect integration of motion information between the auditory and visual pathways (Soto-Faraco, Spence, & Kingstone, 2004; Soto-Faraco, Spence, Lloyd, et al., 2004). As before, if the autism group is less or more affected by audiovisual congruency than the TD group, then that would suggest that the disorder involves reduced or heightened integration between the sense modalities, respectively.

3.1. Method

3.1.1. Stimuli

Throughout each trial, a red fixation point was presented at the center of the monitor display. Two vertically centered disks (diameter = 0.7°) appeared 7.4° to the immediate left and right of fixation. The disks were continually grey, unless there was a flash sequence, in which case they briefly flashed white. The ISI and stimulus onset asynchrony (SOA) of the flashed disks was 100 and 150 ms, respectively (so that each flash lasted 50 ms). The auditory stimulus consisted of two sequential tones ($\sim$500 Hz), one from the left speaker and one from the right speaker. Beep pairs had the same SOA and duration as the flash pairs, and could induce a percept of left–right or right–left apparent motion. To increase the influence of visual signals, the beeps were presented at low volume, and their sound pressure level exceeded that of the background noise only by 1 db. This volume, although quite low, was still detectable by all participants, as confirmed by high performance on the auditory-only trials (see Table 2).

3.1.2. Procedure and design

In the practice block, participants were instructed to press one key if they heard the auditory sequence begin on the left, and press another key, otherwise. At the beginning of each trial, the fixation point appeared. One second later, the beep and/or flash sequence occurred. Participants entered a response, and RT was measured from the onset of the first beep/flash stimulus.

The blocks occurred in the following order: auditory-only (practice), audiovisual, auditory-only, and visual-only. The number of practice trials was determined in the same way as in Experiment 1, except that participants had to receive at least 19 trials of practice to ensure familiarity with the task. Each test block contained 128 non-catch trials. In the audiovisual block, beeps and flashes co-occurred, resulting in two congruent conditions—so that the auditory and visual sequences transitioned in the same direction—and two incongruent conditions, where the orderings of the flash and beep sequences transitioned opposite to one another (see Fig. 2). The four audiovisual pairings were presented pseudo-randomly within a block.

Stimuli in the practice and auditory-only blocks were identical to those of the audiovisual block, with the exception that disks remained grey throughout a trial. Trials in the visual-only block were silent versions of trials in the audiovisual block. Catch stimuli appeared during the entire 100 ms ISI, and occurred as described in the General Methods section. In each test block, the instructions were the same as the practice block, except in the audiovisual block, where participants were told that they would also be shown irrelevant visual information (which was to be ignored), and in the visual-only block, where participants were instructed to respond on the basis of the whether the flash sequence began on the left.

3.2. Results and discussion

As can be seen in Table 2, the two groups performed comparably on a number of measures including: rate of falsely identified catch trials, rate of missed catch trials, accuracy and RT during auditory-only trials, and accuracy in the visual-only trials. The two groups also required a similar number of practice trials to reach performance criterion. The autism group showed a trend of slower response in the visual-only trials ($t(8) = 2.9$, $p = 0.02$), but this effect would not be significant with an accompanying Bonferroni-correction ($p_{crit} = 0.007$).

Accuracy results for Experiment 2 are shown in Fig. 3a. An ANOVA showed that accuracy was greater for congruent than for incongruent trials, $F(1,8) = 189$, $p < 0.001$, $\eta_p^2 = 0.959$. This replicates previous findings and reveals that, overall, the crossmodal dynamic capture was strong. More importantly, the interaction between the group and congruency variables was not significant, $F(1,8) = 2.65$; $p = 0.14$. Even in the incongruent conditions—where the two groups appeared to differ most—there was again no significant difference, $t(8) = 1.77$, $p = 0.11$. No other effects in this analysis were significant ($p > 0.10$).

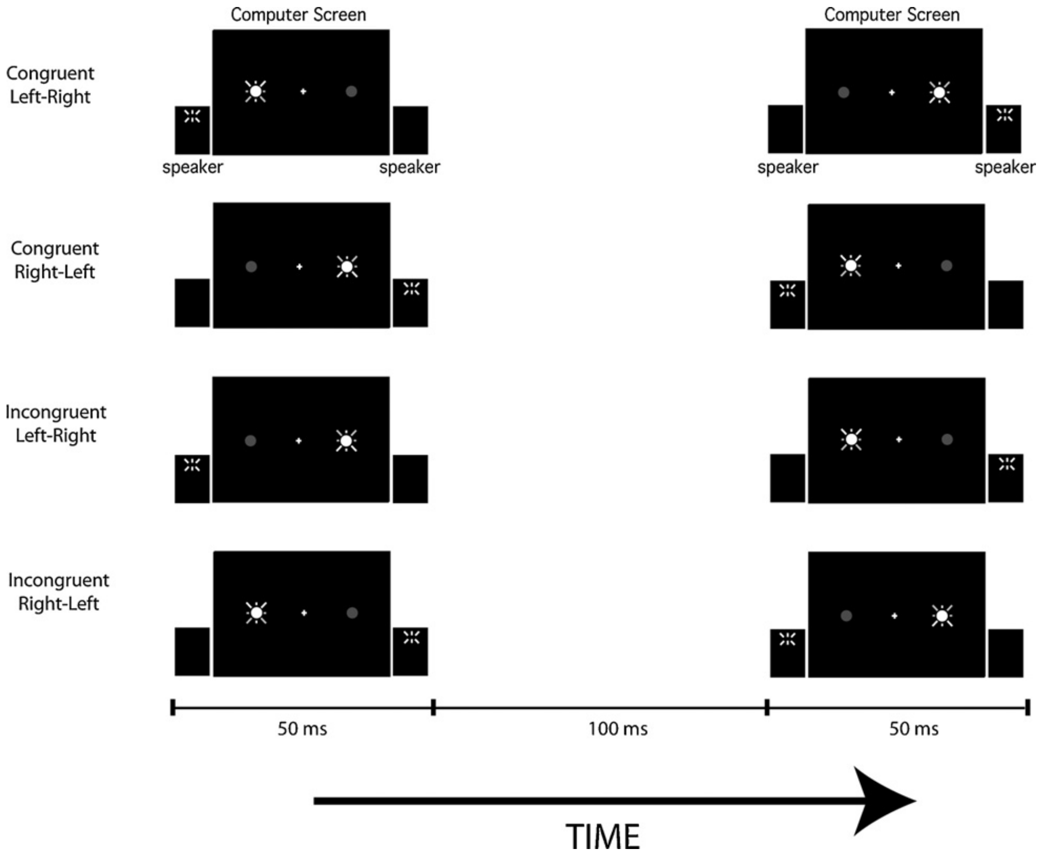

Fig. 2. Depiction of the set-up in Experiment 2. In each trial of the bisensory block of Experiment 2, participants heard left–right or right–left auditory beep sequence, and simultaneously witnessed left–right or right–left visual flash sequence. This produced two kinds of congruent trials, and two kinds of incongruent trials. The primary task was to judge whether the auditory beep sequence began on the left.

An ANOVA applied to the RT data showed results similar to those of the accuracy analysis (see Fig. 3b). There was a main effect of congruency in that responses were faster in congruent than incongruent trials, $F(1,8) = 17.3$; $p < 0.01$, $\eta_p^2 = 0.68$. This underscores the strength of the crossmodal dynamic capture in both groups. The main effect of congruency also shows that participants were not simply responding on the basis of the (dominant) visual modality, as might be suspected from the accuracy data alone (see also, Bonneh et al., 2008); instead, responses were affected by the relation between the auditory and visual information. A final noteworthy finding was that there was a significant three-way interaction, $F(1,8) = 10.6$, $p < 0.05$, $\eta_p^2 = 0.57$. More specifically, in the presence of left-right auditory motion, the autism group tended to be more affected by congruency; otherwise, the TD group tended to be more affected by congruency. The reason for this is unclear, but may have to do with attentional differences between the two groups (e.g., Oruc et al., 2008). For example, the autism group may be

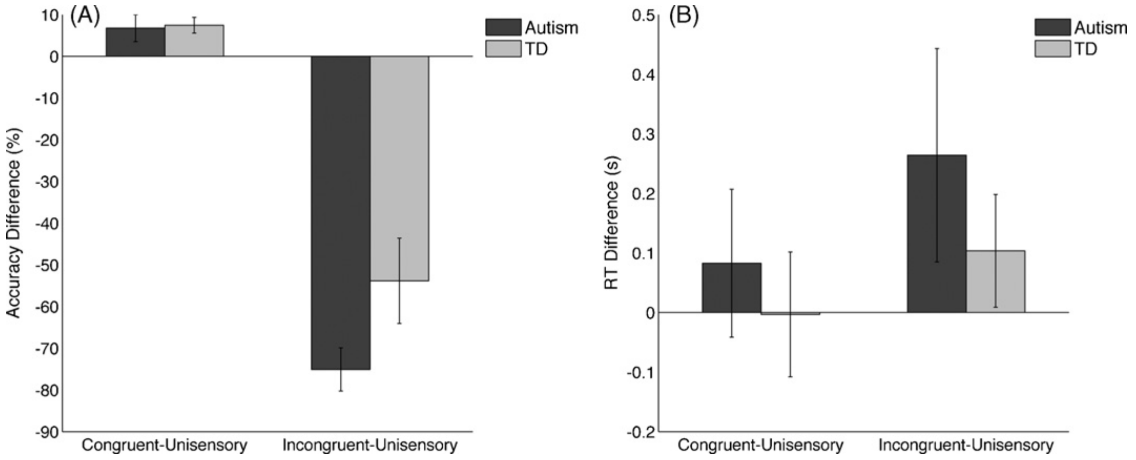

Fig. 3. Results from the crossmodal dynamic capture task in Experiment 2. (a) For the autism and TD groups, average accuracy is shown for the congruent and incongruent conditions, after subtracting the average accuracy in the unisensory (auditory-only) condition. (b) The same differences are shown for each group for the reaction time data. Error bars in both graphs depict ±1 SEM.

more focused on the target (left–right) auditory direction, and this heightened focus, in turn, may accentuate the dynamic capture that occurs with that direction.

4. Experiment 3: temporal numerosity integration (sound-induced flash illusion)

The first two experiments considered whether visual information differentially affected individuals with autism on auditory perceptual tasks. In Experiment 3, we examined the opposite causal relation—whether task-irrelevant auditory stimuli differentially affected visual perception. Observers were presented with one or two flashes paired with one or two beeps. The primary task was to indicate whether or not exactly one flash occurred. Because healthy observers are known to experience sound-induced flash illusions when the beep and flash numbers are incongruent, performance was expected to be better in the congruent than in the incongruent trials (e.g., Shams et al., 2000). As before, if RT or accuracy differences between congruent and incongruent trials are larger or smaller in the autism group, that would suggest heightened or reduced integration, respectively.

4.1. Method

4.1.1. Stimuli

A sketch of the set-up in Experiment 3 is shown in Fig. 4. In all trials, a red fixation cross appeared 5.2° above the center of a black screen. The visual flash stimulus consisted of a white disk (radius = 0.75°) presented 6.8° directly below the fixation point. The duration of each flash was one frame (~13 ms). The flash ISI on two flash trials was either 53, 67, or 79 ms, depending on the difficulty level that the participant could tolerate in the practice session (see Section 4.1.2). The auditory stimulus was a clearly audible 3.5 kHz tone that was presented for 7 ms from both speakers. The first beep occurred 23 ms before the onset of the first flash. Multiple beeps were separated with an ISI of 50 ms. Although the difference between beep ISI and flash ISI could slightly vary between participants, these differences were not expected to have a significant effect on the illusion strength on incongruent trials, since the sound-induced flash illusion occurs within a rather broad 100+ ms window (Shams et al., 2002).

4.1.2. Procedure and design

At the beginning of the experiment, participants were told to respond via a button-press to indicate whether they saw exactly one flash, or more than one flash. At the beginning of each trial, the fixation point appeared. One second later, the beep and/or flash sequence occurred, and participants entered a response. RT was measured from when the offset of the second flash would have occurred. This corresponded to 67, 79, or 92 ms after the offset of the first flash for the 53, 67, and 79 ms ISI, respectively (see below). After a response, participants began the next trial with a key press.

The blocks occurred in the following order: practice visual-only, audiovisual (128 non-catch trials), visual-only (32 non-catch trials), and auditory-only (32 non-catch trials). The number of practice trials was determined in the same way as in Experiment 1, except that participants began with trials of the fastest flash ISI (53 ms). If they failed to respond correctly in 9 out of 10 consecutive trials for a given ISI, then they were given another block that consisted of the next shortest ISI (67 ms). This process was repeated until an ISI yielded at least 9 out of 10 consecutive trials correct. All observers, except for one with autism (79 ms ISI), were able to perform the visual task with an ISI of 67 ms or less.

In the audiovisual block, there were two congruent conditions (1 beep, 2 flashes; 2 beeps, 1 flash) and two incongruent conditions presented quasi-randomly within the block. Trials in the practice and test visual-only blocks were identical to the trials of the audiovisual block, except that no sounds occurred. Trials in the auditory-only block were like the audiovisual block, except that no flashes occurred. The catch stimulus was a white dot (diameter = 0.07°) that appeared in the center of the fixation cross during the ISI.

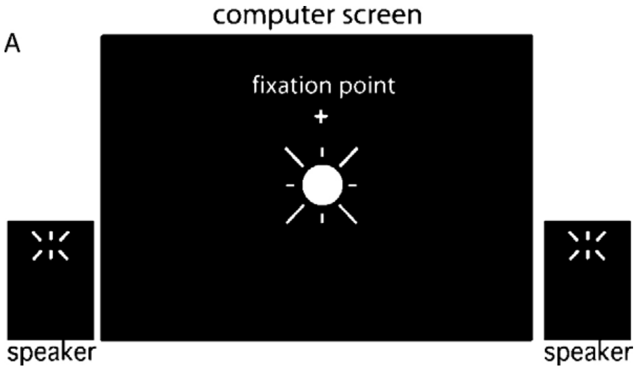

Fig. 4. Depiction of the set-up for Experiment 3. In each trial of the bisensory block, participants focused on a fixation point, saw one or two flashes at a region below fixation, and simultaneously heard one or two beeps emitted from both speakers. The primary task was to judge whether or not there was exactly one flash.

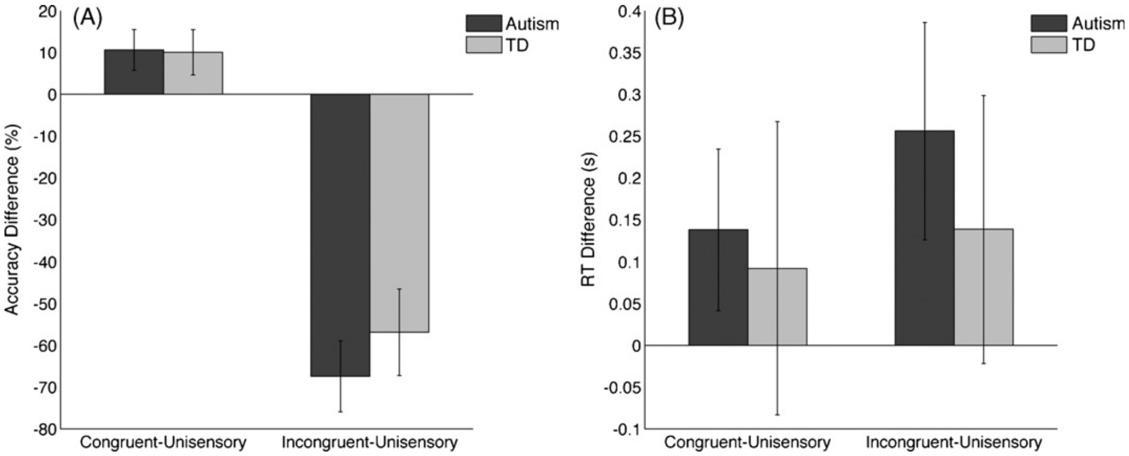

Fig. 5. Results from the audiovisual temporal numerosity integration task in Experiment 3. (a) For the autism and TD groups, average accuracy is shown for the congruent and incongruent conditions, after subtracting the average accuracy in the unisensory (visual-only) condition. (b) The same differences are shown for each group for the reaction time data. Error bars in both graphs depict ±1 SEM.

The instructions were the same in each block, except in the audiovisual block, where subjects were told that they would also be presented with irrelevant auditory information (which was to be ignored), and in the auditory-only block, where participants were instructed to indicate whether they heard exactly one beep.

4.2. Results and discussion

As can be seen in Table 2, the two groups performed similarly on measures including: rate of falsely identified catch trials, rate of missed catch trials, accuracy and RT during visual-only trials, and accuracy in the auditory-only trials. The two groups also required a similar number of practice trials to reach criterion performance, and used a similar flash-frame rate to perform at this criterion. Among the unisensory trials, the only difference that we found was that the autism group tended to respond more slowly in the auditory-only trials, $t(5) = 3.5$, $p = 0.02$. Although this did not reach the Bonferroni-corrected criterion ($p_{crit} = 0.007$), slowed auditory responses in our autism group likely reflect a real characteristic, given that the same trend occurred in Experiment 1 and given also that others have identified the same tendency in autism (Roberts et al., 2008, p. 154).

Accuracy results are shown in Fig. 5a. There was a main effect of congruency such that congruent trials overall produced more accurate responses than incongruent trials, $F(1,5) = 59.4$, $p < 0.01$, $\eta_p^2 = 0.92$. This replicates previous findings and indicates that, overall, the number of beeps strongly influenced the number of flashes reported. Most importantly, the TD and autism participants were comparably influenced by whether the auditory and visual stimuli were congruent, $F(1,5) = 1.6$, $p = 0.27$. No other effects were significant ($p > 0.28$).

Consistent with the accuracy results, congruent trials overall produced faster responses than incongruent trials, $F(1,5) = 9.7$, $p < 0.05$, $\eta_p^2 = 0.66$ (see Fig. 5b). This result indicates, again, that the number of beeps that participants heard strongly influenced the perception of number of flashes. Most importantly, the autism and TD group slowed comparably between the congruent and incongruent trials, $F(1,5) = .90$, $p > 0.30$. No other results of the RT analysis were significant ($p > 0.37$).

Similar to the previous experiment, across both groups there was no statistically significant RT difference between the congruent and unisensory trials ($p > 0.3$). As before, we suggest that the benefit of congruence may obtain only when the unisensory task becomes more challenging so that floor effects can be avoided.

The foregoing accuracy and RT results cohere with and extend the findings of the first two experiments. Members of the autism group were at least as susceptible to the influence of auditory information when performing a visual perceptual task as TD individuals. Even though both groups were well aware that they were to ignore the auditory stimulus, beep numerosity influenced the number of flashes reported. These results are consistent with a recent study that showed that people with autism experience the sound-induced flash illusion normally (van der Smagt et al., 2007).

5. General discussion

The goal of this study was to psychophysically test whether audiovisual integration is atypical in autism. Previous studies have shown that when TD participants simultaneously hear one of two spoken syllables (e.g., /pa or /ta) and visually observe a mouth speaking one of the same two syllables, auditory detection improves when the two modalities match, and deteriorates when they do not (e.g., Kaiser et al., 2005; McGurk and MacDonald, 1976). Previous studies have also shown that when participants are asked to detect the direction of two laterally separated beeps in the presence of concurrent visual apparent motion, performance is superior when the directions of the auditory and visual stimuli match relative to when they are opposite (Sanabria et al., 2004; Soto-Faraco, Spence, & Kingstone, 2004; Soto-Faraco, Spence, Lloyd, et al., 2004). Finally, the ability to identify the number of pulsations of a flash strongly depends on the number of concurrent beeps (Shams et al., 2000). We replicated all of these outcomes for an autism group and a TD group matched on age, sex, and performance IQ. More importantly, in each experiment, the two groups were comparably affected by audiovisual congruency, except in Experiment 1, where the RT of the clinical group was more influenced by congruency. Since the foregoing paradigms plausibly involve mechanisms that operate at different levels of processing, audiovisual integration appears to be unimpaired at a variety of levels.

Our study is important in at least three ways. First, in contrast to previous studies, both accuracy and reaction time were utilized to assess audiovisual integration in autism. Examining both measures allowed us to rule out speed-accuracy tradeoffs and provide stronger evidence for unhindered integration in autism. Another contribution of our study was that it showed that people with autism are at least as susceptible to the well-validated crossmodal dynamic capture effect. Since the paradigm is commonly considered to activate motion processing (Soto-Faraco, Spence, & Kingstone, 2004; Soto-Faraco, Spence, Lloyd, et al., 2004), our data provide the first support for normal crossmodal motion integration in autism. Third, although previous studies already examined the temporal numerosity and speech integration tasks, our results contribute to a literature where a consensus has not yet emerged. The presented data should help tilt the evidence in favor of the view that audiovisual speech and temporal numerosity integration is unimpaired in autism, at least among high-functioning adults.

In the following, we first discuss how results from Experiment 1 bear on previous studies. We then consider implications for theories of autism. We conclude by considering limitations of the current study and directions for future research.

5.1. Speech integration in autism: a comparison to previous studies

Although results from our study primarily reveal comparable audiovisual integration between the control and clinical groups, an interesting (and novel) result is that the autism group exhibited greater RT benefit during congruent audiovisual speech. This outcome should be regarded as preliminary for reasons specified below, but it is worth pointing out that it fits with other recent findings. Klin, Lin, Gorrindo, Ramsay, and Jones (2009) showed that while 2-year olds with autism preferred to look at audiovisually congruent biological motion displays over comparable displays lacking congruence, healthy or developmentally delayed controls did not exhibit a similar preference. Kemner and colleagues similarly pointed out that people with autism fixate on the mouth regions primarily when the faces are moving, but not static (Kemner & van Engeland, 2003; cf. Klin, Jones, Schultz, Volkmar, & Cohen, 2002a). Audiovisual congruence in the mouth area, therefore, may turn out to be especially helpful for increasing attention towards faces (which may be lacking due to face aversion, see below) or for improving language comprehension among those with autism (see also, Williams et al., 2004).

Audiovisual congruency did not affect the accuracy of the two groups differently. This is consistent with the findings of some previous studies (Williams et al., 2004), but not others (de Gelder et al., 1991; Mongillo et al., 2008; Smith & Bennetto, 2007). The apparent discrepancy may owe to attentional differences. Smith and Bennetto (2007) did not monitor eye gaze and did not have a secondary task that required attending to the visual stimulus, making it questionable as to whether the autism group fully attended to the visual display. Likewise, in the study of de Gelder et al., although the experimenters informally observed autistic participants to look toward the face display, there was no measure to ensure that they were. This is important because (a) faces can be aversive, or at least uninteresting, for people with autism to observe (Klin et al., 2002a; Klin, Jones, Schultz, Volkmar, & Cohen, 2002b; Osterling, Dawson, & Munson, 2002; Swettenham et al., 1998); and (b) when attention to the face area is reduced, audiovisual integration will become reduced (Alsius et al., 2005; Talsma, Doty, & Woldorf, 2007; Tiipana, Andersen, & Sams, 2004). In our experiment, by contrast, frequent catch trials required participants to keep constantly vigilant to the screen. Similarity in the performance of the two groups on catch trials provided evidence that both groups attended to the displays comparably. Williams et al. may have also improved their autism group's attention toward the speaking faces by providing the children with frequent encouragement (regardless of accuracy) to attend to the screen. Others have also attributed the diversity of findings in speech integration tasks to differences in visual attention to face stimuli (Mongillo et al., 2008).

5.2. Implications for current theories of autism

To our knowledge, no existing theory makes any explicit predictions regarding multisensory integration in autism, but a few major theories indirectly suggest reduced integration. According to the most recent version (Happé & Frith, 2006) of Happé and Frith's Weak Central Coherence theory (WCC; Frith, 1989; Frith & Happé, 1994), people with autism are more inclined (and often more able) to focus on local details of an auditory or visual stimulus at the expense of global grouping (e.g., Brosnan et al., 2004). A natural extension of this idea would be to predict a heightened capacity to employ information deriving from only a single modality, and to ignore co-occurring stimuli in irrelevant modalities. Plaisted (2001) similarly argued that weak coherence in autism might owe to perceptual tendencies to discriminate better, but to generalize less. If using information from one modality to make a judgment in another modality can be regarded as a form of generalization, then Plaisted's view also appears to predict reduced integration in autism. According to yet another view, autism is associated with an early abnormal brain growth period, which itself favors short-range connections at the expense of long-range ones (e.g., Hughes, 2008). Reduced connectivity may lead to reduced synchronization between different brain areas, which in turn impairs the integration of parts into coherent wholes (Brock et al., 2002). Given that audiovisual interactions is expected to depend on long-range connectivity, any of these under-connectivity views would predict reduced crossmodal interactions in autism. The foregoing theories, although plausible and well-evidenced in other respects, do not readily accommodate our finding of unimpaired audiovisual integration in autism.

5.3. Limitations and future directions

Several limitations of this study are worth noting. First, because the unisensory baseline task was designed to be readily learnable by persons with autism, we were limited in our ability to assess differential accuracy benefits of audiovisual congruence between the two groups. In a similar vein, because the control group tended to respond faster in the unisensory conditions, a smaller degree of reduction in reaction time in the congruent conditions may owe partly to a floor effect, rather than a lower tendency for multisensory integration. Another issue is that patterns of bisensory facilitation may vary depending on the kinds of syllables combined (Mongillo et al., 2008). Future studies will have to consider whether audiovisual integration differentially affects people with autism when the detection task becomes more challenging, or when the syllable combinations become more varied.

Our study, being purely behavioral, is also qualified by the fact that it bears only indirectly on the underlying neurobiological causes. A similar degree of crossmodal interactions in behavior does not necessarily imply that the clinical group utilized the same neural mechanisms for integration (see also, Møller, 2007). Just as people with autism appear to utilize different brain networks for speech and face processing (Kleinhans et al., 2008; Schultz et al., 2000), so too may they employ different multisensory mechanisms to generate approximately normal behavior in the lab.

A final, and perhaps most obvious limitation of our study, is that our results may not apply to other subpopulations of ASD. Any anomaly in integration may become more pronounced with degree of autism and lower mental age (Baranek, Boyd, Poe, David, & Watson, 2007; Leekam, Nieto, Libby, Wing, & Gould, 2007). Also, compensatory mechanisms may develop over time, making perceptual anomalies harder to detect in adults with autism, especially those who are high functioning. Additional research will have to consider how audiovisual integration varies across groups within autism, and how individuals within these groups compare to those who are typically developed and those who are both developmentally impaired and non-autistic.

Acknowledgments

We give special thanks to participating persons with autism and their families for their valuable time. We also thank the Center for Autism Research and Treatment at UCLA, for providing a pilot study award and recruitment/assessment services. We additionally appreciate the valuable input of affiliates of the Visual & Multisensory Perception Laboratory (namely, Aaron Seitz, David Wozny, Robyn Kim, Elizabeth Burke). Finally, we gratefully acknowledge support from: Autism Speaks (formerly Cure Autism Now) and UCLA Graduate Research Mentorship and Dissertation Year Fellowship awards to the first author.

Contributors: OR and LS contributed to the experiment concept and design, OR, NC, and BK to participant recruitment, BK, NC, LS, and OR to data analysis, and BK, LS, OR, and NC to manuscript preparation.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*, 839–843.

American Psychiatric Association. (2000). *Diagnostic and statistic manual of mental disorders (text revision)* (4th ed.). Washington, DC: Author.

Baranek, G. T., Boyd, B. A., Poe, M. D., David, F. J., & Watson, L. R. (2007). Hyperresponsiveness sensory patterns in young children with autism, developmental delay, and typical development. *American Journal on Mental Retardation, 112*, 233–245.

Baumann, O., & Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception. *Cerebral Cortex, 17*, 1433–1443.

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2003). Motion perception in autism: A "complex" issue. *Journal of Cognitive Neuroscience, 15*, 1–8.

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*, 151–157.

Bonneh, Y. S., Belmonte, M. K., Pei, F., Iversen, P. E., Kenet, T., Akshoomoff, N., et al. (2008). Cross-modal extinction in a boy with severely autistic behaviour and high verbal intelligence. *Cognitive Neuropsychology, 25*, 635–652.

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

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Development & Psychopathology, 14*, 209–224.

Brosnan, M. J., Scott, F. J., Fox, S., & Pye, J. (2004). Gestalt processing in autism: Failure to process perceptual relationships and the implications for contextual understanding. *Journal of Child Psychology, 45*, 459–469.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B, 363*, 1001–1010.

Cesaroni, L., & Garber, M. (1991). Exploring the experiences of autism through firsthand accounts. *Journal of Autisms and Developmental Disorders, 21*, 303–313.

Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: Local over-connectivity but long-distance disconnection. *Current Opinion in Neurobiology, 15*, 225–230.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.

Frith, U., & Happé, F. G. E. (1994). Autism: Beyond 'theory of mind'. *Cognition, 50*, 115–132.

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.

Grant, K. W., Greenberg, S., Poeppel, D., & Wassenhove, V. (2004). Effects of spectro-temporal ascynchrony in auditory and auditory-visual speech processing. *Seminars in hearing. All in good time: A tribute to Ira Hirsh, vol. 25* (pp. 241–255).

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in spectrum disorders. *Journal of Autism and Developmental Disorders, 36*, 5–25.

Heaton, P. (2003). Pitch memory, labeling and disembedding in autism. *Journal of Child Psychology and Psychiatry, 44*, 543–551.

Heaton, P, Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant music ability? *Music Perception, 154*, 291–305.

Hughes, J. R. (2008). A review of recent reports on autism: 1000 studies published in 2007. *Epilepsy & Behavior, 13*, 425–437.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex, 15*, 646–653.

Kemner, C., & van Engeland, H. (2003). Autism and visual fixation. *American Journal of Psychiatry, 160*, 1358–1359.

Kern, J. K., Trivedi, M. H., Grannemann, B. D., Garver, C. R., Johnson, D. G., Andrews, A. A., et al. (2007). Sensory correlations in autism. *Autism, 11*, 123–134.

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., et al. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain, 131*, 1000–1012.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. J. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature, 459*, 257–261.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002a). Defining and quantifying the social phenotype in autism. *The American Journal of Psychiatry, 159*, 895–908.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002b). Visual fixation patterns during viewing a naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*, 809–816.

Leekam, S. R., Nieto, C., Libby, S. J., Wing, L., & Gould, J. (2007). Describing the sensory abnormalities of children and adults with autism. *Journal of Autism and Developmental Disorders, 37*, 894–910.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex, 10*, 873–888.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism diagnostic observation schedule* (WPS ed.). Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

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

Magnée, M. J. C. M. , de Gelder, B., van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry, 49*, 995–1000.

Massaro, D. W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *The Journal of Developmental and Learning Disorders, 7*, 111–146.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology & Psychiatry & Allied Disciplines., 43*, 255–263.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound induced flash illusion. *Journal of Neuroscience, 27*, 4120–4131.

Møller, A. R. (2007). Neurophysiologic abnormalities in autism. In B. S. Mesmere (Ed.), *New autism research developments* (pp. 137–158). New York: Nova Science Publishers.

Møller, A. R., Kern, J. K., & Grannemann, B. (2005). Are the non-classical auditory pathways involved in autism and PDD? *Neurological Research, 27*, 625–629.

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

Morsanyi, K., & Holyoak, K. J. (in press). Analogical reasoning ability in autistic and typically-developing children. *Developmental Science*.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility. *Psychological Science, 15*, 133–137.

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 719–730.

Oruc, I., Sinnett, S., Bischof, W. F., Soto-Faraco, S., Lock, K., & Kingstone, A. (2008). The effect of attention on the illusory capture of motion in bimodal stimuli. *Brain Research, 1242*, 200–208.

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology, 14*, 239–251.

Pelli, D. G. (1997). The video toolbox soft ware for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Plaisted, K. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Mahwah, NJ: Lawrence Erlbaum Associates.

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry, 39*, 765–775.

Roberts, T. P. L, Schmidt, G. L., Egeth, M., Blaskey, L., Rey, M. M., Edgar, C. J., et al. (2008). Electrophysiological signatures: Magnetoencephalographic studies of the neural correlates of language impairment in autism spectrum disorders. *International Journal of Psychophysiology, 68*, 149–160.

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

Sanabria, D., Soto-Faraco, S., & Spence, C. (2004). Exploring the role of visual perceptual grouping on the audiovisual integration of motion. *Neuroreport, 15*, 2745–2749.

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., et al. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and asperger syndrome. *Archives of General Psychiatry, 57*, 331–340.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study. *Neuroscience Letters, 378*, 76–81.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*, 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*, 147–152.

van der Smagt, M. J., van Engeland, H., & Kemner, C. (2007). Brief report: Can you see what is not there? Low-level auditory-visual integration in autism. *Journal of Autism and Developmental Disorders, 37*, 2014–2019.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lip reading in autism. *Journal of Child Psychology and Psychiatry, 48*, 813–821.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research, 14*, 139–146.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects of motion perception across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 330–345.

Soto-Faraco, S., Spence, C., Lloyd, D., & Kingstone, A. (2004). Moving multisensory research along: Motion perception across sensory modalities. *Current Directions in Psychological Science, 13*, 29–32.

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

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., et al. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry, 39*, 747–753.

Talsma, D, Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*, 679–690.

Tiipana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology, 16*, 457–472.

Tsermentseli, S., O'Brien, J. M., & Spencer, J. V. (2008). Comparison of form and motion coherence processing in autistic spectrum disorders and dyslexia. *Journal of Autism and Developmental Disorders, 38*, 1201–1210.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *Neuroimage, 37*, 572–578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage, 31*, 1247–1256.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *WASI Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation.

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.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision, 8*(24), 1–11 doi:10.1167/8.3.24 http://journalofvision.org/8/3/24/.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex, 13*, 1034–1043.