

Multisensory Speech Perception in Autism Spectrum Disorder: From Phoneme to Whole-Word Perception

Ryan A. Stevenson , Sarah H. Baum, Magali Segers, Susanne Ferber, Morgan D. Barense, and Mark T. Wallace

Speech perception in noisy environments is boosted when a listener can see the speaker's mouth and integrate the auditory and visual speech information. Autistic children have a diminished capacity to integrate sensory information across modalities, which contributes to core symptoms of autism, such as impairments in social communication. We investigated the abilities of autistic and typically-developing (TD) children to integrate auditory and visual speech stimuli in various signal-to-noise ratios (SNR). Measurements of both whole-word and phoneme recognition were recorded. At the level of whole-word recognition, autistic children exhibited reduced performance in both the auditory and audiovisual modalities. Importantly, autistic children showed reduced behavioral benefit from multisensory integration with whole-word recognition, specifically at low SNRs. At the level of phoneme recognition, autistic children exhibited reduced performance relative to their TD peers in auditory, visual, and audiovisual modalities. However, and in contrast to their performance at the level of whole-word recognition, both autistic and TD children showed benefits from multisensory integration for phoneme recognition. In accordance with the principle of inverse effectiveness, both groups exhibited greater benefit at low SNRs relative to high SNRs. Thus, while autistic children showed typical multisensory benefits during phoneme recognition, these benefits did not translate to typical multisensory benefit of whole-word recognition in noisy environments. We hypothesize that sensory impairments in autistic children raise the SNR threshold needed to extract meaningful information from a given sensory input, resulting in subsequent failure to exhibit behavioral benefits from additional sensory information at the level of whole-word recognition. *Autism Res* 2017, 0: 000–000. © 2017 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: autism spectrum disorders; speech perception; multisensory integration; inverse effectiveness; sensory integration; sensory processing

Introduction

The ability to accurately perceive spoken language is a fundamental ability in social communication. Speech perception, while often conceptualized as an auditory process, is in fact inherently multisensory, with a listener using both auditory speech information as well as visual speech information in the form of oral articulations. Emerging evidence strongly supports the presence of specific deficits in autism spectrum disorder (ASD) that impact how information is integrated across the different sensory modalities [Baum, Stevenson, & Wallace, 2015; Brandwein et al., 2013; Foss-Feig et al., 2010; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011; Ross et al., 2011; Russo et al., 2010; Stevenson, Segers, Ferber, Barense, & Wallace, 2014a; Stevenson et al., 2014b, 2014c, 2014d; Wallace & Stevenson,

2014]. In fact, while ASD is a heterogeneous disorder with complex etiologies, atypical sensory processing is one of the most common symptoms, reported in up to 87% of autistic individuals¹ [Le Couteur et al., 1989; Lord, 1995]. These multisensory differences appear to be related to a reduced ability to integrate individual pieces of sensory information into a coherent unified percept.

¹While researchers and clinicians often feel more comfortable using person-first language such as “individuals with autism,” autistic individuals have endorsed identity-first language that incorporates autism as a component of their identity over person-first language (61% vs. 28%) [Kenny et al., 2015]. While autistic individuals themselves bear the full ability to choose which terminology is used, these preferences also coincide with parents of autistic children (51% vs. 22%) [Kenny et al., 2015] and self-advocates [Sinclair, 1999] I-FA 2016; [Brown & ASAN, 2016]. Given this, we will respect this preference and use this language throughout this manuscript.

From the Department of Psychology, Western University, London, ON, Canada (R.A.S.); Brain and Mind Institute, Western University, London, ON, Canada (R.A.S.); Department of Psychology, University of Washington, Seattle, WA (S.H.B.); York University, Toronto, ON, Canada (M.S.); Department of Psychology, University of Toronto, Toronto, ON, Canada (S.F., M.D.B.); Rotman Research Institute, Toronto, ON, Canada (S.F., M.D.B.); Vanderbilt Brain Institute, Nashville, TN (M.T.W.); Vanderbilt Kennedy Center, Nashville, TN, USA (M.T.W.); Vanderbilt University, Nashville, TN (M.T.W.); Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN (M.T.W.); Department of Psychology, Vanderbilt University, Nashville, TN (M.T.W.)

Received September 14, 2016; accepted for publication February 06, 2017

Address for correspondence and reprints: Ryan A. Stevenson, Westminster Hall, Rm 314e, Western University, London, ON. E-mail: rsteve28@uwo.ca
Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.1776

© 2017 International Society for Autism Research, Wiley Periodicals, Inc.

Impairments in multisensory integration are likely to have direct implications for ASD symptomatology, particularly in the domains of social communication and speech perception. When one both hears and sees a speaker's utterance, speech perception is more accurate [Altieri, Stevenson, Wallace, & Wenger, 2015; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007a; Ross et al., 2011; Stevenson & James, 2009; Stevenson et al., 2014e, 2015; Sumbly & Pollack, 1954] and less effortful [Fraser, Gagne, Alepins, & Dubois, 2010] than when only auditory information is available. These benefits of multisensory integration are greatest in noisy environments where the audible signal is relatively less "effective"; notably these are environments in which autistic children are known to struggle [Foxe et al., 2015]. The concept of multisensory integration being most beneficial when a sensory signal is weak or ambiguous is not unique to speech, and has been collectively subsumed under the concept of "inverse effectiveness." In short, the *less* effective the individual unisensory components are at driving a response, the *greater* the benefit will typically be from processing both sensory inputs in concert [Bishop & Miller, 2009; James, Stevenson, & Kim, 2012; Meredith & Stein, 1983, 1986; Nath & Beauchamp, 2011; Stevenson, Geoghegan, & James, 2007; Stevenson & James, 2009; Stevenson, Kim, & James, 2009; Stevenson et al., 2012a; Wallace, Wilkinson, & Stein, 1996; Werner & Noppeney, 2009].

To our knowledge, three studies to date have measured audiovisual speech-in-noise abilities of autistic individuals [Foxe et al., 2015; Irwin, Tornatore, Brancizio, & Whalen, 2011; Smith & Bennetto, 2007]. One study showed that although autistic adolescents (12–17-year old) demonstrated similar auditory-only abilities relative to their typically-developing (TD) peers but did not benefit from seeing a speaker's mouth as much as their TD peers. This was driven both by a significant reduction in visual lip-reading abilities as well as by a specific multisensory deficit [Smith & Bennetto, 2007]. Another study observed similar results in a closed-set, syllable-discrimination task [Irwin et al., 2011]. Specifically, autistic children showed no difference from TD children (5–15-year old) across a range of signal-to-noise ratios (SNR) in auditory-only performance, but showed a decrease in the visual-only (lip reading) conditions, and a reduced multisensory gain associated with concurrent visual articulations. Finally, in a different study using a word-recognition task with varied SNRs, autistic individuals showed decreased auditory-only and visual-only performance across a wide range of ages (5–17-year old), but a decrease in audiovisual performance only in the younger groups (under the age of 12) [Foxe et al., 2015]. Even for these younger groups, significant differences in the multisensory gain were only observed in the lowest SNRs, and these

differences between autistic and TD individuals were altogether absent in the 13–15-year-old group.

There are a number of consistent findings across these studies. Each study showed at least some deficit unique to multisensory integration of speech. This deficit, even if observed primarily at younger ages, may significantly impact the overall development of individuals on the spectrum [Stevenson et al., 2014a]. Visual lip-reading was consistently found to be reduced in ASD, suggesting that any measure of multisensory benefit in ASD must include a measure of visual speech abilities [Stevenson et al., 2014e]. Speech perception deficits also appear not to be the result of differential gaze patterns, as two of the above-mentioned studies used eye-tracking procedures to account for any such differences [Foxe et al., 2015; Irwin et al., 2011]. There are, however, also considerable differences in the results of these studies (specifically whether auditory, visual, or audiovisual difficulties drive group differences in performance) which could be explained by a number of factors, including differences in the perceptual reports (closed-set syllable discrimination to word recognition within sentences), differences in the levels of SNR used, differences in the age range of the participants, and the differences in the metric used to calculate multisensory gain. Overall these studies point to intriguing changes in audiovisual function subserving speech comprehension in noisy settings in autistic children, the lack of concordance among some of the results illustrates the need for additional research to further clarify the picture.

A number of outstanding questions remain in regards to the audiovisual integrative abilities of autistic children in noisy environments. One of the most important of these revolves around the fact that these prior studies focused on speech comprehension at a single level of analysis (e.g., whole-word [Foxe et al., 2015; Smith & Bennetto, 2007] or syllable level [Irwin et al., 2011]) which provides only a limited view into the nature of the communication challenges in these children. To address this issue, in the current work we examine performance at the level of whole words, but also at more elementary (i.e., phonemic) components of speech, providing a potential window into how changes at one level scaffold changes at other levels. Indeed, our results provide evidence that while autistic children show deficits at higher levels of processing as indexed via whole-word recognition, autistic children show largely intact multisensory processing at lower (i.e., phonemic) levels of speech perception.

Materials and Methods

Experimental protocols were derived from a previously reported study on multisensory integration of speech in a healthy aging population [Stevenson et al., 2015].

Participants

Sixty-two participants (ASD = 25, TD = 37; details below) completed a behavioral speech-in-noise paradigm. Experimental protocols were approved by Vanderbilt University Medical Center's institutional review board and the University of Toronto's Ethics Board. Participants were divided into two diagnostic groups, one with (ASD group) and one without (TD group) a confirmed ASD diagnosis or other neuropsychological disorder. The ASD group included 25 individuals (4 female, mean age = 10.8 years, $sd = 3.4$, range = 6–18) and the TD group included 37 individuals (13 female, mean age = 11.1 years, $sd = 3.4$, range = 6–18). Ages were matched between groups ($P = 0.77$, $t_{(59)} = 0.30$, $d = 0.09$).

Clinical diagnosis of ASD was confirmed by receipt of a child's original diagnostic report, administration by a research-reliable experimenter of the Autism Diagnostic Observation Schedule-1 or -2 [Lord et al., 2012] with an average comparison score of 7.5 ($SD = 1.5$), and completion of the Autism Quotient: Child (AQ-C) [Auyeung, Baron-Cohen, Wheelwright, & Allison, 2008]. Verbal and performance IQ were assessed in the ASD group with the Wechsler Abbreviated Scale of Intelligence-II (WASI-II) [Wechsler, 2011], with an average full scale IQ of 96.4 ($SD = 20$), average standardized verbal score of 99.6 ($SD = 24$), and standardized non-verbal performance score of 100.6 ($SD = 19$). Parents of TD children also completed the AQ-C to ensure they were not at-risk for ASD (no scores above 34/50), as higher scores have been shown to represent a higher expression of overall behaviors characteristic of autistic children [Auyeung et al., 2008]. Additionally, TD children's parents self-reported no psychological or neurological diagnoses.

Stimuli

Stimuli included dynamic, audiovisual (AV) recordings of a single female speaker saying 216 tri-phonemic nouns. Stimuli were selected from a previously published stimulus set which has been normalized and available for public use, The Hoosier Audiovisual Multitalker Database [Sheffert, Lachs, & Hernandez, 1996]. The stimuli selected were monosyllabic English words that were matched across sets for accuracy on both visual-only and audio-only recognition [Lachs & Hernandez, 1998], and were also matched according to lexical neighborhood density [Luce & Pisoni, 1998; Sheffert et al., 1996]. Median age of acquisition for the word set was 5.5-years old, and median frequency was 35.6 per million words [Garlock, Walley, & Metsala, 2001; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012]. This stimulus set of single-word tokens has been used successfully in previous studies of multisensory integration [Stevenson et al., 2007, 2009, 2012b, 2015;

Stevenson & James, 2009; Stevenson, Altieri, Kim, Pisoni, & James, 2010; Stevenson, VanDerKlok, Pisoni, & James, 2011]. Audio signal levels were measured as root mean square (RMS) contrast and equated across all tokens.

All stimuli throughout the study were presented using MATLAB 2012b (MATHWORKS Inc., Natick, MA) software with the Psychophysics Toolbox extensions [Brainard, 1997; Pelli, 1997]. Visual stimuli were 200×200 pixels and subtended $10 \times 10^\circ$ of visual angle. Audio stimuli were presented through two aligned speakers on each side of the monitor. All tokens lasted 2 sec and included all pre-articulatory gestures.

In the visual-only condition, the visual component of each stimulus, or viseme, was presented. Auditory stimuli were all overlaid with 8-channel multitalker babble at 72 dB SPL. The presentation of auditory babble presentation began 500 ms prior to the beginning of the stimulus token and ended 500 ms following token offset. The RMS of the auditory babble was linearly ramped up and down, respectively, during the pre- and post-stimulus 500 ms periods, and was presented with the first and last frames of the visual token, respectively. Auditory stimuli were presented at four separate sound levels relative to the auditory noise. Differences in auditory level of the speech token relative to background babble (SNR) included 0, -6, -12, and -18 dB SPL.

Procedures

Participants sat in a sound- and light-attenuating WhisperRoom™ (Model SE 2000; Whisper Room Inc.) approximately 60 cm from the monitor. Participants were presented with nine separate runs of 24 single-word presentations each: four audiovisual runs (one at each SNR), four auditory-only runs (one at each SNR), and one visual-only run (with auditory multitalker babble). During auditory-only presentations, the first frame of the associated video was presented and remained static throughout the presentations. Run orders were randomized across participants. Within participants, word lists were randomized across runs with no words repeated. Word lists were also counterbalanced between individuals, so words were presented in different modalities and SNRs for each individual.

Experimental procedures were identical for all runs. Participants were instructed to attend to the speaker at all times, and to report the word they perceived by typing the word (on a keyboard placed in front of them). After each trial the experimenter verbally confirmed the participant's report, and then, the next word was presented. Homophones (e.g., seen vs. scene) were not counted as errors, nor were spelling errors when verified verbally. When verbal confirmation was needed, experimenters asked "what did she say?." Similarly, if a

Table 1. Detailed Statistics for Phoneme and Whole-Word Recognition

Modality	SNR	Phoneme recognition					Whole-word recognition				
		TD	ASD	<i>P</i>	<i>t</i> ₍₅₇₎	<i>d</i>	TD	ASD	<i>P</i>	<i>t</i> ₍₅₇₎	<i>d</i>
V	N/A	30.4 ± 13.0	22.9 ± 12.4	0.033	2.18	0.59	6.6 ± 7.2	4.2 ± 6.3	0.190	1.32	0.35
A	0	90.7 ± 8.1	83.0 ± 15.9	0.017	2.46	0.61	80.3 ± 10.7	69.4 ± 19.2	0.007	2.81	0.70
	-6	83.1 ± 8.0	80.4 ± 7.2	0.189	1.33	0.35	66.4 ± 12.2	61.6 ± 12.5	0.15	1.47	0.39
	-12	67.1 ± 10.7	60.3 ± 10.1	0.019	2.42	0.65	45.1 ± 13.0	37.5 ± 11.1	0.024	2.31	0.63
	-18	19.9 ± 8.3	16.8 ± 7.4	0.154	1.44	0.39	5.2 ± 4.3	5.1 ± 4.1	0.934	.084	0.02
AV	0	96.2 ± 5.0	91.5 ± 8.1	0.009	2.72	0.70	91.1 ± 7.3	82.6 ± 12.1	0.001	3.36	0.85
	-6	92.4 ± 8.9	86.2 ± 10.7	0.019	2.42	0.63	82.9 ± 11.9	71.4 ± 14.6	0.002	3.31	0.86
	-12	84.5 ± 11.3	77.4 ± 15.0	0.041	2.09	0.53	69.7 ± 13.5	60.3 ± 18.9	0.031	2.22	0.57
	-18	56.1 ± 13.3	43.7 ± 13.9	0.001	3.45	0.91	30.8 ± 13.8	18.3 ± 9.9	<0.001	3.76	1.04
Multisensory Gain ^a	0	3.2 ± 5.4	5.3 ± 10.5	0.307	1.03	0.25	9.6 ± 10.1	12.3 ± 11.4	0.349	0.94	0.25
	-6	4.8 ± 5.9	1.3 ± 8.9	0.078	1.79	0.46	14.4 ± 12.0	8.0 ± 15.2	0.078	1.79	0.47
	-12	9.6 ± 11.6	8.3 ± 10.2	0.660	0.44	0.12	22.4 ± 14.8	20.3 ± 15.6	0.619	0.50	0.14
	-18	12.3 ± 12.5	8.3 ± 9.2	0.192	1.32	0.36	19.4 ± 13.4	8.8 ± 8.1	0.002	3.33	0.96

P, *t*, and *d* values describe a comparison of accuracies in each experimental condition between TD and ASD groups.

^aValues given for multisensory gain are percent changes between observed and predicted recognition accuracy rates.

participant typed their response phonetically, the response was verbally confirmed and marked as correct when appropriate (e.g., kat vs. cat). No time limit was given for participant responses. Each run lasted approximately 5 min, and all run orders were counterbalanced.

Analysis

Responses were scored in two ways, at the whole-word level and the phoneme level. Word-recognition accuracy was scored as correct if and only if the entire word reported was correctly perceived. Phoneme accuracy allowed for more precise scoring, as participants could be scored as correctly reporting 0–2 phonemes even if the whole word was not correctly identified, or 3 phonemes per word if the whole word was recognized. Scoring at the phoneme level allowed us to test whether multisensory benefits occurred at the lower level of phoneme perception even if the whole word was not accurately recognized. For example, if the word “cat” was presented, an individual may improve from an auditory or visual perception of “con” (1 phoneme correct) to a multisensory perception of “can” (2 phonemes correct)—an improvement that would be overlooked when measuring only at the whole-word level. Mean word accuracy and phoneme accuracy were then calculated for each participant and for each run.

The expected multisensory accuracy predicted by the individual unisensory responses was calculated by:

$$pAV = p(A) + p(V) - [p(A) * p(V)],$$

where *pAV* represents a null hypothesis characterizing what the response will be to audiovisual presentations if the auditory and visual information are processed independently [Stevenson et al., 2014e], and where *p(A)* and *p(V)* represent the individual’s response accuracy to

auditory- and visual-only presentations, respectively. Each participant’s responses to unisensory presentations were used to calculate their individual predicted *pAV* values for each SNR level. These predicted values were then used as a null hypothesis from which we measured multisensory interactions, namely *multisensory gain* which we define here as an increase in performance above and beyond that predicted by non-integrative statistical facilitation [Raab, 1962].

To assess unisensory auditory sensory processing levels, a two (diagnostic group: TD vs. ASD) × four (SNR: 0, -6, -12, -18 dB SPL), mixed-model ANOVA was conducted for both at the word- and phoneme-recognition levels for auditory-only word presentations. To assess unisensory visual processing levels (i.e., lip reading), a between-subject *t*-test was performed across diagnostic groups. To assess multisensory benefits across groups, a 2 (diagnostic group: TD vs. ASD) × 4 (SNR: 0, -6, -12, -18 dB SPL) × 2 (AV-measure: observed AV vs. predicted AV) mixed-model ANOVA was conducted for both word and phoneme accuracy scores. Finally, to directly compare diagnosis-related differences in multisensory gain for word versus phoneme recognition, the percent multisensory gain data were subjected to a 2 (diagnostic group: TD vs. ASD) × 4 (SNR: 0, -6, -12, -18 dB SPL) × 2 (measure: whole-word vs. phoneme) mixed-model ANOVA.

Results

All analyses were run separately for each of the two levels of data scoring: whole-word accuracy and phoneme accuracy. Mean accuracies and standard deviations were calculated for each group, sensory modality, and SNR level (Table 1).

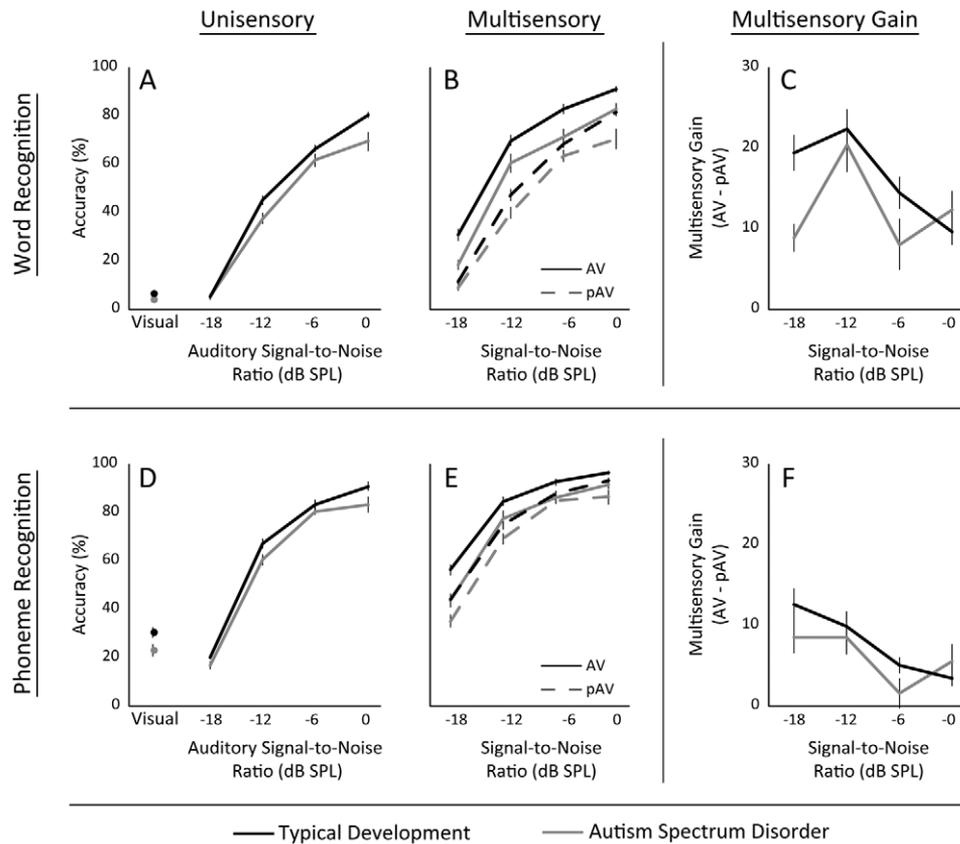


Figure 1. Speech perception. Panels A–C depict whole-word recognition scores, and D–F depict phoneme recognition scores. Unisensory panels (A and D) show accuracies with visual-only presentations as well as for auditory-only presentations at four different SNRs. Multisensory panels (B and E) show accuracies in response to audiovisual presentations at varying SNRs (AV, solid lines) as well as the predicted audiovisual accuracies (pAV, dashed lines) based on unisensory scores using the equation: $A + V - (A \times V)$. Multisensory Gain panels (C and F) depict the differences between audiovisual responses and predicted audiovisual responses as a function of SNR. Error bars depict standard error.

Unisensory Performance

Visual-only whole-word and viseme recognition.

A between-subject *t*-test was used to compare TD and ASD groups' visual-only speech perception accuracies at both the *whole-word and viseme* recognition levels. In both cases, TD individuals showed higher rates of visual-only accuracy. These differences did not reach significance at the whole-word recognition level (Fig. 1A), but were statistically significant at the viseme-recognition level (Fig. 1D). For statistical results, see Table 1.

Auditory-only, whole-word recognition.

For whole-word recognition (Fig. 1A), a main effect of diagnostic group was seen, with the TD group showing higher accuracy than the ASD group ($P < 0.012$, $F_{(1,55)} = 6.71$, $partial-\eta^2 = 0.11$). A main effect of SNR was also observed, with higher SNRs resulting in higher accuracy rates ($P < 0.001$, $F_{(3,55)} = 576.60$, $partial-\eta^2 = 0.91$). Finally, a significant interaction between

diagnostic group and SNR levels was observed, with autistic children showing decreases in accuracy at higher SNR levels as compared to TD children ($P = 0.019$, $F_{(3, 55)} = 3.42$, $partial-\eta^2 = 0.06$). For detailed statistical results, see Table 1. Thus, autistic children showed diminished auditory whole-word recognition abilities across SNR level, and additionally, these impairments increased as SNR decreased.

Auditory-only phoneme recognition.

In the 2-way ANOVA with auditory-only presentations for phoneme recognition (Fig. 1D), a main effect of diagnostic group was also seen, with TD children showing higher accuracy than ASD children ($P < 0.011$, $F_{(1,55)} = 6.90$, $partial-\eta^2 = 0.11$). A main effect of SNR was also seen for phoneme recognition, with higher SNRs leading to higher accuracy rates ($P < 0.001$, $F_{(3,55)} = 980.34$, $partial-\eta^2 = 0.95$). Finally, no interaction between diagnostic group and SNR levels was observed at the phoneme level ($P = 0.20$, $F_{(3, 55)} = 1.59$, $partial-\eta^2 = 0.03$). For detailed

statistical results, see Table 1. Thus, although ASD children performed worse overall, both groups showed similar levels of improvement as SNR level increased.

Multisensory Performance

Audiovisual whole-word recognition. A two-way, mixed-model, repeated-measures ANOVA was conducted with diagnostic group as a between-subjects factor and SNR as a within-subjects factor was conducted on whole-word accuracy observed with audiovisual speech presentations. A significant main effect of diagnostic group was observed with the TD group showing higher whole-word recognition accuracy (Fig. 1B, $P < 0.001$, $F_{(1)} = 15.37$, $partial-\eta^2 = 0.22$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate whole-word recognition responses (Fig. 1B, $P < 0.001$, $F_{(3)} = 506.87$, $partial-\eta^2 = 0.90$). No significant interaction between diagnostic group and SNR was observed ($P = 0.467$, $F_{(3)} = 0.85$, $partial-\eta^2 = 0.02$).

Comparisons of whole-word recognition accuracies were also made between the observed audiovisual response (AV) and the predicted audiovisual (pAV) response based on a given individual's unisensory performance. A three-way, repeated-measures ANOVA was conducted on whole-word recognition scores with diagnostic group as a between-subject factor, and SNR level and observed versus predicted multisensory accuracy (AV and pAV) as within-subject factors. A main effect of observed accuracy versus predicted accuracy for these multisensory conditions was found, with observed audiovisual responses being more accurate than those predicted based on unisensory performance (Fig. 1B, $P < 0.001$, $F_{(1)} = 250.86$, $partial-\eta^2 = 0.82$). A two-way interaction was observed between diagnostic group and observed versus predicted multisensory accuracy with the TD group showing greater behavioral gain ($P = 0.036$, $F_{(1,55)} = 4.64$, $partial-\eta^2 = 0.08$). A significant two-way interaction was also found between observed versus predicted multisensory accuracy and SNR, where a larger difference was seen between actual response accuracy (AV) and predicted (pAV) response accuracy at lower SNRs (Fig. 1C, $P < 0.001$, $F_{(3,55)} = 8.42$, $partial-\eta^2 = 0.13$). This finding is concordant with the principle of inverse effectiveness, in which greater multisensory gain is seen with decreasing SNR, known as inverse effectiveness (i.e., when the unisensory stimulus is degraded). No interaction was found between diagnostic group and SNR level ($P = 0.83$, $F_{(3,55)} = 0.29$, $partial-\eta^2 < 0.01$). A significant three-way interaction was observed, driven by the fact that the ASD children, relative to the TD group, showed diminished multisensory gain as SNR decreased. (Fig. 1F, $P = 0.046$, $F_{(3,55)} = 2.72$, $partial-\eta^2 = 0.05$).

Audiovisual phoneme recognition. A two-way, mixed-model, repeated-measures ANOVA was conducted with diagnostic group as a between-subjects factor and SNR as a within-subjects factor was conducted on phoneme accuracy observed with audiovisual speech presentations. A significant main effect of diagnostic group was observed with the TD group showing higher phoneme recognition accuracy (Fig. 1E, $P = 0.002$, $F_{(1)} = 10.45$, $partial-\eta^2 = 0.16$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate phoneme recognition responses (Fig. 1B, $P < 0.001$, $F_{(3)} = 447.50$, $partial-\eta^2 = 0.89$). A significant interaction between diagnostic group and SNR was observed in audiovisual phoneme recognition ($P = 0.008$, $F_{(3)} = 4.09$, $partial-\eta^2 = 0.07$).

In addition to the simple examination of audiovisual performance, comparisons were also made between the observed AV response and the pAV response based on a given individual's unisensory performance with independent processing assumed (see equation above). A three-way, repeated-measures ANOVA was conducted on phoneme recognition scores with diagnostic group as a between-subject factor, and SNR level and observed versus predicted multisensory accuracy (AV and pAV) as within-subjects' factors. A main effect of observed versus predicted multisensory accuracy was observed, with observed audiovisual responses being more accurate than those predicted based on unisensory performance (Fig. 1B, $P < 0.001$, $F_{(1)} = 99.93$, $partial-\eta^2 = 0.64$). A significant interaction was found between observed versus predicted multisensory accuracy and SNR, which was driven by a larger difference between AV and pAV phoneme-recognition accuracies at lower SNRs (Fig. 1F, $P = 0.001$, $F_{(3,55)} = 8.05$, $partial-\eta^2 = 0.12$). Again, this finding is in concert with the concept of inverse effectiveness. No two-way interactions were observed between diagnostic group and observed versus predicted multisensory performance ($P = 0.22$, $F_{(1,56)} = 1.53$, $partial-\eta^2 = 0.03$), but interactions were seen between diagnostic group and SNR level ($P = 0.015$, $F_{(3,31)} = 3.58$, $partial-\eta^2 = 0.06$). In contrast to results for whole-word recognition, no significant three-way interaction was observed between actual AV and pAV phoneme-recognition accuracies (multisensory gain) at each SNR level and diagnostic group (Fig. 1F, $P = 0.30$, $F_{(3,55)} = 1.22$, $partial-\eta^2 = 0.02$), suggesting that the primary pattern of inverse effectiveness was seen for both groups.

Diagnostic differences in multisensory gain for whole-word and phoneme recognition. As reported above, measurements of multisensory gain revealed a significant difference in the pattern of inverse effectiveness for ASD relative to TD children on

whole-word, but not phoneme recognition. To directly compare multisensory gains in whole-word versus phoneme recognition across diagnostic groups, we conducted a 2 (diagnostic group: ASD vs. TD) \times 4 (SNR: 0, -6, -12, -18 dB SPL) \times 2 (measure: whole-word vs. phoneme) ANOVA on the multisensory gain data shown Figure 1C and F. It is important to note here that comparing across the same data in this way introduces a collinearity issue. Any trial on which a participant accurately perceived the whole word (score of 1) necessarily perceived all phonemes accurately (score of 1). Likewise, any trial on which a participant failed to perceive the whole word correctly (score of 0) necessarily restricted the phoneme accuracy to a less-than-perfect score (score of 0, 0.33, or 0.67). This noted collinearity increases the risk of Type II error and as such, any null findings should be interpreted with caution.

Although not reaching significance, we observed a trend toward a three-way interaction ($P = 0.10$, $F_{(3,171)} = 2.08$, $partial-\eta^2 = 0.04$), indicating that ASD and TD children may exhibit different patterns of inverse effectiveness across whole-word and phoneme accuracies. Follow-up t -tests were conducted to examine the driving factor in this trending interaction (for detailed statistics, see Table 1). In brief, these tests revealed that TD children showed significantly more gain at the lowest SNR (i.e., -18 dB SPL) but only for whole-word recognition. Secondarily, the expected within-subject main effect of SNR was found ($P < 0.001$, $F_{(1,169)} = 6.92$, $partial-\eta^2 = 0.11$), with greater gain seen at lower SNRs. Similarly, a main effect of measure was seen ($P < 0.001$, $F_{(1,171)} = 140.06$, $partial-\eta^2 = 0.71$), with greater gain seen for whole-word recognition performance. When collapsed across whole-word and phoneme recognition, a trend toward a between-subjects effect of diagnostic group was observed ($P = 0.09$, $F_{(1,171)} = 2.98$, $partial-\eta^2 = 0.05$). Significant two-way interactions were found between SNR and measure ($P < 0.001$, $F_{(1,169)} = 6.87$, $partial-\eta^2 = 0.11$). Finally, a trend between SNR and diagnosis was observed ($P = 0.08$, $F_{(1,169)} = 2.28$, $partial-\eta^2 = 0.04$), but there was no observed interaction between diagnostic group and measure ($P = 0.17$, $F_{(1,169)} = 1.98$, $partial-\eta^2 = 0.03$).

Discussion

Sensory challenges are one of the most common complaints of autistic individuals, and have a direct impact on quality of life, including the ability to communicate in noisy, real-world environments. Typically, *seeing* a speaker's mouth can afford substantial behavioral benefits on speech perception [Sumbly & Pollack, 1954]. In ASD, however, decreases in such multisensory benefits have been observed depending on, among other

differences, the level of linguistic complexity. Here, we show that autistic children show significant deficits in speech perception under both unisensory (i.e., auditory alone, visual alone) and multisensory (i.e., audiovisual) conditions and for both whole-word and phoneme recognition. Additionally, autistic individuals showed a decreased level of multisensory gain at low SNR level but only at the level of whole-word recognition. Whereas TD children's performance was largely consistent with the established principle of inverse effectiveness (i.e., where increasing gains in multisensory performance are seen as SNR declines), autistic children failed to show increased multisensory gain as SNR decreased beyond -12db SPL, again particularly during whole-word recognition.

While specific issues with social communication are frequently reported in ASD, sensory processing is also a common issue, and is now included in the most recently updated DSM-5 [APA, 2013]. In our study, these processing differences both within and across the different sensory modalities were readily apparent. Within the senses, both visual and auditory perception of speech were reduced in the ASD relative to the TD group, across both whole-word and phoneme level recognition. This deficit extended to the multisensory domain, where the perception of audiovisual speech was lower in autistic children, again at both the whole-word and phoneme levels. Such deficits likely have a direct impact on children's abilities to communicate in their daily lives, and may also have cascading impacts on cognitive and social processes that rely on successful communication. Indeed, while the current study's N does not allow for a fine-grained age analysis, future studies looking at how this difference emerges throughout development would be beneficial.

In the case of the behavioral benefits specific to multisensory integration, however, a more nuanced pattern emerged. First, both autistic and TD children reliably showed multisensory-mediated enhancements in both whole-word and phoneme recognition. However, at the lowest SNR level (-18 dB), autistic children showed reduced multisensory benefit relative to their TD peers, but only at the level of whole-word recognition, not phoneme recognition. There are two possible explanations for this differential effect.

The first possibility revolves around the principle of inverse effectiveness. Given that autistic individuals show an overall pattern of reduced unisensory responses, these stimuli are, for the autistic group, less effective by definition. Thus, one may predict increased multisensory gain in ASD. However, a number of studies on speech in noise have shown evidence that on falling below a threshold auditory SNR, there is a point at which individuals can no longer extract meaningful information from a given sensory input, and

subsequently fail to exhibit any behavioral benefit from additional sensory information. This results in increasing multisensory gain with less effective stimuli until a threshold is reached, after which multisensory gains decline, creating a distribution peaked at an intermediate SNR. This pattern may be the driver of the differences seen in multisensory gain in word-recognition rates at low SNR, with the ASD group reaching such a threshold before the TD group. Such an interpretation would align with previous observations in non-ASD populations, including in schizophrenia [Ross et al., 2007b] and aging [Gordon & Allen, 2009; Stevenson et al., 2015].

Phoneme-level recognition showed a less differentiated pattern of results. The ASD group showed less multisensory gain at the phoneme level than the TD group, but both groups showed a roughly monotonic increase in multisensory gain as SNR decreased. This finding suggests that while autistic individuals are still showing some preservation in gains associated with integrating the basic auditory (i.e., phonemes) and visual (i.e., visemes) building blocks of speech information, these gains are not sufficient to support an equivalent gain as these building blocks are assembled into whole-word constructs once the SNR reaches a critical point—in the case of this experiment the -18 SNR level. Thus, while attempting to recognize a whole word, an individual may integrate information at the phonemic and visemic level, yet not extract enough information to recognizing the entire word [Stevenson et al., 2015]. By these means, auditory noise (particularly multi-talker babble) may more greatly impact word recognition in ASD relative to phoneme recognition, and using only a binary correct/incorrect scoring of whole-word recognition may overlook intermediate perceptual benefits of integration, especially in more difficult low-SNR conditions. At the same time, it must be noted that word recognition is a more ecologically relevant measure when compared with phoneme/viseme recognition, and thus, the results using this measure should be more heavily weighted. Nonetheless, the differences between performance at these two levels of speech recognition and analysis provide important insights into the mechanisms (and likely networks) subserving these two facets of audiovisual speech integration.

The second possible explanation for the differential patterns of multisensory gain between the two levels of analysis is that there are cognitive differences in language processing between TD and ASD and that these differences manifest much more at the level of whole-word recognition. Perhaps the deficit here emerges as phonemic representations are built into word-based representations. From a more general linguistic and cognitive science aspect, this finding supports theoretical depictions that the architecture of the grammar is separated into different levels of phoneme and word

representation and, critically, that these levels can be differentially impacted in atypical populations. Such a difficulty in combining multiple pieces of information (in this case phonemes) into a coherent perceptual gestalt (whole words) is commonly recognized in several prevailing neurobiological accounts of autism. These include the weak central coherence hypothesis [Burton et al., 2005; Frith & Happe, 1994], the predictive coding hypothesis [Pellicano & Burr, 2012; Sinha et al., 2014; van Boxtel & Lu, 2013; Van de Cruys et al., 2014], and the temporal binding hypothesis [Brock, Brown, Boucher, & Rippon, 2002].

While these data could be explained by either of the two aforementioned hypotheses (or be a cumulative effect of both), these data fit within current models of autism. One of the more germane views of autism for the current work is founded in greater “noise” or variability in neural encoding. This idea posits that less reliable neural response patterns [Perez Velazquez & Galan, 2013] result in less reliable sensory (as well as motor and cognitive) representations [Coskun et al., 2009; Dinstein et al., 2012; Geurts et al., 2008; Haigh, Heeger, Dinstein, Minshew, & Behrmann, 2015; Milne, 2011]. This theory predicts that changes in reliability would not only be seen in early sensory processing streams, but would be amplified as additional neural noise is added at each subsequent level of processing. This model nicely fits the pattern of multisensory benefits seen here, where autistic individuals show greater decrements relative to their TD peers at the word recognition level which is built on the lower level of phoneme perception.

In addition to the differences in multisensory gain itself, as discussed above, it is also important to note the role that differences in visual speech processing may play in the observed pattern of results. All previous studies of audiovisual speech in noise perception in ASD have reported decreased lip reading abilities in ASD [Foxy et al., 2015; Irwin et al., 2011; Smith & Bennetto, 2007]. Despite this noted difference, and despite including measures of lip-reading abilities, only one of these studies incorporated the difference in visual responses into their measurement of multisensory gain [Smith & Bennetto, 2007], whereas the other two compared audiovisual responses to responses with auditory-only speech [Foxy et al., 2015; Irwin et al., 2011]. Given that TD individuals have consistently been found to be able to perceive visual speech more readily than ASD individuals, it is possible that TD individuals will show a greater increase from recognition of audio-only speech to audiovisual speech on account of better visual-only performance *without any difference in multisensory integration*, a concept known as statistical facilitation [Raab, 1962]. Smith and Bennetto [2007] accounted for this using visual performance as a predictor in a hierarchical regression, and in the analysis

Impact of Baseline

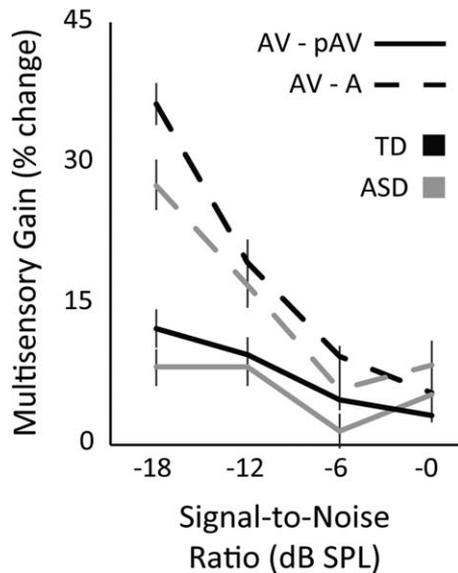


Figure 2. Solid lines represent multisensory gain as calculated in Figure 1, using the predicted audiovisual accuracies (pAV, dashed lines) based on unisensory scores using the equation: $A + V - (A \times V)$. Dashed lines represent multisensory gain calculated using only the auditory responses. Error bars depict standard error.

here, we accounted for this visual difference by modeling the pAV response based on perception of audio-only and visual-only speech in a non-interactive manner [Stevenson et al., 2014e]. The differences between this approach and the comparison to audio-only responses are graphically illustrated in Figure 2. As can be seen in this depiction, the level of gain observed in a given experiment is determined by how one chooses the baseline. Importantly, as shown in Figure 2, calculating multisensory gain relative to auditory *and* visual performance (solid lines) not only results in different absolute levels of multisensory gain relative to using an auditory-only baseline (shaded dashed lines), but has a differential impact relative to the SNR of the presentation, being more impactful at lower SNRs. Given that some studies have shown differential visual lip-reading abilities in ASD, including the current study, this emphasizes the inclusion of visual-only abilities in how baseline measures of integration are calculated in multisensory paradigms [James, Kim, & Stevenson, 2009; James & Stevenson, 2012; James et al., 2012; Stevenson et al., 2009, 2014e].

Conclusion

These data confirm previous findings that autistic children have decreased speech perception abilities relative

to their TD peers in auditory, visual, and audiovisual domains. Furthermore, these data provide novel evidence that autistic children do not benefit from multisensory integration to the same degree as their TD peers do, even when accounting for differences in visual lip reading abilities. This deficit is observed particularly at level of whole-word recognition and at low SNR. Finally, these data also provide a clear example of the importance of accounting for both auditory and visual speech abilities when measuring multisensory integration in autistic children.

Acknowledgments

Western University Faculty Development Research Fund, Banting Postdoctoral Fellowship, NSERC, University of Toronto Dept of Psychology Postdoctoral Fellowship Grant, Autism Research Training Program, Simons Foundation Autism Research Initiative (SFARI), The Wallace Research Foundation, Vanderbilt Kennedy Center MARI/Hobbs Award, Vanderbilt University Kennedy Center, Autism Speaks' Meixner Postdoctoral Fellowship in Translational Research, Grant sponsor: Vanderbilt Institute for Clinical and Translational Research; Grant number: VR7263; Grant sponsor: National Institutes of Health; Grant numbers: F32 DC011993, U54 HD083211, R34 DC010927, R21 CA1834892.

References

- Altieri, N., Stevenson, R.A., Wallace, M.T., & Wenger, M.J. (2015). Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topography*, 28, 479–493.
- APA. (2013). Diagnostic and statistical manual of mental disorders: Dsm-5, 5th ed. Arlington, VA.: American Psychiatric Association.
- Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism spectrum quotient: Children's version (AQ-Child). *Journal of autism and developmental disorders*, 38, 1230–1240.
- Baum, S.H., Stevenson, R.A., & Wallace, M.T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology*, 134, 140–160.
- Bishop, C.W., & Miller, L.M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience*, 21, 1790–1805.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brandwein, A.B., Foxe, J.J., Butler, J.S., Russo, N.N., Altschuler, T.S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex*, 23, 1329–1341.

- Brock, J., Brown, C.C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Development and Psychopathology*, 14, 209–224.
- Brown, L., & Autim Self-Advocacy Network. (2016). Identity-First Language. Retrieved from <http://autisticadvocacy.org/home/about-asan/identity-first-language/>
- Burnette, C.P., Mundy, P.C., Meyer, J.A., Sutton, S.K., Vaughan, A.E., & Charak, D. (2005). Weak central coherence and its relations to theory of mind and anxiety in autism. *Journal of Autism and Developmental Disorders*, 35, 63–73.
- Coskun, M.A., Varghese, L., Reddoch, S., Castillo, E.M., Pearson, D.A., Loveland, K.A., ... Sheth, B.R. (2009). Increased response variability in autistic brains? *Neuroreport*, 20, 1543–1548.
- Dinstein, I., Heeger, D.J., Lorenzi, L., Minshew, N.J., Malach, R., & Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron*, 75, 981–991.
- Foss-Feig, J.H., Kwakye, L.D., Cascio, C.J., Burnette, C.P., Kadivar, H., Stone, W.L., & Wallace, M.T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research Experimentelle Hirnforschung*, 203, 381–389.
- Foxe, J.J., Molholm, S., Del Bene, V.A., Frey, H.-P., Russo, N.N., Blanco, D., ... Ross, L.A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (asd) and their resolution during early adolescence. *Cerebral Cortex*, 25, 298–312.
- Fraser, S., Gagne, J.P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research: JSLHR*, 53, 18–33.
- Frith, U., & Happe, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, 50, 115–132.
- Garlock, V.M., Walley, A.C., & Metsala, J.L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, 45, 468–492.
- Geurts, H.M., Grasman, R.P., Verte, S., Oosterlaan, J., Roeyers, H., van Kammen, S.M., & Sergeant, J.A. (2008). Intra-individual variability in adhd, autism spectrum disorders and tourette’s syndrome. *Neuropsychologia*, 46, 3030–3041.
- Gordon, M.S., & Allen, S. (2009). Audiovisual speech in older and younger adults: Integrating a distorted visual signal with speech in noise. *Experimental Aging Research*, 35, 202–219.
- Haigh, S.M., Heeger, D.J., Dinstein, I., Minshew, N., & Behrmann, M. (2015). Cortical variability in the sensory-evoked response in autism. *Journal of Autism and Developmental Disorders*, 45, 1176–1190.
- Identity-first Autistic. 2016. Retrieved from <http://www.identityfirstautistic.org/>
- Irwin, J.R., Tornatore, L.A., Brancazio, L., & Whalen, D.H. (2011). Can children with autism spectrum disorders “hear” a speaking face? *Child Development*, 82, 1397–1403.
- James, T.W., Kim, S., & Stevenson, R.A. (2009). Assessing multisensory interaction with additive factors and functional MRI. Paper presented at The International Society for Psychophysics, Galway, Ireland.
- James, T.W., & Stevenson, R.A. (2012). The use of fMRI to assess multisensory integration. In *Frontiers in the Neural Bases of Multisensory Processes*. Murray M, Wallace M (Eds.) Taylor & Francis Group, London. Chapter 8, 131–146.
- James, T.W., Stevenson, R.A., & Kim, S. (2012). Inverse effectiveness in multisensory processing. In Stein BE (Ed.), *The new handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.
- Kenny, L., Hattersley, C., Molins, B., Buckley, C.J., Povey, C., & Pellicano, E. (2015). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, 20, 442–462.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods*, 44, 978–990.
- 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. *Front Integrative Neuroscience*, 4, 129.
- Lachs, L., & Hernandez, L.R. (1998). Update: The hoosier audiovisual multitalker database. In Pisoni DB (Ed.), *Research on spoken language processing* (pp. 377–388). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19, 363–387.
- Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry*, 36, 1365–1382.
- Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., Bishop, S. (2012). *Autism diagnostic observation schedule—2nd edition (ADOS-2)*. Los Angeles, CA: Western Psychological Corporation.
- Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Meredith, M.A., & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science* (New York, NY), 221, 389–391.
- Meredith, M.A., & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56, 640–662.
- Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: Evidence from single-trial analysis of evoked EEG. *Frontiers in Psychology*, 2, 51.
- Nath, A.R., & Beauchamp, M.S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience*, 31, 1704–1714.
- Pelli, D.G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.

- Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16, 504–510.
- Perez Velazquez, J.L., & Galan, R.F. (2013). Information gain in the brain's resting state: A new perspective on autism. *Frontiers in Neuroinformatics*, 7, 37.
- Raab, D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24, 574–590.
- 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. *The European Journal of Neuroscience*, 33, 2329–2337.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007a). Do you see what i am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17, 1147–1153.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., & Foxe, J.J. (2007b). Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research*, 97, 173–183.
- Russo, N., Foxe, J.J., Brandwein, A.B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Research*, 3, 253–267.
- Sheffert, S.M., Lachs, L., & Hernandez, L.R. (1996). The hooyer audiovisual multitalker database. In Pisoni DB (Ed.), *Research on spoken language processing* (pp. 578–583). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Sinclair, J. (1999). Why I dislike "person-first" language. Retrieved from <http://autismmythbusters.com/general-public/autistic-vs-people-with-autism/jim-sinclair-why-i-dislike-person-first-language/>
- Sinha, P., Kjelgaard, M.M., Gandhi, T.K., Tsourides, K., Cardinaux, A.L., Pantazis, D., ... Held, R.M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 15220–15225.
- Smith, E.G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 48, 813–821.
- Stevenson, R.A., Altieri, N.A., Kim, S., Pisoni, D.B., & James, T.W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, 49, 3308–3318.
- Stevenson, R., Bushmakina, M., Kim, S., Wallace, M., Puce, A., & James, T. (2012a). Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography*, 25, 308–326.
- Stevenson, R.A., Bushmakina, M., Kim, S., Wallace, M.T., Puce, A., & James, T.W. (2012b). Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography*, 25, 308–326.
- Stevenson, R.A., Geoghegan, M.L., & James, T.W. (2007). Superadditive bold activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*, 179, 85–95.
- Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., ... Wallace, M.T. (2014e). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography*, 27, 707–730.
- Stevenson, R.A., & James, T.W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, 44, 1210–1223.
- Stevenson, R.A., Kim, S., & James, T.W. (2009). An additive-factors design to disambiguate neuronal and areal convergence: Measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Experimental Brain Research*, 198, 183–194.
- Stevenson, R.A., Nelms, C.E., Baum, S.H., Zurkovsky, L., Barense, M.D., Newhouse, P.A., & Wallace, M.T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging*, 36, 283–291.
- Stevenson, R.A., Segers, M., Ferber, S., Barense, M.D., & Wallace, M.T. (2014a). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology*, 5, 379.
- Stevenson, R.A., Siemann, J.K., Schneider, B.C., Eberly, H.E., Woynaroski, T.G., Camarata, S.M., & Wallace, M.T. (2014b). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, 34, 691–697.
- Stevenson, R.A., Siemann, J.K., Woynaroski, T.G., Schneider, B.C., Eberly, H.E., Camarata, S.M., & Wallace, M.T. (2014c). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44, 1470–1477.
- Stevenson, R.A., Siemann, J.K., Woynaroski, T.G., Schneider, B.C., Eberly, H.E., Camarata, S.M., & Wallace, M.T. (2014d). Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44, 3161–3167.
- Stevenson, R.A., VanDerKlok, R.M., Pisoni, D.B., & James, T.W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, 55, 1339–1345.
- Summy, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.
- van Boxtel, J.J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, 4, 19.
- Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121, 649–675.
- Wallace, M.T., & Stevenson, R.A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105–123.
- Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, 76, 1246–1266.
- Wechsler, D., & Hsiao-pin, C. (2011). WASI-II: Wechsler abbreviated scale of intelligence. Pearson.
- Werner, S., & Noppeney, U. (2009). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, 20, 1829–1842.