Published in final edited form as:

Autism Res. 2017 July ; 10(7): 1280–1290. doi:10.1002/aur.1776.

Multisensory speech perception in Autism Spectrum Disorder: From phoneme to whole-word perception

yan A. Stevenson^{1,2}, Sarah H. Baum³, Magali Segers⁴, Susanne Ferber^{5,6}, Morgan D. Barense^{5,6}, and Mark T. Wallace^{7,8,9,10,11}

¹Department of Psychology, Western University, London, ON, Canada

²Brain and Mind Institute, Western University, London, ON, Canada

³Department of Psychology, University of Washington, Seattle, WA, USA

⁴York University, Toronto, ON, Canada

⁵Dept. of Psychology, University of Toronto, Toronto, ON, Canada

⁶Rotman Research Institute, Toronto, ON, Canada

⁷Vanderbilt Brain Institute, Nashville, TN, USA

⁸Vanderbilt Kennedy Center, Nashville, TN, USA

⁹Vanderbilt University, Nashville, TN, USA

¹⁰Dept. of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

¹¹Dept. of Psychology, Vanderbilt University, Nashville, TN, USA

Abstract

Lay Abstract—In a noisy environment it can often be difficult to understand what a speaker is saying, but being able to see the speaker's mouth makes it easier. In fact, the noisier the environment, the more it helps to both see and hear a speaker. This is due to a process known as multisensory integration, where information from two senses, in this case audition and vision, are combined. Autistic children often have difficulties with multisensory integration, which may add to difficulties with social communication. We tested how well autistic and typically developing children integrated auditory speech and visual speech, both at the level of whole words and also at the level of individual speech sounds, or phonemes. At both the whole-word and the phoneme level, autistic children were less able to correctly identify speech sounds, regardless of how they were presented (audio-only, visual-only, audiovisual). Autistic children showed equal benefit in phoneme perception from seeing the mouth movements accompanying the auditory speech. In noisy environments however, these equal gains in phoneme perception did not extend to whole-word recognition, where autistic individuals showed reduced benefit from seeing the mouth movements while hearing auditory speech. Such reductions in speech perception abilities,

including a reduced benefit from the presence of visual speech cues in noisy environments, may contribute to social communication difficulties in autistic children.

Scientific Abstract—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 levels of noise. 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 signal to noise ratios (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.

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 (Foss-Feig et al., 2010; Russo et al., 2010; Kwakye et al., 2011; Ross et al., 2011; Brandwein et al., 2012; Stevenson et al., 2014a; Stevenson et al., 2014b; Stevenson et al., 2014c, d; Wallace and Stevenson, 2014; Baum et al., 2015). 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

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 and ASAN 2016). Given this, we will respect this preference and use this language throughout this manuscript.

differences appear to be related to a reduced ability to integrate individual pieces of sensory information into a coherent unified percept.

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 (Sumby and Pollack, 1954; Ross et al., 2007a; Stevenson and James, 2009; Ross et al., 2011; Stevenson et al., 2014e; Altieri et al., 2015; Stevenson et al., 2015) and less effortful (Fraser et al., 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 (Foxy et al., 2013). 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 (Meredith and Stein, 1983, 1986; Wallace et al., 1996; Stevenson et al., 2007; Bishop and Miller, 2009; Stevenson and James, 2009; Stevenson et al., 2009; Werner and Noppeney, 2009; Nath and Beauchamp, 2011; James et al., 2012; Stevenson et al., 2012a).

To our knowledge, three studies to date have measured audiovisual speech-in-noise abilities of autistic individuals (Smith and Bennetto, 2007; Irwin et al., 2011; Foxe et al., 2013). One study showed that although autistic adolescents (12–17 years old) demonstrated similar auditory-only abilities relative to their 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 and 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 years 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 years old), but a decrease in audiovisual performance only in the younger groups (under the age of twelve) (Foxy et al., 2013). 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 (Irwin et al., 2011;

Foxe et al., 2013). 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 (Smith and Bennetto, 2007; Foxe et al., 2013) 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; Paquette-Smith et al., 2014). 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 et al., 1996). The stimuli selected were monosyllabic English words that were matched across sets for accuracy on both visual-only and audio-only recognition (Lachs and Hernandez, 1998), and were also matched according to lexical neighborhood density (Sheffert et al., 1996; Luce and Pisoni, 1998). Median age of acquisition for the word set was 5.5 years old, and median frequency was 35.6 per million words (Garlock et al., 2001; Kuperman et al., 2012). This stimulus set of single-word tokens has been used successfully in previous studies of multisensory integration (Stevenson et al., 2007; Stevenson and James, 2009; Stevenson et al., 2009; Stevenson et al., 2010; Stevenson et al., 2011; Stevenson et al., 2012b; Stevenson et al., 2015). 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 (MATHEMATICS Inc., Natick, MA) software with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Visual stimuli were 200×200 pixels and subtended 10×10° of visual angle. Audio stimuli were presented through two aligned speakers on each side of the monitor. All tokens lasted two seconds 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 500ms prior to the beginning of the stimulus token and ended 500ms following token offset. The RMS of the auditory babble was linearly ramped up and down, respectively, during the pre- and post-stimulus 500ms 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 60cm 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 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 minutes, 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) $\times$ 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) $\times$ 4 (SNR: 0, -6, -12, -18 dB SPL) $\times$ 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) $\times$ 4 (SNR: 0, -6, -12, -18 dB SPL) $\times$ 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).

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 (Figure 1A), but were statistically significant at the viseme-recognition level (Figure 1D). For statistical results, see Table 1.

Auditory-only, whole-word recognition—For whole-word recognition (Figure 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 (Figure 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 (Figure 1B, $p < 0.001$, $F_{(1)} = 15.37$, $\text{partial-}\eta^2 = 0.22$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate whole-word recognition responses (Figure 1B, $p < 0.001$, $F_{(3)} = 506.87$, $\text{partial-}\eta^2 = 0.90$). No significant interaction between diagnostic group and SNR was observed ($p = 0.467$, $F_{(3)} = 0.85$, $\text{partial-}\eta^2 = 0.02$).

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 vs. predicted multisensory accuracy (AV and pAV) as within-subject factors. A main effect of observed accuracy vs. predicted accuracy for these multisensory conditions was found, with observed audiovisual responses being more accurate than those predicted based on unisensory performance (Figure 1B, $p < 0.001$, $F_{(1)} = 250.86$, $\text{partial-}\eta^2 = 0.82$). A two-way interaction was observed between diagnostic group and observed vs. predicted multisensory accuracy with the TD group showing greater behavioral gain ($p = 0.036$, $F_{(1,55)} = 4.64$, $\text{partial-}\eta^2 = 0.08$). A significant two-way interaction was also found between observed vs. predicted multisensory accuracy and SNR, where a larger difference was seen between actual response accuracy (AV) and predicted (pAV) response accuracy at lower SNRs (Figure 1C, $p < 0.001$, $F_{(3,55)} = 8.42$, $\text{partial-}\eta^2 = 0.13$). 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 ($p = 0.83$, $F_{(3,55)} = 0.29$, $\text{partial-}\eta^2 < 0.01$). 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. (Figure 1F, $p = 0.046$, $F_{(3,55)} = 2.72$, $\text{partial-}\eta^2 = 0.05$).

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 (Figure 1E, $p = 0.002$, $F_{(1)} = 10.45$, $\text{partial-}\eta^2 = 0.16$). A significant main effect of SNR was also observed, with higher SNRs associated with more accurate phoneme recognition responses (Figure 1B, $p < 0.001$, $F_{(3)} = 447.50$, $\text{partial-}\eta^2 = 0.89$). A significant interaction between diagnostic group and SNR was observed in audiovisual phoneme recognition ($p = 0.008$, $F_{(3)} = 4.09$, $\text{partial-}\eta^2 = 0.07$).

In addition to the simple examination of audiovisual performance, comparisons were also made between the observed audiovisual response (AV) and the predicted audiovisual (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 vs. predicted multisensory accuracy (AV and pAV) as within-subjects' factors. A main effect of observed vs. predicted multisensory accuracy was observed, with observed audiovisual responses being more accurate than those predicted based on unisensory performance (Figure 1B, $p < 0.001$, $F_{(1)} = 99.93$, $\text{partial-}\eta^2 = 0.64$). A significant interaction was found between observed vs. predicted multisensory accuracy and SNR, which was driven by a larger difference between audiovisual (AV) and predicted audiovisual (pAV) phoneme-recognition accuracies at lower SNRs (Figure 1F, $p = 0.001$, $F_{(3,55)} = 8.05$, $\text{partial-}\eta^2 = 0.12$). Again, this finding is in concert with the concept of inverse effectiveness. No two-way interactions were observed between diagnostic group and observed vs. predicted multisensory performance ($p = 0.22$, $F_{(1,56)} = 1.53$, $\text{partial-}\eta^2 = 0.03$), but interactions were seen between diagnostic group and SNR level ($p = 0.015$, $F_{(3,31)} = 3.58$, $\text{partial-}\eta^2 = 0.06$). 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 (Figure 1F, $p = 0.30$, $F_{(3,55)} = 1.22$, $\text{partial-}\eta^2 = 0.02$), 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 vs 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 I error and as such, any null findings should be interpreted with caution.

Although not reaching significance, we observed a trend towards a three-way interaction ($p = 0.10$, $F_{(3,171)} = 2.08$, $\text{partial-}\eta^2 = 0.04$), 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. Secondly, the expected within-subject main effect of SNR was found ($p < 0.001$, $F_{(1,169)} = 6.92$, $\text{partial-}\eta^2 = 0.11$), with greater gain seen at lower SNRs. Similarly, a main effect of measure was seen ($p < 0.001$, $F_{(1,171)} = 140.06$, $\text{partial-}\eta^2 = 0.71$), with greater gain seen for whole-word recognition performance. When collapsed across whole-word and phoneme recognition, a trend towards a between-subjects effect of diagnostic group was observed ($p = 0.09$, $F_{(1,171)} = 2.98$, $\text{partial-}\eta^2 = 0.05$). Significant two-way interactions were

found between SNR and measure ($p < 0.001$, $F_{(1,169)} = 6.87$ *partial- η^2* = 0.11). Finally, a trend between SNR and diagnosis was observed ($p = 0.08$, $F_{(1,169)} = 2.28$, *partial- η^2* = 0.04), but there was no observed interaction between diagnostic group and measure ($p = 0.17$, $F_{(1,169)} = 1.98$, *partial- η^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 and 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 upon 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

upon 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 and 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, but, also 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 (Frith and Happe, 1994; Burnette et al., 2005), the predictive coding hypothesis (Pellicano and Burr, 2012; van Boxtel and Lu, 2013; Sinha et al., 2014; Van de Cruys et al., 2014), and the temporal binding hypothesis (Brock et al., 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 and Galan, 2013) result in less reliable sensory (as well as motor and cognitive) representations (Geurts et al., 2008; Coskun et al., 2009; Milne, 2011; Dinstein et al., 2012; Haigh et al., 2014). 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 upon 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 (Smith and Bennetto, 2007; Irwin et al., 2011; Foxe et al., 2013). 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 and Bennetto, 2007), whereas the other two compared audiovisual responses to responses with auditory-only speech (Irwin et al., 2011; Foxe et al., 2013). 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 & Bennetto (2007) accounted for this by using visual performance as a predictor in a hierarchical regression, and in the analysis here, we accounted for this visual difference by modeling the predicted audiovisual 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 et al., 2009; Stevenson et al., 2009; James and Stevenson, 2012; James et al., 2012; Stevenson et al., 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

Grants

Grant sponsor: Western University Faculty Development Research Fund; Grant number: N/A

Grant sponsor: Banting Postdoctoral Fellowship, NSERC; Grant number: N/A

Grant sponsor: University of Toronto Dept of Psychology Postdoctoral Fellowship Grant; Grant number: N/A

Grant sponsor: Autism Research Training Program; Grant number: N/A

Grant sponsor: National Institutes of Health; Grant number: F32 DC011993, U54 HD083211, R34 DC010927, R21 CA1834892,

Grant sponsor: Simons Foundation Autism Research Initiative (SFARI); Grant number:

Grant sponsor: The Wallace Research Foundation; Grant number: N/A

Grant sponsor: Vanderbilt Institute for Clinical and Translational Research Grant; Grant number: VR7263

Grant sponsor: Vanderbilt Kennedy Center MARI/Hobbs Award; Grant number: N/A

Grant sponsor: Vanderbilt University Kennedy Center; Grant number: N/A

Grant sponsor: Autism Speaks' Meixner Postdoctoral Fellowship in Translational Research; Grant number: N/A

Funding for RS was provided by a Western University Faculty Development Research Fund, a Banting Postdoctoral Fellowship administered by the Canadian National Science and Engineering Research Council (NSERC), a University of Toronto Department of Psychology Postdoctoral Fellowship Grant, the Autism Research Training Program via the Canadian Institutes of Health Research (CIHR), and the American National Institutes of Health F32 DC011993. MW was funded by National Institutes of Health U54 HD083211, R34 DC010927, R21 CA1834892, grants from the Simons Foundation Autism Research Initiative (SFARI) and from the Wallace Research Foundation, a Vanderbilt Institute for Clinical and Translational Research grant VR7263, a Vanderbilt Kennedy Center MARI/Hobbs Award, the Vanderbilt Brain Institute, and the Vanderbilt University Kennedy Center. SB was funded by Autism Speaks' Meixner Postdoctoral Fellowship in Translational Research.

References

- Altieri N, Stevenson RA, Wallace MT, Wenger MJ. Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topogr.* 2015; 28:479–493. [PubMed: 24276220]
- APA. Diagnostic and statistical manual of mental disorders: Dsm-5. 5. Arlington, Va: American Psychiatric Association; 2013.
- Baum SH, Stevenson RA, Wallace MT. Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in neurobiology.* 2015; 134:140–160. [PubMed: 26455789]
- Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. *Journal of cognitive neuroscience.* 2009; 21:1790–1805. [PubMed: 18823249]
- Brainard DH. The psychophysics toolbox. *Spatial vision.* 1997; 10:433–436. [PubMed: 9176952]
- Brandwein AB, Foxe JJ, Butler JS, Russo NN, Altschuler TS, Gomes H, Molholm S. The development of multisensory integration in high-functioning autism: High-density electrical mapping and

psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cereb Cortex*. 2012

- Brock J, Brown CC, Boucher J, Rippon G. The temporal binding deficit hypothesis of autism. *Dev Psychopathol*. 2002; 14:209–224. [PubMed: 12030688]
- Brown, L. Autim Self-Advocacy Network. [accessed 13 November 2016] Identity-First Language. 2016. Available at: <http://autisticadvocacy.org/home/about-asan/identity-first-language/>
- Burnette CP, Mundy PC, Meyer JA, Sutton SK, Vaughan AE, Charak D. Weak central coherence and its relations to theory of mind and anxiety in autism. *J Autism Dev Disord*. 2005; 35:63–73. [PubMed: 15796123]
- Coskun MA, Varghese L, Reddoch S, Castillo EM, Pearson DA, Loveland KA, Papanicolaou AC, Sheth BR. Increased response variability in autistic brains? *Neuroreport*. 2009; 20:1543–1548. [PubMed: 19820416]
- Dinstein I, Heeger DJ, Lorenzi L, Minshew NJ, Malach R, Behrmann M. Unreliable evoked responses in autism. *Neuron*. 2012; 75:981–991. [PubMed: 22998867]
- Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. *Experimental brain research Experimentelle Hirnforschung*. 2010; 203:381–389. [PubMed: 20390256]
- Foxe JJ, Molholm S, Del Bene VA, Frey H-P, Russo NN, Blanco D, Saint-Amour D, Ross LA. Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (asd) and their resolution during early adolescence. *Cerebral Cortex*. 2013:bht213.
- Fraser S, Gagne JP, Alepins M, Dubois P. 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*. 2010; 53:18–33.
- Frith U, Happe F. Autism: Beyond “theory of mind”. *Cognition*. 1994; 50:115–132. [PubMed: 8039356]
- Garlock VM, Walley AC, Metsala JL. Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and language*. 2001; 45:468–492.
- Geurts HM, Grasman RP, Verte S, Oosterlaan J, Roeyers H, van Kammen SM, Sergeant JA. Intra-individual variability in adhd, autism spectrum disorders and tourette’s syndrome. *Neuropsychologia*. 2008; 46:3030–3041. [PubMed: 18619477]
- Gordon MS, Allen S. Audiovisual speech in older and younger adults: Integrating a distorted visual signal with speech in noise. *Experimental aging research*. 2009; 35:202–219. [PubMed: 19280447]
- Haigh SM, Heeger DJ, Dinstein I, Minshew N, Behrmann M. Cortical variability in the sensory-evoked response in autism. *J Autism Dev Disord*. 2014
- [accessed 13 November 2016] Identity-first Autistic. 2016. Available at: <http://www.identityfirstautistic.org/>
- Irwin JR, Tornatore LA, Brancazio L, Whalen DH. Can children with autism spectrum disorders “hear” a speaking face? *Child development*. 2011; 82:1397–1403. [PubMed: 21790542]
- James, TW., Stevenson, RA. The use of fmri to assess multisensory integration. In: Murray, MM., Wallace, MT., editors. *The neural bases of multisensory processes*. Boca Raton (FL): 2012.
- James, TW., Kim, S., Stevenson, RA. Assessing multisensory interaction with additive factors and functional mri. *The International Society for Psychophysics*; Galway, Ireland: 2009.
- James, TW., Stevenson, RA., Kim, S. Inverse effectiveness in multisensory processing. In: Stein, BE., editor. *The new handbook of multisensory processes*. Cambridge, MA: MIT Press; 2012.
- Kanner L. Autistic disturbances of affective contact. *Nervous Child*. 1943; 2:217–250.
- Kenny L, Hattersley C, Molins B, Buckley CJ, Povey C, Pellicano E. Which terms should be used to describe autism? Perspectives from the uk autism community. *Autism*. 2015
- Kuperman V, Stadhagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 english words. *Behavior research methods*. 2012; 44:978–990. [PubMed: 22581493]

- Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front Integr Neurosci*. 2011; 4:129. [PubMed: 21258617]
- Lachs, L., Hernandez, LR. Update: The hoosier audiovisual multitalker database. In: Pisoni, DB., editor. *Research on spoken language processing*. Bloomington, IN: Speech Research Laboratory, Indiana University; 1998. p. 377-388.
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J. Autism diagnostic interview: A standardized investigator-based instrument. *Journal of autism and developmental disorders*. 1989; 19:363–387. [PubMed: 2793783]
- Lord C. Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry*. 1995; 36:1365–1382. [PubMed: 8988272]
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear and hearing*. 1998; 19:1–36. [PubMed: 9504270]
- Meredith MA, Stein BE. Interactions among converging sensory inputs in the superior colliculus. *Science (New York, NY)*. 1983; 221:389–391.
- Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology*. 1986; 56:640–662. [PubMed: 3537225]
- Milne E. Increased intra-participant variability in children with autistic spectrum disorders: Evidence from single-trial analysis of evoked eeg. *Frontiers in psychology*. 2011; 2:51. [PubMed: 21716921]
- Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J Neurosci*. 2011; 31:1704–1714. [PubMed: 21289179]
- Pelli DG. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*. 1997; 10:437–442. [PubMed: 9176953]
- Pellicano E, Burr D. When the world becomes ‘too real’: A bayesian explanation of autistic perception. *Trends Cogn Sci*. 2012; 16:504–510. [PubMed: 22959875]
- Perez Velazquez JL, Galan RF. Information gain in the brain’s resting state: A new perspective on autism. *Frontiers in neuroinformatics*. 2013; 7:37. [PubMed: 24399963]
- Raab DH. Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*. 1962; 24:574–590. [PubMed: 14489538]
- Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what i am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex*. 2007a; 17:1147–1153. [PubMed: 16785256]
- Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ. Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia research*. 2007b; 97:173–183. [PubMed: 17928202]
- Ross LA, Molholm S, Blanco D, Gomez-Ramirez M, Saint-Amour D, Foxe JJ. The development of multisensory speech perception continues into the late childhood years. *The European journal of neuroscience*. 2011; 33:2329–2337. [PubMed: 21615556]
- Russo N, Foxe JJ, Brandwein AB, Altschuler T, Gomes H, Molholm S. Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Res*. 2010; 3:253–267. [PubMed: 20730775]
- Sheffert, SM., Lachs, L., Hernandez, LR. The hooiser audiovisual multitalker database. In: Pisoni, DB., editor. *Research on spoken language processing*. Bloomington, IN: Speech Research Laboratory, Indiana University; 1996. p. 578-583.
- Sinclair, J. [accessed 13 November 2016] Why I dislike “person-first” language. 1999. Available at: <http://autismmythbusters.com/general-public/autistic-vs-people-with-autism/jim-sinclair-why-i-dislike-person-first-language/>
- Sinha P, Kjelgaard MM, Gandhi TK, Tsourides K, Cardinaux AL, Pantazis D, Diamond SP, Held RM. Autism as a disorder of prediction. *Proc Natl Acad Sci U S A*. 2014; 111:15220–15225. [PubMed: 25288765]
- Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. *J Child Psychol Psychiatry*. 2007; 48:813–821. [PubMed: 17683453]

- Stevenson R, Bushmakin M, Kim S, Wallace M, Puce A, James T. Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography*. 2012a;1–19.
- Stevenson RA, James TW. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuro Image*. 2009; 44:1210–1223. [PubMed: 18973818]
- Stevenson RA, Geoghegan ML, James TW. Superadditive bold activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*. 2007; 179:85–95. [PubMed: 17109108]
- Stevenson RA, Kim S, James TW. 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*. 2009; 198:183–194. [PubMed: 19352638]
- Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuro Image*. 2011; 55:1339–1345. [PubMed: 21195198]
- Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW. Neural processing of asynchronous audiovisual speech perception. *Neuro Image*. 2010; 49:3308–3318. [PubMed: 20004723]
- Stevenson RA, Segers M, Ferber S, Barense MD, Wallace MT. The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Front Psychol*. 2014a; 5:379. [PubMed: 24904448]
- Stevenson RA, Bushmakin M, Kim S, Wallace MT, Puce A, James TW. Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topogr*. 2012b; 25:308–326. [PubMed: 22367585]
- Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *J Neurosci*. 2014b; 34:691–697. [PubMed: 24431427]
- Stevenson RA, Siemann JK, Woynaroski TG, Schneider BC, Eberly HE, Camarata SM, Wallace MT. Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *J Autism Dev Disord*. 2014c; 44:1470–1477. [PubMed: 24218241]
- Stevenson RA, Siemann JK, Woynaroski TG, Schneider BC, Eberly HE, Camarata SM, Wallace MT. Evidence for diminished multisensory integration in autism spectrum disorders. *J Autism Dev Disord*. 2014d; 44:3161–3167. [PubMed: 25022248]
- Stevenson RA, Nelms CE, Baum SH, Zurkovsky L, Barense MD, Newhouse PA, Wallace MT. Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of aging*. 2015; 36:283–291. [PubMed: 25282337]
- Stevenson RA, Ghose D, Fister JK, Sarko DK, Altieri NA, Nidiffer AR, Kurela LR, Siemann JK, James TW, Wallace MT. Identifying and quantifying multisensory integration: A tutorial review. *Brain Topogr*. 2014e
- Sumbly WH, Pollack I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*. 1954; 26:212–215.
- van Boxtel JJ, Lu H. A predictive coding perspective on autism spectrum disorders. *Front Psychol*. 2013; 4:19. [PubMed: 23372559]
- Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, Wagemans J. Precise minds in uncertain worlds: Predictive coding in autism. *Psychological review*. 2014; 121:649–675. [PubMed: 25347312]
- Wallace MT, Stevenson RA. The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*. 2014; 64C:105–123.
- Wallace MT, Wilkinson LK, Stein BE. Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of neurophysiology*. 1996; 76:1246–1266. [PubMed: 8871234]
- Werner S, Noppeney U. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex*. 2009

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.

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.

Table 1

Detailed statistics for phoneme and whole-word recognition

Phoneme Recognition							Whole-word Recognition				
Modality	SNR	TD	ASD	p	t ₍₅₇₎	d	TD	ASD	p	t ₍₅₇₎	d
V	N/A	30.4±13.0	22.9±12.4	.033	2.18	0.59	6.6±7.2	4.2±6.3	.190	1.32	0.35
	0	90.7±8.1	83.0±15.9	.017	2.46	0.61	80.3±10.7	69.4±19.2	.007	2.81	0.70
	-6	83.1±8.0	80.4±7.2	.189	1.33	0.35	66.4±12.2	61.6±12.5	.15	1.47	0.39
	-12	67.1±10.7	60.3±10.1	.019	2.42	0.65	45.1±13.0	37.5±11.1	.024	2.31	0.63
A	-18	19.9±8.3	16.8±7.4	.154	1.44	0.39	5.2±4.3	5.1±4.1	.934	.084	0.02
	0	96.2±5.0	91.5±8.1	.009	2.72	0.70	91.1±7.3	82.6±12.1	.001	3.36	0.85
	-6	92.4±8.9	86.2±10.7	.019	2.42	0.63	82.9±11.9	71.4±14.6	.002	3.31	0.86
	-12	84.5±11.3	77.4±15.0	.041	2.09	0.53	69.7±13.5	60.3±18.9	.031	2.22	0.57
AV	-18	56.1±13.3	43.7±13.9	.001	3.45	0.91	30.8±13.8	18.3±9.9	<.001	3.76	1.04
	0	3.2±5.4	5.3±10.5	.307	1.03	0.25	9.6±10.1	12.3±11.4	.349	0.94	0.25
	-6	4.8±5.9	1.3±8.9	.078	1.79	0.46	14.4±12.0	8.0±15.2	.078	1.79	0.47
	-12	9.6±11.6	8.3±10.2	.660	0.44	0.12	22.4±14.8	20.3±15.6	.619	0.50	0.14
MultisensoryGain*		-18	12.3±12.5	8.3±9.2	.192	1.32	19.4±13.4	8.8±8.1	.002	3.33	0.96

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

* Values given for multisensory gain are percent changes between observed and predicted recognition accuracy rates.