



# Perception of audio-visual synchrony in infants at elevated likelihood of developing autism spectrum disorder

Kirin N. Suri<sup>1,2</sup> · Margaret Whedon<sup>1</sup> · Michael Lewis<sup>1</sup>

Received: 6 October 2022 / Revised: 5 February 2023 / Accepted: 8 February 2023 / Published online: 23 February 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

The inability to perceive audio-visual speech as a unified event may contribute to social impairments and language deficits in children with autism spectrum disorder (ASD). In this study, we examined and compared two groups of infants on their sensitivity to audio-visual asynchrony for a social (speaking face) and non-social event (bouncing ball) and assessed the relations between multisensory integration and language production. Infants at elevated likelihood of developing ASD were less sensitive to audio-visual synchrony for the social event than infants without elevated likelihood. Among infants without elevated likelihood, greater sensitivity to audio-visual synchrony for the social event was associated with a larger productive vocabulary. **Conclusion:** Findings suggest that early deficits in multisensory integration may impair language development among infants with elevated likelihood of developing ASD.

## What is Known:

- Perceptual integration of auditory and visual cues within speech is important for language development.
- Prior work suggests that children with ASD are less sensitive to the temporal synchrony within audio-visual speech.

## What is New:

- In this study, infants at elevated likelihood of developing ASD showed a larger temporal binding window for a dynamic social event (Speaking Face) than TD infants, suggesting less efficient multisensory integration.

**Keywords** Sensory integration · Autism · Audio-visual speech · Habituation · Language

The ability to use cues from multiple senses in concert (i.e., multisensory integration) is a fundamental aspect of brain function and critical developmental milestone for infants, who must learn to perceive complex, multimodal events in ways that are meaningful and relevant [1, 2]. In general, multisensory integration is accomplished by the detection of *intersensory redundancy*, the spatially coordinated and/or temporally synchronized presentation of the same

information across two or more sense modalities [3]. From this perspective, audio-visual events that occur within close temporal proximity are automatically integrated if they fall within a specific range called the audio-visual temporal binding window (i.e., intersensory temporal contiguity window; see Lewkowicz 2000 [4]). Conceptually, the temporal binding window is a measure of sensitivity to audio-visual temporal synchrony, quantified as the maximum amount of time that auditory and visual sensory inputs can be physically separated and still perceived as unitary or synchronous.

Although infants are sensitive to audio-visual synchrony relations from birth [5–6], the size of the audio-visual temporal binding window has been found to be much larger in infants than in adults [7, 8]. As children acquire perceptual experience with synchronous events, their sensitivity to audio-visual synchrony improves and their temporal binding windows grow smaller [9]. However, there is evidence to suggest this process is disrupted in individuals with autism

Communicated by Peter de Winter.

✉ Margaret Whedon  
margaret.whedon@rutgers.edu

<sup>1</sup> Institute for the Study of Child Development, Rutgers Robert Wood Johnson Medical School, 89 French Street, New Brunswick, NJ 08901, USA

<sup>2</sup> Children's Health at Hackensack Meridian, Hackensack, NJ 07601, USA

spectrum disorder (ASD; [10]). Since very little work in this regard has focused on infancy, the primary purpose of our study was to examine and compare sensitivity to audio-visual asynchrony among infants at elevated likelihood of developing ASD and their typically developing (TD) counterparts.

## Multisensory integration and autism spectrum disorder

Across multiple studies using a variety of paradigms, children with ASD have shown impaired perception of audio-visual relations, especially for social events. Compared to TD children, for instance, who preferentially looked towards a synchronous (vs. asynchronous) audio-visual display of a woman speaking, children with ASD exhibited no clear preference in this regard, suggesting they may not have discriminated between the stimuli [11]. Consistent with this interpretation, older children with ASD have been less accurate than TD children in judging whether auditory and visual speech cues were temporally aligned [12, 13] and have exhibited larger temporal binding windows for audio-visual speech [14, 12, 13]. In general, research suggests that children with ASD are less sensitive to the temporal synchrony of auditory and visual cues within speech, a powerful source of intersensory redundancy for TD infants [5].

Importantly, when non-social events (e.g., hammer tapping a nail, bouncing ball) are depicted, the link between ASD and multisensory integration is less clear: some studies have demonstrated disparate performance in ASD, both enhanced and reduced [15–18], while in other studies, individuals with ASD perform relatively similarly to TD individuals [19, 20, 21, 22]. Of note, in TD individuals, the audio-visual temporal binding window is typically larger for social (e.g., speech) than non-social (e.g., flashes, beeps) events [7, 23–26], suggesting that sensory integration in these contexts are relatively distinct processes in general.

Relatedly, both children and adults with ASD exhibit difficulty perceiving the McGurk illusion [27], in which simultaneously presented (but incongruent) auditory and visual speech cues (e.g., visual “ga” and auditory “ba”) are fused to generate a novel, illusory percept (e.g., “da” or “tha”) [21, 28–30]. Thus, whereas TD individuals apparently integrate the incongruent auditory and visual speech cues in this situation, individuals with ASD do not, suggesting that they may generally rely more on auditory than visual cues to perceive multimodal events. Interestingly, and consistent with this interpretation, children with ASD are reportedly *more* susceptible to a visual flash-beep illusion, in which multiple beeps paired with a single flashing light produce an illusion of having seen multiple light flashes [17]. Thus, when the event is non-social in nature, children with ASD are not necessarily impaired in their capacity for multisensory integration and are possibly even more likely than TD children to do so.

Despite a wealth of research on this topic in children and adults [11, 17, 19, 12, 12, 14, 31, 32, 32], very few studies have examined multisensory processing in children younger than 24 months who are at elevated likelihood of developing ASD. Given the vast literature on multisensory integration in TD infants (see Lewkowicz, 2000, 2014 [4, 33] for reviews), this represents a notable gap in the empirical literature. By examining sensitivity to audio-visual asynchrony among both TD infants and those at elevated likelihood of developing ASD, the current study addresses the question of whether an extended audio-visual temporal binding window is present in infancy as a function of elevated likelihood of developing ASD and for both social and non-social events.

## Implications for language development

Examining these associations in infancy is critical because language develops rapidly across the second year of life and infants’ sensitivity to audio-visual temporal synchrony may contribute to the process. According to the intersensory redundancy hypothesis, multimodal events in which sensory cues are temporally synchronized are highly salient and serve to recruit selective attention and organize perceptual learning in early development [34, 35]. In particular, selective attention to the source of redundant sensory information is thought to allow for “intermodal learning” that captures the salient perceptual dimensions of the cultural world [36]. As a prime example, attending to the visibly moving lips of speaking face (a source of highly redundant sensory information) allows infants to perceptually integrate faces and voices into a coherent whole rather than as a series of disjointed inputs [4, 36], which may be critical for the perception of other information conveyed by faces, such as speaker identity, emotion, and language group [37, 38, 39].

More importantly, however, selective attention to the mouth region of a speaking face may be critical for understanding speech as an intentional action that can be performed by the self. Shortly after they begin to babble (8–10 months), TD infants begin attending more to the mouth (vs. eye) region of a speaking face [40–43], which may support their emerging ability to imitate lip movements associated with speech sounds. Additionally, the detection of audio-visual temporal synchrony is likely critical for establishing relations between spoken words and their visible referents [44, 45]. Thus, by promoting attention to the most important features of an event, an infant’s sensitivity to audio-visual synchrony may serve a foundational role in the acquisition of language, which is often delayed or disrupted in individuals with ASD [46, 47].

Although no previous study has examined whether sensitivity to audio-visual synchrony is associated with language development in infants at elevated likelihood of

developing ASD, there is some support for this idea. Stevenson et al. [48], for instance, found that the size of the audio-visual temporal binding window in children with ASD was related to their speech perception, and this relation was further mediated by their ability to integrate social stimuli like the McGurk Effect. Relatedly, Righi et al. [49] showed that the ability of preschool children with ASD to match synchronous speech with corresponding lip movements predicted both their expressive and receptive language abilities. Finally, Bahrack et al. [50] reported that accuracy of intersensory matching of faces and voices is positively associated with language competence in TD children; more recently, these findings have been extended to infants [51, 52]. Thus, considering that children diagnosed with ASD often exhibit delays in language [12, 14, 17, 30, 32, 48, 53], it is likely that infants at elevated risk for developing the disorder would exhibit impairments in this domain.

## Current study

In sum, the literature suggests that (1) the audio-visual temporal binding window for social stimuli is larger in children who are diagnosed with ASD, but that (2) there are relatively few studies examining the audio-visual temporal binding window in infants less than 24 months. Additionally, (3) although it has never been examined in infants at elevated likelihood of developing ASD, the audio-visual temporal binding window for speech likely impacts subsequent language production. Thus, the primary aim of the current study was to examine and compare the sensitivity to audio-visual synchrony of speech cues among infants at elevated likelihood of developing ASD and in their TD counterparts. Additionally, given the theoretical significance of audio-visual sensory integration for the development of expressive language, we also aimed to examine the associations between infants' sensitivity to audio-visual synchrony and language production.

To assess the size of infants' temporal binding windows, we used the habituation/dishabituation procedure, which is well-established in infants [54–56]. In general, this procedure utilizes looking time to measure infants' attention to a repeated stimulus and subsequent ability to discriminate the repeated stimulus from a novel stimulus. Research has indicated that the number of trials in which habituation occurs is indicative of stimulus encoding and that individual differences are related to later cognitive abilities, such as IQ [56]. Thus, consistent with previous studies [5], before presenting infants with the asynchronous test stimuli, we habituated them to the synchronous stimuli. A speaking face served as the social event and a bouncing ball served as the nonsocial event. Finally, to probe whether a significant relation between the audio-visual temporal binding window for a social

stimulus and language production exists, we assessed the size of infants' productive vocabulary between 17 and 30 months.

In general, we expected that infants at elevated likelihood of developing ASD would exhibit reduced sensitivity to audio-visual synchrony when viewing the social stimulus and hence, their temporal binding window for the speaking face stimulus is expected to be larger compared to TD infants. Given conflicting evidence about the performance of children with ASD when presented with non-social stimuli, we did not expect to find significant group differences in the size of the audio-visual temporal binding window for the nonsocial (bouncing ball) stimulus. Finally, we expect that TD infants will have larger vocabularies than infants at elevated likelihood of developing ASD and a significant positive relation between the social audio-visual temporal binding window and language production was expected for both groups.

## Method

### Participants

Two groups of infants between 4 and 24 months of age were tested. One group was comprised of 35 infants at elevated likelihood of developing ASD ( $M = 12.90$  months,  $SD = 5.49$ , 51% female) and the other of 53 TD infants ( $M = 10.60$  months,  $SD = 5.10$ , 51% female). Approximately half of the infants at elevated likelihood of developing ASD were between 1 and 2 years of age ( $N = 17$ ); the same number of TD infants were between 1 and 2 years of age ( $N = 18$ ). However, there were nearly twice as many TD infants less than 12 months ( $N = 35$ ) as there were infants at elevated likelihood of developing ASD. Thus, because the TD group was slightly younger than the group at elevated likelihood of developing ASD on average,  $t(69) = -2.00$ ,  $p = 0.05$ , age was considered as a factor in the analyses.

Consistent with previous studies, infants were considered at elevated likelihood of developing ASD if they had an older sibling with a confirmed diagnosis of ASD, were born < 36 weeks gestation, or had a birth weight < 2000 g [57–61]. The premature infants ranged between 27- and 36-week gestation and corrected gestational age was accounted for (by using the expected due date as the date of birth to calculate age at the time of the visit). TD infants had no family history of autism, were full-term at birth, had a birth weight of 2000 g or higher, and had a 5-min APGAR score of 7 or higher. All infants were healthy at the time of testing and had no recent history of eye or ear infection. We tested an additional 12 infants but did not include their data due to fussiness ( $n = 6$ ), parental interference ( $n = 2$ ), or equipment failure ( $n = 4$ ). When infants were between 17 and 30 months, their parents were recontacted to fill out a vocabulary assessment.

## Procedures

Infants completed two separate habituation/dishabituation procedures, one that presented a social event (speaking face) followed by one that presented a nonsocial event (bouncing ball); both were presented on a 24-inch Dell computer screen. Testing took place in a quiet, dimly lit room. Infants either sat in a child seat or on their caregiver's lap (about 50 cm from the computer screen); in the latter case, caregivers wore headphones that played white noise. Infants completed the speaking face procedure first as it was the primary measure of interest, followed by the bouncing ball procedure. Between procedures, infants were given a 10-min break during which they were taken out of the test room and encouraged to play with their caregiver. All procedures and materials were approved by the Institutional Review Board and informed consent was obtained prior to data collection.

The speaking face event (see Lewkowicz 2010, [9]) consisted of a woman wearing a neutral expression looking directly into the camera while producing the speech syllable /ba/. The woman opened her mouth, articulated the syllable /ba/, and then closed her mouth every 4 s. The woman's face spanned roughly 1/3 of the computer screen, subtending approximately 19° of visual angle in height and 28° of visual angle in width. An audible /ba/ was synchronous with her lip movements and was presented at 65 dB, A-scale. The bouncing ball event (see Lewkowicz 1996 [8], Minar and Lewkowicz 2018 [62]) consisted of a moving red ball that made an impact sound when it hit the upper/lower bounds of the computer screen. This stimulus was created in Adobe After Effects (Adobe Systems, San Jose, CA). The ball was 2 inches in diameter and subtended approximately 6° in visual angle in height and width. The ball moved at a rate of 10 cm/s (with a 50 ms pause at each endpoint) and was presented in front of a 12 × 16 grid of small white dots against a black background. The change in direction at the upper and lower bounds of the screen was synchronous with the sound of a wooden spoon hitting an empty plastic container, which was presented at 65 db, A-scale.

## Measures

### Habituation/dishabituation

Each habituation trial began when infants attended to the stimulus screen and ended when they disengaged from the screen for a period of 1 s, or until their look duration exceeded the maximum trial length of 60 s [38, 55, 63]. Infants were repeatedly shown the stimulus until the amount of looking between the first and last three habituation trials was decreased by 50%; once this occurred, infants were habituated to the stimulus [56]. To ensure the

appropriate number of habituation trials were administered to each infant, look duration was assessed live by trained coders using a peephole. Observers recorded fixation on an event recorder by observing when the infants' eyes were oriented towards the stimulus. For both event conditions, the number of trials required for the infant to achieve the habituation criterion was calculated.

Once infants were habituated, five test trials depicting the same events were presented at increasing levels of audio-visual asynchrony (333 ms, 500 ms, 666 ms, 833 ms, and 1000 ms), with the sound always preceding its corresponding visual event. The trial in which infants exhibited their longest look duration was taken to indicate the size of their audio-visual temporal binding window for each stimulus condition.

### Vocabulary Production

Toddlers' vocabulary production was assessed between 17 and 30 months using the Toddler form (part IA, "Words Children Use", 680-words) of the MacArthur-Bates Communicative Development Inventory (CDI), a standard vocabulary checklist suitable for children within this age range [64]. CDI data are available for 35 TD infants ( $M = 20.85$ ,  $SD = 3.39$ ) and 18 infants at elevated likelihood of developing ASD ( $M = 21.05$ ,  $SD = 3.35$ ).

## Results

### Habituation/dishabituation

All infants were successfully habituated to the stimuli as indicated by a 50% reduction in looking for both conditions. A paired samples *t*-test revealed that infants required a greater number of trials to achieve habituation criterion for bouncing ball ( $M = 8.93$ ) than for speaking face ( $M = 7.50$ ),  $t(67) = 4.13$ ,  $p < 0.01$ , but there were no other main effects of condition on habituation performance.

Group differences in study variables and correlations with age are displayed in Table 1. For speaking face, there is a significant main effect of group on the initial look duration during the pretest,  $F(1) = 6.47$ ,  $p < 0.05$ , such that TD infants looked significantly longer ( $M = 36$  s) than infants at elevated likelihood of developing ASD ( $M = 26$  s) on average. However, there are no significant group differences with respect to the looking during other phases of habituation or the number of trials required to achieve the habituation criterion. For bouncing all, there are no significant main effects of risk group on habituation performance.

For speaking face, age is significantly negatively correlated with the average look duration across the last three habituation trials and with the number of trials required to achieve habituation, indicating that older infants habituated faster and did not look as long at the stimulus by the end of the procedure as younger infants. For bouncing ball, age is significantly negatively correlated with the average look duration across the first three habituation trials but not with looking during the pretest or the last three habituation trials.

On average, infants’ maximum look during test trials for speaking face ( $M = 18.15$  s) was more than twice as long as their average look duration across the last 3 habituation trials (7.67 s), which represents a 137% increase in looking. A similar proportion was observed for bouncing ball (122% increase). A set of paired samples *t*-tests revealed that these differences in looking were significant across both groups and both conditions (Face/TD:  $M = 8.70$ ,  $t(48) = 7.04$ ,  $p < 0.01$ ; Face/ASD:  $M = 13.10$ ,  $t(32) = 5.20$ ,  $p < 0.01$ ; Ball/TD:  $M = 8.59$ ,  $t(47) = 5.54$ ,  $p < 0.01$ ; Ball/ASD:  $M = 11.10$ ,  $t(24) = 3.91$ ,  $p < 0.01$ ), further suggesting that infants were successfully dishabituated to the asynchronous test stimuli at some point. The proportions of infants in each group who looked longest during each test trial (300 ms, 500 ms, 666 ms, 833 ms, 1000 ms) are displayed in Figs. 1a and b. For the speaking face, most TD infants looked longest during the 500 ms trial whereas most of the infants at elevated likelihood of developing ASD looked longest during the 666 or 833 ms trials. The pattern for bouncing ball is less clear: around a third of infants in both groups looked longest during the 666 ms trial, but most of the at-risk infants looked longest during the 1000 ms trial. The millisecond asynchrony offset value of the trial containing the maximum look duration was conceptualized as the audiovisual temporal binding window.

Across groups, the temporal binding window is significantly smaller for speaking face than for bouncing ball,  $F(153) = 4.48$ ,  $p < 0.05$ ,  $\eta^2 = 0.03$ . Across conditions, the temporal binding window is significantly smaller for TD infants than infants at elevated likelihood of developing ASD,  $F(153) = 7.67$ ,  $p < 0.01$ ,  $\eta^2 = 0.04$ . There was no significant condition  $\times$  group interaction in this regard,  $F(65) = 0.01$ ,  $p = 0.92$ . However, separate independent samples *t*-tests were conducted since not all participants completed both tests. This analysis revealed that TD infants exhibited a significantly smaller temporal binding window than AR infants for the social event,  $t(80) = -2.68$ ,  $p = 0.01$ , Cohen’s  $D = 0.59$ . However, there were no significant group differences in this regard for bouncing ball. A set of paired samples *t*-tests further suggested that the main effect of the condition (Speaking Face < Bouncing Ball) was only significant among TD infants,  $t = -2.13$ ,  $p < 0.05$ . Among the at-risk infants, the size of the temporal binding window for Speaking Face was not significantly smaller than that of Bouncing Ball,  $t = -1.29$ ,  $p = 0.22$ . Thus, unlike TD infants, for at-risk infants, the size of the temporal binding window for a social event was not significantly smaller than it was for a nonsocial event (see Fig. 2).

**Sensitivity to audio-visual synchrony and language production**

Finally, to explore the implications of multisensory processing for language development, we examined whether the size of the audio-visual temporal binding window was associated with vocabulary production. First, because infant age at the time of language testing ( $M = 22.43$ ,  $SD = 6.04$ ) is significantly positively correlated with the CDI vocabulary score ( $R = 0.84$ ), a residualized score (controlling for infant age) was calculated; this variable is

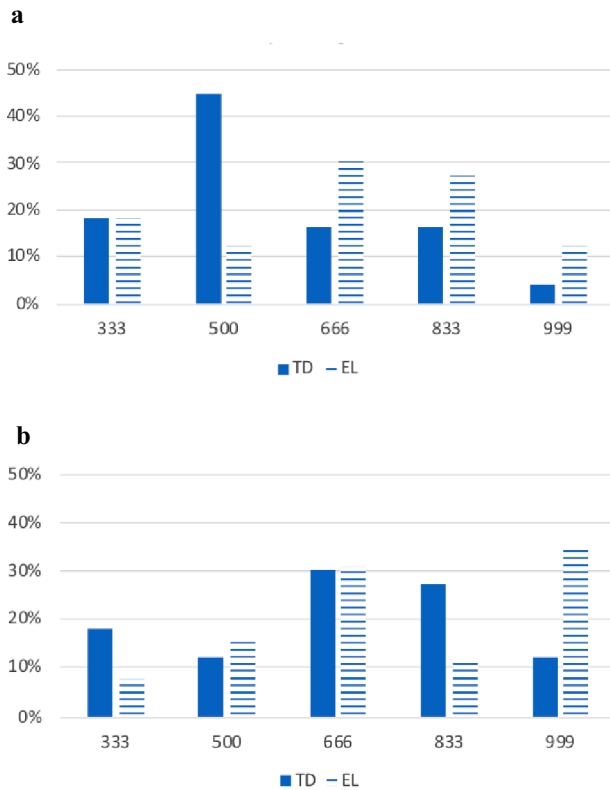
**Table 1** Mean look duration across habituation phases by risk group and condition

|                         | Speaking face |       |        | Bouncing ball |       |        |
|-------------------------|---------------|-------|--------|---------------|-------|--------|
|                         | TD            | EL    | R      | TD            | EL    | R      |
| Pretest                 | 36.01         | 26.24 | -.07   | 36.93         | 35.88 | .01    |
| Average look (hab123)   | 24.13         | 23.13 | -.15   | 24.84         | 22.45 | **-.36 |
| Average look (habxyz)   | 7.77          | 7.53  | **-.34 | 7.65          | 7.92  | -.20   |
| Habituation trials      | 7.47          | 7.48  | *-.24  | 8.98          | 8.85  | -.15   |
| Average look (test)     | 8.62          | 9.74  | .08    | 7.42          | 8.83  | -.04   |
| Maximal look (test)     | 16.47         | 20.63 | *.24   | 16.25         | 19.02 | -.07   |
| Temporal binding window | 575           | 671   | -.07   | 642           | 749   | .01    |
| <i>N</i>                | 49            | 33    | 82     | 48            | 26    | 74     |

*TD* typically developing, *EL* at elevated likelihood of developing ASD, *R* correlation with age, *pre* pretest, *hab123* first three habituation trials, *habxyz* last three habituation trials, *TBW* average offset asynchrony for maximal look

\*\* $p < 0.01$ , \* $p < 0.05$





Note: percentages are respective to each group; TD = typically developing; EL = elevated likelihood of developing ASD

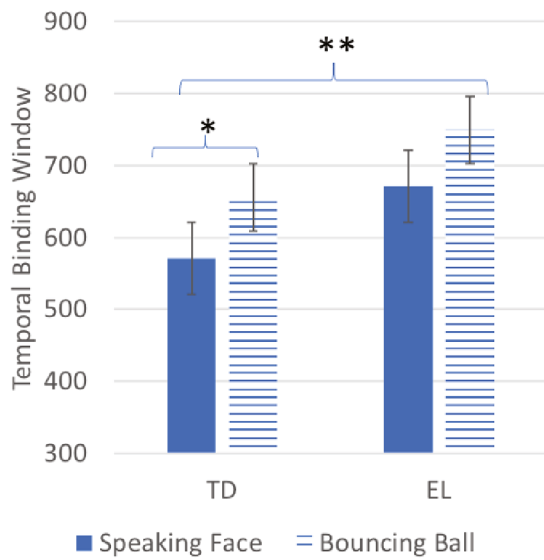
**Fig. 1** **a** Proportion of infants who looked longest during each test trial (speaking face). **b** Proportion of infants who looked longest during each test trial (bouncing ball)

normally distributed ( $M=0$ ,  $SD=116.59$ ,  $Range=636.36$ ,  $Skew=0.37$ ,  $Kurtosis=1.22$ ). Subsequently, the temporal binding window was modeled as a continuous predictor of vocabulary with the infant group (TD vs. elevated likelihood of developing ASD) as a categorical moderator; separate models were conducted for the Speaking Face and Bouncing Ball conditions.

The overall model for Speaking Face was significant,  $F(4, 48)=2.99$ ,  $p<0.05$ ,  $R^2=0.13$ . A main effect of the group was observed as well as a significant interaction between the group and the size of the temporal binding window. To probe the interaction, a simple slopes analysis was conducted. Among TD infants, the slope is not significant ( $p=0.34$ ), but among infants at elevated likelihood of developing ASD, the size of the temporal binding window is significantly positively associated with vocabulary ( $B=0.30$ ,  $p<0.05$ ; see Fig. 3). A similar interaction effect was observed for Bouncing Ball but the model did not exceed the significance threshold,  $F(4, 48)=2.16$ ,  $p=0.09$ .

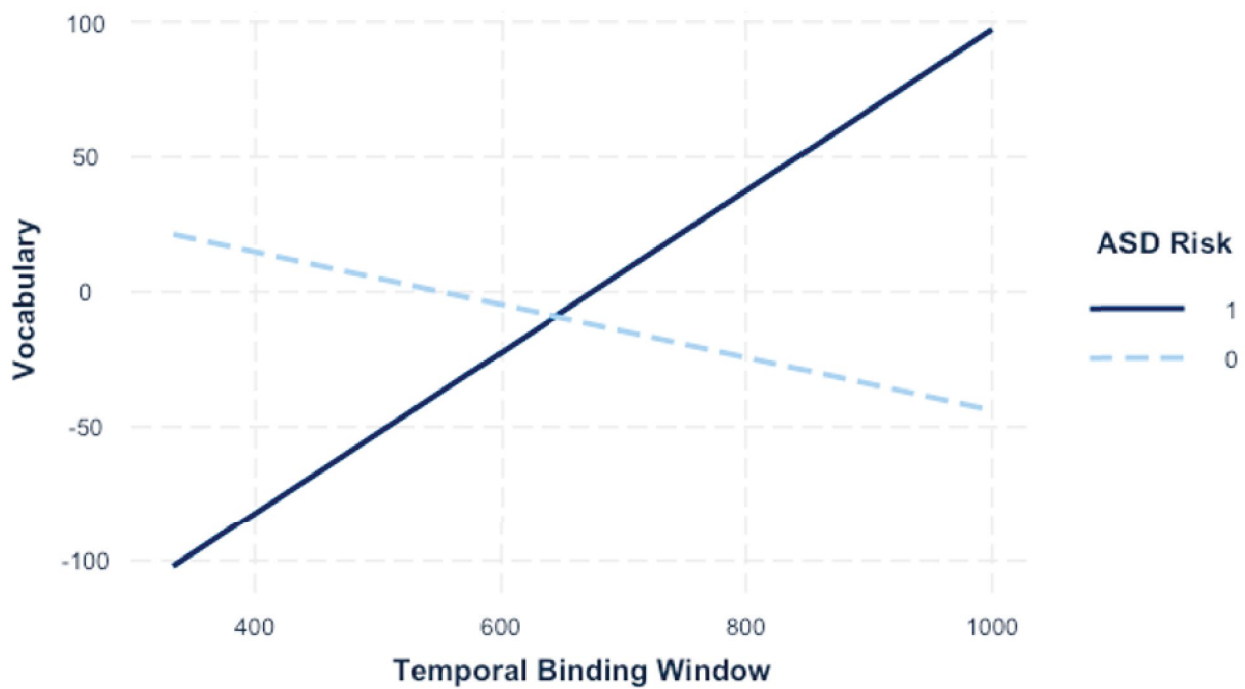
## Discussion

In early development, the perception of temporal synchrony within audio-visual speech is thought to support acquiring language and may be disrupted in infants at elevated likelihood of developing ASD [65]. Despite a large body of research on multisensory integration in TD infants (for reviews see Lewkowicz 2000 [4] or Bahrck and Lickliter 2012 [66]) and children diagnosed with ASD [67–71], few if any studies have investigated multisensory processing in infants at elevated likelihood of developing ASD. This work is crucial to identifying early markers of the disorder because features of atypical development consistent with a broader autism phenotype are often detectable before 1 year of age [59, 72, 73, 74]. Thus, filling a critical gap in the literature, the present study examined and compared sensitivity to temporal asynchrony within audio-visual events among infants at elevated likelihood of developing ASD with their TD counterparts.



Note: \*\*  $p < .01$ , \*  $p < .05$ ; TD = typically developing; EL = at elevated likelihood of developing ASD

**Fig. 2** Average temporal binding window size by group and condition. Note: \*\* $p < 0.01$ , \* $p < 0.05$ ; TD, typically developing; EL, at elevated likelihood of developing ASD



Note: vocabulary is residualized for age

**Fig. 3** Interaction of temporal binding window and infant risk group on vocabulary production. Note: vocabulary is residualized for age

In general, our primary hypothesis was supported: infants at elevated likelihood of developing ASD had a significantly larger audio-visual temporal binding window for a social event than TD infants, although, the size of the temporal binding window for the non-social event (Bouncing Ball) did not significantly differ between groups in this regard. This suggests that TD infants were more sensitive to the asynchrony between auditory and visual speech cues than infants at elevated likelihood of developing ASD, and is consistent with previous research showing larger audio-visual temporal binding windows for social, but not non-social, events in older children with ASD compared to TD children [17, 18, 30]. Additionally, for TD infants, the audio-visual temporal binding window for the social event was significantly smaller than for the non-social event whereas for infants at elevated likelihood of developing ASD, performance in these conditions was relatively similar. In general, these findings suggest that infants who are at elevated likelihood of developing ASD may process multisensory information for social events less efficiently than TD infants [10, 75, 76].

Given language delays are common among children with ASD, we expected that the size of the audiovisual temporal binding window (especially for the social event) would help explain variation in early productive vocabulary. However, our hypothesis was not supported. There was no significant relation between the temporal binding window for Speaking Face and language for TD infants and for ASD infants, a significant positive association was observed such that a larger productive vocabulary was observed among those with a wider temporal binding window for Speaking Face. This is not consistent with the idea that a smaller temporal binding window reflects greater sensitivity to audiovisual asynchrony and a better ability to integrate auditory and visual cues. However, it may suggest that infants at elevated likelihood of developing ASD are relying on different strategies for acquiring language. The size of the temporal binding window for the non-social event (Bouncing Ball) was not significantly associated with language production for either group.

In general, our findings are consistent with prior literature in support of making a distinction between social and nonsocial information processing. Indeed, a number of studies have identified perceptual abnormalities in children with ASD during audio-visual tasks involving human faces and voices but not during tasks involving nonhuman stimuli when compared to their TD counterparts [11, 77–79], [21]). For instance, relative to TD children those with ASD have significantly more difficulty visually orienting to social stimuli such as their name being called, but only slightly more difficulty orienting to non-social stimuli such a musical toy [80]. Perceptual difficulties with social stimuli in individuals with ASD are further highlighted by their impaired performance on the McGurk illusion [27, 81–83]. Individuals with

ASD perceive the McGurk illusion less often than their peers without ASD, often relying instead on the auditory modality to the exclusion of the visual information [21, 12, 32]. Thus, our findings are in agreement with previous research conducted with older children (already diagnosed with ASD) in suggesting a specific deficit in multisensory integration for social information and further suggest this deficit is already present in infancy.

Importantly, our findings are also consistent with Smith et al. [31] and a number of eye-tracking studies [40, 41, 84–86, 87, 88] in which TD children selectively attended to the lip movements of talking faces while ASD children showed reduced attention to this region [89–91]. That is, one potential reason why the infants at elevated likelihood of developing ASD in our study took longer to become dis-habituated in the speaking face condition is because they were not looking at the mouth. By focusing attention to the speaker's mouth, infants may be able to understand the action of speech in relation to their own body; it may allow them to integrate what they are hearing with what they are seeing in order to reproduce the action themselves (i.e., imitation). Although no previous study has examined whether enhanced attention to a speaker's mouth in infancy is associated with greater vocal imitation, it has been shown to predict greater language production at 24 months [43]. More recently, Habayeb et al. [92] reported that mouth-looking in 1- to 2-year-old infants was significantly associated with greater expressive language and that infants at elevated likelihood of developing ASD looked significantly less at the mouth than TD infants. Thus, regardless of why, if infants at elevated likelihood of developing ASD do not selectively attend to the mouth region of a speaking face to the same degree as TD infants, their expressive language development may be impaired.

Neurobiological processes may contribute to the impairments in audio-visual speech processing in ASD. Some studies, for instance, have suggested that the left temporal cortex fails to become specialized for speech processing in individuals with ASD [93], but how this might relate specifically to infants' audio-visual temporal synchrony is unclear. The mirror neuron system (MNS), in addition, has been implicated in language development [94, 95] and is thought to be disrupted in autism [96]. In this system, neurons in the sensorimotor cortex that fire when an action is performed also fire when that same action is observed (i.e., performed by another person). In this way, the MNS is thought to be important for action understanding and “self-other mapping” [97] or the “translation of seeing and hearing into doing” [98]. In developmental EEG studies, greater MNS activation during action observation in infants is associated with better subsequent imitation of the action [99, 100]. Considering that speech is an action that can be observed and imitated, it seems possible that infants could exhibit MNS activation



during audio-visual speech and that it could play a role in the development of expressive language.

Although empirical work on the MNS in infants has largely focused on the perception of manual, object-directed actions (e.g., tapping a block), there is increasing evidence that the MNS is responsive to other types of actions, including communicative gestures (e.g., pointing) and facial movements [101–103]. Given the evidence that TD infants begin attending more to the mouth region of a speaking face towards the end of the first year [41], and that greater attention to the mouth region at this time is associated with better language development [43, 104], it seems highly possible that MNS function plays a role in this process. Whether MNS activation during audio-visual speech is a cause or consequence of heightened visual attention to the speaker's lips is unclear, but it may be disrupted in infants at elevated likelihood of developing ASD. Thus, an important direction for future research involves replication of the present study in conjunction with eye-tracking and neuroimaging methods (e.g., EEG).

## Conclusion

This study is not without its limitations. In addition to the fact that we did not incorporate eye-tracking technology into the assessments, we had a fairly heterogeneous group of infants at elevated likelihood of developing ASD, some having been born premature or low birthweight and others with older siblings that had a diagnosis. Although each of these criteria have been associated with elevated likelihood of developing ASD, they may involve very different etiological pathways. Thus, that we were unable to control for the type of risk factor in our analyses is a limitation. A related limitation is that we were unable to provide information about which infants were ultimately diagnosed with ASD. Finally, although the effect of age on vocabulary was accounted for in our statistical analyses, it is a limitation of the study that vocabulary was not assessed at the same age for all participants.

In summary, results of the current study suggest that the early characteristics of ASD in infants at elevated likelihood of developing the condition also include sensory integration difficulties, specifically with regard to the capacity for audio-visual synchrony. While this notion is still speculative, our findings contribute to a growing body of literature indicating that sub-clinical autistic behaviors may be present in children who might not yet fulfill all the clinical criteria for an ASD diagnosis. Although additional research is needed to understand the link between audio-visual sensory integration and language development in both TD and at-risk infants, this study represents an important first step towards

understanding the nature of attention deficits that contribute to ASD and further suggests that problems in multisensory integration may be present in infants at elevated likelihood of developing ASD long before a clinical diagnosis is usually made.

**Authors' contributions** Suri: writing and data curation; Whedon: writing, review and editing, and data analysis; Lewis: conceptualization, methodology, resources, and funding acquisition.

**Funding** This research was supported by grants (#CAU-T15APL012 #CAUT20APL003) awarded to the Michael Lewis from the New Jersey Governor's Council for Medical Research and Treatment of Autism.

**Availability of data and material** The datasets during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval and consent to participate** All study procedures and materials were approved by the Institutional Review Board at Rutgers (protocol number Pro20150001814).

**Consent for publication** Informed consent was obtained from parents prior to data collection.

**Competing interests** The authors declare no competing interests.

## References

- Gibson JJ (1966) *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston
- Marks LE (1978) *The unity of the senses: Interrelations among the modalities*. Academic Press
- Bahrack LE, Lickliter R (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 36(2):190–201. <https://doi.org/10.1037/0012-1649.36.2.190>
- Lewkowicz DJ (2000) The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull* 126(2):281–308. <https://doi.org/10.1037/0033-2909.126.2.281>
- Lewkowicz DJ (2010) Infant perception of audio-visual speech synchrony. *Dev Psychol* 46(1):66–77. <https://doi.org/10.1037/a0015579>
- Lewkowicz DJ (2000) Infants' perception of the audible visible and bimodal attributes of multimodal syllables. *Child Dev* 71(5):1241–1257. <https://doi.org/10.1111/1467-8624.00226>
- Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9(6):719–721. <https://doi.org/10.1068/p090719>
- Lewkowicz DJ (1996) Perception of auditory–visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform* 22(5):1094–1106. <https://doi.org/10.1037/0096-1523.22.5.1094>
- Lewkowicz DJ, Flom R (2014) The audiovisual temporal binding window narrows in early childhood. *Child Dev* 85(2):685–694. <https://doi.org/10.1111/cdev.12142>

10. Bahrick LE, Todd JT (2012) Multisensory processing in autism spectrum disorders: intersensory processing disturbance as a basis for atypical development. *The New Handbook of Multisensory Processes*, xx–xx
11. Bebko JM, Weiss JA, Demark JL, Gomez P (2006) Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *J Child Psychol Psychiatry* 47(1):88–98. <https://doi.org/10.1111/j.1469-7610.2005.01443.x>
12. Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT (2014) Multisensory temporal integration in autism spectrum disorders. *J Neurosci* 34(3):691–697. <https://doi.org/10.1523/JNEUROSCI.3615-13.2014>
13. de Boer-Schellekens L, Eussen M, Vroomen J (2013) Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Front Integr Neurosci* 7. <https://doi.org/10.3389/fnint.2013.00008>
14. Stevenson RA, Segers M, Ferber S, Barense MD, Camarata S, Wallace MT (2016) Keeping time in the brain: autism spectrum disorder and audiovisual temporal processing. *Autism Res* 9(7):720–738. <https://doi.org/10.1002/aur.1566>
15. Bertone A, Mottron L, Jelenic P, Faubert J (2005) Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain* 128(10):2430–2441. <https://doi.org/10.1093/brain/awh561>
16. Collignon O, Charbonneau G, Peters F, Nassim M, Lassonde M, Lepore F, Mottron L, Bertone A (2013) Reduced multisensory facilitation in persons with autism. *Cortex* 49(6):1704–1710. <https://doi.org/10.1016/j.cortex.2012.06.001>
17. Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res* 203(2):381–389. <https://doi.org/10.1007/s00221-010-2240-4>
18. Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT (2011) Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front Integrat Neurosci* 4:1–11. <https://doi.org/10.3389/fnint.2010.00129>
19. de Boer-Schellekens L, Eussen M, Vroomen J (2012) Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Front Integrat Neurosci* 7:1–8. <https://doi.org/10.3389/fnint.2013.00008>
20. Magnée MJCM, De Gelder B, Van Engeland H, Kemner C (2008) Audiovisual speech integration in pervasive developmental disorder: evidence from event-related potentials. *J Child Psychol Psychiatry* 49(9):995–1000. <https://doi.org/10.1111/j.1469-7610.2008.01902.x>
21. Mongillo EA, Irwin JR, Whalen DH, Klaiman C, Carter AS, Schultz RT (2008) Audiovisual processing in children with and without autism spectrum disorders. *J Autism Dev Disord* 38(7):1349–1358. <https://doi.org/10.1007/s10803-007-0521-y>
22. Van Der Smagt MJ, Van Engeland H, Kemner C (2007) Brief report: Can you see what is not there? Low-level auditory-visual integration in autism spectrum disorder. *J Autism Dev Disord* 37(10):2014–2019. <https://doi.org/10.1007/s10803-006-0346-0>
23. Conrey B, Pisoni DB (2006) Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America* 119(6):4065–4073. <https://doi.org/10.1121/1.2195091>
24. Mégevand P, Molholm S, Nayak A, Foxe JJ (2013) Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS ONE* 8(8). <https://doi.org/10.1371/journal.pone.0071608>
25. Powers AR, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. *In J Neurosci* (Vol. 29, Issue 39). <https://doi.org/10.1523/JNEUROSCI.3501-09.2009>
26. Stevenson RA, Wallace MT (2013) Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res* 227(2):249–261. <https://doi.org/10.1007/s00221-013-3507-3>
27. McGurk H, MacDonald J (1976) Hearing lips and seeing voices (McGurk Effect). *Nature* 264(5588):746–748
28. Saalasti S, Kätsyri J, Tiippana K, Laine-Hernandez M, Von Wendt L, Sams M (2012) Audiovisual speech perception and eye gaze behavior of adults with asperger syndrome. *J Autism Dev Disord* 42(8):1606–1615. <https://doi.org/10.1007/s10803-011-1400-0>
29. Taylor N, Isaac C, Milne E (2010) A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *J Autism Dev Disord* 40(11):1403–1411. <https://doi.org/10.1007/s10803-010-1000-4>
30. Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT (2013) Multisensory speech perception in children with autism spectrum disorders. *J Autism Dev Disord* 43(12):2891–2902. <https://doi.org/10.1007/s10803-013-1836-5>
31. Smith E, Zhang S, Bennetto L (2017) Temporal synchrony and audiovisual integration of speech and object stimuli in autism. *Research in Autism Spectrum Disorders* 39(March):11–19. <https://doi.org/10.1016/j.rasd.2017.04.001>
32. Stevenson RA, Segers M, Ferber S, Barense MD, Wallace MT (2014) The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Front Psychol* 5(MAY):1–4. <https://doi.org/10.3389/fpsyg.2014.00379>
33. Lewkowicz DJ (2014) Early experience and multisensory perceptual narrowing. *Dev Psychobiol* 56(2):292–315. <https://doi.org/10.1002/dev.21197>
34. Bahrick LE, Lickliter R (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 36(2):190–201. <https://doi.org/10.1037/0012-1649.36.2.190>
35. Bahrick LE, Lickliter R (2002) Intersensory redundancy guides early perceptual and cognitive development. *Adv Child Dev Behav* 30:153–187. [https://doi.org/10.1016/s0065-2407\(02\)80041-6](https://doi.org/10.1016/s0065-2407(02)80041-6)
36. Bahrick LE, Lickliter R, Flom R (2004) Intersensory redundancy guides the development of selective attention perception and cognition in infancy. *Curr Dir Psychol Sci* 13(3):99–102. <https://doi.org/10.1111/j.0963-7214.2004.00283.x>
37. Bahrick LE, McNew ME, Pruden SM, Castellanos I (2019) Intersensory redundancy promotes infant detection of prosody in infant-directed speech. *J Exp Child Psychol*:183295–183309. <https://doi.org/10.1016/j.jecp.2019.02.008>
38. Lewkowicz DJ, Minar NJ, Tift AH, Brandon M (2015) Perception of the multisensory coherence of fluent audiovisual speech in infancy: its emergence and the role of experience. *J Exp Child Psychol* 130:147–162. <https://doi.org/10.1016/j.jecp.2014.10.006>
39. Vaillant-Molina M, Bahrick LE (2012) The role of intersensory redundancy in the emergence of social referencing in 5½-month-old infants. *Dev Psychol* 48(1):1–9. <https://doi.org/10.1037/a0025263>
40. Hillairet de Boisferon A, Tift AH, Minar NJ, Lewkowicz DJ (2017) Selective attention to a talker’s mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Dev Sci*. <https://doi.org/10.1111/desc.12381>
41. Lewkowicz DJ, Hansen-Tift AM (2012) Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc Natl Acad Sci USA* 109(5):1431–1436. <https://doi.org/10.1073/pnas.1114783109>
42. Pons F, Bosch L, Lewkowicz DJ (2015) Bilingualism modulates infants’ selective attention to the mouth of a talking face.

- Psychol Sci 26(4):490–498. <https://doi.org/10.1177/0956797614568320>
43. Tenenbaum EJ, Sobel DM, Sheinkopf SJ, Malle BF, Morgan JL (2015) Attention to the mouth and gaze following in infancy predict language development. *J Child Lang* 42(6):1173–1190. <https://doi.org/10.1017/S0305000914000725>
  44. Lakshmi J., Gogate Lorraine E., Bahrack (1998) Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *J Exp Child Psychol* 69(2) 133–149. <https://doi.org/10.1006/jecp.1998.2438>
  45. Gogate LJ, Bahrack LE. Intersensory redundancy and 7-month-old Infants' memory for arbitrary syllable-object relations. *Infancy* 2(2):219–231. [https://doi.org/10.1207/S15327078IN0202\\_7](https://doi.org/10.1207/S15327078IN0202_7)
  46. Luyster RJ, Kadlec MB, Carter A, Tager-Flusberg H (2008) Language assessment and development in toddlers with autism Spectrum disorders. *J Autism Dev Disord* 38(8):1426–1438. <https://doi.org/10.1007/s10803-007-0510-1>
  47. Tager-Flusberg H. On the nature of a language acquisition disorder: the example of autism. In: *The development of language and language researchers*. Psychology Press, pp 261–280
  48. Stevenson RA, Segers M, Ncube BL, Black KR, Bebko JM, Ferber S, Barense MD (2018) The cascading influence of multisensory processing on speech perception in autism. *Autism* 22(5):609–624. <https://doi.org/10.1177/1362361317704413>
  49. Righi G, Tenenbaum EJ, McCormick C, Blossom M, Amsó D, Sheinkopf SJ (2018) Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Res* 11(4):645–653. <https://doi.org/10.1002/aur.1918>
  50. Bahrack LE, Todd JT, Soska KC (2018) The multisensory attention assessment protocol (MAAP): characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *Dev Psychol* 54(12):2207–2225. <https://doi.org/10.1037/dev0000594>
  51. Elizabeth V, Torrence EJ, Todd LE. Bahrack Intersensory processing of faces and voices at 6 months predicts language outcomes at 18 24 and 36 months of age. *Infancy:infa.12533*. <https://doi.org/10.1111/infa.12533>
  52. Edgar EV, Todd JT, Bahrack LE (2022) Intersensory matching of faces and voices in infancy predicts language outcomes in young children. *Dev Psychol* 58(8):1413–1428. <https://doi.org/10.1037/dev0001375>
  53. Brandwein AB, Foxe JJ, Butler JS, Frey HP, Bates JC, Shulman LH, Molholm S (2014) Neurophysiological indices of atypical auditory processing and multisensory integration are associated with symptom severity in autism. *J Autism Dev Disord* 45(1):230–244. <https://doi.org/10.1007/s10803-014-2212-9>
  54. Brooks-Gunn J, Lewis M (1981) Infant social perception: responses to pictures of parents and strangers. *Dev Psychol* 17(5):647–649. <https://doi.org/10.1037/0012-1649.17.5.647>
  55. Cohen LB (1969) Observing responses, visual preferences, and habituation to visual stimuli in infants. *J Exp Child Psychol*. [https://doi.org/10.1016/0022-0965\(69\)90004-6](https://doi.org/10.1016/0022-0965(69)90004-6)
  56. Lewis M, Goldberg S, Campbell H (1969) A Developmental study of information processing within the first three years of life: response decrement to a redundant signal. *Monogr Soc Res Child Dev*. <https://doi.org/10.2307/1165696>
  57. Grønberg TK, Schendel DE, Parner ET (2013) Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. *JAMA Pediatr* 167(10):947–953. <https://doi.org/10.1001/jamapediatrics.2013.2259>
  58. Limperopoulos C, Bassan H, Sullivan NR, Soul JS, Robertson RL, Moore M, Ringer SA, Volpe JJ, Plessis AJD (2008) Positive screening for autism in ex-preterm infants: prevalence and risk factors. *Pediatrics* 121(4):758–765. <https://doi.org/10.1542/peds.2007-2158>
  59. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, Bryson S, Carver LJ, Constantino JN, Dobkins K, Hutman T, Iverson JM, Landa R, Rogers SJ, Sigman M, Stone WL (2011) Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics* 128(3). <https://doi.org/10.1542/peds.2010-2825>
  60. Pinto-Martin JA, Levy SE, Feldman JF, Lorenz JM, Paneth N, Whitaker AH (2011) Prevalence of autism spectrum disorder in adolescents born weighing <2000 grams. *Pediatrics* 128(5):883–891. <https://doi.org/10.1542/peds.2010-2846>
  61. Schendel D, Bhasin TK (2008) Birth weight and gestational age characteristics of children with autism, including a comparison with other developmental disabilities. *Pediatrics* 121(6):1155–1164. <https://doi.org/10.1542/peds.2007-1049>
  62. Minar NJ, Lewkowicz DJ (2018) Overcoming the other-race effect in infancy with multisensory redundancy: 10-12-month-olds discriminate dynamic other-race faces producing speech. *Dev Sci* 21(4):e12604. <https://doi.org/10.1111/desc.12604>
  63. Möhring W, Liu R, Libertus ME (2017) Infants' speed discrimination: effects of different ratios and spatial orientations. *Infancy* 22(6):762–777. <https://doi.org/10.1111/infa.12196>
  64. Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, Tomasello M, Mervis CB, Stiles J (1994) Variability in early communicative development. *Monogr Soc Res Child Dev*. <https://doi.org/10.2307/1166093>
  65. Bahrack LE (2010) Intermodal perception and selective attention to Intersensory redundancy: implications for typical social development and autism. *The Wiley-Blackwell Handbook of Infant Development* 1:120–166
  66. Bahrack LE, Lickliter R (2012) The role of intersensory redundancy in early perceptual, cognitive, and social development. In: Bremner AJ, Lewkowicz DJ, Spence C (eds) *Multisensory development*. Oxford University Press, pp 183–206. <https://doi.org/10.1093/acprof:oso/9780199586059.003.0008>
  67. Feldman JI, Kuang W, Conrad JG, Tu A, Santapuram P, Simon DM, Foss-Feig JH, Kwakye LD, Stevenson RA, Wallace MT, Woynarowski TG (2019) Brief report: differences in multisensory integration Covary with sensory responsiveness in children with and without autism Spectrum disorder. *J Autism Dev Disord* 49(1):397–403. <https://doi.org/10.1007/s10803-018-3667-x>
  68. Stevenson RA, Baum SH, Segers M, Ferber S, Barense MD, Wallace MT (2017) Multisensory speech perception in autism spectrum disorder: from phoneme to whole-word perception. *Autism Res* 10(7):1280–1290. <https://doi.org/10.1002/aur.1776>
  69. Baum SH, Stevenson RA, Wallace MT (2015) Behavioral perceptual and neural alterations in sensory and multisensory function in autism spectrum disorder. *Prog Neurobiol*:134140–134160. <https://doi.org/10.1016/j.pneurobio.2015.09.007>
  70. Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynarowski TG, Camarata SM, Wallace MT (2014) Multisensory temporal integration in autism Spectrum disorders. *J Neurosci* 34(3):691–697. <https://doi.org/10.1523/JNEUROSCI.3615-13.2014>
  71. Beker S, Foxe JJ, Molholm S (2018) Ripe for solution: delayed development of multisensory processing in autism and its remediation. *Neurosci Biobehav Rev*:84182–84192. <https://doi.org/10.1016/j.neubiorev.2017.11.008>
  72. Elsabbagh M, Gliga T, Pickles A, Hudry K, Charman T, Johnson MH (2013) The development of face orienting mechanisms in infants at-risk for autism. *Behav Brain Res* 251:147–154. <https://doi.org/10.1016/j.bbr.2012.07.030>
  73. Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, Rogers SJ, Rozga A, Sangha S, Sigman M, Steinfeld MB,



- Young GS (2010) A prospective study of the emergence of early behavioral signs of autism. *J Am Acad Child Adolesc Psychiatry* 49(3):256–266.e2. <https://doi.org/10.1016/j.jaac.2009.11.009>
74. Ozonoff S, Young GS, Belding A, Hill M, Hill A, Hutman T, Johnson S, Miller M, Rogers SJ, Schwichtenberg AJ, Steinfeld M, Iosif AM (2014) The broader autism phenotype in infancy: when does it emerge? *J Am Acad Child Adolesc Psychiatry* 53(4):398–407.e2. <https://doi.org/10.1016/j.jaac.2013.12.020>
  75. Brandwein AB, Foxe JJ, Russo NN, Altschuler TS, Gomes H, Molholm S (2011) The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cereb Cortex* 21(5):1042–1055. <https://doi.org/10.1093/cercor/bhq170>
  76. Falck-Ytter T, Von Hofsten C, Gillberg C, Fernell E (2013) Visualization and analysis of eye movement data from children with typical and atypical development. *J Autism Dev Disord* 43(10):2249–2258. <https://doi.org/10.1007/s10803-013-1776-0>
  77. Dawson G, Meltzoff AN, Osterling J, Rinaldi J, Brown E (1998) Children with autism fail to orient to naturally occurring social stimuli. *J Autism Dev Disord* 28(6):479–485. <https://doi.org/10.1023/A:1026043926488>
  78. Irwin JR, Tornatore LA, Brancazio L, Whalen DH (2011) Can children with autism spectrum disorders “hear” a speaking face? *Child Dev* 82(5):1397–1403. <https://doi.org/10.1111/j.1467-8624.2011.01619.x>
  79. Klin A (1991) Young autistic children’s listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *J Autism Dev Disord*. <https://doi.org/10.1007/BF02206995>
  80. Dawson G, Webb SJ, McPartland J (2005) Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. In *Developmental Neuropsychology*. [https://doi.org/10.1207/s15326942dn2703\\_6](https://doi.org/10.1207/s15326942dn2703_6)
  81. Bahrick LE, Walker AS, Neisser U, Walker-Andrews AS, Van Atteveldt N, Murray MM, Thut G, Schroeder CE, Bahrick LE, Lickliter R, McNew ME, Pruden SM, Castellanos I, Todd JT, Soska KC, Lickliter R, Flom R, Bebko JM, Weiss JA, Amso D (2014) The Multisensory Attention Assessment Protocol (MAAP): characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *J Autism Dev Dis* 39(1):233–241. <https://doi.org/10.1177/1084713807307409>
  82. Bebko JM, Schroeder JH, Weiss JA (2014) The McGurk effect in children with autism and asperger syndrome. *Autism Res* 7(1):50–59. <https://doi.org/10.1002/aur.1343>
  83. de Gelder B, Vroomer J, Van der Heide L (1991) Face recognition and lip-reading in autism. *Eur J Cogn Psychol* 3(1):69–86. <https://doi.org/10.1080/09541449108406220>
  84. Chawarska K, MacAri S, Shic F (2012) Context modulates attention to social scenes in toddlers with autism. *J Child Psychol Psychiatry* 53(8):903–913. <https://doi.org/10.1111/j.1469-7610.2012.02538.x>
  85. Grossman RB, Schneps MH, Tager-Flusberg H (2009) Slipped lips: onset asynchrony detection of auditory-visual language in autism. *J Child Psychol Psychiatry* 50(4):491–497. <https://doi.org/10.1111/j.1469-7610.2008.02002.x>
  86. Rice K, Moriuchi JM, Jones W, Klin A (2012) Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *J Am Acad Child Adolesc Psychiatry* 51(3):238–248. <https://doi.org/10.1016/j.jaac.2011.12.017>
  87. Smith EG, Bennetto L (2007) Audiovisual speech integration and lipreading in autism. *J Child Psychol Psychiatry* 48(8):813–821. <https://doi.org/10.1111/j.1469-7610.2007.01766.x>
  88. Sterling L, Dawson G, Webb S, Murias M, Munson J, Panagiotides H, Aylward E (2008) The role of face familiarity in eye tracking of faces by individuals with autism spectrum disorders. *J Autism Dev Disord* 38(9):1666–1675. <https://doi.org/10.1007/s10803-008-0550-1>
  89. Jones W, Klin A (2013) Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*. <https://doi.org/10.1038/nature12715>
  90. Norbury CF, Brock J, Cragg L, Einav S, Griffiths H, Nation K (2009) Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *J Child Psychol Psychiatry* 50(7):834–842. <https://doi.org/10.1111/j.1469-7610.2009.02073.x>
  91. Tenenbaum EJ, Amso D, Abar B, Sheinkopf SJ (2014) Attention and word learning in autistic, language delayed, and typically developing children. *Front Psychol* 5(MAY):1–9. <https://doi.org/10.3389/fpsyg.2014.00490>
  92. Habayeb S, Tsang T, Saulnier C, Klaiman C, Jones W, Klin A, Edwards LA (2021) Visual traces of language acquisition in toddlers with autism spectrum disorder during the second year of life. *J Autism Dev Disord* 51(7):2519–2530. <https://doi.org/10.1007/s10803-020-04730-x>
  93. Eyster LT, Pierce K, Courchesne E (2012) A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain* 135(3):949–960. <https://doi.org/10.1093/brain/awr364>
  94. Salo VC, Ferrari PF, Fox NA (2018) The role of the motor system in action understanding and communication: evidence from human infants and non-human primates. *Dev Psychobiol* 61(3):390–401. <https://doi.org/10.1002/dev.21779>
  95. Rizzolatti G, Arbib MA (1998) Language within our grasp. *Trends Neurosci* 21(5):188–194. [https://doi.org/10.1016/S0166-2236\(98\)01260-0](https://doi.org/10.1016/S0166-2236(98)01260-0)
  96. Oberman LM, Hubbard EM, McCleery JP, Altschuler EL, Ramachandran VS, Pineda JA (2005) EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cogn Brain Res* 24(2):190–198. <https://doi.org/10.1016/j.cogbrainres.2005.01.014>
  97. Marshall PJ, Meltzoff AN (2014) Neural mirroring mechanisms and imitation in human infants. *Philos Trans R Soc B Biol Sci* 369(1644):20130620. <https://doi.org/10.1098/rstb.2013.0620>
  98. Pineda JA (2005) The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”. *Brain Res Rev* 50(1):57–68. <https://doi.org/10.1016/j.brainresrev.2005.04.005>
  99. Warreyn P, Ruyschaert L, Wiersema JR, Handl A, Pattyn G, Roeyers H (2013) Infants’ mu suppression during the observation of real and mimicked goal-directed actions. *Dev Sci* 16(2):173–185. <https://doi.org/10.1111/desc.12014>
  100. Filippi CA, Cannon EN, Fox NA, Thorpe SG, Ferrari PF, Woodward AL (2016) Motor system activation predicts goal imitation in 7-month-old infants. *Psychol Sci* 27(5):675–684. <https://doi.org/10.1177/09567976166632231>
  101. Ferrari PF, Vanderwert RE, Paukner A, Bower S, Suomi SJ, Fox NA (2012) (2012) distinct EEG amplitude suppression to facial gestures as evidence for a Mirror mechanism in newborn monkeys. *J Cogn Neurosci* 24(5):1165–1172. [https://doi.org/10.1162/jocn\\_a\\_00198](https://doi.org/10.1162/jocn_a_00198)
  102. Quandt LC, Marshall PJ, Shipley TF, Beilock SL, Goldin-Meadow S (2012) Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: an EEG study of action production and gesture observation. *Neuropsychologia* 50(12):2745–2751. <https://doi.org/10.1016/j.neuropsychologia.2012.08.005>

103. Rayson H, Bonaiuto JJ, Ferrari PF, Murray L (2016) Mu desynchronization during observation and execution of facial expressions in 30-month-old children. *Dev Cogn Neurosci*:19279–19287. <https://doi.org/10.1016/j.dcn.2016.05.003>
104. Tsang T, Atagi N, Johnson SP (2018) Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *J Exp Child Psychol*:16993–16109. <https://doi.org/10.1016/j.jecp.2018.01.002>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.