Infant Behavior and Development 76 (2024) 101973

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Sensitivity to temporal synchrony and selective attention in audiovisual speech in infants at elevated likelihood for autism: A preliminary longitudinal study

Itziar Lozano [a,b,*,1], Mercedes Belinchón [a,2], Ruth Campos [a,3]

[a] Department of Basic Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain
[b] Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Key-words:
Audiovisual speech perception
Autism
Infant siblings
Selective attention
Temporal synchrony
Longitudinal

ABSTRACT

Autism Spectrum Disorder is a highly heritable condition characterized by sociocommunicative difficulties, frequently entailing language atypicalities that extend to infants with a familial history of autism. The developmental mechanisms underlying these difficulties remain unknown. Detecting temporal synchrony between the lip movements and the auditory speech of a talking face and selectively attending to the mouth support typical early language acquisition. This preliminary eye-tracking study investigated whether these two fundamental mechanisms atypically function in infant siblings. We longitudinally tracked the trajectories of infants at elevated and low-likelihood for autism in these two abilities at 4, 8, and 12 months ($n = 29$). We presented two talking faces (synchronous and asynchronous) while recording infants' gaze to the talker's eyes and mouth. We found that infants detected temporal asynchronies in talking faces at 12 months regardless of group. However, compared to their typically developing peers, infants with an elevated likelihood of autism showed reduced attention to the mouth at the end of the first year and no variations in their interest to this area across time. Our findings provide preliminary evidence on a potentially atypical trajectory of reduced mouth-looking in audiovisual speech during the first year in infant siblings, with potential cascading consequences for language development, thus contributing to domain-general accounts of emerging autism.

1. Introduction

Autism Spectrum Disorder (ASD) is an early-onset neurodevelopmental condition. Although it is primarily defined based on sociocommunicative deficits, and restricted and repetitive behaviours (American Psychiatric Association, 2013), it often entails language impairments. Not all individuals with ASD show language difficulties (e.g., Howlin, 2003) and, when present, are highly diverse

---

Abbreviations: ADOS-2, Autism Diagnostic Observational Schedule – second edition; AOI, Area of Interest; ASD, Autism Spectrum Disorder; EL-infants, Infants at elevated likelihood for autism; *IIH*, *Intersensory Impairment Hypothesis*; LL-infants, infants with a low likelihood of developing this condition; LEAT, Language Exposure Assessment Tool; MSEL, Mullen Scales of Early Learning; PTLT, Proportion of Total Looking Time.

* Correspondence to: Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Jaracza, 1, 00–378 Warsaw, Poland.
E-mail addresses: ilozano@psych.pan.pl (I. Lozano), mercedes.belinchon@gmail.com (M. Belinchón), ruth.campos@uam.es (R. Campos).

[1] ORCID: 0000–0003-4280–4747

[2] ORCID: 0000–0001-5367–279X

[3] ORCID: 0000–0003-0209–6002

https://doi.org/10.1016/j.infbeh.2024.101973
Received 31 January 2024; Received in revised form 13 June 2024; Accepted 14 June 2024
Available online 27 June 2024
0163-6383/© 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

(Charman et al., 2003), but language atypicalities often occur across the lifespan (e.g., Miniscalco et al., 2012). For instance, language delay is one of the first concerns of parents of toddlers with ASD (Herlihy et al., 2015), who indeed show reduced speech vocalizations at 18–24 months (Plumb & Wetherby, 2013). Despite its frequency, the developmental mechanisms underlying language atypicalities in ASD remain unknown.

1.1. *Toward Targeting Early Candidate Mechanisms*

A specific hypothesis has been proposed to account for the early development of atypicalities in language and other sociocommunicative skills observed in individuals with ASD. The *Intersensory Impairment Hypothesis* (*IIH*; Bahrick, 2010; Bahrick & Todd, 2012) posits that reduced sensitivity to detect temporal asynchronies in intersensory redundancy would cascade into deficits in audiovisual integration and, ultimately, in the social-communication impairments core of ASD. ‘Redundancy’ here refers to the simultaneous availability and temporal synchronization of the same information across two or more sensory modalities (Bahrick & Lickliter, 2014) and, more specifically, to the visual and auditory modalities. Crucially, atypical temporal synchrony detection would also account for subclinical atypicalities linked to autism, such as those in language acquisition. From this approach, a further candidate early mechanism that may account for sociocommunicative and language atypicalities in ASD would be selective attention (Bahrick, 2010). Selective attention is the ability to focus on specific elements or events while ignoring others (Bahrick & Lickliter, 2014). In the context of audiovisual speech, it refers to preferential attention toward different facial features (i.e., the eyes and mouth). An early disruption in detecting temporal synchrony would affect how redundancy guides and directs infants’ selective attention to relevant events, thus losing both the perceptual benefits of perceiving in a unified way and the sociocommunicative developmental gains (Bahrick, 2010).

Increasing evidence supports that atypical detection of temporal asynchronies in audiovisual events in individuals with ASD precedes and accounts for impairments in the audiovisual integration of these same events (Feldman et al., 2018; Stevenson et al., 2018). Atypical temporal asynchrony detection is not limited to social events (see, e.g., Beker et al., 2018) — or audiovisual modalities (see Murat Baldwin et al., 2022 for a review) — but has mainly been observed when processing audiovisual speech (Stevenson et al., 2014). Individuals with ASD show atypicalities in detecting temporal asynchronies in talking faces uttering speech, from childhood and across the lifespan (Bebko et al., 2006; Beker et al., 2018). Furthermore, these difficulties concurrently relate to lower vocabulary skills in children with ASD (Righi et al., 2018; Todd & Bahrick, 2022). Selective attention to the speaker’s articulating mouth is also affected in ASD. Toddlers with ASD look less than typically developing controls at the face and mouth regions only when a talker is speaking (Shic et al., 2020). Reduced mouth-looking predicts weaker audiovisual speech integration in children with ASD (Feng et al., 2021), and increasing their attention to this area by cueing it improves audiovisual speech integration (Feng et al., 2022). Although reduced selective attention to the mouth may underlie atypical temporal synchrony detection in ASD, supporting evidence remains preliminary (Grossman et al., 2015). As anticipated by the *IIH,* these two perceptual and attentional mechanisms seem to atypically function in ASD, which may reduce the learnability of audiovisual speech and impact language development (e.g., Righi et al., 2018). However, the very early emergence in life of atypicalities in these two mechanisms in ASD remains underexplored, as they are likely to start during the first year of life while the condition is not reliably diagnosed until age 3.

1.2. *Prospective infant siblings’ design approach*

ASD is a highly heritable condition (Tick et al., 2016). Given this high heritability, the language atypicalities linked to the condition usually extend to first-degree relatives of diagnosed individuals. Infants at elevated likelihood for autism (EL-infants, henceforth) present an increased likelihood for developing the condition themselves, with a recurrence risk of ~20 %, compared to ~1.5 % in the general typically developing population (Ozonoff et al., 2011). This familial likelihood contributes to developmental atypicalities in EL-infants even in those with no ASD outcome. For example, McDonald et al. (2020) found that, among unaffected EL-infants, infants from multiplex risk families (who have more than one sibling diagnosed with ASD) show several lower cognitive abilities than infants from single-incidence families (who have one single diagnosed sibling with ASD). This finding suggests that multiplex family seem to show the strongest atypicalities in infant siblings.

Crucially, regardless of ASD likelihood status, EL-infants commonly show early atypicalities in language development (Belteki et al., 2022; Hudry et al., 2014), including lower consonant production, fewer speech-like vocalizations (Paul et al., 2011), and delayed babbling, production, and comprehension onset (Iverson & Wozniak, 2007). Earlier susceptibilities may also be present in this population in the mechanisms typically underlying language development.

Prospective longitudinal studies with EL-infants offer a unique opportunity to explore the underlying developmental mechanisms atypically operating early in life before observable language atypicalities emerge. More specifically, EL-infants allow longitudinally tracking the trajectories of sensitivity to temporal synchrony and selective attention — hypothesized as atypical by the *IIH* — from the very early months of development, when we know these mechanisms emerge in typical development. Although evidence is mixed (see Cox et al., 2022 for a meta-analysis), overall, it suggests that the timing of the typical trajectories of changes of both mechanisms overlap in development across the first postnatal year. Increased mouth-looking in speaking faces is observable at the same time-point in development as infants start to detect asynchronies in audiovisual speech (i.e., 8 months; Hillairet de Boisferon et al., 2017; Pons & Lewkowicz, 2014, Experiment 1). Both mechanisms are driven by infants’ level of expertise in a language during the first year, as they enter perceptual narrowing with audiovisual speech and become more attuned to their phonological native categories (Lewkowicz & Hansen-Tift, 2012; Pons & Lewkowicz, 2014; Werker & Tees, 1984).

From domain-general approaches to ASD, the first year of life is a crucial period for the emergence of early atypicalities in EL-infants, including perceptual, attentional, or motor (Campos et al., 2019; Elsabbagh & Johnson, 2016). We argue that the linguistic difficulties observed in individuals with ASD at different stages of life and in EL-infants may be expressed earlier as atypicalities in the *mechanisms* underlying language acquisition. Given the functional role of sensitivity to temporal synchrony (Edgar et al., 2022) and selective attention (Bastianello et al., 2022) in efficiently processing audiovisual speech during the first year of life in typically developing infants, and their predictive role in later language outcomes (Edgar et al., 2023; Lozano et al., 2022; Viktorsson et al., 2023; Young et al., 2009), here we propose to investigate them as two candidate atypical mechanisms in EL-infants as a whole group. The underlying reasoning is that the familial likelihood is shared by EL-infants and therefore these two developmental mechanisms may be affected even in those not diagnosed with ASD (see, e.g., Szatmari et al., 2016 for a similar approach).

If EL-infants show a reduced detection of temporal synchrony compared to typically developing infants with a low likelihood of developing this condition (LL-infants, henceforth), this may constrain their efficient processing of their language experiences. As audiovisual *fluent* speech is a highly complex, unpredictable, and dynamic signal, a diminished sensitivity to temporal synchrony could prevent EL-infants from reducing the perceptual uncertainty of this event, ultimately impacting their ability to perceive visual and auditory speech as a unitary event (as found by Guiraud et al., 2012). If EL-infants show atypical selective attention to audiovisual speech in the first postnatal year, this may lead them to reduced access to the relevant visual cues on which language acquisition typically relies.

To our knowledge, only one published study has tested whether EL-infants show atypical sensitivity to temporal synchrony in audiovisual speech (Suri et al., 2023). This work found that EL-infants aged 4 to 24 months old showed less sensitivity to detect audiovisual asynchronies in a talking face uttering syllables than LL-infants, but no group differences in non-social events (a bouncing ball). Furthermore, higher sensitivity to temporal asynchronies during the first two years of life predicted larger vocabulary production in toddlerhood only in LL-infants. These results support that atypical sensitivity to temporal synchrony in audiovisual speech is a plausible candidate mechanism shared by EL-infants as a whole group. However, the cross-sectional design used and the wide heterogeneity of the ages tested made it difficult to establish the onset of the emergence of atypical sensitivity to temporal synchrony in EL-infants. Moreover, the type of stimuli used —syllables— prevents generalizing these findings to audiovisual *fluent* speech, a more ecological event closer to infants' natural social experiences. Selective attention in audiovisual speech also seems to not functionally work in EL-infants. Although EL-infants aged 6–18 months do not differ from LL-infants in their pattern of selective attention to the articulating mouth (Santapuram et al., 2022), increased attention to the mouth at the end of the first year only supports later language outcomes in LL-infants (Chawarska et al., 2022). However, these studies cross-sectionally but not longitudinally tracked this mechanism, which may have hindered targeting the onset of atypicalities in the earliest key periods of the first year.

1.3. The current preliminary study

The aims of the current preliminary study were twofold. First, we investigated whether EL-infants differed from LL-infants in their sensitivity to detect temporal synchrony in audiovisual speaking faces during the first year. We longitudinally tracked these abilities at 4, 8, and 12 months to align with the timing of typical perceptual narrowing for audiovisual speech (Lewkowicz & Hansen-Tift, 2012) and the emergence of domain-general atypicalities in EL-infants (Campos et al., 2019; Elsabbagh & Johnson, 2016). Based on the *IIH* (Bahrick & Todd, 2012), we hypothesize that while LL-infants would discriminate temporal synchrony in audiovisual speech during the first year, EL-infants would show a reduced sensitivity to this cue across this period. Previous limited research on audiovisual *fluent* speech suggests that infants start to discriminate temporal asynchronies at 8 months (Pons & Lewkowicz, 2014, Experiment 1), later than for audiovisual syllables (at 4 months; Lewkowicz, 2010, Experiment 1). This may seem surprisingly late, as one may argue that audiovisual fluent speech offers richer cues than syllables, like head prosody and multiple visual articulators. However, silences in audiovisual syllables are more systematic and predictable than in audiovisual fluent speech, potentially helping infants to detect temporal asynchrony. Therefore, we predict significant group differences between EL-infants and LL-infants in their ability to detect temporal asynchrony in audiovisual fluent speech at 8 and 12 months, but not at 4 months. We expect that LL-infants would detect temporal synchronies at 8 and 12 months, but not at 4 months. Due to the lack of prior studies on EL-infants regarding this ability in audiovisual *fluent* speech, it is challenging to make specific predictions regarding the onset of discrimination of temporal asynchrony in this group — and, therefore, the beginning of group differences. Based on domain-general accounts for ASD (Campos et al., 2019; Elsabbagh & Johnson, 2016), we anticipate that EL-infants would show less sensitivity to this perceptual cue than typically developing infants during the first year.

Second, the current study tested whether EL-infants differ from LL-infants in their pattern of selective attention to the mouth versus the eyes of a talking face in the first year. Testing this research question entailed investigating two subquestions longitudinally. First, we tested the trajectory of changes in selective attention to the mouth relative to the eyes of a talking face within each group across the first year (i.e., whether LL-infants and EL-infants change their pattern of selective attention to audiovisual speech across the first year). Second, we tested potential group differences across time-points (i.e., whether the trajectory of selective attention to the eyes and mouth in audiovisual speech followed by LL-infants changes during the first year of life differently than in EL-infants). For testing each prediction, we will use an absolute (proportion of total looking time to the mouth; PTLT Mouth) and a relative measure of preference (PTLT Eyes-Mouth; eyes preference *minus* mouth-preference, where positive scores mean greater eyes-looking while negative scores mean greater mouth-looking). Importantly, these metrics assume that an increase in infants' eyes-looking does not necessarily penalize an increase in mouth-looking and vice versa (see more on this in Section 2.4.).

Beginning with our predictions of within-group changes over time-points, building upon cross-sectional and longitudinal research in typically developing infants (e.g., Lewkowicz & Hansen-Tift, 2012; Lozano et al., 2022; Pons et al., 2015; Tsang et al., 2018), we hypothesize that, when measuring absolute preference to the mouth (PTLT Mouth), LL-infants will show an increase in attention to this facial area across the first year, reflected by significantly greater attention to the mouth between 4 and 8 months (as in Lewkowicz & Hansen-Tift, 2012) and between 8 and 12 months (reflecting a continuous increase in mouth-preference at the second half of the first year, as in previous longitudinal studies; e.g., Tsang et al., 2018; Lozano et al., 2022). When measuring relative preference to the mouth (PTLT Eyes-Mouth), LL-infants will show a *u*-shaped pattern of changes across the first year — more looking at the eyes at 4 months (reflected by positive scores at this time-point), followed by a shift to the mouth at 8 months (reflected by significant differences in this measure between 4 and 8 months, that will go from positive to negative scores), and an equivalent preference to the eyes and mouth at 12 months (reflected by significant differences on PTLT Eyes-Mouth between 8 and 12 months, and scores close to 0, which will indicate both the onset of the attentional shift back to the eyes and the trajectory of increased absolute mouth-looking predicted above between 8 and 12 months) — . In contrast, we expect EL-infants to maintain constant mouth-preference during the first year, thus not varying their attention to the mouth over time, indicated by no significant differences in absolute or relative mouth preference between time-points (4 and 8 months, nor 8 and 12 months).

Finally, regarding our prediction on potential group-differences across time-points, based on domain-general accounts for the emergence of ASD (Campos et al., 2019; Gliga et al., 2014), we hypothesize that EL-infants will show reduced preference for the mouth of a talking face compared to LL-infants at 12 months, but not earlier (i.e., at 4 and 8 months). This will be indicated by EL-infants showing less preference for the mouth than LL-infants at 12 months, but not at 4 nor at 8 months (i.e., in absolute mouth-preference, this will be indicated by EL-infants showing significantly lower scores in PTLT Mouth than LL-infants; in relative mouth-preference, by EL-infants showing significantly greater scores in PTLT Eyes-Mouth than LL-infants). This means that we expect that the between-group differences in mouth-preference at 12 months will be driven by the LL-infants having increased their preference to the mouth towards the end of the first year, while the EL-infants would have not.

2. Methods

2.1. Participants

The data included in this paper are part of a prospective longitudinal study (TRABERITEA Project), in which infant siblings of children with autism were assessed in a larger protocol and followed up into early toddlerhood. This project investigated the developmental trajectories on several relevant human abilities (face, music, auditory speech, and audiovisual speech processing) and

Table 1

Descriptive statistics (mean, standard deviation, and range) of final sample participant characteristics and group comparisons on age, sex and general development.

Elevated likelihood for ASD (*N*=14; 8 males)				**Low likelihood for ASD (*N*=15; 9 males)**			**Group comparisons**[a]
	4 m (*n* = 11) *M* (*SD*) Range	8 m (*n* = 13) *M* (*SD*) Range	12 m (*n* = 10) *M* (*SD*) Range	4 m (*n* = 14) *M* (*SD*) Range	8 m[b] (*n* = 13) *M* (*SD*) Range	12 m (*n* = 13) *M* (*SD*) Range	[c] *t*-test/[d] Mann-Whitney (*p* value)
Age (days)	133.63 (13.05) 107-159	250.15 (11.83) 223-270	368.10 (9.13) 348-378	134.64 (11.67) 109-155	249.57 (7.41) 235-260	375.30 (9.05) 356-388	[c] $p = .84$; $p = .87$; $p = .07$
MSEL_VR	42.18 (4.87) 34-47	51.00 (7.63) 37-64	53.30 (9.46) 41-74	49.14 (6.38) 34-58	52.23 (10.00) 37-80	58.76 (6.66) 47-74	[d] $p = .005^{**}$; $p = .97$; $p = .06$
MSEL_FM	46.09 (6.13) 35-54	53.00 (12.78) 32-68	65.00 (8.48) 55-77	44.85 (4.58) 35-54	55.76 (7.36) 38-68	62.30 (7.53) 49-74	[d] $p = .56$; $p = .83$; $p = .50$
MSEL_GM	47.27 (8.87) 37-65	43.69 (12.09) 23-65	46.70 (14.06) 31-73	47.78 (8.84) 30-61	44.92 (7.75) 30-56	37.38 (9.56) 20-53	[c] $p = .88$; $p = .76$; $p = .07$
MSEL_RL	46.90 (10.80) 28-66	40.53 (9.28) 23-61	48.00 (8.81) 35-64	46.57 (12.47) 20-66	49.00 (5.14) 38-56	46.92 (6.88) 35-60	[d] $p = 1.0$; $p = .004^{**}$; $p = .87$
MSEL_EL	48.18 (5.05) 42-55	46.38 (6.04) 33-55	49.30 (10.79) 37-66	51.50 (6.58) 34-55	46.76 (7.35) 33-60	50.61 (7.50) 40-62	[d] $p = .07$; $p = .81$; $p = .63$
MSEL_ELC	91.81 (8.19) 77-104	95.53 (11.58) 70-111	107.70 (12.96) 89-127	96.07 (8.44) 82-108	102.38 (7.01) 93-117	109.38 (8.71) 97-122	[d] $p = .21$; $p = .08$; $p = .71$

Notes: m= months; MSEL: Mullen Scales of Early Learning. VR: Visual Receptive; FM: Fine Motor; GM: Gross Motor; RL: Receptive Language; EL: Expressive Language; ELC: Early Learning Composite scores.

$^{*}p \leq .05$, $^{**}p \leq .01$

[a] Independent samples *t*-test were conducted in variables that met assumptions of normality and homogeneity of variance, while Mann-Whitney was used when one of them or any of them were met.

[b] Notice sample size in this measure slightly differs from total sample size due to missing values and that one participant in the LL-infants did not complete MSEL.

parent-child interactions.

Infants in the EL-group had at least one older sibling with ASD by the time they joined our study, whose diagnosis was verified by either inspection of prior medical records or via assessment with Autism Diagnostic Observation Schedule-2 by clinician experts of our team (ADOS-2; Lord et al., 2015). They were recruited through our website and associations of individuals with ASD and their families at the Community of Madrid and surrounding cities under a collaboration agreement (ALANDA, AMITEA Program of Gregorio Marañón University Hospital, and other entities belonging to The Autism Spain Confederation-AE). Infants from the LL-group had at least one older sibling without a diagnosis of ASD or other developmental conditions. They were recruited via incidental sampling by our research team members, word of mouth, advertisements, and our project's website.

Thirty-one infants were initially recruited to participate in the current study at three time-points: 4, 8, and 12 months. Out of those, one participant of the LL-group was excluded for having an older sibling under clinical assessment due to a possible behavioural disorder. Further, one participant in the EL-group was excluded due to a change in the diagnosis of his older sibling. Two participants from the EL-group were twins (one female and one male).

The final sample after exclusions consisted of 29 infants (14 EL-infants; 15 LL-infants; 58.6 % males; see Table 1 for final samples) that contributed with useable eye-tracking data. Our final sample size is consistent with previous research with infant siblings investigating early mechanisms of language acquisition, ranging from 8 to 264 EL-infants and 9 to 104 LL-infants (see Morrel et al., 2023 for a systematic review). Particularly, it is in line with sample sizes in similar longitudinal eye-tracking studies that densely follow the first year of life, which ranged from 11 to 47 EL-infants and 16 to 25 LL-infants (e.g., Droucker et al., 2013; Jones & Klin, 2013; Thorup et al., 2016). Not all infants contributed to the three time-points; hence, there were missing values. Among the tested infants, 4 joined the study at 8 months ($n = 3$ EL-infants and $n = 1$ LL-infants), thus not contributing to measures at 4 months. Further, 2 infants missed visits in-between ($n = 1$ EL-infant and $n = 1$ LL-infants). In addition, one LL-infant only contributed to 4 months, and 4 EL-infants did not contribute to data at 12 months because they were not this age yet when we had to finish this study. A further reason for missing values was not contributing enough valid eye-tracking data (see data loss for each measure in Supplementary Analysis 1 and Table A.1).

According to parent report, all infants had White ethnicity, were born at term (>36 weeks gestation age), and had no hearing or visual impairments or any neurodevelopmental or genetic disorder. They were Spanish monolingual (i.e., mostly exposed to Spanish from birth), according to (1) parent report of overall infants' language experience (monolingual or bilingual) and (2) parents' systematic interview to quantify language exposure with an in-house adapted version of the Language Exposure Assessment Tool (LEAT; De Anda et al., 2016). This study was approved by the Ethics Committee of the Universidad Autónoma de Madrid as part of the TRABERITEA project and conducted in accordance with APA ethical standards in the treatment of human study sample. All families gave informed consent to take part at each time-point. All infants received a diploma and a toy at the end of each visit for their participation. Data collection was pre-pandemic (2016–2019). Data are not publicly accessible, but code and materials will be made available upon reasonable request to the authors.

We tested whether groups differed in global developmental level (see Table 1) with an in-house Spanish translation of the Mullen Scales of Early Learning (MSEL; Mullen, 1995) administered by trained TRABERITEA team members. Groups did not differ significantly in general cognitive development measured by the Early Learning Composite (ELC) scores at any time-point (all *ps* >.08). However, Mann-Whitney U tests indicated higher scores in LL-infants than EL-infants in the Visual Receptive subscale at 4 months, $U = 27.00$, $p = .005$, and in the Receptive Language subscale at 8 months, $U = 30.00$, $p = .004$. There were no significant group differences in sex (LL-infants: 9 males vs. 6 females; EL-infants: 8 males vs. 6 females; $\chi 2 = 0.24$; $p = .87$), or age (all $ps > .07$).

2.2. Eye-tracking procedure

At all time-points, the infants participated in a bigger protocol involving other eye-tracking tasks, infant-parent interactions, parent-reported questionnaires, and standardized developmental scales. We counterbalanced the order of task presentation across participants. Gaze data were collected using a Tobii TX300 eye tracker with a 23″monitor (1920 × 1080 pixels), a sampling rate of 300

Fig. 1. Screenshot of exemplar trial of the experimental task.

Hz, and 0.5°monocular accuracy. The testing session started with a five-point infant-friendly calibration routine. The experiment began after infants were successfully calibrated in at least 4 points per eye (and, at minimum, one at the top, one at the bottom, and one central). During the task, infants sat on their parents' lap, about 65 cm from the screen. Tobii Studio 3.4.5 software was used for stimuli presentation and eye-gaze recording. Infants' behaviour during the task was monitored and recorded using a hidden camera.

An audiovisual preferential looking paradigm was used. In each trial, two speaking faces of the same female were simultaneously presented at both sides of the screen as two displays (see Fig. 1 and an example clip in the Supplementary Materials): 'Synchronous'— where auditory speech and the visual articulatory lip movements were temporarily aligned — and 'Asynchronous' — where auditory and visual speech were temporally misaligned. In the Asynchronous display, the auditory stream was planned to precede the visual one by 666 ms. to ensure that our 4-month-old infants already perceived this level of misalignment (as it occurs, e.g., in syllables; Lewkowicz, 2010). However, due to a technical error, the level of asynchrony across trials was 500 and 666 ms. (see full trials' specifications in Table A.3), with 59.6 % trials of 500 ms and 40.4 % trials of 666 ms. Crucially, this did not impact our main results (see Supplementary Analysis 2). We counterbalanced the presentation side of the synchronous display across trials and participants by limiting consecutive same-sided to two, maintaining an equal trial number of synchronous and asynchronous videos on each side. The trial presentation order was fixed to keep the story's meaning coherent.

Two dynamic 5-sec audiovisual attentional getters were shown in the centre of the screen, one at the beginning of the task to ensure infants' initial attention and one at the end to control for fatigue effects. Further, 1-sec animations were presented between trials to redirect infants' attention before each trial's onset. Each trial started and ended with a completed word uttered to avoid giving additional cues to infants. The entire eye-tracking procedure lasted ~2 min. Videos were edited in Premiere Pro CS6, exported at a rate of 23 frames per second, and converted into .AVI files.

2.3. *Stimuli*

Stimuli consisted of ten 10-sec-videos of a female native Spanish actress uttering a story about a crew of animals (see full transcription in Table A.14). To maximize the closeness of stimuli to infants' experiences, we instructed the actress to use auditory and visual infant-directed speech (i.e., higher pitch and prosodic exaggeration, direct gaze, exaggerated eyes and lip movements; Kitamura et al., 2014). We recorded the videos without a script so the actress performed naturally.

2.4. *Data pre-processing*

Eye-tracking data was pre-processed using Tobii Studio software 3.4.5. For each trial and talking face (synchronous and asynchronous), we created four dynamic Areas of Interest (AOIs; see Figures A.1 and A.2): redundancy region (rectangular shape of maximum dimension 847 × 736 pixels), eyes region (rectangular shape, 335 × 185 pixels), mouth region (oval shape, 335 × 185 pixels), and face region (oval shape, 447 × 595 pixels). The medium region between the two talking faces (421 × 736 pixels) was excluded from the analyses.

After extracting AOIs, the Total Visit Duration (without zeros) was computed for each. We decided to use this measure for being less sensitive than others (e.g., Total Fixation Durations) to individual differences in data quality. Next, we calculated three dependent measures of preference for each trial, participant, and time-point. First, the Percentage of Total Looking Time (%) to the Synchronous and Asynchronous displays, calculated as % of total looking time to each display relative to % of total looking time to both displays collapsed. Second, the Proportion of Total Looking Time for the mouth (PTLT Mouth), calculated as the total looking time to the mouth relative to the whole face (as in Hillairet de Boisferon et al., 2017). Third, a differential measure of preference to the eyes relative to the mouth (PTLT Eyes-Mouth), calculated as the total looking time to the eyes relative to the whole face (PTLT Eyes) *minus* the total looking time to the mouth relative to the whole face (PTLT Mouth), as in Morin-Lessard et al. (2019). Note that the whole face included looking at the eyes and mouth. We included these latter two measures of preference for being complementary but different. While PTLT Mouth reflects absolute selective attention to this region, PTLT Eyes-Mouth captures the attentional competition between these areas when infants explore talking faces. Of note, by using separate metrics of preference (PTLT Mouth and PTLT Eyes) we do not assume that reduced mouth-looking necessarily implies increased eyes-looking, as some studies taking this approach have shown that preferences for these regions can change independently in development (i.e., Lozano et al., 2022). Furthermore, unlike other differential measures of preference (e.g., the EMI-eyes-mouth index; the mean amount of preference to the eyes, relative to both the eyes and mouth) PTLT Eyes-Mouth captures that infants may preferentially attend to other facial regions (e.g., the chin or nose; Constantino et al., 2017; Viktorsson et al., 2023) rather than the eyes and mouth when exploring talking faces.

We excluded infants with valid gaze sample below 25 % (according to Tobii Studio) *and* not contributing with at least 2 valid trials out of 10 (as in Falck-Ytter et al., 2018). We also excluded trials in which (1) the participant was fussy, crying, or the caregiver talked, (2) total looking time at the screen was below 15 % of the trial length (as in Kleberg et al., 2019), or (3) there was bias side (i.e., infants looked only at one screen side during the trial).

Watching infants' looking behaviour offline during the task revealed several periods of eye-tracking data loss while infants were actually attentive to the screen, possibly due to suboptimal calibrations or infants' head or body movements. Thus, we manually coded data on Percentage of Total Looking Time (%) measure to maximize the data available for analyses, since recruiting and testing EL-infants is highly difficult and costly (Jones et al., 2019) — see a full report in Supplementary Analysis 3. After coding, we applied the same exclusion criteria above to ensure good data quality, which did not significantly differ between groups (Table A.1).

2.5. Statistical analyses

All analyses were run in IBM SPSS Statistics 27 (IBM Inc., 2020). We used linear mixed-effect models (LMMs) to analyse our whole longitudinal dataset despite having missing values (Field, 2018). All analyses included the following predictors (see Tables A.4-A.6): Time-point (4 months vs. 8 months vs. 12 months), Group (EL-infants vs. LL-infants) and Redundancy (Synchronous vs. Asynchronous), and all possible two-way and three-way interactions. We kept all predictors fixed and simultaneously entered them. Time-point and Redundancy were entered as within-subject factors and Group as a between-subject factor. We specified random intercepts for subjects to control for individual variability and used REML (Restricted Maximum Likelihood) as the model estimator, since not all measures met the assumptions of normality (significant Shapiro-Wilk's test and inspection of Q-Q plots). The outcome variables differed between analyses: Percentage of Total Looking Time (in %), PTLT Mouth, or PTLT Eyes-Mouth. Alpha level was .05 (two--tailed) throughout.

3. Results

3.1. Sensitivity to temporal asynchrony

We ran a 3 × 2× 2 LMM on the Percentage of Total Looking Time (%), with Time-point (4 months vs. 8 months vs. 12 months) and Redundancy (Synchronous vs. Asynchronous) as within-subject predictors and Group (EL-infants vs. LL-infants) as between-subject predictors (see full descriptive statistics in Table A.4 and full results in Table A.7). There was a significant Time-point x Redundancy interaction, $F(2, 138) = 4.5$, $p = .01$, $r = 0.19$, indicating that, regardless of group, infants' pattern of preference for the synchronous and asynchronous displays differed across time-points (see Figs. 2a and 2b). Follow-up analyses (Bonferroni-corrected post hoc tests) showed that infants preferred looking at the synchronous display over the asynchronous one at 12 months, $F(1, 138) = 13.78$, $p < .001$, but equivalently looked at these two displays at 4 months, $F(1, 138) = .07$, $p = .79$, and 8 months, $F(1, 138) = .33$, $p = .56$ (see Figure A.3). Our results indicate that, regardless of group, we did not find evidence of infants discriminating temporal asynchronies before 12 months. No other main effects or interactions were found (all $ps \geq .14$).

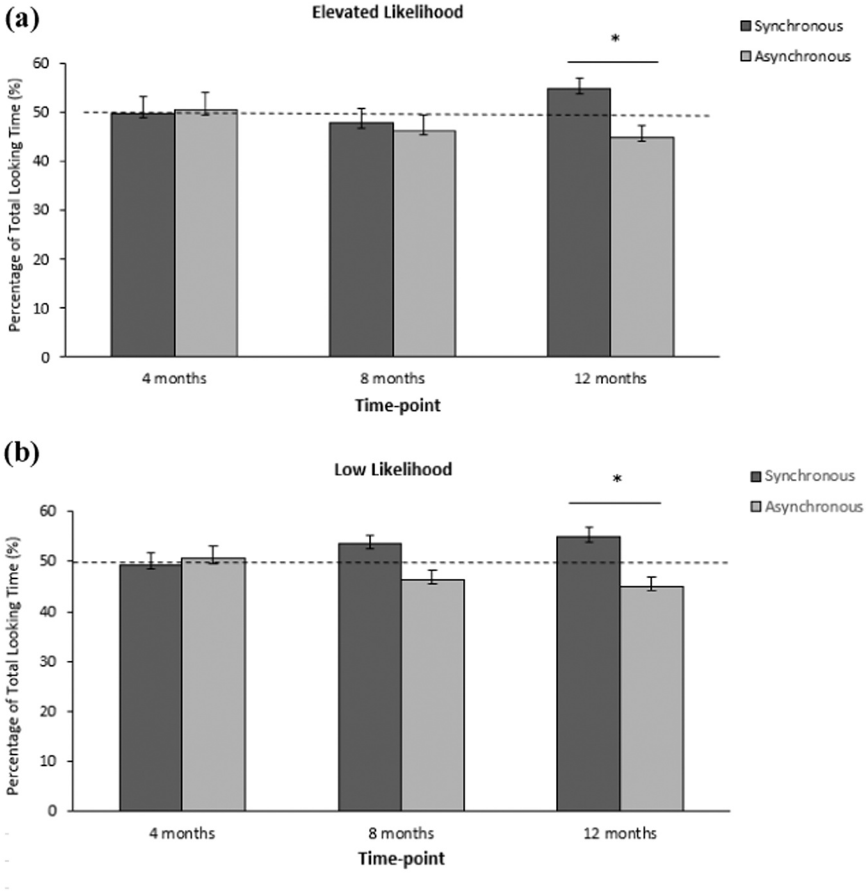

Fig. 2. Mean Percentage of Total Looking Time (%) to the synchronous and asynchronous displays by group and time-point. (a) EL-infants; (b) LL-infants. Dashed lines represent preference at chance. Error bars indicate standard error of the mean. *p < .05.

To further investigate the significant Time-point × Redundancy interaction and the null Redundancy × Group × Time-point interaction, we conducted a series of one-sample *t*-tests (directional) comparing the Percentage of Total Looking Time to the synchronous event to chance (50 %) within each group and time-point, as in Righi et al. (2018). This statistical test avoids the possibility that synchronous and asynchronous preferences are reciprocal, despite not allowing to properly test the longitudinal changes we observed in the LMM. We found that EL-infants did not prefer the synchronous event vs. chance at 4 months, $t(10) = -.02$, $p = .492$, $d = -.006$, nor at 8 months, $t(12) = -.694$, $p = .250$, $d = -.192$, but they did at 12 months, $t(9) = 2.20$, $p = .027$, $d = .699$. LL-infants also did not prefer the synchronous event over chance at 4 months, $t(13) = -.248$, $p = .404$, $d = -.066$, but they did at 8 months, $t(13) = 2.14$, $p = .026$, $d = .573$, and 12 months, $t(12) = 2.69$, $p = .011$, $d = .735$. Thus, results suggest that both EL-infants and LL-infants detected temporal synchrony at 12 months, but only LL-infants did so at 8 months. To test the robustness of these findings despite our small and underpowered sample size (see Supplementary Analysis 4 for power analyses), we replicated the previous analyses but conducted Bayesian one-sample *t*-tests (directional). We found that, overall, all results held, except for those of LL-infants at 8 months, which were inconclusive. The full report is available in Supplementary Analyses 5.

3.2. *Selective attention to facial features*

We conducted a 3 × 2 × 2 LMM on the proportion of total looking time to the mouth, with Time-point (4 months vs. 8 months vs. 12 months) and Redundancy (Synchronous vs. Asynchronous) as within-subject predictors and Group (EL-infants vs. LL-infants) as between-subject predictors (see descriptive statistics in Table A.5 and full results in Table A.8). We found a significant Group × Time-point interaction (Fig. 3), $F(2, 97.09) = 4.00$, $p = .02$, $r = 0.25$, indicating that groups differed in their looking time to the mouth depending on the time-point. Follow-up analyses (Bonferroni-corrected post-hoc tests) showed no indication that the groups differed in their preference for the mouth at 4 months, $F(1, 63.9) = 0.80$, $p = .37$, and 8 months, $F(1, 48.24) = 1.7$, $p = .19$. However, groups significantly differed at 12 months, $F(1, 57.2) = 4.81$, $p = .03$, when LL-infants looked to the mouth more than EL-infants.

Our second hypothesis, nonetheless, focused not only on investigating potential group differences in the patterns of changes in preference for the mouth across time-points, but also potential within-group differences across time-points. To test this, we ran pairwise comparisons with Bonferroni correction between time-points separately for each group (see the full report in Tables A.10 and A.11). We found that LL-infants significantly changed their preference for the mouth between 4 and 8 months ($p = <.001$), and between 4 and 12 months ($p = .003$), while infants' mouth-looking remained constant between 8 and 12 months ($p = .81$). In contrast, EL-infants did not significantly change their preference for the mouth between any time-point (all $ps > .07$). No other main effects or interactions were found (all $ps \geq .29$).

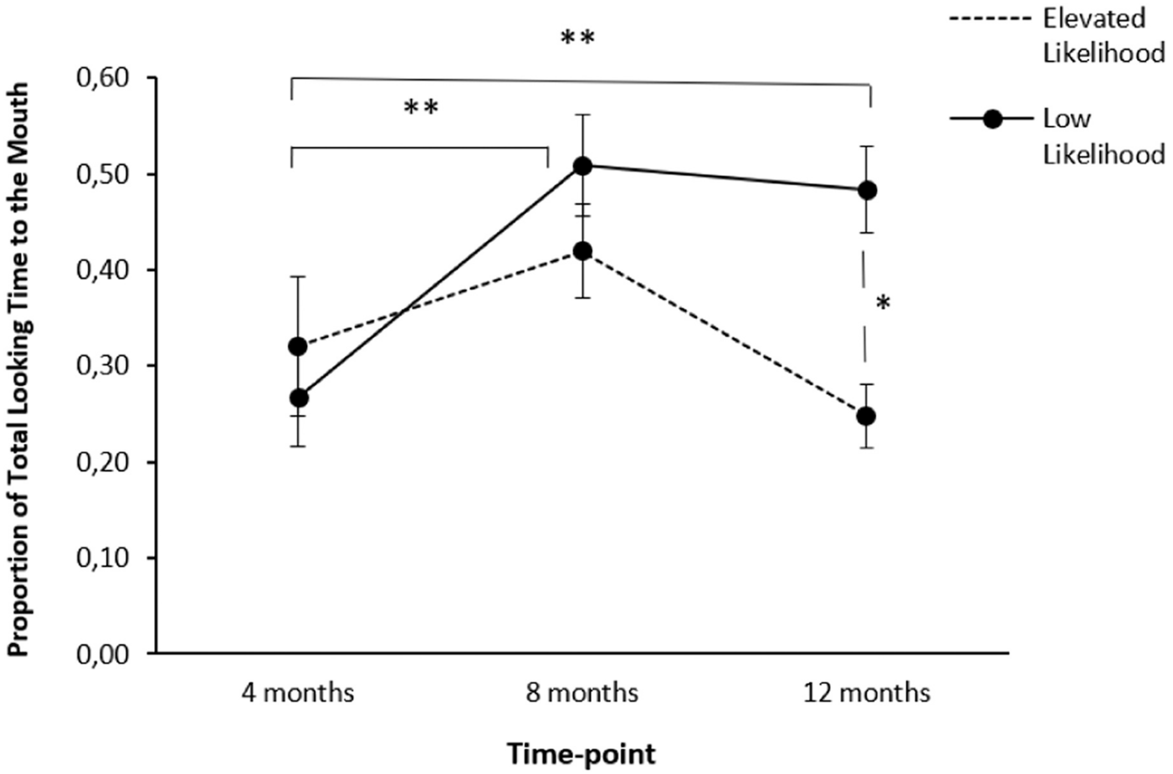

Fig. 3. Mean Proportion of Total Looking Time to the synchronous and asynchronous mouths by group and time-point. Bars indicate standard error of the mean. $^{*}p < .05$, $^{**}p < .01$.

We next replicated the latter analysis but on PTLT Eyes-Mouth to better capture the attentional competition between these regions. Results were consistent with those for PTLT Mouth (see full descriptive statistics in Table AAppendix A.6 and full results in Table A.9). The significant Group × Time-point interaction (Fig. 4), $F(2, 101.27) = 5.83$, $p = < .01$, $r = 0.29$, was again accounted by groups significantly differing in their relative preference for the eyes and mouth at 12 months ($p < .01$). Pairwise comparisons (with Bonferroni correction) between time-points separately for each group showed, again, that the pattern of changes in preference across time-points was different within each group (see full report in Tables A.12 and A.13). LL-infants significantly increased their looking time for the mouth relative to the eyes between 4 and 8 months ($p = .002$) and between 4 and 12 months ($p = .001$), suggesting an attentional shift from the eyes (indicated by positive scores) to the mouth (indicated by negative scores) that became stabilized between 8 and 12 months ($p = 1.00$). EL-infants did not significantly change their looking time to the mouth relative to the eyes across any time-point (all $ps \geq .09$), thus suggesting a lack of changes in their relative preference to these areas across the first year. No other main effects or interactions were found (all $ps \geq .09$).

4. Discussion

4.1. No evidence of reduced temporal asynchrony detection in audiovisual speech in EL-infants

Contrary to our prediction that EL-infants would show a reduced sensitivity to temporal asynchrony in audiovisual fluent speech during the first year, we found no evidence for group differences in temporal synchrony detection. Infants significantly preferred the synchronous over the asynchronous talking face at 12 months regardless of group, but showed no preference for either at 4 and 8 months, indicating that there was no evidence that infants detect temporal asynchronies before the end of the first year. Bayesian analyses — whose results will prevail for being more conservative — replicated the results from frequentist analyses at 4 and 12 months, showing moderate support for the null hypothesis at 4 months regardless of group, and moderate evidence for the hypothesis that infants preferred looking at the synchronously talking face over chance at 12 months, regardless of group. At 8 months, we found inconclusive evidence for LL-infants and moderate support for no temporal synchrony detection in EL-infants. Overall, this suggests no evidence that sensitivity to temporal synchrony is a mechanism affected in EL-infants at the group level during the first postnatal year, at least under the conditions of our study. Therefore, we found no evidence in support of the prediction by the *IIH* (Bahrick, 2010; Bahrick & Todd, 2012) positing an atypicality in this perceptual ability in infants with a familial history of ASD.

Our study provides, to our knowledge, the first longitudinal preliminary evidence of the development of sensitivity to temporal

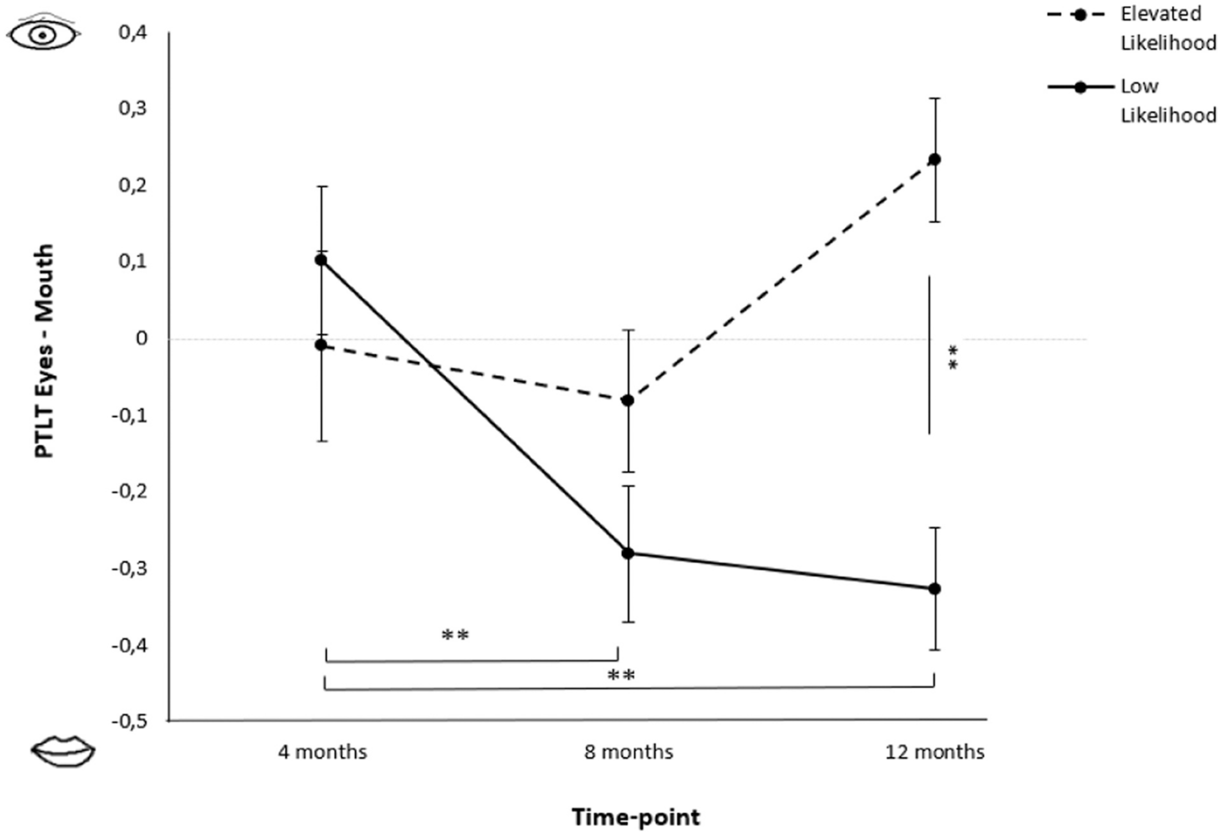

Fig. 4. Mean PTLT Eyes-Mouth by group and time-point. Bars indicate standard error of the mean. $^{**}p < .01$.

asynchrony in audiovisual *fluent* speech in EL-infants in the first year (for cross-sectional, see Santapuram et al., 2022; Suri et al., 2023). This contrasts with the reduced ability to match audiovisual speech information found by Guiraud et al. (2012) in 9-month-old EL-infants as a whole group. If diminished temporal asynchrony detection leads to reduced audiovisual speech integration, as in children with ASD (Stevenson et al., 2018), our results are unexpected. One possible explanation is that group differences in matching auditory and visual information may be more evident when tested in audiovisual syllables (e.g., Guiraud et al., 2012) than in the audiovisual *fluent* speech we used. Silences between syllables, absent in continuous stream in fluent speech, might facilitate LL-infants to detect temporal asynchronies, thus making potential atypicalities in same-aged EL-infants — and, therefore, group differences — more noticeable at 8 months. In contrast, the complexity of fluent speech may protract LL-infants' onset to detect temporal asynchronies in this event to 12 months (vs. 8 months in syllables; Pons & Lewkowicz, 2014, Experiment 1), when EL-infants already succeed too, thus reducing the chances of finding group differences at this time-point. Alternatively, our small sample size (compared to $n = 31$ in Guiraud et al., 2012) may have reduced our chances of detecting group differences that could still occur. Thus, we should cautiously interpret our null results and do not rule out the possibility that the detection of asynchronies in audiovisual speech may be a mechanism affected at a group level in EL-infants. Our results should be considered as preliminary. Future better-powered studies should reexamine whether they replicate our null results or observe reduced sensitivity to temporal synchrony in EL-infants.

Recent cross-sectional findings support this latter possibility. Using a habituation paradigm, unlike us, Suri et al., (2023) found that 4 to 24-month-old EL-infants ($n = 35$) were less sensitive than LL-infants to detect audiovisual asynchronies in syllables. EL-infants showed a significantly larger temporal binding window than LL-infants only in the audiovisual speech events but equivalent in the nonspeech social events (a bouncing ball). Although their wide range of ages tested and collapsed in the analyses does not allow to establish the onset of reduced sensitivity to temporal synchrony, these results support the presence of diminished sensitivity to temporal synchrony in audiovisual speech *syllables* in EL-infants as a whole group. Future well-powered longitudinal studies using *fluent* audiovisual speech instead of syllables — closer to infants' natural social experiences — are needed to establish the exact onset time of this reduced sensitivity in development.

Our study also provides the first preliminary longitudinal report on the development of sensitivity to temporal asynchrony in audiovisual *fluent* speech in typically developing infants across the first year (see Edgar et al., 2022 for a longitudinal study during the second year), partially replicating cross-sectional findings showing that it emerges in the second half of the first year (Hillairet de Boisferon et al., 2017; Pons & Lewkowicz, 2014, Experiment 1), although slightly later than previously found. In contrast with our results (inconclusive at 8 months but moderately supporting temporal synchrony detection at 12 months), Pons and Lewkowicz (2014) found that infants detected asynchronies of 500 and 666 ms (but not 366 ms) at 8 instead of 12 months. However, they used habituation instead of a preferential looking paradigm, potentially impacting task demands. In our study, infants had to compare two audiovisual talking faces, which might be more challenging than detecting a change in asynchrony in a single speaking face. Alternatively, infants could need a protracted period of experience with audiovisual *fluent* speech to learn to detect temporal synchrony in this event, thus pointing to this property as a not as 'early' amodal cue as thought to be. However, note that inconclusive results indicate a lack of evidence either way, so we cannot rule out that LL-infants are detecting temporal synchrony at 8 months. These results may reflect a transition period in learning to detect temporal asynchronies, high within-group variability, or an underpowered sample size, but we cannot draw firm conclusions. Our findings, however, align with the time-point when perceptual narrowing in audiovisual *fluent* speech emerges — 12 months (Lewkowicz & Hansen-Tift, 2012; Lewkowicz et al., 2015) —, suggesting that learning to detect temporal asynchrony in audiovisual speech could be especially useful for infants when they become attuned to the categories of their native language (see Lewkowicz, 2014). Detecting temporal synchrony may boost learning the perception of audiovisual unity regardless of the specific nature of speech (native and non-native) immediately before infants can rely more on language-specific phonemic speech cues.

4.2. Preliminary evidence of reduced selective attention to the articulating mouth in EL-infants at the end of the first year

Overall, our findings provide preliminary evidence supporting our prediction based on domain-general accounts of emerging ASD (Campos et al., 2019; Gliga et al., 2014) that posited group differences in preference for the mouth only at the end of the first year. At 12 months, EL-infants showed reduced preference for the articulating mouth compared to LL-infants. This result was consistent across both absolute and relative preference measures. However, due to the small sample size, these results and our interpretation below should be considered preliminary until further replication in larger samples.

The reduced mouth-looking of EL-infants at the end of the first year might potentially indicate that, unlike LL-infants, they may not find visual speech cues that beneficial for language acquisition. Audiovisual speech is a highly complex, dynamic, and difficult-to-predict event (Elsabbagh & Johnson, 2016). Audiovisual redundancy in speech may not support face scanning in EL-infants but make the processing of audiovisual speech more difficult than facilitating it (as suggested by Shic et al., 2014). Unlike LL-infants, they might find the redundancy in the mouth overly complex, potentially directing their attention toward a simpler, non-redundant area — the eyes — as an adaptation. This potential deviation of the attentional trajectory may lead EL-infants to progressively reduce the attunement to the mouth necessary to benefit from audiovisual speech cues for language acquisition. This interpretation aligns with findings showing no group differences in mouth-looking at 12 months between EL and LL-infants but associations between this ability and language outcomes in toddlerhood only in LL-infants (Chawarska et al., 2022). Perhaps at the extreme of this continuum, toddlers with ASD show reduced mouth-looking relative to control peers and no association with expressive language (Habayeb et al., 2021).

Alternatively, the reduced mouth-looking of EL-infants at 12 months might potentially reflect their differential experiences with audiovisual speech redundancy earlier in life. The first postnatal year is critical for revealing them, especially when more complex multisensory integration and high temporal resolution processing become important for age-appropriate typical behaviours

(Elsabbagh & Johnson, 2016). In line with Neuroconstructivism (Campos et al., 2019; Gliga et al., 2014; Karmiloff-Smith, 1998), we suggest that EL-infants' early accessibility to audiovisual redundancy may have been constrained early in their natural context of interactions by internal factors (i.e., atypicalities in infants' neurocognitive system across different domains: motor, attentional, perceptual) and/or internal-external factors (e.g., experiencing different social communicative inputs in caregiver-child dyadic interactions, like having less conversational turns; e.g., Swanson et al., 2019). As a result, as observed in other infants underexposed to this event (e.g., hearing infants with deaf caregivers; Mercure et al., 2019), we speculate that a reduced early audiovisual redundancy may have potentially led EL-infants to a trajectory of diminished attention to the mouth.

Note that reduced mouth-preference in EL-infants did not necessarily entail increased eyes-preference. Although one may interpret their positive scores in PTLT Eyes-Mouth at 12 months as a preference for the eyes, the lack of evidence regarding changes in their relative preference for the mouth across time-points does not allow us to conclude this. Within-group variability may have hindered a clear directionality in either preference for the eyes or the mouth. This interpretation goes according to the mixed literature on eye gaze in infant siblings, with most evidence finding no group differences compared to LL-infants (e.g., Chawarska et al., 2022; Droucker et al., 2013; Elsabbagh et al., 2014) or reduced eyes-preference only in the subset of EL-infants with ASD (e.g., Shic et al., 2014).

Crucially, although we did not find evidence of group differences before 12 months, the trajectory of changes between time-points was different within each group from earlier, as we hypothesized. We suggest that these differing developmental changes across the first year may possibly account for the group differences found at the end of this period. The trajectory of mouth-looking observed in LL-infants only partially replicates the *u*-shaped pattern of changes we predicted based on prior cross-sectional research using free-viewing tasks (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). While we also found more looking to the eyes at 4 months and a shift to the mouth at 8 months, we did not observe an equivalent preference for the eyes and mouth at 12 months (but see Morin-Lessard et al., 2019 for similar results to ours). Instead, infants maintained a preference for the mouth. In the context of our task, we interpret that at this time-point infants may still need to focus on the mouth to seek visual speech cues beneficial for detecting temporal synchrony. Supportively, we found moderate evidence that our infants discriminated temporal asynchrony precisely at 12 months, possibly pointing to these two abilities as being potentially closely related language learning mechanisms. Furthermore, when using tasks similar to ours in children, mouth-looking and temporal synchrony detection seem to be related at a group level (Grossman et al., 2015; Righi et al., 2018).

In contrast to the trajectory followed by LL-infants, EL-infants showed no evidence of changes in relative or absolute preference for the mouth across the first year. This finding aligns with our prediction of constant mouth-preference during the first year, perhaps suggesting that the audiovisual redundancy provided by this facial feature may not become differentially salient for EL-infants in this period but instead remains stable. Our results provide preliminary evidence supporting the *IIH* prediction that EL-infants as a whole group would show an early atypicality in the mechanism of selective attention (Bahrick, 2010; Bahrick & Todd, 2012), but they need replication before being considered robust. If confirmed, this finding could constitute a unique contribution of our study since previous cross-sectional (Santapuram et al., 2022) and longitudinal research (Elsabbagh et al., 2014; Jones & Klin, 2013; Shic et al., 2014) on selective attention to talking faces in EL-infants between 6 and 18 months thus far found no evidence of group differences in mouth-preference. These null results could be perhaps due to not densely tracking this ability at multiple time-points across relevant developmental periods of the first year, excluding EL-infants with ASD from the analyses (e.g., Chawarska et al., 2022), or including only male EL-infants (Jones & Klin, 2013).

4.3. *Theoretical implications*

Our study has several theoretical implications. Contrary to the *IIH* (Bahrick & Todd, 2012), we found no evidence that sensitivity to temporal synchrony is affected in infants at familial likelihood for ASD. However, as hypothesized by the *IIH*, selective attention to the mouth during audiovisual speech might potentially be a mechanism atypically functioning in infant siblings at the end of the first year. This is an important preliminary contribution of our study, as we may have targeted one widespread domain-general atypicality still underemphasized by the *IIH* and other domain-general approaches for ASD (Campos et al., 2019; Gliga et al., 2014).

Reduced mouth-looking may be a potential risk factor for later language atypicalities particularly relevant during the first postnatal year. Given the crucial role of increased mouth-looking for typical language acquisition, we suggest that atypicalities in this mechanism may underlie and be an early expression of the language atypicalities affecting EL-infants as a whole (Belteki et al., 2022; Hudry et al., 2014). What remains unknown, however, is which *developmental processes* underlie the trajectory between the initial attentional atypicality and the later affected language outcomes. If our results are replicated, we speculate that EL-infants' reduced attention to the mouth may potentially reduce their chances of benefiting from accessing audiovisual redundant native phonemic cues, reduce the amount of exposure to this event, and perhaps diminish the fidelity with which they process audiovisual speech. Ultimately, we speculate that this alternative trajectory could potentially increase the likelihood of later cascading atypicalities in different subcomponents of language acquisition (Chita-Tegmark et al., 2015; Hudry et al., 2014). In fact, this hypothesis has been predicted by Neuroconstructivist accounts. From this view, the lack of fit between EL-infants' abilities and the properties of audiovisual speech (i.e., dynamic, unpredictable, and highly variable) would account for infants' reduced attunement to the properties of their linguistic environment (Gliga et al., 2014). We suggest that an initial genetic vulnerability in selective attention to the mouth may possibly be shared early by the majority of EL-infants but its effects when interacting with their social and linguistic experience across development could vary between individuals, leading to heterogeneous language and ASD outcomes influenced by multiple other cumulative risk factors (Elsabbagh, 2020). Supportively, studies with infant and toddler twins indicate genetic influences on preferential attention to the eyes and mouth across both typical and ASD-affected individuals that relate to individual differences in later language outcomes (Constantino et al., 2017; Viktorsson et al., 2023).

4.4. Limitations and future research

Our study presents limitations. Our small sample size (underpowered for our LMM on temporal synchrony detection, as the achieved power was close to ~35 %; see 'Power Analysis' in Supplementary Analysis 4), yet a general issue in infant sibling studies due to the prevalence of ASD (Jones et al., 2019), warrants the need for future better-powered studies with more reliable effect sizes (ours ranged from small to medium) for more robust conclusions. We used asynchronies of 500 and 666 ms instead of the planned 666 ms. Our control analysis showed that this technical error did not impact our main results, but our preliminary findings should be cautiously considered when compared with those of previous studies. Unlike us, future research should measure selective attention to the eyes and mouth separately from detecting temporal asynchrony in a free-viewing paradigm to provide more reliable measures.

Our study opens new research avenues. First, the role of alternative pathways in selective attention to the mouth in EL-infants relative to language outcomes still needs to be addressed. While we interpreted reduced mouth-preference as a risk factor, one may argue that it could be protective. EL-infants may reduce their attention to the mouth to deal with a perceptually overwhelming event and instead follow a more auditory pathway of attunement with audiovisual speech. Second, the specificity of the predictive value of atypical trajectories in selective attention to the mouth to diagnostic ASD outcomes and language atypicalities of the phenotype (e.g., see Shic et al., 2020) requires investigation. Third, whether selective attention and detection of temporal synchrony are independent or intertwined mechanisms in both LL and EL-infants and whether their relationship varies across development remains unclear. Finally, if our findings are extensible to real contexts of infant-caregiver interactions when using new methods as head-mounted eye-trackers (e.g., Zhao et al., 2023), selective attention to the mouth in a very specific time-point of development (end of 1st year) would constitute a promising area for early intervention in language acquisition in infant siblings.

5. Conclusions

Our findings suggest preliminary evidence that reduced selective attention to the mouth in audiovisual speech at age one year might be an early candidate mechanism to account for the language atypicalities usually developed by infant siblings (Hudry et al., 2014). If further research replicates our tentative results, a differential *timing* in this attentional trajectory relative to typical counterparts could be interpreted as an early expression of later language difficulties. An initial subtle attentional atypicality may perhaps increasingly interfere with how EL-infants learn from their social environments and constrain the development of complex abilities learned through interaction with others, particularly language. The continuity of the affected mechanisms in populations at genetic increased likelihood for the condition and individuals with ASD (Righi et al., 2018; Zhao et al., 2023) places reduced mouth-looking as a potential atypical candidate mechanism of the language difficulties commonly linked to the condition.

Funding

This work was funded by a predoctoral contract awarded to IL by the Spanish Ministry of Education, Culture and Sports [Grant number FPU13/03508; 2015–2018] and two mobility grants for international stays [Grants FPU 2016 and UAM 2018 mobility subprograms]. It also benefited from a grant awarded to RC by the Spanish Ministry of Economy and Competitiveness and by the European Regional Developmental Fund (ERDF) (TRABERITEA Project – Grant number PSI2015–66509-P MINECO/FEDER; 2016–2019), and the project TRABERITEA: "Developmental trajectories in infants at risk for autism: processes of neurocognitive specialization and early interaction and phenotypic outcomes" [Grant number PID2020–117087 GB-I00] – financed by the Spanish Ministry of Science and Innovation, and the State Research Agency, of which RC is PI. Additional work of this research has been funded by a National Science Centre of Poland grant SONATINA 7 [Grant number 2023/48/C/HS6/00264, to IL]. Any views expressed are those of the author(s) and not necessarily those of the funders. The funders had no role in study design, data collection and analysis, the decision to publish or manuscript preparation.

CRediT authorship contribution statement

Itziar Lozano Sánchez: Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mercedes Belinchón:** Writing – review & editing, Supervision, Resources, Conceptualization. **Ruth Campos:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of Competing Interest

None.

Data availability

Data are not publicly accessible, but code and materials will be made available upon reasonable request to the authors.

Acknowledgements

The authors thank all the infants and their families for participating in our study, ALANDA, AMITEA Program of Gregorio Marañón University Hospital, and other entities belonging to The Autism Spain Confederation-AE for contributing to contacting the families, and the TRABERITEA team members for contributing to recruiting, protocol setting-up, scheduling visits, and testing: Dr. Pastora Martínez-Castilla, Dr. Rubén Palomo, Dr. Carmen Nieto, Dr. María Sotillo, Ruth Vidriales, Laura Escribano, Margarita Quesada, Beatriz Cavero, and María Verde. IL also thanks Dr. Tomalski and Prof. Bahrick for their generous eye-tracking training and insightful theoretical discussions at their Labs, Dr. López Pérez and Anna Malinowska-Korczak for their lessons on using eye-tracking with babies, and Dr. Joana del Hoyo Bilbao for acting on stimulus recording.

Declarations of permissions

Written permission has been obtained from the person in Fig. 1 for the use of her image.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.infbeh.2024.101973.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5^{TM} (5th ed.). Washington, DC: Author.

Bahrick, L. E. (2010). Intermodal Perception and Selective Attention to Intersensory Redundancy: Implications for Typical Social Development and Autism. In J. G. Bremner, & T. D. Wachs (Eds.), *The Wiley-Blackwell Handbook of Infant Development* (pp. 120–166). Wiley Blackwell. https://doi.org/10.1002/9781444327564.ch4.

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science, 23*(6), 414–420. https://doi.org/10.1177/0963721414549187

Bahrick, L. E., & Todd, J. T. (2012). Multisensory Processing in Autism Spectrum Disorders: Intersensory Processing Disturbance as a Basis for Atypical Development. In B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 657–674). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/8466.003.0058.

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development, 67*, Article 101709. https://doi.org/10.1016/j.infbeh.2022.101709

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry, 47*(1), 88–98. https://doi.org/10.1111/j.1469-7610.2005.01443.x

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience & Biobehavioral Reviews, 84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Belteki, Z., van den Boomen, C., & Junge, C. (2022). Face-to-face contact during infancy: How the development of gaze to faces feeds into infants' vocabulary outcomes. *Frontiers in Psychology, 13*, Article 997186. https://doi.org/10.3389/fpsyg.2022.997186

Campos, R., Nieto, C., & Núñez, M. (2019). Research domain criteria from neuroconstructivism: A developmental view on mental disorders. *WIREs Cognitive Science, 10*(3), Article e1491. https://doi.org/10.1002/wcs.1491

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*(1), 213–236. https://doi.org/10.1017/s0305000902005482

Chawarska, K., Lewkowicz, D., Feiner, H., Macari, S., & Vernetti, A. (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. *Journal of Child Psychology and Psychiatry, 63*(12), 1466–1476. https://doi.org/10.1111/jcpp.13595

Chita-Tegmark, M., Arunachalam, S., Nelson, C. A., & Tager-Flusberg, H. (2015). Eye-tracking measurements of language processing: Developmental differences in children at high risk for ASD. *Journal of Autism and Developmental Disorders, 45*(10), 3327–3338. https://doi.org/10.1007/s10803-015-2495-5

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., Gillespie, S., Klaiman, C., Klin, A., & Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature, 547*(7663), 340–344. https://doi.org/10.1038/nature22999

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2022). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. *Infancy, 27*(1), 67–96. https://doi.org/10.1111/infa.12436

De Anda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016a). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research, 59*, 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234

Droucker, D., Curtin, S., & Vouloumanos, A. (2013). Linking infant-directed speech and face preferences to language outcomes in infants at risk for autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 56*(2), 567–576. https://doi.org/10.1044/1092-4388(2012/11-0266

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2022). Intersensory matching of faces and voices in infancy predicts language outcomes in young children. *Developmental Psychology, 58*, 1413–1428. https://doi.org/10.1037/dev0001375

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2023). Intersensory processing of social events in 6-month-olds predicts language outcomes at 18, 24, and 36 months of age. *Infancy, 28*(3), 569–596. https://doi.org/10.1111/infa.12533

Elsabbagh, M. (2020). Linking risk factors and outcomes in autism spectrum disorder: Is there evidence for resilience? *BMJ, 368*, l6880. https://doi.org/10.1136/bmj.l6880

Elsabbagh, M., Bedford, R., Senju, A., Charman, T., Pickles, A., Johnson, M. H., & BASIS Team. (2014). What you see is what you get: Contextual modulation of face scanning in typical and atypical development. *Social Cognitive and Affective Neuroscience, 9*(4), 538–543. https://doi.org/10.1093/scan/nst012

Elsabbagh, M., & Johnson, M. H. (2016). Autism and the social brain: The first-year puzzle. *Biological Psychiatry, 80*(2), 94–99. https://doi.org/10.1016/j.biopsych.2016.02.019

Falck-Ytter, T., Nyström, P., Gredebäck, G., Gliga, T., Bölte, S., & the EASE Team. (2018). Reduced orienting to audiovisual synchrony in infancy predicts autism diagnosis at 3 years of age. *Journal of Child Psychology and Psychiatry, 59*(8), 872–880. https://doi.org/10.1111/jcpp.12863

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Feng, S., Lu, H., Wang, Q., Li, T., Fang, J., Chen, L., & Yi, L. (2021). Face-viewing patterns predict audiovisual speech integration in autistic children. *Autism Research, 14*(12), 2592–2602. https://doi.org/10.1002/aur.2598

Feng, S., Wang, Q., Hu, Y., Lu, H., Li, T., Song, C., Fang, J., Chen, L., & Yi, L. (2022). Increasing audiovisual speech integration in autism through enhanced attention to mouth. *Developmental Science, 26*, Article e13348. https://doi.org/10.1111/desc.13348

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. SAGE.

Gliga, T., Jones, E. J. H., Bedford, R., Charman, T., & Johnson, M. H. (2014). From early markers to neuro-developmental mechanisms of autism. *Developmental Review, 34*(3), 189–207. https://doi.org/10.1016/j.dr.2014.05.003

Grossman, R. B., Steinhart, E., Mitchell, T., & McIlvane, W. (2015). Look who's talking!" Gaze patterns for implicit and explicit audio-visual speech synchrony detection in children with high-functioning autism. *Autism Research, 8*(3), 307–316. https://doi.org/10.1002/aur.1447

Guiraud, J. A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., Elsabbagh, M., Johnson, M. H., & the BASIS Team. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PLOS ONE, 7*(5), Article e36428. https://doi.org/10.1371/journal.pone.0036428

Habayeb, S., Tsang, T., Saulnier, C., Klaiman, C., Jones, W., Klin, A., & Edwards, L. A. (2021). Visual traces of language acquisition in toddlers with autism spectrum disorder during the second year of life. *Journal of Autism and Developmental Disorders, 51*, 2519–2530. https://doi.org/10.1007/s10803-020-04730-x

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism, 19*(1), 20–28. https://doi.org/10.1177/1362361313509731

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science, 20*(3), Article e12381. https://doi.org/10.1111/desc.12381

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 33*(1), 3–13. https://doi.org/10.1023/a:1022270118899

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., Johnson, M. H., & Charman, T. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(1), 154–167. https://doi.org/10.1007/s10803-013-1861-4

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders, 37*(1), 158–170. https://doi.org/10.1007/s10803-006-0339-z

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature, 504*(7480), 427–431. https://doi.org/10.1038/nature12715

Jones, E. J. H., Mason, L., Begum Ali, J., van den Boomen, C., Braukmann, R., Cauvet, E., Demurie, E., Hessels, R. S., Ward, E. K., Hunnius, S., Bolte, S., Tomalski, P., Kemner, C., Warreyn, P., Roeyers, H., Buitelaar, J., Falck-Ytter, T., Charman, T., Johnson, M. H., & Eurosibs Team. (2019). Eurosibs: Towards robust measurement of infant neurocognitive predictors of autism across Europe. *Infant Behavior & Development, 57*, Article 101316. https://doi.org/10.1016/j.infbeh.2019.03.007

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2*(10), 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS One, 9*(10), Article e111467. https://doi.org/10.1371/journal.pone.0111467

Kleberg, J. L., Nyström, P., Bölte, S., & Falck-Ytter, T. (2019). Sex differences in social attention in infants at risk for autism. *Journal of Autism and Developmental Disorders, 49*(4), 1342–1351. https://doi.org/10.1007/s10803-018-3799-z

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology, 46*, 66–77. https://doi.org/10.1037/a0015579

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*(2), 292–315. https://doi.org/10.1002/dev.21197

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology, 130*, 147–162. https://doi.org/10.1016/j.jecp.2014.10.006

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S. L., & Schedule, A. A. D. O. (2015). *ADOS-2 Manual (Part I): Modules*, 1–4.

Lozano, I., López Pérez, D., Laudańska, Z., Malinowska-Korczak, A., Szmytke, M., Radkowska, A., & Tomalski, P. (2022). Changes in selective attention to articulating mouth across infancy: Sex differences and associations with language outcomes. *Infancy, 27*(6), 1132–1153. https://doi.org/10.1111/infa.12496

McDonald, N. M., Senturk, D., Scheffler, A., Brian, J. A., Carver, L. J., Charman, T., … Jeste, S. S. (2020). Developmental trajectories of infants with multiplex family risk for autism: A baby siblings research consortium study. *JAMA Neurology, 77*(1), 73–81. https://doi.org/10.1001/jamaneurol.2019.3341

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science, 22*(1), Article e12701. https://doi.org/10.1111/desc.12701

Miniscalco, C., Fränberg, J., Schachinger-Lorentzon, U., & Gillberg, C. (2012). Meaning what you say? Comprehension and word production skills in young children with autism. *Research in Autism Spectrum Disorders, 6*(1), 204–211. https://doi.org/10.1016/j.rasd.2011.05.001

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology, 55*, 1640–1655. https://doi.org/10.1037/dev0000750

Morrel, J., Singapuri, K., Landa, R. J., & Reetzke, R. (2023). Neural correlates and predictors of speech and language development in infants at elevated likelihood for autism: A systematic review. *Frontiers in Human Neuroscience, 17*, Article 1211676. https://doi.org/10.3389/fnhum.2023.1211676

Mullen, E. M. (1995). *Mullen Scales of Early Learning manual (AGS ed.)*. American Guidance Service.

Murat Baldwin, M., Xiao, Z., & Murray, A. (2022). Temporal synchrony in autism: A systematic review. *Review Journal of Autism and Developmental Disorders, 9*(4), 596–617. https://doi.org/10.1007/s40489-021-00276-5

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., … Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. *Pediatrics, 128*(3), e488–e495. https://doi.org/10.1542/peds.2010-2825

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52*(5), 588–598. https://doi.org/10.1111/j.1469-7610.2010.02332.x

Plumb, A. M., & Wetherby, A. M. (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research: JSLHR, 56*(2), 721–734. https://doi.org/10.1044/1092-4388(2012/11-0104)

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science, 26*(4), 490–498. https://doi.org/10.1177/0956797614568320

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica, 149*, 142–147. https://doi.org/10.1016/j.actpsy.2013.12.013

Righi, G., Tenenbaum, E. J., McCormick, C., Blossom, M., Amso, D., & Sheinkopf, S. J. (2018). Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Research, 11*(4), 645–653. https://doi.org/10.1002/aur.1918

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., Kim, S. Y., Keceli-Kaysili, B., Bottema-Beutel, K., Lewkowicz, D. J., Wallace, M. T., & Woynaroski, T. G. (2022). Mechanisms by which early eye gaze to the mouth during multisensory speech influences expressive communication development in infant siblings of children with and without autism. *Mind, Brain, and Education, 16*(1), 62–74. https://doi.org/10.1111/mbe.12310

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological Psychiatry, 75*(3), 231–237. https://doi.org/10.1016/j.biopsych.2013.07.009

Shic, F., Wang, Q., Macari, S. L., & Chawarska, K. (2020). The role of limited salience of speech in selective attention to faces in toddlers with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 61*(4), 459–469. https://doi.org/10.1111/jcpp.13118

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism, 22*, 609–624. https://doi.org/10.1177/1362361317704413

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience, 34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Suri, K. N., Whedon, M., & Lewis, M. (2023). Perception of audio-visual synchrony in infants at elevated likelihood of developing autism spectrum disorder. *European Journal of Pediatrics*. https://doi.org/10.1007/s00431-023-04871

Swanson, M. R., Donovan, K., Paterson, S., Wolff, J. J., Parish-Morris, J., Meera, S. S., & IBIS Network. (2019). Early language exposure supports later language skills in infants with and without autism. *Autism Research, 12*(12), 1784–1795. https://doi.org/10.1002/aur.2163

Szatmari, P., Chawarska, K., Dawson, G., Georgiades, S., Landa, R., Lord, C., … Halladay, A. (2016). Prospective longitudinal studies of infant siblings of children with autism: Lessons learned and future directions. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(3), 179–187. https://doi.org/10.1016/j.jaac.2015.12.014

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., Falck-Ytter, T., & EASE Team. (2016). Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular Autism, 7*, 1–10. https://doi.org/10.1186/s13229-016-0069-9

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 57*(5), 585–595. https://doi.org/10.1111/jcpp.12499

Todd, J. T., & Bahrick, L. (2022). Individual differences in multisensory attention skills in children with autism spectrum disorder predict language and symptom severity: Evidence from the multisensory attention assessment protocol (MAAP). *Journal of Autism and Developmental Disorders, 53*(12). https://doi.org/10.1007/s10803-022-05752-3

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology, 169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Viktorsson, C., Portugal, A. M., Li, D., Rudling, M., Siqueiros Sanchez, M., Tammimies, K., Taylor, M. J., Ronald, A., & Falck-Ytter, T. (2023). Preferential looking to eyes versus mouth in early infancy: Heritability and link to concurrent and later development. *Journal of Child Psychology and Psychiatry, 64*(2), 311–319. https://doi.org/10.1111/jcpp.13724

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science, 12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Zhao, Z., Tang, H., Zhang, X., Zhu, Z., Xing, J., Li, W., Tao, D., Qu, X., & Lu, J. (2023). Characteristics of visual fixation in chinese children with autism during face-to-face conversations. *Journal of Autism and Developmental Disorders, 53*(2), 746–758. https://doi.org/10.1007/s10803-021-04985-y