**Abstract**

Although the pattern of visual attention towards the region of the eyes is now well-established for infants at an early stage of development, less is known about the extent to which the mouth attracts an infant’s attention. Even less is known about the extent to which these specific looking behaviours towards different regions of the talking face (i.e., the eyes or the mouth) may impact on or account for aspects of language development. The aim of the present systematic review is to synthesize and analyse (i) which factors might determine different looking patterns in infants during audio-visual tasks using dynamic faces and (ii) how these patterns have been studied in relation to aspects of the baby’s development. Four bibliographic databases were explored, and the records were selected following specified inclusion criteria. The search led to the identification of 19 papers (October 2021). Some studies have tried to clarify the role played by audio-visual support in speech perception and early production based on directly related factors such as the age or language background of the participants, while others have tested the child’s competence in terms of linguistic or social skills. Several hypotheses have been advanced to explain the selective attention phenomenon. The results of the selected studies have led to different lines of interpretation. Some suggestions for future research are outlined.

*Keywords:*selective attention, audio-visual perception, infants, language development

**1. Introduction**

Studies investigating patterns of selective attention - a key measure of visual attention which implies that the child voluntarily attends to specific areas of a visual scene (Ruff & Rothbart, 1996) - have highlighted a particular visual interest in faces in early infancy (e.g., Amso, Haas, & Markant, 2014; DiGiorgio, Turati, Altoè, & Simion, 2012; Frank, Amso, & Johnson, 2014; Frank, Vul, & Johnson, 2009; Libertus, Landa, Haworth, 2017) and greater looking to hands and instrumental actions on objects with increasing age (Aslin, 2009; Fausey, Jayaraman, & Smith, 2016; Frank, Vul, & Saxe, 2012; see also Nelson & Oakes, 2021). Looking at the face has long been held to be a factor in children’s early language and social development (Brooks & Meltzoff, 2002, 2005), given that interactions are generally based on face-to-face situations (Atagi & Johnson, 2020; Kuhl, 2007; Masek, McMillan, Paterson, Tamis-LeMonda, Golinkoff, & Hirsh-Pasek, 2021; Stern, 1974). The child’s looks towards internal features of the face (i.e., their ability to selectively attend to these characteristics), with its dynamic properties, could affect audio-visual perception of speech (Guellaï et al., 2014) and this has been related to the child’s social, communication and language development (Bahrick & Todd, 2012; Choi, Black, & Werker, 2018).

We see two reasons to conduct a systematic analysis of studies investigating children’s selective attention to internal features of the face. The first is that most of the studies in this field have revealed the existence of a “temporal coincidence” between the time children start looking towards the mouth and certain communication milestones. The second reason is that some specific attentional shifts (e.g., from the eyes to the mouth) may serve as predictors of the child’s language development or the child’s ability to engage or disengage with their parents during play. Thus, the study of such a phenomenon could be relevant for clinical programs designed for populations at risk for language disorders and to promote interventions able to support development in such children.

Among the studies showing a greater interest in the human face, some have sought to understand where an infant voluntarily looks when exposed to talking faces (i.e., selective attention). These studies agree that infants show particular interest in the eyes in early infancy (Jones, & Klin, 2013; Maurer & Salapatek, 1976). Haith, Bergman, and Moore (1977) reported that between 9 and 11 weeks infants’ attraction to the eyes increases as a function of both age and type of stimulus (i.e., talking faces vs auditory stimulus only). This suggests that audio-visual speech situations increase the level of engagement, focusing children’s attention on the eyes. Indeed, the eyes have a well-established role in communication and in human interactions since they provide socio-emotional information from a very young age. Communication expressed through eye-gaze can engage an infant in a task or a conversation as it conveys emotion. In addition, at a later point in the child’s development this will be part of Joint Attention episodes, thus also supporting their vocabulary learning some months later (Brooks & Meltzoff, 2005) as well as their social development (Pons, Bosch, & Lewkowicz, 2019; Wagner, Luyster, Yim, Tager-Flusberg, & Nelson, 2013).

Since the attraction to the eyes is a well-established pattern in early infancy, the question that needs to be answered is, when and to what extent does the mouth attract infant attention? As suggested by previous studies, information from the mouth can be a source of phonetic or linguistic cues and attraction to the mouth could be explained by individual factors, such as language-related-experiences (such as bilingualism) or being exposed to particular conditions (such as noise). For example, it has been widely demonstrated, for adults, that the redundant effects of the simultaneous presence of auditory and visual information (in particular, information provided by the mouth) facilitate language comprehension when speech is difficult to understand, as in the presence of noise or a non-native language (Jerger, Damian, Tye-Murray, & Abdi, 2014; Kròl, 2018; Reisberg, McLean, & Goldfield, 1987), or when there are physical barriers, as in the case of people with hearing loss (Bernstein, Tucker, & Demorest, 2000; Summerfield, 1992). Infants might also benefit from multimodal information (sight of mouth+auditory experience of voice) under such conditions. Additionally, in the case of bilinguals, for whom language acquisition is more demanding (Ayneto & Sebastian-Galles, 2017), visual cues provided by the mouth may make it easier to recognise and understand speech (Bahrick & Lickliter, 2000; Bahrick & Pickens, 1988). The perceptual and motor cues provided by the mouth could enhance a child’s retention of information regarding articulatory movements as a basis for speech-related learning, both in the first year of life and later (once they have begun to produce speech themselves). In fact, it has been suggested that a ‘perceptuo-motor link’ (Imada, Zhang, Cheour, Tauli, Ahonen, & Kuhl, 2006, p. 957) establishes sensorimotor maps for vocal production. Interestingly, in their brain imaging study of babies aged 6 to12 months Imada and colleagues found activation of the speech motor areas during a listening task (i.e., in response to hearing speech) only after the child had begun to accumulate adult-like production experience, that is, once they had begun to produce canonical babble, or babble in adult-like syllables (a finding anticipated in Vihman, 2002). Moreover, access to the visual information obtainable from the articulatory movements of speech promotes lexical recognition (Fort, Kandel, Chipot, Savariaux, Granjon, & Spinelli, 2013).

**Reasons for studying selective visual attention towards talking faces in infancy**

Investigation of the child’s selective visual attention towards specific regions of the face (eyes or mouth) while exposed to dynamic/talking faces and the relationship between the looking pattern towards these regions and specific aspects of the child’s development have only recently begun to receive serious attention. Most of the studies in this field attempt to identify the individual factors (i.e., age or language background) that might explain different looking behaviours (towards the eyes or the mouth) when children are exposed to dynamic talking faces on a screen; thismay tend to underestimate what happens outside the laboratory setting. There are several reasons why the study of selective attention towards talking faces has become a shared interest in the fields of developmental and educational psychology. The main one is that the acquisition of language is multi-dimensional (Vigliocco, Perniss, & Vinson, 2014) and it requires integrated modalities (involving both perception and production, for example: Keren-Portnoy, DePaolis & Vihman, under review; Majorano, Vihman, & DePaolis, 2014). The child’s looking behaviour (towards adult’s faces or towards object in the room) and its relationship with developmental outcomes has been studied in natural settings. For example, Gogate and colleagues (2006) tested how children learn specific word-object referents by analysing gaze switching (mother-object). They found that infant attention, together with maternal use of temporal synchrony, explained success in learning target words during an interactive session with parents. This supports the proposal of a link between selective visual attention and a child’s language learning or development in a real-life situation (Cetincelik, Rowland et al., 2021; Gogate & Hollich, 2010; Gogate, Bolzani, & Betancourt, 2006; Suanda et al., 2019; Yu & Smith, 2012, 2013). These studies consider not only what the child does and looks at in a natural and spontaneous setting but also the influence of the audio, visual and tactile aspects that are engaged in language learning, as well as the influence of the input.

Most of the studies adopting eye-tracking to investigate selective visual attention have attempted to clarify three aspects. Firstly, they have considered how early experience can modulate or affect visual looking behaviour in the first two years of life (Hillairet De Boisferon, Tift, Minar, & Lewkowicz, 2017; Mercure et al., 2017). As has emerged from recent findings, multiple aspects of experience contribute to determining the child’s face scanning (Oakes, DeBolt, Beckner, Voss, & Cantrell, 2021). However, findings relating to the factors that might influence looking patterns in infancy are somewhat contradictory and a common line of interpretation is difficult to identify.

Secondly they have considered how looking behaviour affects later language (Imafuku & Myowa, 2016; Kushnerenko, Tomalski, Ballieux, Potton, Birtles, Frostick, & Moore, 2013; Pons, Bosch, & Lewkowicz, 2019; Tenenbaum, Sobel, Shenkopf, Malle, & Morgan, 2015; Tsang, Atagi, & Johnson, 2018). Indeed, although the relationship between specific patterns of visual attention and language development is now well established (Akthar & Gernsbacher, 2007; Beuker, Rommelse, Donders, & Buitelaar, 2013; Brooks & Meltzoff, 2005, 2015; Carpenter, Nagell, Tomasello, Butterworth, 1998; Morales et al., 1998), less is known about the extent to which specific looking behaviours towards different regions of the talking face (i.e., the eyes or the mouth) may impact on or account for aspects of language development.

Thirdly, they have considered how factors such as exposure to one or more languages may influence a child’s behaviour (Ayneto & Sebastian-Galles, 2017; Fort, Ayneto-Gimeno, Escrichs, & Sebastian-Galles, 2018; Pons, Bosch, & Lewkowicz, 2015). Two main hypotheses have been advanced to explain differences in the looking patterns of typically developing children: the *language expertise hypothesis* (affecting both bilinguals and monolinguals) and the *language distance hypothesis* (affecting bilinguals only), based on the distance between the two languages. The language expertise hypothesis highlights a correspondence between the child’s gaze pattern and their linguistic level, showing a move from a preference for the eyes in pre-babbling infants, through increased interest in the mouth when infants begin to produce canonical babbling, and then a decrease in looks to the mouth in favour of the eyes when the infants become more linguistically advanced, toward the end of the first year (Hillairet De Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015) and a maintenance of attention towards the mouth for bilingual children only until the end of the first year of life (Lewkowicz & Hansen-Tift, 2012). The language distance hypothesis is based on the finding that bilingual infants learning two closely related languages (such as Spanish and Catalan) look more toward the mouth than do their bilingual peers learning two more distantly related languages (e.g., Spanish and Russian/German) (Birules et al., 2019).Thus, it is not simply the bilingual condition per se that leads to more looking at the mouth but also the phonological, rhythmic, phonotactic, morphological and lexical properties of the languages the child is learning. According to this proposal, infants learning two similar languages should stand to gain more from redundant audio-visual cues, given the challenge of distinguishing or discriminating them, than bilinguals learning two distant languages.

The present review aims to provide a description of current experimental studies that adopt eye-tracking to investigate infants’ selective visual attention towards the mouth or the eyes. The review focuses on this methodology because it is the only one that provides a clear measure of selective attention towards internal areas of a talking face; moreover, it offers the opportunity to make fair comparisons among studies that have adopted the same technology. We also review findings that test the relationship between the child’s selective visual attention and developmental outcomes. In particular, the goal of this review is to present a systematic analysis of studies that have considered selective visual attention to different areas of the face in the first two years of life, when children are acquiring the basic elements of their language.

The two main goals of this review are as follows:

1. To describe how the child’s looking behaviour towards a specific region of the face has been found to change based on different endogenous (i.e., age, monolingual/bilingual) and exogenous factors (i.e., native vs non-native, congruent vs incongruent, synchronised vs de-synchronised language). We first provide a methodological description of the studies and then focus on the interaction between the factors that may determine a child’s looking patterns.

2. To report the findings on how looking behaviour is related to concomitant and future language and social skills. In particular, we analyse the relationship between the infant's looking behaviour and other competencies, with a particular focus on language.

The discussion will outline several interpretations and speculative hypotheses based on the findings reported from a developmental perspective. We will identify possible future lines of research to investigate the processes underlying audio-visual speech perception and their relation to speech production.

**2. Method**

**2.1 Information sources and search criteria**

This systematic review is based on the PRISMA method, following an update of the QUOROM guidelines for reporting systematic reviews and meta-analyses (Moher, Liberati, Tetzlaff, & Altman, 2009). Four key databases were selected: 1) Psychinfo, 2) Pubmed, 3) Web of Science and 4) Scopus. A systematic search was conducted of all four. Basically, three areas of interest were covered: 1) “selective attention”, 2) “language development”; 3) the population of interest: infants. More specifically, for each area, specific keywords were used (Table 1), with the exclusion of terms referring to atypical development or the McGurk effect (which are outside the scope of this paper). Each of the keywords selected for the present study was individually inserted in the data sources string and then Boolean operators were used to find the papers of interest (Example of Search String: ((((((((ALL=(sensory OR multisensory OR multimodal OR scanning OR gaze OR attention))) AND ALL=((integration OR perception OR speech OR face))) AND ALL=(phonological OR lexical OR language )) AND ALL=(INFAN\*)) NOT ALL=(atypical )) NOT ALL=(autism))) NOT ALL=(McGurk)).

More than three thousand papers result from the search. After duplications (n = 1404) were removed, 2488 papers were screened (see Study selection).

|  |  |
| --- | --- |
| Areas of interest | Terms used |
| Selective Visual Attention | selective visual attention/ audiovisual speech perception/multisensory integration (sensory OR multisensory OR multimodal OR scanning OR gaze OR attention) AND (integration OR perception OR speech OR face) |
| Language Development | (phonological OR lexical OR language) |
| Infancy | infan\* |

**Table 2.** *Search Terms used*

**2.2 Eligibility criteria**

Inclusion and exclusion criteria were adopted to limit the review to relevant items. English was required as the language of publication; studies in other languages were excluded. Only articles published in peer-reviewed journals were considered (excluding purely theoretical papers, conference papers, PhD dissertations, etc.). The participants’ age range (0-2 years, the period when children acquire the basic elements of their native language and start to develop basic selective attentional skills: Garon, Bryson, & Smith, 2008) was also a criterion, as was typical (not atypical) development.

**2.3 Protocol**

The search strategy resulted in the identification of 19 records (see the PRISMA flow diagram in Figure 1).

Records identified through database searching (n = 3892)

Pubmed (n = 913)

PsychInfo (n = 905)

Scopus (n = 955)

WOS (n = 1119)







Records screened (title)  
(n = 2488 )













Studies included in the review from the systematic search

**N= 19**

**Figure 1.** *PRISMA Flow Diagram (Moher et al., 2009).*

**2.4 Study selection**

The methodology for the selection of the papers to be reviewed involved three phases, following different objectives (Gough, Oliver, & Thomas, 2017): I) screening based on paper title and type of publication only; II) screening of abstract and III) full text screening.

***2.4.1 Phase I: Title screening***

Preliminary screening based on title alone led to the exclusion of 2323 papers for the following reasons:

1. Conference papers, commentaries, dissertations, meta-analyses, book chapter, theoretical papers and systematic reviews not automatically excluded by the system.
2. Direct indication in title that the focus is on an area outside the scope of the present review.
3. Focus on adults or older children or on atypical development.

Papers whose aim was unclear from their title were set aside for consideration in Phases II and III. Thus, 165 were screened in the next phases.

***2.4.2 Phase II and III: abstract and full text screening***

The second and third screenings based on the abstract and the full text analysis addressed the specific inclusion and exclusion criteria, which led to the exclusion of 145 items. The remaining 19 papers were analysed in the next phase.

**3. Results**

The records analysing selective attention in infancy were published between 2012 and 2021 (last search run in October 2021). The literature search led to the identification of 19 papers which study the patterns of infants’ looking behaviour towards specific Areas of Interest (hereafter, AOIs) (the eyes, the mouth) of a talking face (Table 2).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nr** | **Authors**  **(First, year)** | **Age of Participants** | | **Monolingual**  **(n, language)** | **Bilingual**  **(n, languages)** | **Eye-tracking model** | **Assessment** | **Visual Stimuli**  **(Features)** | **Auditory Stimuli**  **(Features)** | **Index**  **(to tally the looking time)** |
| 1 | Birules et al. 2019 | 15 months to 4-6 years |  | | 47, Spanish/Catalan | Tobii X120 |  | Female Talking Face; Upright and Facing forward | Recited monologue in native and non-native language (IDS) | PTLT |
| 2 | Imafuku et al. 2019 | 6 months | i) 46, Japanese ii)33, Japanese | |  | Tobii TX300 | Record of child vocal responses | Female Talking Face; Manipulation of direction (Upright/Inverted) and of speaker’s gaze (Direct/Averted) | Vowels (/a/, /u/) recited in natural manner | PTLT |
| 3 | Morin-Lessard et al. 2019 | 5, 9, 12, 14 months; 2, 3, 4 & 5 years and adults | 156, English or French | | 136, babies and 129 adults, English/French | Tobii T60-XL or a Tobii TX300 | MB-CDI | Female Talking Face; Upright and Facing forward | Recited monologues in native or non-native language (IDS) | PTLT |
| 4 | Pons et al. 2015 | 4, 8 & 12 months | i) 60, Catalan or Spanish | | ii) 63, Spanish/Catalan | Tobii X120 | Language questionnaire (exposure) at 8 months and MB-CDI at 12 months | Female Talking Face; Upright and Facing forward | Recited monologues in native or non-native language (IDS) | PTLT |
| 5 | Tsang et al. 2018 | 6-12 months | 27, English | | 33, English/another language | EyeLink 1000 | Mullen scale of early learning (non-verbal cognitive and verbal skills, expressive and receptive language) | Female Talking Face; Upright and Facing forward | Sentences in native language  (IDS) | Eyes-Mouth-Index |
| 6 | Lewkowicz et al. 2012 | 4, 6, 8, 10, 12 months & adults | 80 babies, 21 adults (exp 1); 90 babies, 19 adults (exp 2), English | |  | Eye-trac Model 6000 |  | Female Talking Face; Upright and Facing forward | Recited monologue in a native and non-native language; (IDS and ADS) | PTLT |
| 7 | Pons et al. 2019 | 12 months | 34, Catalan or Spanish | |  | Tobii X120 | Adaptive Behavior Questionnaire from Bayley Scales of Infant and Toddler Development (BSID-III) | Female Talking Face;  Upright and Facing forward | Recited monologues in native and non-native language; (IDS) | PTLT |
| 8 | Ayneto et al. 2017 | 8 &12 months | A) 22 (8 months), 22 (12 months) B) 12 (8 months) and 12 (12 months), Catalan or Spanish | | A) 22, bilinguals (8 months), 22 (12 months) B) 12 (8 months) and 12 (12 months). Catalan/Spanish bilinguals infants | Tobii 60XL |  | Dynamic faces of adults and babies representing emotional expressions | - | PTLT |
| 9 | Boisferon et al. 2018 | 14 & 18 months | 44 (14 months) and 47 (18 months), English | |  | Eye-trac Model 6000 | MB-CDI | Female Talking Face; Upright and Facing forward | Recited monologues in native and non-native language (IDS and ADS) | Eyes-Mouth-Index |
| 10 | Kushnerenko er al. 2013 | 6. 9, 14-16 months | 37, English | |  | Tobii TX300 and ERP | Preschool language scale - 4 (Auditory Comprehension and Expressive Communication) and the MB-CDI | Female Talking Face; Upright and Facing forward | Syllables (ba/ga) presented in incongruent manner (i.e., Audio “ba”; video “ga”) | PTLT |
| 11 | Fort et al. 2018 | 15, 18 months | i) 40, Catalan or Spanish | | i) 40, Spanish/Catalan bilinguals, ii) 20 Spanish/Catalan bilinguals | Tobii TX300 |  | Female Talking Face; Upright and Facing forward; a non-verbal signal appeared after the speech stimuli, an eyebrow raising or lip protrusion of the same talking face | Six-syllable-long sentences in native language (ADS) | PTLT |
| 12 | Boisferon, et al. 2017 | 4, 6, 8, 10,12 months | i) 93, English ii) 81, English | |  | Eye-trac Model 6000 |  | Female Talking Face Upright and Facing forward | Recited monologues (IDS/ADS) in native and non-native desynchronised manner | PTLT |
| 13 | Imafuku et al. 2016 | 6-12 months & adults | 21 (6 months), and 21 (12 months) and 14 (adults), Japanese | |  | Eye-trac Model 6000 | MB-CDI | Female Talking Face; Upright and Facing forward | Short story in native synchronised or de-synchronised manner  (IDS) | PTLT |
| 14 | Mercure et al. 2019 | 4-8 months | 73 (4 months), 28 (8 months), English | | 22, English/another language; 23 bimodal bilingual with deaf mothers | Functional near infrared spectroscopy task and TobiiT120 | Mullen scale of early learning | Female Talking Face; Upright and Facing forward | Syllables (ba/ga) presented in congruent, incongruent manner (i.e., Audio “ba”; video “ga”) | PTLT |
| 15 | Tenenbaum et al. 2015 | 12, 18 & 24 months | 61, English | |  | Applied Science Laboratories (ASL) Pan-Tilt | MB-CDI | Female Talking with two objects; gaze shift towards object or head turn towards object | Short sentences in native speech  (IDS) | Eyes-Mouth-Index |
| 16 | Kròl, 2018 | 17-35 months | 40, Poland | |  | SMI RED250Mobile | Looking while listening | Female Talking Face; Upright and Facing forward | Nursery rhymes with increasing levels of noise (no noise, medium and high noise) in native speech.  (IDS) | PTLT |
| 17 | Sekiyama et al. 2021 | 6-36 months & adults | 120 (infants) and 12 (adults), Japanese | |  | Tobii X120 eye-tracker | MB-CDI | Female talking face; upright and facing forward | Short sentences in native speech  (IDS) + signal-to-noise ratio | PTLT |
| 18 | Oakes et al., 2021 | 7.5 – 10.5 months | 66, English  6, Other than English | | 23, English-other languages | SMI-RED M eye-tracker | - | White and Asian American  women reciting a nursery rhyme (rhymes was replaced with music) | Music | PTLT |
| 19 | Tomalsky et al. 2012 | 6-9 months | 32, English | |  | Tobii T120 eye-tracker | - | Female native English speaker | Syllables (ba/ga) presented in a congruent and/or redundant versus mismatched and non-redundant manner |  |

*ET = Eye-Tracker; MB-CDI = MacArthur‑Bates Communicative Development Inventories; IDS= Infant Directed Speech; ADS = Adult Directed Speech; PTLT = Proportion of Total Looking Time.*

**Table 2.** *Characteristics of each study*

**3.1 Methodological characteristics**

The studies considered infants of different ages, with different linguistic backgrounds, in cross-sectional (n = 17) or semi-longitudinal designs (n = 2: Kushnerenko et al., 2013; Tenenbaum et al., 2015): See Table 1. All were experimental studies that used an eye-tracker to identify looking behaviours, in line with our stated purposes. In one case non-invasive neuroimaging (Event-Related Potential) was also used, but the results of this technique are not included here as they go beyond the scope of the review (Kushnerenko et al., 2013). All the studies included followed similar experimental procedures, with minor differences, for instance, in the calibration process or in the type of audio and visual stimuli or apparatus (i.e., the eye-tracking model).

*Age and number of participants.* All studies investigated selective attention in infants aged 4 to 12 months (n = 13 studies, 27-179 participants); over 13 months (n = 5, 28-91 participants); or aged 5-24 months (n = 1, 209 participants). Some studies also included a control group of older children (3-6-year-old children: n = 2, 29 and 83 participants) or adults (n = 4, 12-129 participants), but the results for these groups are not included here.

*Language background.* Participants were reported to be exposed at home to one language (monolinguals, English: Hilairet de Boisfeiron et al., 2017, 2018, Kushnerenko et al., 2013, Lewkowicz et al., 2012, Tenenbaum et al., 2015, Tomalsky et al., 2012; Japanese: Imafuku et al., 2016, 2019; Sekiyama et al. 2021; Catalan or Spanish: Pons, Bosch, & Lewkowicz, 2019; Polish: Kròl, 2018) or two (Catalan-Spanish: Birules et al., 2019). Seven studies included both monolingual and bilingual infants: learning Catalan and/or Spanish (Ayneto & Sebastian‐Galles, 2017; Fort et al., 2018; Pons et al., 2015), English only or English and another language (Mercure et al., 2019; Oakes et al., 2021Tsang et al., 2018), English and/or French (Morin-Lessard et al., 2019).

*Types of stimuli presented.* The auditory and visual characteristics of the stimuli used are reported in Table 1. Each child was exposed to dynamic trials showing a female face reciting short sentences, monologues (12 studies), rhymes with an increasing level of noise (1), syllables (3), vowels (1) or faces with silent emotional expressions (1) or music (1). The stimuli are presented differently across studies but in general, for each experimental trial, a single face appears on the screen and the child’s looking time towards that face is tallied, using the Proportion of Total Looking Time (PTLT): Total time child spends looking at each AOI (eyes, mouth), divided by time spent looking at entire face (including also other portions of face). Three studies used an alternative to the PTLT, the eyes-mouth index (EMI) (e.g., eyes/[eyes + mouth]). The EMI index, unlike the PTLT, disregards looks to other parts of the face or to the screen.

In most studies the stimuli are presented in an AV congruent and synchronised modality (that is, what participants see is congruent and synchronised with what they are hearing). However, four studies used both synchronised and de-synchronised AV stimuli (2) or congruent and incongruent syllables (3). The stimuli were produced in the participants’ native/non-native language, in adult-directed speech (ADS) or infant-directed speech (IDS), and the visual stimulus was an upright face. In two studies dynamic faces expressing emotional states (Ayneto & Sebastian-Galles, 2017) or accompanied by music (Oakes et al., 2021) appeared on the screen in the absence of any auditory/speech stimulus.

*Developmental outcomes.* Three studies carried out analyses of the extent to which gaze preference at an early age (6 or 9 or 12 months) predicts a child’s language skills measured some months later (at 12, 14-16 or 18-24 months). In another seven studies measures of the child’ receptive and expressive vocabulary and communicative or social abilities were collected at the time the infant’s eye gaze was recorded through the use of questionnaires administered to parents; to assess word recognition proficiency, Kròl and colleagues (2018) also used a looking-while-listening procedure. Tsang et al. (2018) measured verbal skills and other non-verbal cognitive skills (object permanence, shape matching and imitative skills). Pons et al. (2019) assessed social and communicative functioning in addition to making a language assessment. Imafuku et al. (2019) considered the number of vowel imitations the infants produced during the experimental task. Due to the differences in the types of outcomes considered, we report the results separately, following two specific hypotheses, i.e., the language hypothesis and the social hypothesis (see paragraph 3.3).

**3.2 Interactions between factors that determine children's looking**

To achieve our first goal, i.e., to understand how the child’s looking behaviour towards specific regions of the face changes based on different endogenous and exogenous factors, we begin by describing each of the papers selected. Specifically, each paper will be classified based on characteristics of the participants (monolingual vs bilingual) and the type of stimulus used (e.g., native vs non-native language; congruent vs incongruent language). That is, we will describe how monolingual babies respond when exposed to the native language (3.2.1) or a non-native one (3.2.2); how bilingual children’s looking behaviour changes in response to a native (3.2.3) or a non-native (3.2.4) language; how the looking patters of monolingual or bilinguals babies change when they listen to synchronized vs. desynchronised, native vs, non-native short stories (3.2.5), congruent vs. incongruent syllables (3.2.6), a story presented in an ADS vs IDS manner (3.2.7) or when they are looking to faces expressing non-speech information (3.2.8).

***3.2.1 Monolingual children listening to a native language (IDS)***

Of the studies of monolingual infants exposed to native-language IDS with congruent AV stimuli in the first two years of life (n = 10), only five explored attentional shifts in relation to infant age by including more than one age group. In these studies the proportion of looking towards eyes and mouth was considered for infants under the age of 12 months (Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; Pons, Bosch, & Lewkowicz, 2015), for older infants (Hillairet De Boisferon et al., 2018; Morin-Lessard et al., 2019; Sekiyama et al. 2021) or for a range of ages from 5 months to 5 years (cross-sectionally) (Morin-Lessard et al., 2019). The remaining studies (Imafuku et al., 2019; Kròl, 2018; Pons et al., 2019; Tenenbaum et al, 2015; Tsang et al., 2018) considered only a single age, but their findings contribute to the overall picture of looking patterns at specific points in time.

When age is treated as a grouping factor, two main attentional shifts emerge: The first shift from eyes to mouth was found between 4 and 8 months. At 4 months monolingual babies exposed to their native language looked more at the eyes than at the mouth (Lewkowicz & Hansen Tift, 2012; Pons et al., 2015). In accord with Lewkowicz et al. (2012), Imafuku et al. (2019) and Sekiyama et al. (2021) confirmed greater attention to the eyes over the mouth in 5-6-month-old monolingual infants, while other authors found that between ages 4-6 months babies spent equal time looking at mouth and eyes (Morin-Lessard et al., 2019). At around 8 months infants generally display a preference for the mouth over the eyes (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), even if that is not always the case (Sekiyama et al., 2021, still found a preference for the eyes at that age, for example).

The second attentional shift was identified when the children begin to lose interest in looking toward the mouth, spending the same amount of time looking towards mouth and eyes by around 12 months (Morin-Lessard et al., 2019; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015; Pons et al., 2019). However, Tenenbaum, and colleagues (2015) found that 12-month-olds preferred to look more towards the mouth as compared to the eyes when presented with native-language sentences. Tsang and colleagues (2018) failed to find significant differences between mouth-lookers and eye-lookers based on age in monolingual babies aged 6-12 months, although they found a tendency toward increasing looking towards the mouth with age, which goes against the language expertise hypothesis.

The most discussed and still unresolved issue involves participants older than 12 months: Two studies have reported that monolingual children aged 14-18 months or older showed greater looking time towards the mouth region while listening to their native language presented in IDS (Hillairet De Boisferon et al., 2018; Morin-Lessard et al., 2019); in another study monolingual toddlers (17-35 months) exposed to their native language, with different levels of competing noise (Kròl, 2018), showed – as the authors expected – longer looking towards the mouth in the high- or medium- than in the no-noise condition, but with no differences between time spent looking at mouth vs. eyes in the no-noise condition.

To sum up, eight of ten studies agree in their findings regarding infant looking patterns in the first year of life. Two main attentional shifts have been identified: The first, from eyes to mouth, was found at 8-9 months; Morin-Lessard et al. (2019) report equal looking to mouth and eyes and Sekiyama et al. (2021) a bias for the eyes. The second attentional shift, at 12 months, is back towards the eyes. Only two studies disagree with this shift (Tenenbaum et al., 2015; Tsang et al., 2018), finding instead that 12-month-old infants, like younger ones, show more looking towards the mouth. The emergence of a third shift after the age of 12 months was highlighted in two out of three studies. However, the issue of ‘language expertise’ (i.e., pre-babbling infants should look towards the eyes; babbling infants should look towards the month and then, when they have begun producing words, they should come back to the eyes) that arose in these studies remains to be resolved: At the point when the infants have become more linguistically advanced, a decrease in looking towards the mouth is expected, in favour of more looking towards the eyes.

***3.2.2 Monolingual children listening to a non-native language (IDS)***

Within the studies exploring selective visual attention in monolingual babies exposed to non-native speech a common consensus is that, in this condition, attention towards the mouth is maintained at 12 months. In particular, a shift from the eyes to the mouth was identified between 4 and 8 months. However, unlike the case of native language stimuli, this preference is seen even when babies are older (12 months). Lewkowicz and Hansen-Tift (2012) were the first to identify this attentional shift when exposing English monolingual babies to Spanish stimuli. They noticed that, as with exposure to a native language, 4-to-8-month-old English-learning babies exposed to Spanish shift their attention towards the mouth, but at around 12 months, unlike the response to the native language, they continue to look longer towards the mouth region. This is interpreted as a sign of the infant’s need for redundancy for an unfamiliar stimulus. This looking pattern has been confirmed by two more recent studies (Pons, Bosch & Lewkowicz, 2015, 2019); similarly, Tsang, Atagi and Johnson (2018) showed that 6-to-12-month-old babies looked longer toward the mouth when exposed to non-native speech. Some more recent studies considered babies older than 12 months (e.g., Hillairet De Boisferon et al., 2018; Morin-Lessard et al., 2019). In the Hillairet De Boisferon study (2018), 14- and 18-month-old babies exhibit a preference for the mouth, regardless of the language of exposure (native or non-native). A different finding emerged in Morin Lessard et al. (2019), which found no greater looking towards the mouth at any age (not at 5-, 9-,12-, 14- or 18-months) but did find such a tendency for 3- to 4-year-old children. The findings with regard to infants under 18 months of age go against the language expertise hypothesis; that hypothesis is presumably irrelevant when testing with non-native languages as the children would have had no experience with the language of the stimuli at any age.

***3.2.3 Bilingual children listening to a native language (IDS)***

The looking behaviour of bilingual babies exposed to native-language IDS speech stimuli is different from that of monolingual babies: Only one significant attentional shift towards the mouth is expected, at around 8 months. Indeed, Pons et al. (2015) reported that 4-month-olds looked equally at eyes and mouth, but at 8 months infants started to look longer at the mouth than the eyes, and they continue to do so at 12 months (Pons et al., 2015) and 15 months (Birules et al., 2019). Tsang et al. (2018) found increasing interest in the mouth from age 6 to 12 months; these results are very similar to those reported for monolingual babies in the presence of non-native speech. Morin-Lessard and colleagues (2019) failed to find that 12-month-old bilinguals looked longer towards the mouth than to the eyes when exposed to their (dominant) native language, finding instead that bilingual children allocated equal attention towards the eyes and the mouth toward the end of the first year of life.

To summarise, three (Birules et al., 2019; Pons et al., 2015; Tsang t al., 2018) of the four studies that tested bilingual babies on native speech stimuli in the first year of life found greater interest in the mouth in babies aged from 4- to 15-months. This deployment of visual attention is different from the pattern found in monolingual babies exposed to native speech and instead resembles that of monolingual babies exposed to non-native speech.

***3.2.4 Bilingual children listening to a non-native language (IDS)***

Both Pons et al. (2015) and Tsang and colleagues (2018) found that at 12 months or younger bilingual infants look longer at the mouth than the eyes when non-native speech is presented, like the older babies (15 months) in Birules and colleagues (2019), but unlike the babies in Morin-Lessard et al. (2019). Indeed, Morin-Lessard and colleagues (2019) failed to find that 12-month-old bilingual babies learning English and French looked more at the mouth when exposed to a non-native control language (Russian), finding instead an equal proportion of looking time towards both AOIs. On the other hand, these authors found that 14-month-old babies exposed to non-native speech displayed marginally greater interest in the mouth.

Agreement among three out of four studies also emerged in relation to bilingual infants exposed to a non-native language: They show a preference for the mouth over the eyes. The authors of the one study that goes in a different direction (Morin-Lessard et al., 2019) interpret their unexpected result with reference to the fact that they used different speakers for the different languages. Finally, to delve more deeply into the findings relating to bilingual infants, a recent study tested the language distance hypothesis, finding that the closeness of the two languages to which the child is exposed in everyday life may affect the child’s looking preference. Birules et al. (2019) argued that selective attention to a talker’s mouth is modulated by language proximity and speculated that the distance between languages could continue to affect children’s preferential attention to mouth over eyes up to age 4-6 years if the two languages are closely related phonologically. This hypothesis is in need of further study.

***3.2.5 Monolingual babies listening to synchronized-desynchronised, native and non-native short stories***

Only two studies investigated how AV synchrony could explain specific looking patterns towards face’s regions. Studies agree that younger infants (4-6 months) and older infants (12 months) show equal looking towards the two regions of the face when presented with desynchronised native speech. Imafuku and Myowa (2016) found no main effect of condition (synchronised vs desynchronised), although they did find a significant interaction between condition and participant age: At 6 months infants spent more time looking at the mouth than the eyes in the synchronized native-language condition, but less preference for the mouth in the desynchronised condition, where babies spent an equal proportion of looking time towards the eyes and the mouth; in contrast, at 12 months babies exposed to synchronized-desynchronised speech displayed equal looking time toward the mouth and the eyes in the two conditions. Hillairet de Boisferon and colleagues (2017) explored synchrony using desynchronised native and non-native speech. They found that infants exposed to desynchronised stimuli, whether at under 8 months (at 4 and 6 mos.) or over (at 10 and 12 mos.), and whether the speech was native or non-native, showed no preference for eyes or mouth. An exception was the 12-month olds, who exhibit a preference for the mouth over the eyes with non-native stimuli. At 8 months, when exposed to desynchronised speech, infants looked longer towards the mouth than the eyes, regardless of whether the speech was native or non-native.

***3.2.6 Monolingual and bilingual babies listening to congruent vs. incongruent syllables***

When the audio stimulus (e.g., the syllable /ba/) is not congruent with what the infant is perceiving visually (articulation of a contrasting syllable, such as /ga/), infants could be expected to attempt to integrate the two modalities by looking more at the mouth, which supplies more articulatory information than the eyes. This effect was shown in three studies that tested the looking pattern of babies exposed to isolated syllables in their native language (monolingual – incongruent only: Kushnerenko et al., 2013, Tomalsky et al., 2012; monolingual and bilingual – congruent and incongruent: Mercure et al., 2019). Kushnerenko et al. (2013) found that at around 6-9 months infants tend to look more at the mouth than the eyes in an incongruent AV condition. Mercure and colleagues report an increase with age in looking time toward the mouth rather than the eyes for monolingual and bilingual babies from 4 to 8 months when speech is presented in an incongruent way, and also an increase in sensitivity towards incongruencies with age. More specifically, they found that younger infants (4 - 6.5 months) showed no difference in time spent looking at each AOI in either congruent or incongruent AV conditions, while olderinfants (6.6 - 8 months) showed a preference for mouth over eyes in the incongruent condition only (as also found by Tomalsky et al., 2012).

To summarise, the three studies that considered exposure to noncongruent AV speech syllables have shown that babies listening to and watching AV incongruent native stimuli look more at the mouth than the eyes: The older the child, the more they preferentially looked towards the mouth, in the incongruent condition only. This is taken as a sign of surprise or interest in the speech.

***3.2.7 Children listening to ADS vs IDS***

Investigation of IDS vs. ADS speech style is an additional perspective through which to analyse differences in infant looking patterns. The preference for the mouth over the eye region during exposure to IDS or ADS is related to the perceptual attributes of those speech styles, with IDS being more exaggerated and attractive to younger infants (words are clearly pronounced and easier to segment: Fernald, 1985). Contradictory results have been reported in relation to the interaction between the IDS vs. ADS conditions and the age of the participants.

Of these studies, two investigated the looking pattern in infants younger than 12 months listening to either IDS or ADS: The infants looked longer to the eyes in response to ADS and to the mouth in response to IDS (Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012). Greater looking to the mouth over the eyes in IDS was found up to the age of 14 months but not in ADS where they showed no clear preferences (Hillairet de Boisferon et al., 2018). Looking to the eyes in response to ADS in these younger babies might simply reflect a lack of engagement with speech that is relatively unfamiliar.

One study (Fort et al., 2018) analysed the same pattern in older babies exposed to ADS sentences. Fort and colleagues (2018) reported a general preference for mouth over eyes in 15- and 18-month-old monolinguals and bilinguals. These results together can be interpreted in light of the language expertise hypothesis: The younger the baby, the more they preferentially look at the mouth when exposed to IDS, due to its auditory and visual characteristics.

However, we cannot compare younger and older infants here since the older babies were tested using ADS only. However, Fort and colleagues’ results could be interpreted in light of the fact that infants of that age can distinguish and extract information through knowledge of the perceptual properties of their language (e.g., intonation), even when the speech is not directed towards them (Werker & Fennel, 2004).

***3.2.8 Children listening to non-speech information***

It is useful to consider a situation where babies are exposed to nonspeech facial information alone, in order to understand possible biases in looking towards the eyes or the mouth. Three studies sought to test whether the mouth preference displayed by bilingual and monolingual infants depends on speech alone or whether non-speech stimuli will also elicit that effect (Ayneto & Sebastian-Galles, 2017; Oakes, DeBolt, Beckner, Voss, & Cantrell, 2021; Fort et al., 2018). Ayneto et al. (2017) explored the looking patterns of monolingual and bilingual 8- and 12-month-olds exposed to dynamic faces expressing emotion. All the infants showed a preference for mouth over eyes, with a stronger preference for the mouth for the bilinguals in the presence of an infant face expressing emotion. In the presence of an adult face expressing emotion only 8-month-old bilinguals looked more to the mouth than to the eyes; instead, the 12-month-olds looked an equal proportion of time to eyes and mouth. Using a different design, Fort et al. (2018) analysed the looking behaviour of 15-month-old monolinguals and bilinguals and 18-month-old bilinguals presented with speech followed by non-speech (eyebrow raise or lip protrusion): A general preference for the mouth region over the eyes when looking at the non-speaking face (i.e., where looking behaviour is measured during the non-speech event) was confirmed for 15-month-old monolinguals and bilinguals exposed to lip protrusion. When exposed to the eyebrow condition, equal looking time towards the mouth and the eyes emerged for 15-month-old monolingual and bilingual infants only, with more looking towards the eyes for the bilingual group at 18 months.

Finally, Oakes and colleagues (2021) found that when 7.5- to 10.5-month-old children were visually exposed to White and Asian American women reciting a nursery rhyme and auditorily exposed to music, they preferred to look toward the lower part of the face (i.e., the mouth), regardless the type of stimulus(i.e., the ethnicity of the face).

To sum up, these studies agree in finding that 8-, 12-, and 15-month-old infants spend more time looking towards the mouth than the eyes, even when exposed to non-speech. These results support the findings of a significant difference between monolingual and bilingual looking patterns in response to speech, suggesting different strategies in speech processing for the bilingual group, even when presented with non-talking faces.

**3.3 Infants’ looking behaviour and developmental outcomes**

Of the selected studies, not many have compared selective attention with developmental outcomes (i.e., expressive or receptive vocabulary or social skills). However, seven studies directly investigated children’s linguistic skills at the time of the experiment (linguistic hypothesis, Imafuku et al., 2019; Hillairet De Boisferon et al., 2018; Kròl, 2018; Morin-Lessard et al., 2019; Pons, et al., 2019; Sekiyama et al., 2021; Tsang et al., 2018) and three studies did so at a later point in time (Imafuku & Myowa, 2016; Kushnerenko et al., 2013; Tenenbaum et al., 2015). Only two studies attempted to find a relationship between the child’s social skills and their looking behaviors (social hypothesis, Pons et al., 2019; Tsang et al., 2018). Many of the studies measured raw looking time towards eyes or mouth rather than a preference for one over the other. The remaining studies (not included in the next sections) investigated the child’s looking pattern without relating it to any other variable, although they did speculate as to the child’s future linguistic development (by speculating on a “temporal coincidence” between the time children showed attentional shifts from one region of the face and another, and the acquisition of important developmental milestones).

***3.3.1 The linguistic hypothesis (with language skills tested at the same time of the experiement)***

Expressive vocabulary has been found to be positively related to the time spent looking at the mouth at the time of the experiment. Two studies reported that infants who preferentially looked more towards the mouth between 6 and 12 months (Tsang et al., 2018), or at 9, 12 and 14 months (Morin-Lessard et al., 2019) showed higher expressive language skills (preverbal vocalizations) as measured through parental questionnaires. Note that, unlike Tsang et al., who found the same for bilinguals, Morin-Lessard et al. found this pattern in the monolingual group only. However, no relationship with receptive skills was observed in either study: Those children who preferentially looked more towards the eyes or the mouth scored similarly on receptive vocabulary, probably because children’s looking towards specific parts of the face is more robustly linked to production processes and elicits more production related responses (for example, looking towards the mouth may elicit imitation). More recently, Imafuku et al. (2019) identified a significant negative correlation between those infants who look more at the eyes than the mouth and the mean proportion of vowel imitations at 6 months (based on spontaneous production at the time of the experiment, r (44) = − .45, p = 0.002). That is, the more infants looked at the eyes, the fewer vocal imitations they produced. In older children (mean age 2 years), Kròl (2018) reported a significant positive relationship between looking towards the mouth and proficiency in comprehension (receptive skills): Children who preferentially looked toward the mouth rather than the eyes also had higher receptive vocabulary scores and thus greater linguistic proficiency. And Sekiyama et al. (2021) reported a significant partial correlation between children’s expressive vocabulary and looking towards the eyes in the mild-noise condition (p = .409, p =.10).

Finally, Pons et al. (2019) and Hillairet de Boisferon and colleagues (2018) failed to find any relationships – at 12 and 18 months, respectively – between the children’s looking behaviours and their linguistic level (in either comprehension or production), based on a parental questionnaire.

***3.3.2 The linguistic hypothesis (with language skills tested at a later point in time)***

Three studies investigated the longitudinal relationship between selective visual attention at the time of the experiment and language skills, at a later point in time: Imafuku et al. (2016) found the child’s 6-month looking preference toward the mouth to be significantly positively related to 12-month receptive (but not expressive) vocabulary when exposed to either synchronised or desynchronised speech (r = .72). Tenenbaum and colleagues (2015) found that raw attention to the mouth at 12 months predicts expressive language skills at 18 months (R2expr= .00-.20) and 24 months (R2expre= .04-.28): Attention to the mouth at 12 months explained the variance in productive and receptive lexicons six months later and in productive lexicons twelve months later (at which age receptive lexicon was not measured). Exploring infant gaze in response to incongruent AV syllables and comparing the findings longitudinally with linguistic skills, Kushnerenko and colleagues (2013) found a significant relationship between expressive language and looking preference towards the eyes (partial-r = .42): Those children who looked longer to eyes than mouth between 6 and 9 months obtained higher comprehension scores between 14 and 16 months.

***3.3.3 The social hypothesis***

The processes underlying AV perception were also studied in relation to the child’s social development, as eye contact is widely known to be essential for initiating communication. Pons and colleagues (2019) found that at 12 months scores on a questionnaire evaluating social development (social interaction and joint attention) are positively related to the proportion of attention to the eyes at that age (r = 0.437), suggesting that the infants were able to benefit from social cues: The higher the proportion of total looking towards the eyes, the higher the child’s social competence. This is the only study to directly test social competence in relation to the child’s gaze. Tsang and colleagues (2018) hypothesized a possible relationship between looks to the mouth and the ability to discriminate emotional faces in bilingual babies but failed to find any such relationship.

**4. Discussion**

Our systematic search led to the identification of 19 studies published between 2012 and 2021. These studies have investigated children’s looking patterns towards the eyes or mouth of a dynamic face in the first three years of life. All of the studies selected in the present systematic review were conducted in a lab-based situation using an eye-tracker designed for infant research, thus disregarding the role of the social context for the determination of these visual behaviours. For this reason, the reference to other methodologies and other studies in this research field is worth considering in the discussion of the results. Most of the studies can readily be compared, as they used similar paradigms and experimental procedures, but for others, the comparison is more challenging because of different approaches to analysing data as well as different research questions and hypotheses. Thus, it is sometimes difficult to find a common outcome in this literature, due to the large number of variables considered (different ages, different stimuli or types of experiences). However, addressing the separate elements that combine into comparable measures has helped us to gain a picture as to how these factors impact on infant linguistic or emotional development.

**Interactions between factors that determine children's looking: emergent topics and unresolved issues**

*The “temporal coincidence”*. From a developmental perspective, a looking bias towards the mouth at around 8-9 months (for monolingual and bilingual babies exposed to native and non-native speech) can be related to two important skills – the linguistic developments that typically occur around those ages (such as the emergence of canonical babbling: Oller, 2000) and the emergence of endogenous selective attention, or the ability to orient and allocate attentional resources (attentional shifts) towards specific stimuli of interest (Ruff & Rothbart, 1996). This implies that there may be a “temporal coincidence” between the child’s shifts towards a specific region of the face, their synchronous language development and their ability to voluntarily attend to what they want look at. Indeed, it has been hypothesised that when a child starts to babble, they become more interested in the mouth region because of the visual redundancy that the mouth affords and because their interest in speech is increasing sharply and they have begun to appreciate the relevance of the information provided by the mouth. In short, the mouth can be considered a visual cue for language learning. All this may be made possible by the fact that around this age children become better able to direct their attention endogenously, rather than only respond to whatever attracts their attention exogenously (Ruff & Rothbart, 1996). However, this interpretation remains speculative in view of the many variables in these studies, which make comparison difficult. In addition, no individual analyses of the participants involved have been carried out to assess whether there really is such a “coincidence”.

*Monolingual vs bilinguals.* The first important finding concerns what happens over the first year in monolingual and bilingual babies. From the various results taken together (3.2.1-3.2.4) we can conclude that, generally speaking, from one year of age, monolingual and bilingual infants differ in their looking pattern in response to a native but not to a non-native language (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015): While monolinguals deploy their attention to the eyes when presented with their native language, bilinguals continue to look more at the mouth. These studies ascribe that preference to the difficulties of learning two languages, or to different attentional strategies, especially when the languages the bilinguals are exposed to are similar (Birules et al., 2019). Indeed, these babies are in the process of building phonological representations for each language, which might require more effort for phonologically similar than for distant languages or a single language.

*The language expertise hypothesis at the end of the first year of life.* The second important finding has to do with what happens when monolingual children reach the end of the first year of life: When they start to produce their first words, monolingual babies become somewhat more ‘independent’ and no longer require the visual redundancy cue provided by the mouth; accordingly, they stop showing such a preference in the presence of native speech (Lewkowicz & Hansen-Tift, 2012) and look more towards the eyes. According to the language expertise hypothesis, attentional preferences also relate to the degree of the infant’s own language knowledge or likely discriminatory abilities. For example, 12-month-old monolingual infants are expected to be more advanced in the perceptual (auditory/visual), phonological and rhythmic aspects of their native language as compared with younger infants. However, individual differences in language competencies have never been tested or even checked against parental report to support that hypothesis. This seems to be a point that should be considered in future studies. Older infants are hypothesized to need fewer redundant visual cues and therefore to stop attending more to the mouth than to the eyes. This could explain why some studies have found an attentional shift to the eyes in monolingual children at the end of the first year of life in response to native speech – but not to non-native speech, in the presence of which they continue to look more towards the mouth (Lewkowicz & Hansen-Tift, 2012). This points to the functional role of the mouth as a visual cue for language acquisition: children can positively benefit from visual speech (lipreading or speechreading – that is, the speech that children can detect from seeing adults speak: Caldognetto & Zmarich, 2000, see also Boysson-Bardies & Vihman, 1991 for related evidence)*.* However, some studies have reported a return to a preference for the mouth in children older than 12-months, which goes against the language expertise hypothesis. One explanation has been provided but not yet investigated (Hillairet De Boisferon et al., 2018; Morin-Lessard et al., 2019): It could be that when the child begins to produce words, they redeploy their attention to the mouth because they have now reached a new developmental level and at the same time they are more motivated and interested in speech than younger children are. This may mean that multisensory redundancy continues to play a central role even after language use has begun to be established.

*Noise, ADS, incongruency and non-speech information.* As already well-established in previous studies, being exposed to noise, incongruent stimuli or non-speech visual information (3.2.5-3.2.8) could also elicit looking towards the mouth. These results further point to the role of the mouth as a visual cue for perceiving and processing both speech and non-speech information in the laboratory setting. Finally, being exposed to ADS, elicits greater looking towards the mouth (over the eyes) only in children older than 15 months (before that age children showed a greater looking towards the mouth over the eyes only when exposed to IDS). Thus, ADS is more accessible to older than to younger babies in native speech, but it is still harder to process, so looks to the mouth, with the accompanying articulatory cues, should be beneficial.

*Some unresolved issues.* In general, the first part of this review shed light on the role of endogenous and exogenous factors in shaping the child’s looking preference when exposed to dynamic faces. Different explanations for the prevalent preference for the mouth have been provided, though the evidence is to some extent contradictory: Looking at the mouth suggests a need for a visual cue to support language learning, but it might also reflect a more advanced stage in language learning or specific features of the language being learned (for example, Sekiyama et al., 2021, speculated that there is less reliance on visual support in the case of the Japanese language) or cultural or experiential differences (Oakes et al., 2021). Increased interest in the mouth is presumably due to the fact that the co-occurrence of visual and auditory information (i.e., the redundancy hypothesis, which carries the information presented in multiple sensory modalities, “selectively recruits attention” and “facilitates perceptual differentiations”, Bahrick & Lickliter, 2000) supports the production and articulation of the sounds of language, but it is perhaps also due to babbling infants becoming more attracted to speech and experiencing more of an interest in the speech source (Vilain, Dole, Loevenbruck, Pascalis, & Schwartz, 2019). A redundant stimulus may be useful at this stage because although speech can be perceived auditorily, visual information regarding the articulatory movements that produce it can further clarify it. However, there are notable individual differences in looking preferences (Morin-Lessard et al., 2019), as has been demonstrated for atypically developing babies (Young, Merin, Rogers, & Ozonoff, 2009). Some studies have suggested a possible role for specific language stages (e.g., babbling), but none have actually measured the extent of babble, based on production data, in relation to the experimental findings. Moreover, no longitudinal studies have yet been carried out and individual differences among babies have not been considered, as the experimental studies are all based on group results. In addition, most of the studies test the relationship between language and looking behaviour at 12 months only, without testing the child’s later competencies, at a more advanced linguistic or social stage.

**Infants’ looking behaviour and developmental outcomes**

*Language and social hypotheses.* Even though the relationship between the child’s attention and language development is now well-established in naturalistic studies (see Cetincelik, Rowland et al., 2021 for a review), no such pattern is clearly evident in studies using eye-tracking to investigate the child’s scanning patterns for specific areas of the face (i.e., selective visual attention). Contradictory results and explanations have emerged across the studies, but two main lines of interpretation can be identified. The first assumes a direct relationship between eye gaze and the child’s linguistic competence (*the linguistic hypothesis*), while the other supports a relationship of gaze with social skills (*the social hypothesis*). In summary, the time spent looking at the mouth is generally associated with an infant’s early expressive language skills, even after the end of the first year, sustaining the ongoing learning process – and this is in accord with all of the studies. Infant preference for mouth over eyes when presented with non-native speech could mean that babies recognise a novel linguistic pattern and deploy their attention to the mouth to maximally profit from the redundancy of the articulatory movements (Bahrick & Lickliter, 2014). Looking towards the mouth is not linked to receptive skills in this research, since the context in which these abilities were tested was limited and does not reflect real-life situations. Indeed, a larger context with objects and people (with face, hands, and bodies in motion) may help the child to understand and contextualise specific lexical referents (see also Rader & Zukow-Goldring 2010; 2012; Gogate et al., 2006; Matatyaho et al., 2008). On the other hand, the specific and unnatural context of the face appearing on a screen could elicit more production skills (e.g., imitation), as also confirmed by the significant correlations reported. Recently a few studies have advanced hypotheses relating the child’s looking preferences to their social development. Tsang et al. (2018) found no such relationship, but Pons et al. (2019) found that looking time toward the eyes is positively correlated with social development, i.e., the more a child looks at the eyes, the more socially advanced they are.

*Mediators and predictors.* Although no clear, consistent resultshaveyetemerged as regards the relation of looking behaviour to language skills, some conclusions can be attempted. Within the studies covered here, a distinction can be made between contemporaneous language skills as mediators of looking behaviour and looking behaviour as a possible predictor of language and social skills some months later. Two studies that have taken the former tack reported a positive relationship between a preference for the mouth over the eyes at the time of the experiment and concurrent expressive vocabulary. Also, Imafuku et al. (2019) tallied spontaneous imitative vocal responses produced throughout the experiment at 6 months and found a positive relationship with looks to the mouth. The relationship between looking behaviour and receptive vocabulary was investigated in three studies, but in two of them no such relationship was identified, while one study found a significant positive relationship between a preference for the mouth and receptive lexicon size. This suggests that the more the child looks at the mouth, the more advanced they are in lexical acquisition. Looking behaviour tested as a *predictor* of language skills was investigated in three studies. These studies go in different directions because they considered children of different ages and they tested looking behaviour under different conditions, making comparison difficult. However, a trend can be inferred from the results of two of the three studies: Looking preference towards the mouth during the experiment (Imafuku et al., 2013), but also absolute looking time towards the mouth (Tenenbaum et al., 2015), was found to predict a larger receptive vocabulary some months later.

*Unresolved issues.* The question remains open: Is there a relationship between a child’s looking behaviour and their linguistic skills? And in particular, how and why might such a relationship arise? Is language knowledge guiding or shaping the looking behaviour, or is it influenced by what the child is looking at, or both? Furthermore, the role played by looking at the mouth could be considered to be grounded in a child’s intrinsic need for more cues or as a result of the child’ experience. That is, the child may search for the visual cue because they require the redundancy provided by articulatory movements to reproduce them (i.e., the intersensory redundancy hypothesis, Bahrick and Lickliter, 2000), or the child may have understood the utility of attention to the mouth once they have begun to have experience with word production. Attention to a talking face is considered to have a powerful predictive role in the social sphere (Brooks & Meltzhoff, 2002; 2005; Young et al., 2009), but this issue has yet to be extensively explored in selective attention studies.

**Limitations, conclusions and future directions**

The exclusion of all studies investigating selective visual attention towards faces in semi-experimental or natural settings or studies testing children with atypical development constitutes a limitation here. Indeed, as stated in the Introduction, learning a language cannot be reduced to what is observed in an experimental setting in which the child is presented with a dynamic face. Moreover, in a natural context children are exposed to stimuli presented multimodally, through multiple senses (auditory, visual, tactile, motor) and with different types of input which play a crucial role in determining developmental outcomes. However, so far no studies have been conducted integrating spontaneous language data (or measures of prelinguistic vocal complexity: see, for example Santapuram et al., 2022, regarding an atypical population) and face scanning patterns to answer the question of the role of the internal features of the face as a support for the acquisition of language, even though a preponderant role for language background or type of stimulus is highlighted in almost all the studies reviewed. The present review suggests a need for further work of this kind to increase our understanding of how and to what extent visual selective attention affects language acquisition in the first years of a baby’s life. Future studies could embrace multi-method approaches in which ad hoc tasks are specifically built to replicate a situation in which children not only attend to a screen with an isolated talking face but are engaged in the task in their homes with their principal caregiver. For example, recently several studies have tried to find an innovative way of studying this phenomenon in a real-life situation, using head-mounted cameras (Braddik & Atkinson, 2011; Jayaraman, Fausey, & Smith, 2017). The use of a head-mounted camera is becoming more and more widespread in child development studies (see, for example, Jayaraman, Fausey, & Smith, 2017; Yu & Smith, 2013), because it provides a more precise measure of the child’s looking behaviour or gaze movements in a naturalistic environment, without the use of an eye-tracking device.

In addition, special attention needs to be paid to the effects of the widespread occurrence of SARS-CoV-2 as regards the use of face masks (see Grieco-Calub, 2021). These days wearing masks is common practice, both in daily life and in educational contexts. As Lewkowitz wrote in his opinion article ‘Masks can be detrimental to babies’ speech and language development’ (2021), visible articulations coming from the mouths of people talking to children play a key role in language acquisition and communication development. Lacking access to motor information or having a degraded signal affects the perception of speech. It impedes learning a new language because audiovisual speech processing supports language acquisition and development, especially in the first years of life (Fortin, 2020, see also Green, Staff, Bromley, Jones, & Petty, 2021). All this, again, supports the idea that the audio and visual integration of speech signals in natural situations is a basic process that helps the child to access all the necessary information to process and learn a language.

**5. References**

Ayneto, A., & Sebastian‐Galles, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Dev. Sci.*, *20*, 1–11. https://doi.org/10.1111/desc.12446

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS one*, *9*(1), e85701. https://doi.org/10.1371/journal.pone.0085701

Akhtar, N., & Gernsbacher, M. A. (2007). Joint Attention and Vocabulary Development: A Critical Look. *Language and linguistics compass, 1*(3), 195–207. https://doi.org/10.1111/j.1749-818X.2007.00014.x

Aslin R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and vision science*, 86(6), 561–565. https://doi.org/10.1097/OPX.0b013e3181a76e96

Atagi, N., & Johnson, S. P. (2020). Language Experience Is Associated with Infants’ Visual Attention to Speakers. *Brain Sciences, 10*(8), 550. MDPI AG. http://dx.doi.org/10.3390/brainsci10080550

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology, 36*(2), 190–201. https://doi.org/10.1037//0012-1649.36.2.190

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., & Pickens, J. N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant Behavior & Development, 11*(3), 277–296. https://doi.org/10.1016/0163-6383(88)90014-8

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. 1453–1508). Cambridge, MA: MIT Press.

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. (2019). Inside bilingualism:. *Dev. Sci.*, *22*, 1–11. https://doi.org/10.1111/desc.12755

Bernstein, L.E., Tucker, P.E., & Demorest, M.E., (2000). Speech perception without hearing.

*Percep. Psychophys*, *62*, 233–252. <https://doi.org/10.3758/BF03205546>

Beuker, K. T., Rommelse, N. N., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant behavior & development, 36*(1), 71–83. https://doi.org/10.1016/j.infbeh.2012.11.001

Boysson-Bardies, B. de & Vihman, M. M. (1991). Adaptation to Language: Evidence from Babbling and First Words in Four Languages. *Language, 67*(2), 297–319. https://doi.org/10.1353/lan.1991.0045

Brooks, R., & Meltzoff, A. (2002). The Importance of Eyes: How Infants Interpret Adult Looking Behavior. *Dev. Sci.*, *38*(6), 958–966. https://doi.org/10.1037/0012-1649.38.6.958

Brooks, R., & Meltzoff, A. (2005). The development of gaze following and its relation to language. *Dev. Sci.*, *8*, 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: a longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of experimental child psychology, 130*, 67–78. https://doi.org/10.1016/j.jecp.2014.09.010

Caldognetto, E., & Zmarich C. (2001). Phonological information from the visible articulatory movements: Italian data on lipreading for speech therapy. Linguistic Theory, Speech and Language Pathology, Speech Therapy, 3, CNR, Padova

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the SRCD*, *63*. https://doi.org/10.2307/1166214

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development. *Frontiers in psychology, 11*, 589096. https://doi.org/10.3389/fpsyg.2020.589096

Choi, D., Black, A. K., & Werker, J. F. (2018). Cascading and Multisensory Influences on Speech Perception Development. *Mind, Brain and Education, 12*(4), 212–223. https://doi.org/10.1111/mbe.12162

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: an eye-tracking study with 3- and 6-month-old infants and adults. *Journal of experimental child psychology*, *113*(1), 66–77. https://doi.org/10.1016/j.jecp.2012.04.012

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development, 8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fort, M., Ayneto‐Gimeno, A., Escrichs, A., & Sebastian‐Galles, N. (2018). Impact of bilingualism on infants’ ability to learn from talking and nontalking faces. *Lg. Learning*, *68*(S1), 31–57. https://doi.org/10.1111/lang.12273

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes, 28*(8), 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Fortin, J. (2020, June 23). *Masks Keep Us Safe. They Also Hide Our Smiles.* The New York Times. <https://www.nytimes.com/2020/06/23/style/face-mask-emotion-coronavirus.html>

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of experimental child psychology, 118*, 13–26. https://doi.org/10.1016/j.jecp.2013.08.012

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy, 17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to Maternal Multimodal Naming by 6- to 8-Month-Old Infants and Learning of Word-Object Relations. *Infancy, 9*(3), 259–288. https://doi.org/10.1207/s15327078in0903\_1

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review, 117*(2), 496–516. https://doi.org/10.1037/a0019049

Gough, D., Oliver, S., & Thomas, J. (2017). *An Introduction to Systematic Reviews*. NY: SAGE.

Green, J., Staff, L., Bromley, P., Jones, L., & Petty, J. (2021). The implications of face masks for babies and families during the COVID-19 pandemic: A discussion paper. *Journal of neonatal nursing, 27*(1), 21–25. <https://doi.org/10.1016/j.jnn.2020.10.005>

Grieco-Calub, T. M., & Olson, J. (2015). Individual differences in real-time processing of audiovisual speech by preschool children. *The Journal of the Acoustical Society of America, 137*, 2375–2375. doi: 10.1121/1.4920629

Guellaï, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: some critical questions. *Frontiers in Psychology, 5*, 812–812. https://doi.org/10.3389/fpsyg.2014.00812

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853–855. https://doi.org/10.1126/science.918670

Hillairet De Boisferon, A., Tift, A., Minar, N., & Lewkowicz, D. (2017). Selective attention to a talker’s mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Dev. Sci.*, *20*, e1238. https://doi.org/10.1111/desc.12381

Hillairet de Boisferon, A., Tift, A., Minar, N., & Lewkowicz, D. (2018). The redeployment of attention to the mouth of a talking face during the second year of life*. J. Exp. Child Psych.*, *172*, 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research, 53*(2), 298–310. <https://doi.org/10.1044/1092-4388(2009/08-0243)>

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. (2006). Infant speech perception activates Brocaʼs area. *NeuroReport*, *17*, 957–962. https://doi.org/10.1097/01.wnr.0000223387.51704.89

Imafuku, M., & Myowa, M. (2016). Developmental change in sensitivity to audiovisual speech congruency and its relation to language in infants. *Psychologia*, *59*, 163-172. https://doi.org/10.2117/psysoc.2016.163

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation. *Dev. Sci.*, *22*. https://doi.org/10.1111/desc.12825

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are Faces Denser in the Visual Experiences of Younger Than Older Infants? *Developmental Psychology, 53*(1), 38–49. https://doi.org/10.1037/dev0000230

Jerger, S., Damian, M., Tye-Murray, N., & Abdi, H. (2014). Children use visual speech to compensate for non-intact auditory speech. *J. Exp. Child Psych.*, *126*, 295–312. https://doi.org/10.1016/j.jecp.2014.05.003

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.

Keren-Portnoy, T., DePaolis, R. A. & Vihman, M. M. (under review). Dynamic interactions between production and perception build the foundation for word learning. *Child Development.*

Król, M. (2018). Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price. *PloS One*, *13*(3), e0194491–. https://doi.org/10.1371/journal.pone.0194491

Kuhl, P.K. (2007). Is speech learning “gated” by the social brain? *Dev. Sci.*, *10*, 110–120. <https://doi.org/10.1111/j.1467-7687.2007.00572.x>

Kushnerenko, E.V. Tomalski, P., Eballieux, H., Epotton, A., Ebirtles, D., Efrostick, C., & Moore, D.G. (2013). Brain responses and looking behaviour during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, *4*, 432. https://doi.org/10.3389/fpsyg.2013.00432

Lewkowicz, D. J. (2021, February 11). Masks Can Be Detrimental to Babies’ Speech and Language Development. Scientific American. <https://www.scientificamerican.com/article/masks-can-be-detrimental-to-babies-speech-and-language-development1/>

Lewkowicz, D., & Hansen-Tift, A. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *PNAS*, *109*, 1431–1436. <https://doi.org/10.1073/pnas.1114783109>

Libertus, K., Landa, R. J., & Haworth, J. L. (2017). Development of Attention to Faces during the First 3 Years: Influences of Stimulus Type. *Frontiers in psychology, 8*, 1976. https://doi.org/10.3389/fpsyg.2017.01976

Majorano, M., Vihman, M.M., De Paolis, R.A. (2014). The relationship between infants’ production experience and their processing of speech. *Language learning and development, 10*, 179–204. <https://doi.org/10.1080/15475441.2013.829740>

Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review, 60*, Article 100961. https://doi.org/10.1016/j.dr.2021.100961

Matatyaho, D. J., and Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy* 13, 172–184. doi: 10.1080/15250000701795655

Maurer, D., & Salapatek, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Development, 47*(2), 523–527. https://doi.org/10.2307/1128813

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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses. *PLoS Med*, *6*: e1000097. <https://doi.org/10.1371/journal.pmed1000097>

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior & Development, 21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

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. *Dev. Psy.*, *55*(8), 1640–1655. <https://doi.org/10.1037/dev0000750>

Nasim, O., Fahad, M.S., Ahmad, K., Khan, S., & Shah, D. (2017). Lip reading as reinforcement for speech reproduction in deaf children with hearing aids. *International Journal of Medical Students, 3*(1-2):7-12.

Nelson, C. M., & Oakes, L. M. (2021). "May I Grab Your Attention?": An Investigation Into Infants' Visual Preferences for Handled Objects Using Lookit as an Online Platform for Data Collection. *Frontiers in psychology, 12*, 733218. https://doi.org/10.3389/fpsyg.2021.733218

Oakes, L. M., DeBolt, M. C., Beckner, A. G., Voss, A. T., & Cantrell, L. M. (2021). Infant Eye Gaze While Viewing Dynamic Faces. *Brain Sciences, 11*(2), 231. http://dx.doi.org/10.3390/brainsci11020231

Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum Associates Publishers.

Pons, F., Bosch, L., & Lewkowicz, D. (2015). Bilingualism modulates infants’ selective attention to the mouth of a talking face. *Psych. Sci.*, *26*, 490–498. <https://doi.org/10.1177/0956797614568320>

Pons, F., Bosch, L., & Lewkowicz, D. (2019). Twelve-month-old infants’ attention to the eyes of a talking face is associated with communication and social skills. *IBAD*, *54*, 80–84. <https://doi.org/10.1016/j.infbeh.2018.12.003>

Rader, N. d. V.& Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture, 10*(2-3), 202–221. https://doi.org/10.1075/gest.10.2-3.05rad

Rader, N. d. V., & Zukow-Goldring, P. (2012). Caregivers’ gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences, 34*(5), 559–568. https://doi.org/10.1016/j.langsci.2012.03.011

Ruff, H.A., & Rothbart, M.K. (1996). *Attention in early development: Themes and Variations*. Oxford University Press: New York.

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

Sekiyama, K., Hisanaga, S., & Mugitani, R. (2021). Selective attention to the mouth of a talker in Japanese-learning infants and toddlers: Its relationship with vocabulary and compensation for noise. *Cortex, 140*, 145–156. https://doi.org/10.1016/j.cortex.2021.03.023

Stern, D. N. (1974). *Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors*. In M. Lewis & L. A. Rosenblum, The effect of the infant on its caregiver. Wiley-Interscience.

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The Signal in the Noise: The Visual Ecology of Parents' Object Naming. Infancy: the official journal of the *International Society on Infant Studies, 24*(3), 455–476. https://doi.org/10.1111/infa.12278

Tenenbaum, E., Sobel, D., Sheinkopf, S., Malle, B., & Morgan, J. (2015). Attention to the mouth and gaze following in infancy predict language development. *JChLg.*, *42*, 1408. <https://doi.org/10.1017/S0305000914000725>

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology, 10*(5), 611–624. https://doi.org/10.1080/17405629.2012.728076

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

Vihman, M. M. (2002). The role of mirror neurons in the ontogeny of speech. In Stamenov, M., & Gallese, V. (eds.), *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369*(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292

Vilain, D., Dole, M., Lœvenbruck, H., Pascalis, O., & Schwartz, J. L. (2019). The role of production abilities in the perception of consonant category in infants. *Dev. Sci.*, *22*, e12830–n/a. <https://doi.org/10.1111/desc.12830>

Wagner, J. B., Luyster, R. J., Yim, J. Y., Tager-Flusberg, H., & Nelson, C. A. (2013). The role of early visual attention in social development. *International journal of behavioral development, 37*(2), 118–124. https://doi.org/10.1177/0165025412468064

Werker, J. F., & Fennell, C. T. (2004). Listening to sounds versus listening to words. In D. G. Hall & S.R. Waxman (Eds.), *Weaving a Lexicon* (pp. 79–109). Cambridge, MA: MIT Press.

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months. *Dev. Sci.*, *12*, 798–814. <https://doi.org/10.1111/j.1467-7687.2009.00833.x>

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu C, Smith LB (2013) Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE, 8*(11): e79659. https://doi.org/10.1371/journal.pone.0079659