UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO DI SCIENZE UMANE

SCUOLA DI DOTTORATO DI SCIENZE UMANISTICHE

DOTTORATO DI RICERCA IN SCIENZE UMANE

CICLO XXXIV°/ANNO 2018

TITOLO DELLA TESI DI DOTTORATO

MULTI-SENSORIALITY IN LANGUAGE ACQUISITION: THE RELATIONSHIP BETWEEN SELECTIVE VISUAL ATTENTION TOWARDS THE ADULT'S FACE AND LANGUAGE SKILLS

S.S.D. M-PSI/04 – PSICOLOGIA DELLO SVILUPPO E DELL'EDUCAZIONE

Coordinatore: Prof.ssa MANUELA LAVELLI

Tutor: Prof.ssa MARINELLA MAJORANO
Firma

Co-Tutor: Prof.ssa MARILYN VIHMAN
Firma

Co-Tutor: Prof.ssa TAMAR KEREN-PORTNOY
Firma

Dottorando: TAMARA BASTIANELLO
Firma

Multi-Sensoriality In Language Acquisition: The Relationship Between Selective Visual Attention Towards The Adult's Face And Language Skills

Verona

ISBN pending

TABLE OF CONTENTS

SUMMARY (Italian)

Introduzione

Le componenti uditive e visive del linguaggio offrono al bambino informazioni cruciali per il processamento del parlato. L'abilità del bambino di integrare informazioni da diverse fonti multimodali (audio e visive) e di focalizzare l'attenzione sui segnali rilevanti presenti nell'ambiente circostante (*selective visual attention*) sono aspetti importanti che influenzano le prime fasi di acquisizione di una lingua.

Alcuni recenti studi hanno ipotizzato e testato la relazione tra attenzione selettiva visiva verso specifiche aree del volto parlante (occhi o bocca) e le abilità linguistiche di bambini nei primi anni di vita. Molti ricercatori hanno speculato su come questa relazione potesse essere mediata dal livello di *expetise* del bambino, a livello linguistico (*language expertise hypothesis),* ma nessuno studio, fin ad ora, ha cercato di approfondire questa ipotesi, andando ad investigare le abilità linguistiche dei bambini usando misure di linguaggio spontaneo.

Altri studi, hanno cercato di esplorare come diversi comportamenti attentivi verso specifiche aree del volto (occhi o bocca) fossero correlati alle abilità linguistiche concomitanti o longitudinali dei partecipanti. In molti casi, i risultati di questi studi hanno confermato l'esistenza di relazioni significative tra attenzione visiva selettiva e abilità linguistiche al tempo dell'esperimento o qualche mese dopo.

Obiettivi

L'obiettivo generale di questa tesi è quello di esaminare il fenomeno dell'attenzione selettiva visiva verso il volto e la sua relazione con lo sviluppo del linguaggio sia in un setting di laboratorio sia in un contesto naturalistico. In particolare, tre sono gli obiettivi specifici:

- il **primo obiettivo specifico** è quello di sintetizzare e analizzare i fattori individuati dalla letteratura di riferimento che possono determinare diversi *patterns* di attenzione selettiva visiva nei bambini durante un compito audiovisivo. Ed in particolare, descrivere come la letteratura spiega questi *patterns* in relazione agli aspetti dello sviluppo del linguaggio;

- il **secondo obiettivo specifico** è quello di analizzare sperimentalmente l'attenzione selettiva visiva del bambino verso specifiche aree del volto (occhi e bocca) durante un compito di esposizione audiovisivo. In particolare, lo studio è volto ad indagare due aspetti. Il primo aspetto riguarda l'età e la condizione linguistica (esposizione ad una lingua nativa *vs* una lingua non nativa) dei partecipanti e come queste influenzano l'attenzione selettiva visiva verso specifiche aree del volto. Il secondo aspetto riguarda l'esplorazione dell'esistenza di una correlazione tra comportamento attentivo dei bambini la produzione vocale al tempo dell'esperimento e all'ampiezza del vocabolario tre mesi dopo;
- il **terzo obiettivo specifico** è quello di capire se l'attenzione a volti o altre parti della scena visiva (oggetto, altre parti della stanza) è influenzato o spigato dalle abilità vocali del bambino al tempo del *task* e se gli episodi di fissazione al volto adulto possono essere predetti da specifiche proprietà fonologiche e semantiche del parlato del bambino.

Metodo

Per quanto concerne il **primo studio**, una rassegna sistematica della letteratura è stata condotta esplorando quattro fonti bibliografiche e usando specifici criteri di inclusione per selezionare la letteratura scientifica di interesse.

Per quanto riguarda il **secondo studio**, i movimenti oculari verso un volto parlante la lingua nativa (Italiano) e non-nativa (Inglese) di 26 bambini tra i 6 e i 14 mesi sono stati tracciati usando l'eye tracker. Due gruppi sono stati creati sulla base dell'età (G1, M = 7 mesi, N = 15 bambini; G2, M = 12 mesi, N = 11 bambini). Ogni competenza linguistica del bambino è stata valutata due volte, al tempo dell'esperimento, attraverso l'osservazione diretta e tre mesi dopo, attraverso il MB-CDI. Due gruppi sono stati creati sulla base della produzione vocale dei bambini (vocalizzi pre-canonici, babbling, parole) attraverso un *latent class cluster analysis*: una classe vocale "*alta*" (percentuale di babbling e parole più alta) e una classe vocale "*bassa*" (percentuale maggiore di produzioni pre-canoniche).

Per quanto concerne il **terzo studio**, il comportamento attentivo di 29 bambini tra i 12 e i 19 mesi è stato esplorato utilizzando sia una videocamera stazionaria (posizionata di fronte alla diade) e una go-pro (posizionata sulla fronte del caregiver di riferimento) durante un semplice task linguistico (*single object task*). Durante il task i bambini sono stati esposti ad un set di stimoli audiovisivi, parole vere e non parole, scelte sulla base dei report dei genitori e sulle risposte al MB-CDI. Il comportamento attentivo dei bambini è stato codificato offline, secondo per secondo per un totale di 116 sessioni. La codifica ha riguardato specifiche aree di interesse (il volto, l'oggetto, o altre parti della stanza). La produzione vocale per ogni bambino è stata quantificata usando LENA e le produzioni del bambino (vocalizzi pre-canonici, babbling, parole) durante un periodo di gioco con la mamma sono state trascritte foneticamente.

Risultati

La rassegna sistematica della letteratura (**Capitolo 2**) ha portato all'identificazione di 19 articoli. Alcuni dei quali volti a chiarire il ruolo giocato da diversi fattori nel spiegare diversi *patterns* attentivi. Altri interessati ad indagare la correlazione tra l'attenzione selettiva visiva verso specifiche aree del volto alle competenze linguistiche o sociali dei partecipanti, aprendo le porte a diverse linee interpretative.

Il primo studio empirico (**Capitolo 3**) ha messo in luce che i bambini italiani con età superiore ai 12 mesi, mostrano maggiore interesse verso l'area della bocca, specialmente quando esposti alla lingua nativa. Questo è in accordo con la recente letteratura, ma contrasta con la *language expertise hypotesis* (secondo la quale bambini attorno all'anno di età dovrebbero spostare il focus attentivo dalla bocca agli occhi). Il secondo risultato emerso in questo lavoro empirico riguarda l'interesse verso l'area della bocca per i bambini che hanno maggiori livelli di produzione in termini di babbling e parole al tempo dell'esperimento. Il terzo risultato riguarda l'associazione positiva tra il comportamento attentivo verso la bocca ed il vocabolario espressivo dei bambini misurato tramite questionario (MB-CDI) tre mesi dopo l'esperimento.

Dal secondo studio empirico (**Capitolo 4**) emerge una differenza significativa in termini di tempo attentivo verso il volto adulto tra i bambini del gruppo linguistico "*alto*" e *"basso"* durante un task condotto in un contesto naturalistico.

In particolare, da questo studio emergono due risultati interessanti: il primo è che i bambini che producono forme vocaliche più avanzate (babbling e parole) guardano di più verso il volto adulto, specialmente quando esposti alle non-parole. Il secondo riguarda l'esistenza di una relazione significativa tra gli episodi di fissazione al volto e le abilità vocaliche del bambino al tempo del task (vocalizzi pre-canonici, babbling e parole). In particolare, emerge che la quantità di babbling prodotto ha un ruolo nel predire gli episodi di fissazione al volto durante il task, sia per le parole sia per le non parole.

Conclusioni

Diverse ipotesi linguistiche e sociali sono state avanzate per spiegare le differenze emerse dalla rassegna della letteratura in relazione al fenomeno dell'attenzione selettiva visiva. Gli studi empirici presentati in questa tesi hanno portato due contributi originali in quest'ambito di ricerca. Da un lato, i nostri risultati confermano l'idea che la bocca e, più in generale, il volto forniscono segnali visivi cruciali nelle prime fasi di acquisizione del linguaggio. Dall'altro lato, i risultati hanno messo in luce che la conoscenza linguistica e le abilità linguistiche dei partecipanti aiutano a spiegare diversi comportamenti attentivi. In altre parole, è possibile dire che l'attenzione selettiva ai volti, o a specifiche aree di questi, è spiegata dalle conoscenze e abilità linguistiche attuali dei partecipanti.

ABSTRACT (English)

Introduction

Speech is the result of multimodal or multi-sensorial processes. The auditory and visual components of language provide the child with information crucial to the processing of speech. The language acquisition process is influenced by the child's ability to integrate information from multimodal (audio and visual) sources and to focus attention on the relevant cues in the environment; this is selective visual attention. This dissertation will explore the relationship between children's selective visual attention and their early language skills.

Several recent studies with infant populations have hypothesised or tested the relationship between children's selective visual attention towards specific regions of the talking face (i.e., the eyes or the mouth) and their language skills. These studies have tried to show how concomitant or longitudinal language skills can explain looking behaviours. In most cases, these studies have speculated on how this relationship is mediated by the child's level of language expertise (this is known as the language expertise hypothesis). However, no studies until now, to the best of our knowledge, have investigated the child's linguistic skills using spontaneous language measures.

Aims

The dissertation has one broad aim, within which there are three particular aims. The broad aim is to examine the phenomenon of selective visual attention toward the face in both a laboratory and a naturalistic setting, and its relationship with language development. The three particular aims are as follows.

The first aim is to synthesise and analyse the factors that might determine different looking patterns in infants during audiovisual tasks using dynamic faces; it describes how the literature explains these patterns in relation to aspects of language development.

The second aim is to experimentally investigate the child's selective visual attention towards a specific region of the adult's face (the eyes and the mouth) in a task using the eye-tracking method. In particular, the study will explore two

questions: First, how do age and language condition (exposure to native vs non-native speech) affect looking behaviour in children? Second, are a child's looking behaviours related to vocal production at the time of the experiment and to vocabulary rates three months later, and if so, how?

The third aim is to understand whether selective attention towards the face or other parts of the visual scene (i.e. the object or elsewhere) is influenced or explained by the child's vocal skills at the time of the task. And can the episodes of fixation towards the adult's face be predicted by specific phonological and semantic properties (i.e., pre-canonical vocalisations, babbling, words) of the child's speech?

Method

For the first study, a systematic review of the literature was conducted, exploring four bibliographic databases and using specific inclusion criteria to select the records.

For the second study, eye movements towards a dynamic face (on a screen), speaking in the child's native language (Italian) and a non-native language (English), were tracked using an eye-tracker in 26 infants between 6 and 14 months. Two groups were created based on age (G1, M = 7 months, N = 15 infants; G2, M = 12 months, N = 11 infants). Each child's language skill was assessed twice: at the time of the experiment (through direct observation, Time 1) and three months later (through MB-CDI, Time 2). Two groups were created, based on the child's vocal production (Time 1, latent class cluster analysis): a high class (higher percentage of babbling and words) *vs* a low class (higher percentage of pre-canonical vocalisations).

For the third study, the looking behaviour of the same 29 children between 12 and 19 months was tracked, using both a stationary video camera and a head-mounted camera on the mother's head during a single object task. During the task, children were exposed to a set of audiovisual stimuli, real words and non-words, chosen based on the parents' reports and their MB-CDI answers. The child's looking behaviour was coded offline second-by-second for a total of 116 sessions. The coding relates to specific areas of interest, i.e., the face, the object or elsewhere. The vocal production of each child was quantified using a LENA device, and their speech during a play period with their mothers was transcribed phonetically.

Results

The systematic search of the literature (**Chapter 2**) identified 19 papers. Some tried to clarify the role played by audiovisual factors in support of speech perception (provided by looking towards the eyes or the mouth of a talking face). Others related selective visual attention towards specific areas of the adult's face to the child's competence in terms of linguistic or social skills, this leads to correspondingly different lines of interpretation.

The first empirical study (**Chapter 3**) shows that Italian children older than 12 months displayed a greater interest in the mouth area, especially when they were exposed to their native language. This accords with the more recent literature but contrasts with the language expertise hypothesis. The second significant result of Chapter 3 is that children who had a higher level of production in terms of babbling and words at the time of the experiment looked more towards the mouth area. The study reported in Chapter 3 also demonstrated a positive association between the child's looking to the mouth and their expressive vocabulary as measured (using the MB-CDI) three months after the experiment

The second empirical study (**Chapter 4**) shows a significant difference in the looking time towards the adult's face between children with low- and high-vocal production in a naturalistic setting. More specifically, from this study, we find two things. Firstly, we found that the children who produced more advanced vocal forms (higher amount of babbling and word production) looked more towards the adult's face, especially when exposed to non-words. Secondly, that a significant relationship exists between the episodes of fixation towards the adult's face and the child's vocal skills (i.e., pre-canonical vocalisations, babbling, words); babbling productions predicted the episodes of face fixation in the task as a whole, for both words and non-words.

Conclusion

Linguistic and social-based hypotheses attempting to explain the differences in the selective visual attention phenomenon emerged from the literature review. The empirical studies presented in this thesis bring two original contributions to this research field. First, our findings reinforce the idea that the mouth and, more generally the face, provide crucial visual cues when acquiring a language. Secondly, our results demonstrate that language knowledge and language skills at the time the child was observed significantly help to explain different looking behaviours. In other words, we can conclude that each child's attention to faces is shaped by their own linguistic characteristics.

CHAPTER 1

General Introduction

Learning a language is a complex and dynamic process that requires the integration of several mechanisms. The acquisition of language is an experience-driven process that requires the child to actively select, extrapolate and integrate information from several sensory channels, especially the auditory and the visual one. Through face-to-face interactions, children experience the multisensory nature of speech from the early stages of their development.

Selective attention is one of the mechanisms that support language acquisition; it is, along with "sampling" (i.e., collecting information) and "learning" (i.e., "extracting structures and integrating information"), a significant precursor of language acquisition (D'Souza, D'Souza, & Karmiloff-Smith, 2017).

The present dissertation investigates the phenomenon of selective visual attention towards faces in children in the first two years of life. The adult face with its salient visual characteristics (i.e., the eyes and the mouth) is one of the most notable attractors in the child's environment from the very first hours of life. The child's attention to human faces was investigated here together with the child's language development.

Mechanisms supported by looking towards faces in infancy

> *For optimal learning, not only must an infant orient to a speaking face, they must also look at the relevant parts of the face (e.g., the eyes [for gaze-following] or mouth [for disambiguating speech sounds] and integrate visual and auditory information*. (D'Souza, D'Souza, & Karmiloff-Smith, 2017, p. 608)

Human faces capture babies' attention from a very young age because of their dynamic characteristics that make them attractive to children. Such a preference for faces is also known as a "bias for faces or face-like stimuli" (Frank, Vul, & Johnson, 2009; but see also Cassia, Turati, & Simion, 2004; Farroni, Johnson, Menon, Zulian, Faraguna, & Csibra, 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991; Simion, Macchi Cassia, Turati, & Valenza, 2001). Attention to faces in infancy has mainly been studied in the first weeks of a baby's life (Guellaï, Coulon & Streri, 2011; Guellaï, Mersad & Streri, 2015). From these studies, some clear patterns or preferences have emerged. For example, infants prefer their mother's face rather than other women's faces (Bushnell et al., 1989; Pascalis et al., 1995; Bushnell, 2001), especially when these faces are accompanied by speech (Sai, 2005). Newborns also prefer to look at human rather than non-human faces (Goren, Sarty, Wu, 1975; Johnson, Dziurawiec, Ellis, Morton, 1991), and at talking rather than non-talking faces (Bahrick et al., 2016). Less is known, however, about what happens later, during the first two years of life, when children start to accumulate linguistic, social and emotional experiences, and about how their looking preferences or behaviours are linked to particular developmental milestones (especially language development).

Why is it important to study attention to faces in infancy? Looking at a face plays a crucial role in infancy for several reasons. One of the main reasons is that, through faces, it is possible to communicate emotional states and to establish joint attunement (through social signals, i.e., eye contact, Gredebäck, Fikke, & Melinder, 2010; Hains & Muir, 1996; Parise, Reid, Stets, & Striano, 2008), establishing a basis for the development of attachment. Moreover, the voice derives from the mouth, as do cues to the motor properties of speech (i.e., articulatory movements or routines), which are relevant for children learning to speak. In other words, studying selective infant attention to faces is crucial in developmental psychology because looking towards specific areas of the face optimises language learning (Ayneto & Sebastian-Galles, 2017). However, the developmental hypotheses that have been advanced so far have not yet been tested to explain shifts or changes in the looking of children experimentally exposed to faces in the first year of life (Ayneto & Sebastian-Galles, 2017).

Several mechanisms are supported by looking towards the mouth and the eyes. The dynamic properties of the face (Guellaï, Streri, & Yeung, 2014) affect the audiovisual perception of speech (see Chapter 1). Among the studies showing a greater interest in the human face, most agree on infants' interest in **the eyes** in early infancy (Jones & Klin, 2013; Maurer & Salapatek, 1976). Haith, Bergman, and Moore (1977) reported that in infants between 9 and 11 weeks, attraction to the eyes increases as a function of both the age and the type of stimulus (i.e., talking faces vs auditory-only stimulus). This finding suggests that in audiovisual speech the level of engagement is higher, focusing children's attention on the eyes. The eyes of the person speaking attract the attention of the addressee. The eyes play an important role in communication, in human interactions, since they provide socio-emotional information. Furthermore, the eyes, the gaze, can engage children in a task or a conversation, as they convey emotion (Johnson et al., 2015). In addition, at a later point in the child's development, this helps to establish joint attunement with the speaker (such as in Joint Attention episodes), thus also supporting their vocabulary knowledge some months later (Brooks & Meltzoff, 2005) and their social development (Pons, Bosch, & Lewkowicz, 2019; Wagner, Luyster, Yim, Tager-Flusberg, & Nelson, 2013).

Looking towards **the mouth** is necessary when children need to distinguish two similar word forms (Maurer & Werker, 2014), since looking towards the mouth facilitates word recognition (Weatherhead & White, 2017). The retention and recall of articulatory sequences from the speaker's mouth is supported by exposure to perceptual and motor cues/signals present in the orofacial cavity, both before the end of the first year (e.g., Lewkowitz & Hansen-Tift, 2012) and after the first year (e.g., Tenenbaum, Sobel, Shenkopf, Malle, & Morgan, 2015; Tsang, Atagi, & Johnson, 2018). Visual (articulatory) and auditory (voice) information coming from the mouth help in discriminating words from pseudowords in acoustically normal conditions but especially in noise speech conditions (Fort, Spinelli, Savariaux, & Kandel, 2012; Grieco-Calub, & Olson, 2015; Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009; Lalonde & Holt, 2015).

But mouth looking also helps to distinguish different languages, because it is possible to extrapolate useful phonological indicators from the mouth (Berdasco-Muñoz, Nazzi, & Yeung, 2019, p. 1354; see also Navarra, Soto-Faraco, & Spence, 2014; Peña, Langus, Gutiérrez, Huepe-Artigas, & Nespor, 2016). For children aged around 11 to 15 months, access to information from the mouth articulating words unfamiliar to them could be more attractive and provide a more robust lexical representation. This interest directly supports the learning process since children are learning new words in their native language at that age (Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015). Indeed, seeing and observing mouth movements (i.e., lipreading) provides information about temporal and phonetic properties of the speech and helps listeners to decode vocal signals (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Grant & Greenberg, 2001; Yehia, Rubin, & Vatikiotis-Bateson, 1998).

Looking towards the mouth could also be a mirroring mechanism: as children become more specialised, they are more likely to look towards movements that they can reproduce themselves. Experimental studies have shown, for example, that children are better at matching heard and seen sounds that they have already mastered in their phonological repertoire (Streri, Coulon, Marie, & Yeung, 2016). At a neural level, there is some evidence that "concurrent visual stimuli modulate activity in the auditory pathways (Reale et al., 2007; Van Wassenhive, Grant, & Poeppel, 2004) and that both auditory and audiovisual speech activate the speech motor system in the perceiver (Okada & Hickok, 2009; Skipper, van Wassenhove, Nusbaum, & Small, 2007). There is also evidence that self-produced articulatory movements can alter adult perceptions of speech produced by others (Sams, Möttönen, & Sihvonnen, 2005; Scott, Yeung, Gick, & Werker, 2013), as can disrupting articulator-specific areas in the premotor cortex (Möttönen, Dutton, & Watkins, 2013)." (Choi, Bruderer, & Werker, 2019, p. 1389).

The literature, then, reinforces the idea of a strong relationship between the perception and production of speech, especially in the very first years of life, when children are specialising in their native speech (DePaolis, Vihman, & Keren-Portnoy, 2011; Majorano, Vihman, & Depaolis, 2014). Furthermore, when a language is not familiar, children continue to focus their attention towards specific and informative regions of the face, i.e., the mouth (Pons, Bosch, & Lewkowitz, 2015). Finally, having access to mouth movements support the baby's processing of language when they have to deal with more than one language (i.e., in bilinguals whose language acquisition seems to be more demanding, Ayneto & Sebastian-Galles, 2017).

Audiovisual modalities involved in the language learning process

> *Integrating information across sensory modalities enables the brain to benefit from both redundant and complementary information* (Atilgan & Bizley, 2021)

Learning a language in social contexts is a complex, dynamic and multi-modal process (Vigliocco, Perniss, & Vinson, 2014), which requires the child to select, extrapolate, retain and retrieve speech sequences from the environment around them. In children, the process entails the simultaneous activation of several mechanisms or abilities: perceptual, cognitive, articulatory/motor, attentional and mnemonic (Dietrich, Swingley, & Werker, 2007). The integration of different sensory modalities (especially the auditory and visual modalities) in language learning is now receiving greater attention, although inconsistencies remain among different theoretical approaches. Notwithstanding the fact that auditory information alone is not sufficient for recognising spoken words or learning a language (Havy, Foroud, Fais, & Werker, 2017; Havy & Zesiger, 2017), some models have been used to explain the recognition process through the auditory or visual input alone (Marslen-Wilson & Welsh 1978; McClelland & Elman; Norris, 1994; Norris, Mcqueen, & Cutler, 2000). For example, Teinonen, Aslin, Alku, and Csibra (2008) showed that exposure to visual information alone (face exposure) enhances 6-month-old infants' phoneme discrimination (without considering the auditory input). In addition, visual speech information can also influence word recognition and learning when children are listening to mispronounced words (Weatherhead & White, 2017).

However, this kind of information alone (only visual or only auditory) is not enough to build the lexical representations of speech (Havy et al., 2017; Havy & Zesiger, 2017). More recently, Weatherhead, Arredondo, Nácar Garcia and Werker (2021) have shown that visual information from talking faces "does augment an auditory-only signal, providing additional linguistic information in support of learning and recognition of novel words" (p. 15). In a previous study, Weatherhead and White (2017) showed that 11-15-month-old infants exposed to auditory and visual information paid more attention to unfamiliar words than to familiar ones. Weatherhead and White's findings contrasted with previous studies (Hallé & deBoysson-Bardies, 1994; Swingley, 2005), which found that children displayed greater attention towards familiar words. Weatherhead and White explained these contrasting results by reference to the methodological approach they adopted: while previous studies exposed children to the auditory information alone, they provided both audio (sound of the word) and visual (face) exposure.

In line with this study, a more comprehensive approach is needed to explain the integration of auditory and visual/facial information/input as a supportive mechanism for the identification or recognition of phonetic features or for the activation of the lexical candidates for the consonant target (Gogate, Bahrick, & Watson, 2000; Jouanjean-L'Antoune, 1997). Some studies have speculated on the relevance of variations in visual attention towards the mouth, or specific regions of the face, in various circumstances. These include specific points and stages of development (Lewkowicz & Hansen-Tift, 2012), multiple language exposures at birth (e.g., in children learning phonologically close vs distant languages, Birulés, Bosch, Brieke, Pons, & Lewkowicz, 2019), different levels of noise (Jerger, Damian, Tye-Murray, & Abdi, 2014; Reisberg, McLean, & Goldfield, 1987), and the presence of physical impairments (Summerfield, 1992). Moreover, recent evidence has highlighted the idea that observing the articulatory movements coming from the face and more specifically from the mouth of a talking face (visual stimuli) provides relevant information about the temporal and phonetic properties of the speech and of the acoustic signals (Lewkowicz & Hansen-Tift, 2012).

Face Scanning and language development

Pascalis and colleagues (2014) supported the idea that attention to the face and language "(is) essential for communication, but they have been studied traditionally as separate areas with little interaction between the two domains, even when their links are acknowledged" (p. 65). But although the relationship is now well established between specific patterns of visual attention (e.g., Joint Attention or the face-hand-shift) and language development (Akthar &

Gernsbacher, 2007; Beuker, Rommelse, Donders, & Buitelaar, 2013; Brooks & Meltzoff, 2005, 2015; Carpenter, Nagell, & Tomasello, 1998; Morales, Mundy, & Rojas, 1998), less is known about the extent to which looking behaviours and scanning patterns towards particular regions of the adult's talking face may impact on or explain language development.

It is now well-established that integrating auditory and visual information from a talking face support or explain the child's later language comprehension (Kròl, 2018) or production (Morin-Lessard, Poulin-Dubois, Segalowitz, & Byers-Heinlein, 2019) in the first years of life. As also recently reported by Çetinçelik and colleagues (2021), eye gaze towards adult faces seems to facilitate the learning process by enhancing attention and memory encoding. It is indeed through gaze following or gaze orientation that babies access some of the meanings of the world around them and start to share communicative intentions with their principal caregiver (Cribsa & Gergely, 2009).

Other (mostly experimental) studies have tried to clarify the relationship between attention towards specific parts of the face and the child's language skills. For example, studies have attempted to test the relationship between the child's looking preference for the mouth region and their vocal skills, i.e., the imitation rate at the time of the experiment (Imafuku, Kanakogi, Butler, & Myowa, 2019) and later lexical knowledge (Imafuku et al., 2016). These studies have found a positive relationship. Other studies have reported that the visual preference for the eyes is related to the child's social abilities (Pons, Bosch, & Lewkowicz, 2019). Thus, the scientific evidence agrees that audiovisual integration mechanisms are related to, and sustain, facilitate and enhance language acquisition. This applies especially when the child's language system is not yet specialised, that is, in the very first years of a baby's life (Rosenblum, Schmuckler, & Johnson, 1997), before the emergence of the first words.

Many studies have demonstrated this integration process in static, controlled or experimental conditions, potentially underestimating the multi-modal and dynamic events that happen in the baby's daily interactions with their caregiver. Moreover, in the literature considered, word recognition mechanisms or lexical knowledge are not tested directly (most of the linguistic measures come from self-report questionnaires completed by parents). These studies do not depict the authentic situation of the children's everyday life in which they encounter and learn new lexical items.

Children's attention to adult faces in interactive and laboratory settings: open issues

Speech production is the natural result of multi-sensorial mechanisms occurring together (visual-auditory and motor mechanisms) and of the child's accumulated experience. Infants are active actors in their language acquisition since they gather information by moving and interacting with the world and people around them (Lewkowicz & Lickliter, 1994; Thelen & Smith, 1994; Yoshida & Burling, 2011). But they also play a passive role since they receive (with no control over it) an unquantifiable number of inputs from different caregivers and different cues from the environment around them. To understand how a child approaches words and new lexical items in the first two years of life we need to consider all these aspects from a dynamic perspective, embracing the concept of complexity (D'Suoza, D'Suoza, & Karmiloff-Smith, 2017). For this reason, it is appropriate to adopt a multi-modal approach.

Allocating attentional resources towards specific aspects of a visual scene is part of the learning process. In visually selecting the stimuli, children activate their attentional resources and cognitive skills along with their previously acquired knowledge. In learning situations, multisensoriality – the simultaneous presence of both the bimodal signal coming from the face/voice and the visual, tactile, or gestural sign – plays a role in facilitating "the detection of arbitrary syllable–object relations or (the) mapping of syllables onto objects by preverbal infants" (in Gogate & Bahrick, 2000, p. 879). For example, empirical evidence shows that the practice of adults of naming words linked to target objects on which children are focusing their interest does elicit faster word-learning (Akhtar, Dunham, & Dunham, 1991; Olsen-Fulero, 1982; Tamis-Lemonda & Bornstein, 1989). This means that linking the visual properties of the object with the vocal target that refers to that object makes it easier for the child to retain that word in memory ("multi-modal motherese", Gogate & Bahrick, 2000). It has also been shown that the child's ability to allocate their visual attentional resources towards the parent's eyes or face, in order to establish eye contact with them (a precursor of triadic interaction, e.g., joint attention), is closely related to the child's later vocabulary development (Akthar & Gernsbacher, 2007; Beuker et al., 2013; Brooks & Meltzoff, 2005, 2015; Carpenter et al., 1998; Morales et al., 1998).

Lab-based experiments cannot provide evidence on how infants deal with input containing different structures and cues and how children retain statistical regularities from speech around them before comprehending and producing a word. Moreover, studies that have investigated selective visual attention towards the face, or specific parts of it, have typically shown babies photographs or short video clips with talking faces or objects (Gredebäck, Eriksson, Schmitow, Laeng, & Stenberg, 2012) which does not reflect reality. Children need to capture much more, and more varied, information every day because they are contained in complex situations.

To cope with this gap in laboratory studies, some observational research has investigated which parts of their environment children explore more during their second year of life, using a head camera placed on the child's or the caregiver's head. For example, Schmitow and Stenberg (2015) showed that children tend to focus on hand actions and that this preference is related to language development. Other studies have shown that when two adults are chatting in real-life situations, children are not primarily interested in attending to the persons speaking to each other. This result contrasts partially with some eye-tracking studies (Augusti, Melinder, & Gredebäck, 2010; von Hofsten et al., 2009), which found that children preferred to look more towards people speaking than towards non-speakers. Very young children seem not to focus on the faces of people talking; this has been explained by a lack of interest in language, since these children are still in a preverbal phase. They are, however, more attracted by the object or other stimuli in the room, as shown in studies by Yu and colleagues (2013) or by Yoshida and Smith (2008). Moreover, studies conducted with a head-mounted camera (see Smith, Yu, Yoshida, & Fausey, 2015 for a review) have shown that babies younger than four months frequently look at faces

(Jayaraman, Fausey, & Smith, 2015; Sugden, Mohamed-Ali, & Moulson, 2014), while older children seldom looked towards the mother's face during free play (Yu and Smith, 2013) or when she talked to them (Franchak, Kretch, Soska, & Adolph, 2011).

It is interesting to notice that the results obtained from a controlled experimental eye-tracking procedure are somewhat different from those from observation in a natural situation (Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Franchak et al., 2011; Yu & Smith, 2013). In the latter case, children can select and explore the environment around them and attend only to stimuli that interest them. It is vital to understand the ecological mechanisms underlying language acquisition at this stage.

To the best of our knowledge, it is still unclear which aspects of a visual scene contribute to language development, and how they do so (Libertus, Landa, & Haworth, 2017). Nor is it clear how basic-level attentional abilities are used to construct higher-level language skills (D'Suoza, D'Suoza, Johnson, & Karmiloff-Smith, 2015). It is an open question whether the child's current language skills shape and drive their looking behaviour, or whether it is rather their attentional skills and looking behaviours that drive language development.

While laboratory and controlled settings permit the collection of precise measures, the method of observation in the home setting provides researchers with a more ecological and realistic view of what acquiring a language in everyday life means. But very few studies have simultaneously considered the child's visual exploration and its relationship with language, so the relation between what the child observes and what the child can learn is still unknown (Roy et al., 2006). Moreover, not many studies have considered the role of the child's current level of phonological and lexical development and how these abilities can explain or influence their gaze and looking behaviour or vice-versa. A few studies have used both experiments and observations to investigate the relationship, if any, between the presence of multi-modal cues and the child's production skills. And very few studies in this field have also taken the linguistic input into account or considered the long-term effects of a looking preference towards some parts of the face, or of the environment, on the child's language development. For these reasons, further research is needed on these missing areas.

Outline of the dissertation

By conducting three studies, both in a laboratory and in a spontaneous setting, this dissertation aims to examine the phenomenon of selective visual attention toward faces (or areas of the face: mouth and eyes) in infants and its relationship with language development.

The first study consists in a systematic review of the literature (Chapter 2) and has two aims. The first is to describe how infant looking behaviour towards specific regions of the talking face (i.e., the eyes or the mouth) varies with several endogenous (e.g., age, monolingual/bilingual) and exogenous factors (e.g., language background, type of condition). The second is to report and clarify the findings of studies using eye-tracking technology on the relationship between the child's looking behaviour (towards the mouth or the eyes) and their current or future language and social skills.

Chapters 3 and 4 present two empirical studies involving Italian children with typical development. In Chapter 3, we obtain a measure of selective visual attention towards a specific region of the talking face (the eyes or the mouth) by using eye-tracking while the children were listening to a dynamic face talking in their native or non-native language. In Chapter 4, a measure of the child's looking towards human faces in a natural/spontaneous setting was obtained by using two cameras: a head-mounted camera and a stationary video camera positioned in front of the dyad during a simple task. A measure of vocal production was obtained by means of both direct measures of language development (observations and LENA measures) and indirect self-report measures (questionnaires).

The first empirical study aims at describing the looking behaviour (i.e., selective visual attention) of 26 Italian children at different ages (G1 = around 7 months; G2= around 12 months) (Chapter 3, Study 1) and its relationship with their language skills at the time of the experiment and three months later (Chapter

3, Study 2). In this chapter, unlike previous studies that used indirect measures of language skills (i.e., questionnaires), this study obtains spontaneous vocal productions from direct observation of the mother-child interactions during play. This study is a pilot study to test the replicability of previous results (e.g., Lewkowicz & Hansen-Tift, 2012). But its main aim is to explore in depth the results of those studies integrating the language measures retrieved from questionnaires and the child's selective visual attention towards specific areas of the talking face. Until now, to the best of our knowledge, no one has considered the child's linguistic skills at the time of the experiment and three months later using spontaneous language measures.

Chapter 4 reports on an additional empirical study, which took into account the social nature of language and communication: it investigates the same phenomenon but with the children interacting with their mothers during a simple task. The findings have been obtained through a methodologically innovative study conducted with 29 children between 12 and 19 months, an age at which they were expected to produce some words. The children's looking behaviour was tracked by the combined use of a head camera (on the mother's head) and a stationary video camera (in front of the dyad). For this study, the children's language was studied and assessed by two parallel methods. These are the LENA device (Language ENvironment Analysis system), an innovative tool to collect quantitative measures of speech, and the analysis of spontaneous data from direct observation during a play section.

The final chapter (Chapter 5, Conclusions) summarises the main research findings, defines some possible future research perspectives on this research topic, and explains the practical impact of the present work.

References

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

Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language, 18*(1), 41–49. https://doi.org/10.1017/S0305000900013283

Atilgan, H., & Bizley, J. K. (2021). Training enhances the ability of listeners to exploit visual information for auditory scene analysis. *Cognition, 208*, 104529. https://doi.org/10.1016/j.cognition.2020.104529

Augusti, E.M., Melinder, A., & Gredebäck, G. (2010). Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Frontiers in Psychology, 1*(1), 161–161. https://doi.org/10.3389/fpsyg.2010.00161

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

Bahrick, L. E., Todd, J. T., Castellanos, I., & Sorondo, B. M. (2016). Enhanced attention to speaking faces versus other event types emerges gradually across infancy. *Developmental psychology, 52*(11), 1705–1720. https://doi.org/10.1037/dev0000157

Berdasco-Muñoz, E., Nazzi, T., & Yeung, H. H. (2019). Visual scanning of a talking face in preterm and full-term infants. *Developmental Psychology, 55*(7), 1353–1361. https://doi.org/10.1037/dev0000737

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

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2019). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental science, 22*(3), e12755. https://doi.org/10.1111/desc.12755

Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. In M. H. Bornstein (Ed.), *Maternal responsiveness: Characteristics and consequences* (pp. 49–61). Jossey-Bass.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental science, 8*(6), 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

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development, 10*(1-2), 67–74. https://doi.org/10.1002/icd.248

Bushnell, I. W., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. British *Journal of Developmental Psychology, 7(*1), 3–15. https://doi.org/10.1111/j.2044-835X.1989.tb00784.x

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i–143.

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

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Choi, D., Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychonomic bulletin & review, 26*(4), 1388–1399. https://doi.org/10.3758/s13423-019-01601-0

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences, 13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

d'Souza, D., d'Souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *Journal of Child Language, 44*(3), 591–627. https://doi.org/10.1017/S030500091700006X

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. *Developmental science, 17*(2), 270–281. https://doi.org/10.1111/desc.12122

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior & Development, 34*(4), 590–601. https://doi.org/10.1016/j.infbeh.2011.06.005

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native Language Governs Interpretation of Salient Speech Sound Differences at 18 Months. *Proceedings of the National Academy of Sciences, 104*(41), 16027–16031. https://doi.org/10.1073/pnas.0705270104

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' Preference for Face-Relevant Stimuli: Effects of Contrast Polarity. *Proceedings of the National Academy of Sciences, 102*(47), 17245–17250. https://doi.org/10.1073/pnas.0502205102

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2012). Audiovisual vowel monitoring and the word superiority effect in children. *International Journal of Behavioral Development, 36,* 457–467. doi: 10.1177/0165025412447752

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking. *Child Development, 82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

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

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child development, 71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544–549.

Grant, K.W., & Greenberg, S. (2001). Speech intelligibility derived from asynchronous processing of auditory-visual information. *Proceedings of the Workshop on Audio-Visual Speech Processing*.

Gredebäck, G., Eriksson, M., Schmitow, C., Laeng, B., & Stenberg, G. (2012). Individual differences in face processing: Infants' scanning patterns and pupil dilations are influenced by the distribution of parental leave. *Infancy, 17*(1), 79–101. https://doi.org/10.1111/j.1532-7078.2011.00091.x

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science, 13*(6), 839–848.

Grieco-Calub, T. M., and 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., Coulon, M., & Streri, A. (2011). The role of motion and speech in face recognition at birth. *Visual Cognition, 19*(9), 1212–1233. https://doi.org/10.1080/13506285.2011.620578

Guellaï, B., Mersad, K., & Streri, A. (2015). Suprasegmental information affects processing of talking faces at birth. *Infant Behavior and Development, 38*, 11–19.

Guellaï, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. F*rontiers in Psychology, 5*, Article 812.

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67*(5), 1940–1951. https://doi.org/10.2307/1131602

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

Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior & Development, 17*(2), 119–129. https://doi.org/10.1016/0163-6383(94)90047-7

Havy, M., & Zesiger, P. (2017). Learning spoken words via the ears and eyes: Evidence from 30-month-old children. *Frontiers in Psychology, 8*, Article 2122. https://doi.org/10.3389/fpsyg.2017.02122

Havy, M., Foroud, A., Fais, L., & Werker, J. F. (2017). The Role of Auditory and Visual Speech in Word Learning at 18 Months and in Adulthood. *Child development, 88*(6), 2043–2059. https://doi.org/10.1111/cdev.12715

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. *Developmental Science, 22*, e12825. https://doi.org/10.1111/desc.12825

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The Faces in Infant-Perspective Scenes Change over the First Year of Life. *PloS One, 10*(5), e0123780–e0123780. https://doi.org/10.1371/journal.pone.0123780

Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: a new multimodal picture–word task. *Journal of Experimental Child Psychology, 102,* 40–59. doi: 10.1016/j.jecp.2008.08.002

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

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience and Biobehavioral Reviews, 50*, 169–179. https://doi.org/10.1016/j.neubiorev.2014.10.009

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

Jouanjean-L'Antoene, A. (1997). Reciprocal interactions and the development of communication and language between parents and children. n. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 312-327). Cambridge, UK: Cambridge University Press.

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

Lalonde, K., and Holt, R. F. (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research, 58*, 135–150. doi: 10.1044/2014_JSLHR-H-13-0343

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

Lewkowicz, D. J., & Lickliter, R. (Eds.). (1994). *The development of intersensory perception: Comparative perspectives*. Lawrence Erlbaum Associates, Inc.

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*, Article 1976. https://doi.org/10.3389/fpsyg.2017.01976

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204. https://doi.org/10.1080/15475441.2013.829740

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

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

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: a comparison of language and faces. *Developmental psychobiology, 56*(2), 154–178. https://doi.org/10.1002/dev.21177

McClelland, J. L., Elman, J. L. (1986). The TRACE model of speech perception *Cognitive Psychology, 18,* 1–86.

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. *Developmental Psychology*, *55*(8), 1640–1655. https://doi.org/10.1037/dev0000750

Möttönen, R, Dutton R, & Watkins KE (2013). Auditory-Motor Processing of Speech Sounds. *Cerebral Cortex, 23*(5), 1190–1197. 10.1093/cercor/bhs110

Navarra, J., Soto-Faraco, S., & Spence, C. (2014). Discriminating speech rhythms in audition, vision, and touch. *Acta psychologica, 151*, 197–205. https://doi.org/10.1016/j.actpsy.2014.05.021

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *The Behavioral and brain sciences, 23*(3), 299–370. https://doi.org/10.1017/s0140525x00003241

Okada, K., & Hickok G (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters, 452*(3), 219–223. 10.1016/j.neulet.2009.01.060

Olsen-Fulero, L. (1982). Style and stability in mother conversational behaviour: A study of individual differences. *Journal of Child Language, 9*(3), 543–564. https://doi.org/10.1017/S0305000900004906

Parise, E., Reid, V. M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social neuroscience, 3*(2), 141–150. https://doi.org/10.1080/17470910701865458

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior & Development, 18*(1), 79–85. https://doi.org/10.1016/0163-6383(95)90009-8

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the Links Among Face Processing, Language Processing, and Narrowing During Development. *Child Development Perspectives, 8*(2), 65–70. https://doi.org/10.1111/cdep.12064

Peña, M., Langus, A., Gutiérrez, C., Huepe-Artigas, D., & Nespor, M. (2016). Rhythm on your lips. *Frontiers in Psychology, 7,* Article 1708.

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

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

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, ... Brugge, JF. (2007). Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience, 145*(1), 162–184. 10.1016/j.neuroscience.2006.11.036

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Lawrence Erlbaum Associates, Inc.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59*(3), 347–357. https://doi.org/10.3758/BF03211902

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., …Gorniak, P. (n.d.). The Human Speechome Project. In *Symbol Grounding and Beyond* (pp. 192–196). Springer Berlin Heidelberg. https://doi.org/10.1007/11880172_15

Sai, F.Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development, 14*(4), 29–50. https://doi.org/10.1002/icd.376

Sams, M, Möttönen R, & Sihvonen T (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*(1–3), 429–435. 10.1016/j.cogbrainres.2004.11.006

Schmitow, C., & Stenberg, G. (2015). What aspects of others' behaviors do infants attend to in live situations? *Infant Behavior & Development, 40*, 173–182. https://doi.org/10.1016/j.infbeh.2015.04.002

Scott M, Yeung H, Gick B, & Werker JF (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America, 133*(4), EL286–EL292. 10.1121/1.4794932

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development, 10,* 59–65. DOI: 10.1002/icd.247

Skipper, JI, van Wassenhove V, Nusbaum HC, & Small SL (2007). Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral Cortex, 17*(10), 2387–2399. 10.1093/cercor/bhl147

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

Streri, A., Coulon, M., Marie, J., & Yeung, H. H. (2016). Developmental change in infants' detection of visual faces that match auditory vowels. *Infancy, 21*(2), 177–198. https://doi.org/10.1111/infa.12104

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: typical, daily exposure to faces documented from a first-person infant perspective. *Developmental psychobiology*, *56*(2), 249–261. https://doi.org/10.1002/dev.21183

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. In V. Bruce, A. Cowey, A. W. Ellis, & D. I. Perrett (Eds.), *Processing the facial image* (pp. 71–78). Clarendon Press/Oxford University Press.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., Morgan, J. L., & Shah, R. J. (2015). "Attention to the mouth and gaze following in infancy predict language development": Corrigendum. *Journal of Child Language, 42*(6), 1408. https://doi.org/10.1017/S0305000915000501

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

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

Van Wassenhove, V, Grant KW, & Poeppel D (2004). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. 10.1073/pnas.0408949102

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

Viola Macchi, C., Turati, C., & Simion, F. (2004). Can a Nonspecific Bias Toward Top-Heavy Patterns Explain Newborns' Face Preference? *Psychological Science, 15*(6), 379–383. https://doi.org/10.1111/j.0956-7976.2004.00688.x

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*, *3*(2), 556–569. https://doi.org/10.1016/j.rasd.2008.12.003

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

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition, 160*, 103–109. https://doi.org/10.1016/j.cognition.2017.01.002

Weatherhead, D., Arredondo, M. M., Nácar Garcia, L., & Werker, J. F. (2021). The Role of Audiovisual Speech in Fast-Mapping and Novel Word Retention in Monolingual and Bilingual 24-Month-Olds. *Brain sciences, 11*(1), 114. https://doi.org/10.3390/brainsci11010114

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*(1), 23–43. https://doi.org/10.1016/S0167-6393(98)00048-X

Yoshida, H., & Burling, J. M. (2011). A new perspective on embodied social attention. *Cognition, Brain, Behavior: An Interdisciplinary Journal, 15*(4), 535–552.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy, 13,* 229– 248. doi:10.1080/15250000802004437

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 10.1371/journal.pone.0079659

CHAPTER 2

Infant Looking Preferences Towards Dynamic Faces: A Systematic Review

Introduction

Studies investigating patterns of selective attention - a key measure of visual attention (Ruff & Rothbart, 1996) - towards the social scene have highlighted a 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; Stern, 1974). The child's looks towards internal features coming from the face, with its dynamic properties, could affect the 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).

Among the studies showing a greater interest in the human face, some have sought to understand where an infant looks when exposed to talking faces. These studies agree on infants' 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 attraction to the eyes increases as a function of both the age and the type of stimulus (i.e., talking faces vs auditory only stimulus). This finding 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 since it conveys emotion. In addition, at a later point in the child's development this could help to establish joint attunement with the speaker (as in Joint Attention episodes), thus also supporting a child's vocabulary knowledge 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 resolved is, when and to what extent does the mouth attract infant attention? As suggested by previous studies, information from the mouth could be a source of phonetic or linguistic cues; thus, 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 afforded by the simultaneous presence of auditory and visual information (in particular, information provided by the mouth) facilitates language comprehension when speech is difficult to understand, as in the presence of noise or a non-native language, for example (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, for example (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 & Licktier, 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 before the end of 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).

Reasons for studying selective visual attention towards talking faces in infancy

The investigation of the child's selective visual attention towards specific regions of the human face (e.g., the eyes or the 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 when children are exposed to dynamic talking faces on a screen, which-may 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 language acquisition is a multi-dimensional (Vigliocco, Perniss, & Vinson, 2014) process that 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 and its relationship with developmental outcomes have been studied in natural settings, within the complexity of the interaction. For example, Gogate and colleagues (2006) tested how children learn specific word-object referents by analysing gaze switching (mouth-hand coordination). They found that infant attention, together with maternal use of temporal synchrony, explained success in learning the target words during an interactive session with parents. This finding 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 the process of language learning, as well as the influence of the input. However, in the naturalistic studies, to the best of our knowledge, no one has investigated exactly what happens when children look at faces and what can explain gaze shifts from one region to another during infancy.

Most of the studies adopting an eye-tracking technology to investigate the selective visual attention phenomenon have attempted to clarify three aspects. Firstly, they 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, the findings relating to the factors that might influence looking patterns during infancy are somewhat contradictory and a common line of interpretation is difficult to identify.

Secondly, they 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 considered how factors such as exposure to one or more languages may influence the 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). 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 offer a description of those current experimental studies that adopt eye-tracking methodology in order to investigate the child's selective visual attention towards the mouth or the eyes and to 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 about dynamic areas of the face in the first two years of life, when children are acquiring most of 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 (language development, social development), 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.

Method

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, a systematic review on the studies exploring the McGurk effect already exists, see Tomalsky, 2015). 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).

Table 1

Terms used for running the search in the four databases

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*

Eligibility criteria

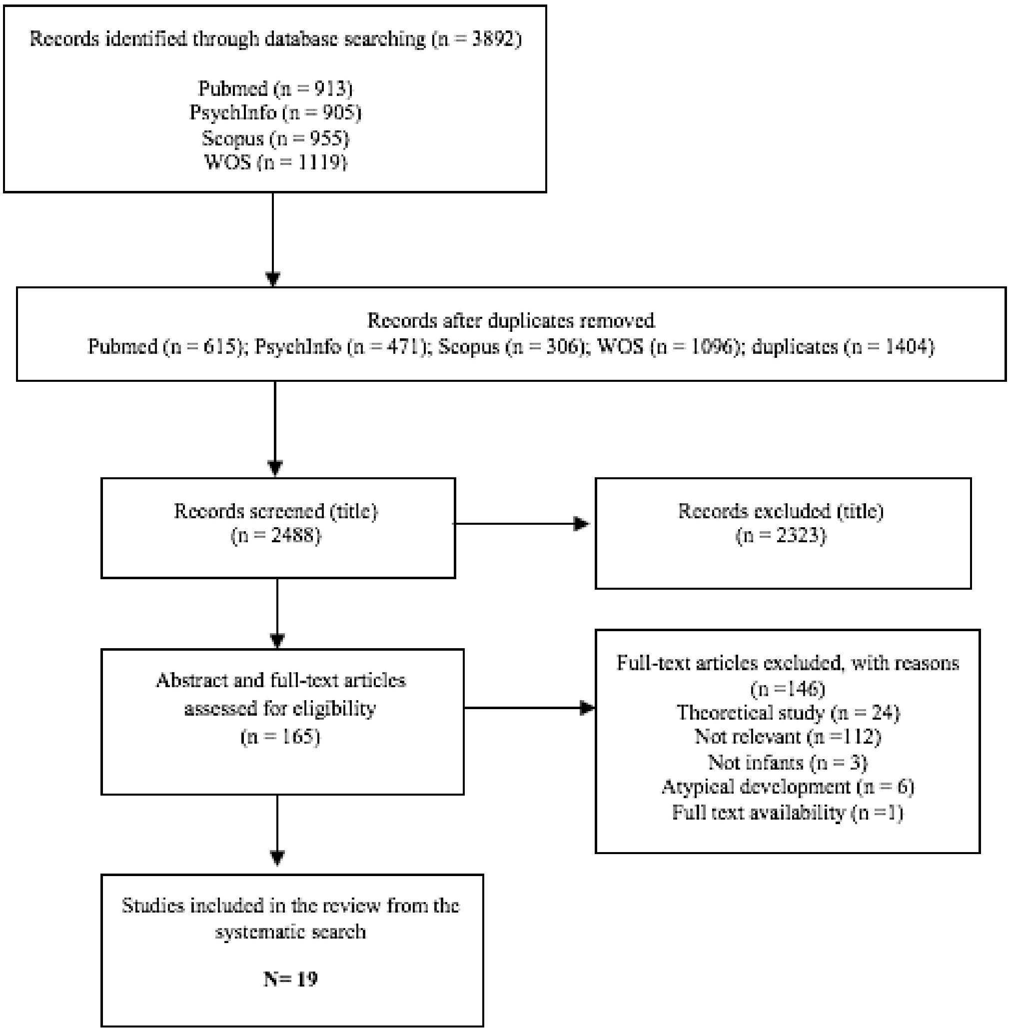
Inclusion and exclusion criteria were adopted to limit the review to relevant items. English was required as the language of publication; studies in other languages were excluded. Only articles published in peer-reviewed journals were considered (excluding purely theoretical papers, conference papers, PhD dissertations, etc.). The participants' age range of (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 development (not atypical).

Protocol

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

Figure 1

PRISMA Flow Diagram (Moher et al., 2009)



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.

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.

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.

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).

Table 2

Characteristics of each study

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
				33		Mullen scale of early learning (non-verbal	Female Talking Face;	Sentences in	Eyes
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	Król, 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-REDM 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	

Note. For each study, the participants’ age, their linguistic background (monolingual or bilingual), the Eye Tracker (ET) model, the assessment they took part in (MB-CDI = MacArthur-Bates Communicative Development Inventories) and the type of stimulus which they were exposed to (IDS= Infant Directed Speech; ADS = Adult Directed Speech) are reported. The the type of index of selective visual attention used is also reported (PTLT = Proportion of Total Looking Time).

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 2. 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 purposes 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., eye-tracker 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., 2021; Tsang 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 2. 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 or congruent and incongruent syllables. 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 speech stimulus.

Language and other developmental experiences. 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.

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 of 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 or a non-native one; how bilingual children's looking behaviour changes in response to a native or a non-native language; how the looking patters of monolingual or bilinguals babies change when they listen to synchronized vs. desynchronised, native vs, non-native short stories, congruent vs. incongruent syllables, a story presented in an ADS vs IDS manner or when they are looking to faces expressing non-speech information.

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 attentional 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 at around 12 months (Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; 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.

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.

Bilingual children listening to a native language (IDS)

The looking behaviour of bilingual babies exposed to native-language IDS speech stimulus 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 continued doing 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 of the four studies that tested bilingual babies on native speech stimuli in the first year of life (Birules et al., 2019; Pons et al., 2015; Tsang et al., 2018) 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.

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.

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 regions of the face. 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, whether the speech was native or non-native.

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 but, more interestingly, they found that the looking preference toward a specific region of the face is correlated with language development some months later (see section ‘Infants’ looking behaviour and developmental outcomes’). 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 older-infants (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.

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 in IDS was found up to the age of 14 months (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. Taken together, these results 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. Nevertheless, 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). 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.

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 the presence of 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; Fort et al., 2018; Oakes, DeBolt, Beckner, Voss, & Cantrell, 2021). 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 prefer to look toward the lower part of the face (i.e., the mouth), regardless of the type of stimulus.

To sum up, these studies agree in finding that 8-, 12-, and 15-month-old bilinguals 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.

Infants' looking behaviour and developmental outcomes

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*).

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 (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). 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.

The linguistic hypothesis (with language skills tested at the same time of the experiment)

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 as measured through parental questionnaires. Note that, unlike Tsang and colleagues, who found the same for bilinguals, Morin-Lessard and colleagues 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 (r = .409, p =.10).

Finally, Pons et al. (2019) and Hillairet de Boisferon and colleagues (2018) failed to find any relationship – 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.

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 (R^2_{expr}= .00-.20) and 24 months (R^2_{expr}= .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.

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.

Discussion

Our systematic search led to the identification of 19 studies published between 2012 and 2021. These studies investigated children's looking patterns towards the eyes or mouth of a dynamic face in the first three years of life. All 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 here. 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

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 or first words: Vihman, 2014) 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). 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.

The first important speculation concerns the question of what happens over the first year in monolingual and bilingual babies. From the various results taken together we can conclude that, generally speaking, from one year of age, monolingual and bilingual infants differ in their looking pattern in response to a non-native but not a 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 second important result 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. 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. 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 another developmental level, and at the same time they are more motivated and interested in their 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.

Thirdly, being exposed to noise, to ADS (rather than IDS), incongruent stimuli or non-speech visual information could also elicit looking towards the mouth, which thus further points to the role of the mouth as a visual cue for perceiving and processing both speech and non-speech information in the laboratory setting.

In general, the first part of the systematic 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 & Licktier, 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

Although no clear, consistent results have yet emerged as regards the relation of looking behaviour to language skills, some conclusions can be attempted. We can say that 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 of life, sustaining the ongoing learning process – and this is in accord with all the studies. Infant preference for mouth over eyes when presented with non-native speech could mean that babies recognise a different 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 Gogate et al., 2006; Matatyaho et al., 2008; Rader & Zukow-Goldring 2010; 2012). 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.

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. The question remains open: Is there a relationship between a child’s looking behaviour and their linguistic skills? And, in particular, what are the mechanisms that link the two, if any? Language knowledge can guide or shape the looking behaviour itself, but it can also be determined by what the child is looking at. That is, it remains unclear whether perception influences production or vice versa. Furthermore, the role played by looking at the mouth could be considered to be grounded in a child's intrinsic need or as the 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 & 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.

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.

Limitations and conclusion

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. In the 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 and face scanning patterns to answer the question of the role of the internal features of the face as a support or learning mechanism 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 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.

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

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

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

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

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., 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.

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

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. (2006). Infant speech perception activates Broca's 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

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., & 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

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

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

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.

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.

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

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.(2015). Developmental Trajectory of Audiovisual Speech Integration in Early Infancy. A Review of Studies Using the McGurk Paradigm. *Psychology of Language and Communication,19(*2) 77-100. https://doi.org/10.1515/plc-2015-0006

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.

Vihman, M. M. (2014). *Phonological Development. The First Two Years*. WILEY Blackwell.

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, L.B. (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

CHAPTER 3

Do Language Skills Shape What An Infant Looks At In A Talking Face? An Exploratory Study In Italian Infants

Introduction

Infants' selective visual attention towards human faces has been much investigated as a perceptual mechanism underlying language acquisition in the first two years of the life (Ayneto & Sebastian-Galles, 2017; Fort, Ayneto-Gimeno, Escrichs, & Sebastian-Galles, 2018; Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012). Increasingly in the last few years, though, studies have started to investigate the role of selective visual attention towards specific regions of the human face and to consider its relationship with the child's linguistic skills as reported by parents at the time of the experiment [for a review see Chapter 2]. Longitudinal perspectives, looking at the children language some months after the experiment, have also been undertaken (Imafuku & Myowa, 2016; Kushnerenko et al., 2013; Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2015). Despite the several hypotheses that have been advanced, a clear pattern of infant looking behaviour has not yet emerged. Moreover, no studies have investigated the relationship between the child's looking behaviour and their language skills with reference to their current vocal production (pre-canonical babbling, babbling or words).

Selective visual attention

The investigation of the child's selective visual attention towards specific regions of the human face (e.g., the eyes or the mouth) while exposed to dynamic/talking faces has only recently received much attention. The phenomenon of selective visual attention was experimentally explored in studies in which the child's looking behaviour - a key measure of visual attention (Ruff & Rothbart, 1996) - was tracked while the child listened to adults' talking faces appearing on the screen. Looking at the caregiver's face could be a sign of both social engagement and interest in language. In other words, the articulatory movements of the mouth could attract children's attention, thus enhancing language learning (Tenenbaum et al., 2015) or social abilities (Pons, Bosch, & Lewkowicz, 2019). The mouth could represent a visual cue that supports language acquisition and development in the first years of the child's life and at later points in time. Also, the redundant stimuli provided by the articulatory movements and the acoustic signals/sounds coming from the mouth can facilitate and support language acquisition and learning. Indeed, looking at the mouth, or the articulatory movements coming from the mouth, could enhance the perceptual saliency and encoding of information, according to the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000; Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012). The eyes play an important role in communication and human interactions too since they provide socio-emotional information. In addition, they might help establish joint attunement with the speaker (as in Joint Attention episodes), thus also supporting children's vocabulary knowledge some months later (Brooks & Meltzoff, 2005).

Lewkowicz and Hansen-Tift's study (2012) produced the first empirical evidence of selective visual attention as a proportion of children's total looking time (henceforth, PTLT, i.e., the “total amount of looking directed at each Areas Of Interest (AOI)” divided by the “total amount of looking at any portion of the face”, p. 1432) while listening to speech., Lewkowicz and Hansen-Tift identified two attentional shifts when children were exposed to their native speech, based on their age. The first attentional shift was from the eyes to the mouth, between 4 and 8 months. The second was a shift back towards the eyes, at 12 months. Some speculations were advanced to explain these shifts, based on the *language expertise hypothesis* (see Chapter 1 for details) (Hillairet De Boisferon et al., 2017; Lewkowitz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowitz, 2015). The language expertise hypothesis proposes that there is an effect of the child's linguistic level (pre-babbling, babbling, words) or developmental stage on the child's gaze patterns. Lewkowitz and Hansen-Tift (2012) speculated on the two attentional shifts identified between 4 and 8 months and then around 12 months. At these ages two developmental achievements in the child's language acquisition typically emerge: first the onset of canonical babbling, second, the emergence of the first words.

Looking behaviour and phonological development

Studies investigating selective visual attention have suggested that the period related to babbling may correspond to increased interest in the mouth and then, when children start to produce words, a decline in interest in the mouth a return to looking more to the eyes. Canonical babbling is a fundamental milestone in typical language acquisition (Vihman, 2014, for a recent systematic review see Morgan & Wren, 2018), and it appears in the period before the emergence of the first words (Fagan, 2009) when children are between 6 and 8 months (Oller, 2000). Babbling is motoric practice, during which children produce adult-like consonant (C) + vowel (V) syllables that can easily be identified or recognised by the principal caregiver. Babbling onset is characterised by the repetition of multiple syllables (Fagan, 2009; Vihman 2014). Based on this last characteristic, babbling can be classified into reduplicated (e.g., mama) and variegated babbling (e.g., beba, daba). Before the emergence of canonical babbling, children produce pre-canonical forms, defined by Lang and colleagues (2019) as quasi-resonant sounds, cooing sounds, fully resonant sounds or raspberries. Babbling precedes words (i.e., productions with a sound-meaning correspondence) by several months (Fagan, 2009). Some evidence has also shown that there is continuity between babbling production and early words (Vihman, Macken, Miller, Simmons, & Miller, 1985; for Italian, Majorano & D'Odorico, 2011). More specifically, the first words produced contain sequences that children are able to reproduce because these are part of a stable phonological repertoire (Keren- Portnoy, Majorano, & Vihman, 2009, see also the concept of Vocal Motor Scheme in McCune & Vihman, 2001).

The interest in babbling in language acquisition studies has been receiving considerable attention for several reasons. First, the idea of a link between babbling forms and a child's first words (Paul & Jennings, 1992; Stoel-Gammon, 1985, 1989; Vihman, et al., 1985) is now well-established. When children start to produce frequent and stable motoric routines, they are also building a solid basis for later lexical development. Second, the extent of variegation predicts expressive vocabulary (Keren-Portnoy, Majorano, & Vihman, 2009; McCune & Vihman, 2001; Stoel- Gammon, 1989). Children showing a delay, or a lack of canonical babbling production, have a smaller expressive vocabulary at 18, 24 and 30 months of age (Oller, Eilers, Neal, & Schwartz, 1999). For this reason, babbling can be considered a clinical marker that suggest atypical development.

For example, Eilers and Oller (1994) showed that children with hearing difficulties do not reach this milestone until 11 months at the earliest. This is related to the idea that external vocal input plays a primary role in the determination of the babbling milestone. Another interesting aspect is that when children start to babble and build stable vocal patterns, the caregivers around them tend to support this behaviour by repeating or providing a response to what the child is training on (Gros-Louis, West, Goldstein, & King, 2006).

No studies have empirically verified the hypothesis that it is the emergence or the production of CV syllables without identifiable target that affects the child's preferential looking to the mouth because none have attempted to relate the child's looking behaviour to a particular language level. Many studies instead confirmed, or failed to confirm, some of these looking patterns. For example, Pons and colleagues (2015) replicated the finding that children look longer towards the mouth between 4 and 8 months. However, in contrast with what Lewkowitz and Hansen-Tift found, recent evidence has shown that children exposed to their native language continue to look more towards the mouth region beyond 12 months old (Hillairet De Boisferon et al., 2018). Shifts in gaze were then hypothesised to depend also upon the child's age or some characteristics of the stimuli (e.g., native vs non-native speech). For example, when children are exposed to a non-native language, the shift from eyes to mouth is identified between 4 and 8 months, but it does not change when children are 12 months old (Lewkowitz and Hansen-Tift, 2012). Indeed, unlike the case of native language stimuli (with several exceptions in the literature, e.g., see Hillairet De Boisferon et al., 2018 or Morin-Lessard et al., 2019), this longer looking towards the mouth during non-native speech exposure is seen even when the infants are older (12 months). This is interpreted as a sign of a need for audio and visual redundancy for an unfamiliar stimulus (Pons, Bosch & Lewkowitz, 2015, 2019). But not all studies agree: Morin Lessard and colleagues found that children at 5, 9,12, 14, and 18 months did not spend more time looking towards the mouth than in the eyes when exposed to a non-native stimulus, but they did find this preferential looking pattern in 3- to 4-year-old children. At this stage, more replication studies are needed to understand whether there is a common or universal pattern of looking behaviour among children or to explain which individual features are relevant in the definition of such patterns.

In the study of infant's language development, the integration of multiple sensory modalities, i.e., the use and integration of several senses or modalities (e.g., visual, auditory) in the perception of speech (Bremner, Lewkowick & Spence, 2012), represents one of the mechanisms underpinning language acquisition (Michon, Lopez, & Aboitiz, 2019). The importance of looking towards the mouth has appeared in perception studies considering the relationship between motor skills and production abilities. It has emerged that the sensory-motor information provided by the articulatory movements of the mouth, together with the auditory information coming from the talker' speech, can influence infants' speech perception (Choi, Bruderer, & Werker, 2019; Vilain, Dole, Lœvenbruck, Pascalis, & Schwartz, 2019). A study using neuroimages revealed a 'perceptuo-motor link' in children while conducting a listening task (Imada, Zhang, Cheour, Tauli, Ahonen, & Kuhl, 2006, p. 957). The link consists of the activation of speech motor areas in response to hearing speech. Imada and colleagues demonstrated that only when children had begun to accumulate adult-like production experience (i.e., once they had started to produce canonical babble or babble in adult-like syllables) is there an activation of the speech motor areas (a finding anticipated in Vihman, 2002). Such evidence confirmed a link between perception and the production of speech.

Hypotheses of an association between the child's current language skills and their gaze behaviour have been proposed and tested (Imafuku et al., 2016; Imafuku et al., 2019; Hillairet De Boisferon et al., 2018; Kròl, 2018; Morin-Lessard et al., 2019; Pons et al., 2019; Tsang et al., 2018). Several studies have based their results on the answers to parental questionnaires about language and socio-emotional development. However, exploration of the relationship between the child's looking behaviour and their language and social skills at the time of the experiment have been based on different looking measures (e.g., PTLT, EMI index), so firm conclusions cannot be drawn. For example, as concerns language skills, Tsang and colleagues (2018) found that relative attention to the mouth (i.e., *ME Index*, “similar to the proportion total looking time (PTLT) metric used by Lewkowicz and Hansen-Tift (2012) except that the numerator used in our metric subtracts dwell time to the eyes from dwell time to the mouth”, p. 99), is associated with higher scores in expressive vocabulary in monolingual and bilingual children between 6 and 12 months of age. Morin-Lessard and colleagues found that PTLT difference scores (“which subtracted the PTLT for the mouth from the PTLT for the eyes”, p. 1645) are significantly negatively correlated with expressive vocabulary scores in monolingual babies between 9 months and 2 years. That is, children with higher vocabulary scores spent more time looking towards the mouth (because a PTLT difference score above zero indicates greater interest in the eyes, and below zero indicates greater interest in the mouth). Hillairet de Boisferon and colleagues (2018), however, failed to find any association when computing an eyes-mouth index (*EMI index*, eyes/eyes+mouth), having classified infants both on their vocabulary scores (higher than 50 words and less than 50 words) and on the proportions of their attention towards the eyes and the mouth in children aged 18 months. As concerns socio-emotional development, Pons and colleagues (2019) found a significant positive correlation between the PTLT (as defined by Lewkowicz & Hansen Tift, 2012) for the eyes and the child's social abilities at 12-months of age (Pons, Bosch, & Lewkowicz, 2019). But no studies have yet considered the relationship between the child's looking towards regions of the face and spontaneous language recorded at the time of the experiment. Nor have any studies considered how indices of vocal advancement based on spontaneous speech might be related to such looking patterns. Imafuku et al. (2019) made the first attempt, considering the record of 6-month-old children's vocal responses (i.e., vowel production) during the experiment as a linguistic outcome while also controlling for the child's looking behaviour (as a PTLT towards the eyes and the mouth). They confirmed that the more the children produced imitative speech-like sounds (vowels), the more they looked towards the mouth. However, Imafuku et al.'s results have not been supported by other studies.

A few longitudinal studies have also found a relationship between the child's visual attention towards the mouth and their linguistic outcomes, with different results (Imafuku & Myowa, 2016; Kushnerenko et al., 2013; Tenenbaum et al., 2015). For example, in Imafuku et al. (2016), 6-month-old infants showed a significant, positive relationship between the PTLT towards the mouth and their receptive skills when tested six months later. Kushnerenko (2013) found that those children who scored higher in a comprehension test at 14-16 months had looked longer towards the eyes and less towards the mouth when they were between 6-and 9-months. Tenenbaum and colleagues (2015) found that the amount of time spent looking towards the mouth, calculated as a mouth-eyes index ("proportion of attention to the mouth/proportion of attention to the mouth + eyes", p. 1182) at 12 months, predicts language skills at 18 and 24 months, thus suggesting a predictive role of the child's looking towards the mouth for later language skills. However, these results are too different and fragile for a conclusion to be drawn.

The present study

The present study investigates the relationship between children's looking behaviour between 6 and 14 months, their vocal skills at the time of the experiment and their vocabulary size three months after the experiment. The present contribution is composed of two studies.

The first study analysed infants' selective visual attention towards a talking face in two groups of infants between 6 and 10 months (G1) and between 11 and 14 months (G2) when exposed to their native (i.e., Italian) and a non-native (i.e., English) language.

The second study focuses on the relationship between the child's looking behaviour and their language skills: Children's linguistic profiles were determined from their production, based on direct observation, in the week of the experiment. In addition, all families received the MB-CDI three months later, measuring the child's vocabulary size in production. Data from the observation and the questionnaire were considered together with the data from the experiment.

Generally, with this study, we wanted to investigate how the child's looking behaviour could be explained by several individual factors. We considered the child's age and language knowledge and other environmental factors, such as the type of language to which the children were exposed during the experiment (native vs non-native). We also wanted to test the nature of the relationship, if any, between the child's looking behaviour and their linguistic developmental stage and whether specific looking patterns are related to the child's current vocal production and later vocabulary. The specific research questions are as follows:

Study 1. Do infants at different ages follow particular looking patterns when exposed to dynamic faces talking in their native and non-native language? We expected that children's looking behaviour would vary with age (G1 *vs* G2). More specifically, between the two shifts reported by Lewkowitz & Hansen-Tift (2012), children between 6 and 10 months (G1) were expected to spend the same proportion of looking time towards the eyes as towards the mouth, both in their native and non-native language, as they are in a transition period, not yet having entered the first-word phase. Children aged between 11 and 14 months (G2) were expected, in line with the recent literature, to look more towards the mouth than the eyes when listening to their native language, due to the emergence of their first words (Frank et al., 2012; Hillairet de Boisferon et al., 2018; Morin-Lessard et al., 2019; Pons et al., 2018). We expected that children's looking behaviour would vary with the type of stimuli they are listening to / looking at during the experiment (native *vs* non-native speech). Children were expected to be more attracted by the two areas of the face (as a general index of attention PTLT eyes+ PTLT mouth) when listening to the native language than when listening to their non-native language, due to the familiarity of the language of the story. In addition, a greater allocation of attentional resources towards the mouth area (PTLT mouth) than towards the eyes is expected when children are exposed to their native language, due to a recognition or matching mechanism between what they hear, what they see and what they know.

Study 2. Research Question 1. Is there an association between the child's language skills (vocal production) at the time of the experiment and their selective visual attention towards a part of the face (i.e., the mouth)? We expected to find that children with more advanced linguistic skills would allocate more attentional resources towards the mouth region, in line with those studies arguing that when infants start to acquire their lexicon they pay more attention to highly salient audiovisual speech cues which are located in the talker's mouth (e.g., Hillairet de Boisferon, et al., 2018).

Research Question 2. Is there an association between the child's selective visual attention towards a region of the face (i.e., the mouth) at the time of the experiment and their language skills three months later? We expected, in line with the literature (Imafuku, Kawai, Niwa, Shinya, & Myowa, 2019; Tenenbaum et al., 2015), that those children who allocate more attentional resources towards the mouth at the time of the experiment will display higher expressive vocabulary scores three months later: That is because once children start improving their lexical skills, they seek or are attracted to multisensory redundancy cues, which are provided by the combination of auditory and visual signals, thus also showing more interest in language.

Study 1

Method

Participants

A total of thirty-four infants aged between 6 and 14 months took part in the study. The final sample was composed of 26 infants. Eight infants were excluded due to a failure in data extraction or in the calibration procedure at the time of the experiment. The children were split into two groups based on age (see Table 1 for more details on the sample characteristics). The first group (G1) comprised 15 children (12 males), tested at around 7 months (*M* = 7.67; *SD* = 1.47, 6-10 months) and the second group (G2), 11 children (5 males), tested at around 12 months (*M* = 12.4, *SD* = .894, 11-14 months). No parents reported developmental delays or problems at the time of their child's birth. Children's mean weight at birth was 3232 kg (*SD* = 378) for G1 and 3150 kg. (411) for G2. All infants were born in Italy. The mean number of years of schooling for mothers of the G1 children was 16.5 years (*SD* = 3.5) and for mothers of the G2 children, 18.5 years (SD = .934); for the G1fathers it was 13.5 years (*SD* =6.03) and for the G2 fathers, 15.1 years (*SD* = 5.22). No significant differences emerged between G1 and G2 in the children's birth weight or in parental level of education (see Table 1 for more details on the *non-parametric Mann-Whitney tests*). The families were recruited for the study through local services for infants and joined voluntarily.

Table 1

Description of the sample (children and parents) for G1 and G2; comparisons between groups in terms of gender (Chi-Square Test), Weight, Age, Parental Age, and Level of Education (Mann-Whitney tests)

	G1 (*n* = 15)	G2 (*n* = 11)	G1 vs G2 (χ^2 and Mann-Whitney U p values)
Children's Characteristics			
Male (*N*)	12	5	$\chi^2 = 3.35$, $p = .067$
Female (*N*)	3	6	
Weight (grams) — M (SD)	3231.67 (378.38)	3104.64 (382.70)	*p* = 471
Age (months) — M (SD)	7.67 (1.47)	12.4 (.894)	$p < .001$
Parents' Characteristics			
Age (Mother) — M (SD)	33.93 (3.97)	33.29 (3.77)	*p* = .895
Age (Father) — M (SD)	37.4 (6.03)	36.93 (5.65)	*p* = .687
Level of Education (Mother) (years) — M (SD)	16.57 (3.50)	17.36 (3.18)	*p* = .407
Level of Education (Father) (years) — M (SD)	13.53 (4.03)	14.64 (5.06)	*p* = .446

Instruments

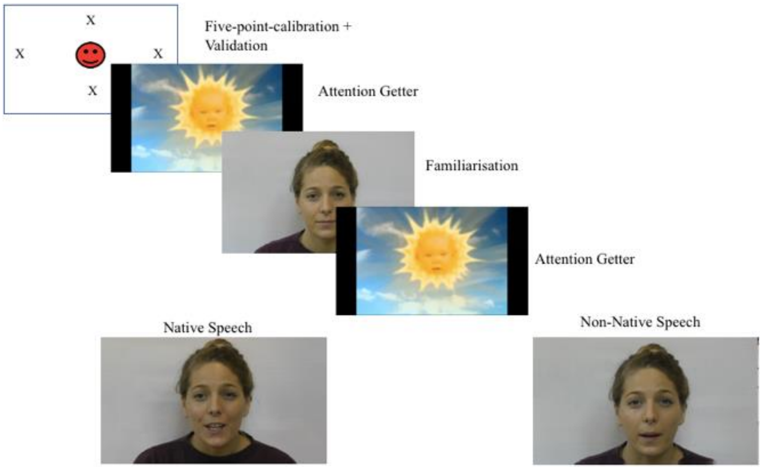
Apparatus

The infants' eye movements were tracked in a sound-attenuated, dimly illuminated room, 60 cm away from the screen of an eye tracker Eye-Link1000 (SR Research) while sitting on their mother's lap (Hessel, Andersson, Hooge, Nyström, & Kemner, 2015; Wass, Forssman, & Leppänen, 2014).

The experimenter ran the 5-point calibration and the validation before starting the test (see Figure 1 for a graphic representation of the design). A visual familiarisation trial appeared on the screen with the photograph of the woman whose speaking face was used for the entire experiment. An attention-getter appeared on the screen to attract the child's attention before the first trial and between the different speech conditions (native or non-native speech). Two trials (native and non-native speech) were run for each child (Figure 1). The child's eye movements were coded offline using Data Viewer Software for the duration of each monologue. Two AOIs were considered: the eye region and the mouth region (Borgi, Cogliati-Dezza, Victoria Brelsford, Meints, & Cirulli, 2014).

Figure 1

Experimental Procedure of the selective visual attention task



Note. A five-point calibration phase and a validation phase (smiling face appearing on five points of a screen) were followed by a familiarisation phase (a static image of the actress of the test phase appears on the screen, accompanied by no sound). An attention-getter appears on the screen. When the child's gaze was directed at the screen (and the child was not attracted by other stimuli around them), the experimenter started the test phase. Each child was presented with two videos, one in their native language, the other in their non-native language (English). The presentation order of the two conditions (native and non-native) was random across participants.

Measures

Two measures were taken for the Study 1:

1) the *proportion of looking time* (PTLT) that participants spent looking at each AOI. The PTLT was calculated by dividing the total looking directed at each AOI by the total looking at the entire screen (different from Lewkowicz and Hansen-Tift, who considered amount of looking at any portion of the face). -

The proportion of attention towards the two AoIs was also calculated by summing the proportion of looking time towards the eyes and the proportion of looking time towards the mouth.

2) the *PTLT difference score* (i.e., PTLT eyes - PTLT mouth). A positive value indicates a longer looking time towards the eyes; a negative value indicates a longer looking time towards the mouth.

Stimuli

Two videos were shown to the children. The videos consist of an Italian female who also speaks native-like English telling two short stories in a child-directed manner, one in the child's native language (Italian) and one in their non-native language (English). The duration of each story was 45 seconds. The presentation of the stimuli was counterbalanced across participants: for half the children in each age group the native stimulus was run first, then the non-native one, and the reverse order for the other half. Before the experiment started, a static image of the actress's face appeared on the screen (this familiarisation phase lasted less than one minute and was used to familiarise the child with the actress's face).

Procedure

The experimental procedure to evaluate the child's selective visual attention towards a talking face with an eye-tracker device was conducted in the laboratory setting. The study was approved by the local ethical committee of the University of Verona.

Data analyses

All statistical analyses for Study 1 were conducted using the Jamovi software (version 1.2, 2020).

Descriptive statistics on looking time towards the two AOIs for each condition (native vs non-native). Descriptive statistical tests (mean and standard deviation) were conducted to investigate the child's eye movements while exposed to their native and non-native language. Both *PTLT* towards the eyes and the mouth (during the entire trial) and a *PTLT difference score* were reported.

Non-parametric tests were used since the data did not meet the normality criteria of distribution (Shapiro Wilk tests, $ps < .05$).

Comparison between age groups in the two conditions (native vs non-native). Two *Mann-Whitney t-tests* (one for each condition) were run, with age as a grouping variable, to explore differences in the *PTLT* towards the two AOIs between children at different ages (G1 and G2), both in the native and non-native language. We also checked whether significant differences had emerged for the general index of attention (a sum of attention towards the two AOIs) between children at different ages, in different conditions.

Comparison between the language conditions (native vs non-native). Two non-parametric paired sample t-tests (Wilcoxon, one for each AOI; 1) PTLT-eyes-engl *vs* PTLT-eyes-ita; 2) PTLT-mouth-engl *vs* PTLT-mouth-ita) were run to test the differences between the *PTLT* towards the eyes and the mouth in the two language conditions for the entire group of children.

Results

Descriptive statistics. The distribution of the child's visual attention towards the eyes and the mouth for each age group (G1 and G2) and each condition (native vs non-native) are reported in Table 2 (PTLT) and in Figure 2 (*PTLT difference score*).

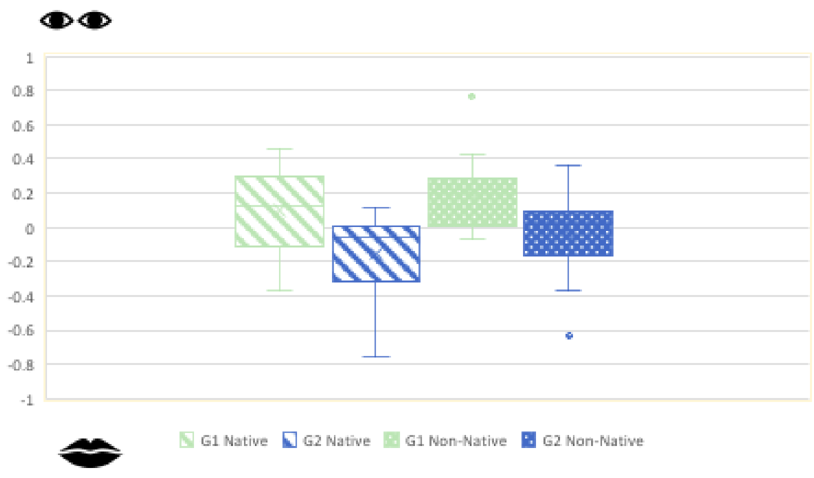
Table 2

The PTLT time (Mean and SD) towards the eyes and the mouth based on child age, in the native and non-native condition, and the general index of attention towards the two AOIs in the two conditions

	Eyes		Mouth		Attention to the two AoIs (eyes+mouth)	
	Native	Non-Native	Native	Non-Native	Native	Non-native
G1	0.281 (0.195)	0.356 (0.179)	0.183 (0.132)	0.163 (0.170)	0.465 (0.219)	0.520 (0.222)
G2	0.171 (0.126)	0.202 (0.133)	0.326 (0.183)	0.235 (0.152)	0.497 (0.199)	0.437 (0.258)

Figure 2

Proportion of looking time difference scores in the two groups (G1 vs G2) in response to the Italian and English languages.



Note. Negative scores indicate more attention towards the mouth; positive scores indicate more attention towards the eyes.

Comparison between age groups. A first significant difference emerged between the two age groups in relation to the proportion of time spent looking towards the mouth in the native condition (*Mann-Whitney* = 39, *p* = .024): Older children looked significantly more towards the mouth (*M* = 32.6%) than younger children (*M* = 18.3%) in the native speech condition. However, older and younger children spent similar amounts of time looking towards the eyes when exposed to the story in the native language (*p* = .203). The second significant difference emerged between the younger and older children in relation to the PTLT towards the eyes in the non-native language condition (*Mann-Whitney* = 41, *p* = .033). The younger group looked significantly more towards the eyes (35.6%) than the older group (20.2 %). However, older and younger children spent similar amounts of time looking towards the mouth when exposed to the story in the non-native language (*p* = .604).

When the timespan of attention towards the two AOIs is considered as a sum of the attention towards the eyes and the mouth, no statistically significant differences emerge between the two age groups, either in native (*Mann-Whitney* = 75, *p* =.716) or non-native speech (*Mann-Whitney* = 66, *p*= .406). This means that the younger and older children in our sample displayed similar attention times towards the two AOIs (as a sum), both in the native and non-native language conditions.

Comparisons between language conditions. Generally, when the *PTLTs* *t*owards the two AOIs in the two language conditions were compared for the entire group (*paired sample t-test*) (Table 3), a significant difference emerged for the PTLT towards the mouth only (*W* = 94.5, *p* = .041).

Table 3

The PTLT towards the eyes and the mouth in the native and non-native condition for the entire group of children; in the table mean and standard deviation were reported

	Non-Native		Native	
	Eyes	Mouth	Eyes	Mouth
G1+G2	0.291 (0.176)	0.193 (0.179)	0.235 (0.175)	0.244 (0.168)

The children looked significantly longer towards the mouth in the native speech condition (24.4% of the time) than in the non-native speech condition (19.3%). However, no such difference was found for the time spent looking towards the eyes. At a descriptive level, children seemed to look more towards the eyes in the non-native speech condition (29.1% of the time) than in the native speech condition (23.5% of the time). However, this was not statistically confirmed (*W* = 243, *p* = .089).

Discussion

The first study aimed to assess the children's looking pattern behaviour by controlling for the age and the language they were listening to during the experiment.

When considering the *PTLT* towards the eyes and the mouth in the two language conditions, a significant difference between the two age groups emerged for the native speech condition. The older children looked significantly more towards the mouth than the younger children (between 6 and 10 months), which is partially in line with previous evidence (for example, Lewkowicz & Hansen Tift, 2012 and Pons et al., 2015 found such a pattern in children younger than 8 months; Imafuku et al., 2019 and Sekiyama et al., 2021, found such a pattern in 6-month-old children). In line with the literature, children at 12 months were expected to process speech differently from children aged around 4 and 8 months, since they have accumulated more experience (Lewkowicz & Hansen-Tift, 2012). Recent evidence has shown that children at around 12 months or later seemed to still be interested in the mouth as compared to the eyes when presented with native language monologues (Hillairet de Boisferon et al., 2018; Morin-Lessard et al., 2019; Pons et al., 2015; Pons et al., 2019; Tsang et al., 2018), in line with our results. This finding suggests that children older than 12 months still need the redundancy of the visual cue produced by the mouth. This could be linked to their specific linguistic stage or, more generally, to their developmental stage (see Chapter 2). Indeed, at the end of the first year, children were not expected to have mastered all the skills needed to talk and process language sufficiently well, and thus they cannot be "independent" of visual cues. Also, as proposed by Hillairet de Boisferon (2018), by the second year of life children were found to be more attracted by the mouth because the mechanism underlying novel word learning implies different mechanisms from those for learning sounds. Furthermore, as an alternative explanation, these patterns could be language-specific (Berdasco-Muñoz, Nazzi, & Yeung, 2019). More studies are needed to confirm either hypothesis.

In the non-native speech condition, the younger children (around seven months) spent a larger amount of time looking towards the eyes than the older group. The study by Lewkowicz and Hansen-Tift (2012) partially investigated such a pattern, but in children no older than 12 months. They found that monolingual children up to 8 months displayed equal looking time towards the eyes and the mouth when exposed to non-native speech, while children older than eight months displayed longer looking time towards the mouth when exposed to non-native speech. Other studies of monolingual children exposed to a non-native language reported longer looking time towards the mouth, independently of age, which partially contrasts with our finding. Children around seven months seemed to be more interested in the eyes when listening to a non-native language. Eye-looking could indicate social advance, as reported by Tsang and colleagues (2018) or Pons et al. (2019). However, very few studies have yet tested such a social hypothesis. The language expertise hypothesis is probably irrelevant when testing children with a non-native language because they would have had no experience with the language of the stimuli, whatever their age.

Overall, children in the two age groups spent similar times looking towards the two AOIs (considered as a general index of attention, i.e., the sum of looking towards the eyes and mouth) when exposed to both their native and non-native speech. We expected that children would spend longer looking towards the two AOIs when the speaker was talking in their non-native language, an "unfamiliarity effect". However, they showed similar looking time towards or interest in the two faces, probably because the two stimuli are similarly attractive.

When considering the effect of the language condition on the *PTLT* (regardless of age), a significant difference emerged for the *PTLT* towards the mouth in the native language condition. They looked significantly longer towards the mouth region when they were listening to native speech. Thus, when they are familiar with the language they are listening to, they allocate most of their attentional resources towards the mouth because they can exactly match what they hear and see to what they are familiar with. However, at this point of the study, it seems worth investigating whether phonological and language skills may account for such behaviour.

One of the limits of the present study is the small sample size and the high individual variability among children. Another weakness is in the construction of the experiment: we did not pre-test the level of "attractiveness" of the two stimuli. In addition, in the native language video, the woman's shoulders are visible, while in the non-native speech video they are not. This might problematically affect our ability to identify the areas of interest for the child and how low the child has to look to be scored as looking toward the mouth. However, we did not detect any difference between the participants in their general attention between the two conditions, so this should not be a real problem. But we are aware of it, and we need to adjust it for future studies. From a developmental perspective, the social and language explanations are only speculation at this stage. The two hypotheses need more studies to be confirmed. The second study investigates the relationship between the child's looking behaviour during the selective visual attention task and their language development at the time of the task and three months later.

Study 2

The children's vocal production was tested by conducting a direct observation at Time 1. The same children were also assessed on their vocabulary rate three months after the experiment (Time 2) using the MB-CDI. The child's selective visual attention at Time 1 was then related to the child's preverbal and verbal production at the time of the experiment and to the child's vocabulary size three months later.

Method

Participants

The same 26 infants whose gaze was tracked between 6 and 14 months (Study 1) were observed and video-recorded while interacting with their mother at home, at Time 1, in the week of the experiment. Their language skills were measured at Time 2, three months later ($M_{age}G1$ = 10.7, SD = 1.35; $M_{age}G2$ =

15.4, *SD* = 1.03) by using a questionnaire (i.e., the MB-CDI). The characteristics of the sample are the same as in Study 1.

Instruments

Mother-child interactions (Time 1)

Infants were video recorded for around 20 minutes during spontaneous interaction with their mother while playing with sets of toys provided by the experimenter (duration of the video, *M* = 20.4, *SD* = 2.43).

In each play session, four sets of toys were provided to the mothers to stimulate as many spontaneous production as possible: 1) a food set, 2) a farm set, 3) a transport set, and 4) a nurturing set. Mothers were required to interact with their children as they usually do, to make the situation as natural and spontaneous as possible. The video observations were conducted at the infant's home, a familiar context for sustaining spontaneous production and reducing distractions.

Each child's preverbal and verbal production were phonetically transcribed by using ELAN (Version 6.0) [Computer software] (2020) and coded in CHAT of CHILDES (Codes for the Human Analysis of Transcripts, MacWhinney, 2000). Crying, vegetative sounds and shouts were not transcribed.

The number of vocalisations produced by the child was tallied. Three classes of production were identified, based on their phonological and semantic properties: 1) pre-canonical vocalisations (i.e., the percentage of quasi-resonant sounds, cooing sounds, fully-resonant sounds or raspberries not containing phonological structures or CV sounds), 2) babbling (i.e., the percentage of syllables consisting of at least one consonant and one vowel; reduplicated babbling was counted as one production) 3) words (i.e., the percentage of productions that have a sound-meaning correspondence) (Stoel-Gammon, 1989; Vihman and McCune, 1994; Vihman et al., 1985; see also Lang et al., 2019).

Reliability. A second independent transcriber transcribed 20% of the video recordings. Reliability index based on the child's amount of vocal production was good (94% of agreement).

MacArthur-Bates Communicative Development Inventory (Time 2)

The Italian short version of the MacArthur-Bates Communicative Development Inventory – MB-CDI Words and Gestures – (Primo Vocabolario del Bambino - PVB, Caselli, Bello, Rinaldi, Stefanini, & Pasqualetti, 2015) was administered to all infants three months after the experiment. The Italian short version of "Words and Gesture" is usually used for children from 9 to 24 months. The number of words produced by each child was counted (in line with the studies exploring the longitudinal relationship between selective attention al vocabulary rate, see Kushnerenko et al., 2013; Tenenbaum et al., 2015).

Measures

Selective visual attention. Time 1. Following Hillairet de Boisferon and colleagues (2018), Merin et al., (2007), Tenenbaum et al., (2015) and Young et al., (2009) (they studied the relationship between selective attention and language skills), infants' relative attention to the eyes and mouth was calculated, i.e., eyes–mouth index (EMI-index). EMI-index was calculated by dividing the amount of gaze to the eyes (sec) by the total amount of time the infant looked to either the eyes or the mouth (seconds) (i.e., Eyes/[Eyes + Mouth]). This index, unlike the PTLT, disregards looks to other parts of the face or to the screen. An EMI < .50 means that the infant looked more at the mouth, whereas an EMI > .50 means they looked more at the eyes. PTLT was also considered here (see Study 1).

Vocal production. Time 1. Amount of 1) preverbal vocalisations, 2) babbling, and 3) advanced forms produced by the child during interaction with their caregiver.

Vocal production. Time 2. Number of words produced as measured through the MB-CDI.

Data analyses

Two sets of analyses were run to test the relationship between selective visual attention and language development. Mplus (version 7, Muthén & Muthén, 2007) and Jamovi (version 1.2, 2020) was used to run the statistical tests.

Latent class cluster analysis and classes differences (direct observation). We used latent class cluster analyses (Vermunt & Magidson 2000) to classify children according to their vocal production (Time 1) into two groups: high (children producing more babbling and words) and low (more pre-canonical vocalisations). Three linguistic behaviours were categorically coded into high vs low categories (Everit, 1980): the number of pre-canonical vocalisations, the number of CV syllables that had no identifiable target (i.e., the word that could be coded as babbling) and the number of words. Since the sample was small, we prefer to identify two linguistic profiles, we ran only models with two classes. Following DiStefano and Kamphaus (2006), probabilities and entropy values were calculated and reported. Entropy indicates how well the model predicts class memberships. It ranges from 0 to 1 (the higher the entropy value is, the better it predicts the model, Vermunt & Magidson, 2002). Once the two linguistic classes were determined (at Times 1 and 2, separately), children's looking behaviour (as the dependent variable) was analysed based on these two classes. More specifically, we ran a series of Mann-Whitney U tests with the two groups (high vs low class) as grouping variable. The dependent variables were relative amount of attention to the eyes (EMI-index), PTLT towards the eyes and the mouth, and the general index of attention.

Correlations and T1 language measures effect. Non-parametric correlations were conducted (Spearman's rho) in order to identify the relationship between language measures at Time 1 and Time 2. Non-parametric correlations were conducted (Spearman's rho) in order to identify the relationship between on the one hand, selective visual attention (EMI-index and PTLT) towards the mouth or the eyes and, on the other hand, later vocabulary. Finally, the contribution/ the effect of the language measures at T1 will be controlled for the significant relationships emerging between selective attention at T1 and language measures at T2.

Results

Latent class cluster analysis and classes differences (direct observation, Time 1). Table 4 shows the distribution of the population in the classification as having a low or high linguistic profile based on the combination of pre-canonical vocalisations, canonical babbling and words at the time of the experiment. It also shows the characteristics of each sub-group based on the linguistic class created. The entropy value of the model is .839, which is considered good.

To better understand the features of each linguistic group, the mean and the standard deviation are reported in Table 4, and the differences among groups are tested (Mann-Whitney). The children in the two linguistic groups differed significantly on the percentage of pre-canonical vocalisations, CV syllables and words produced. Children in the high-group produced fewer pre-canonical vocalisations but more CV syllables and advanced forms than children in the low linguistic group. The two groups also differed significantly when compared by age (M_{age}-HIGH = 11, SD = 1.5; M_{age}-LOW = 7.59, SD = 2.15).

Table 4

Sample distribution (Population share) in the two classes. Description of the preverbal and verbal characteristics (percentage, M, and SE) in the two classes and comparisons among classess (Mann-Whitney)

	Population Share (percentage)	Pre-canonical vocalisations (%)	*p*	CV syllables (%)	*p*	Words (%)	*p*
Class 1(High)	0.35	27.8 (8.56)	<.001	52.6 (8.30)	<.001	19.6 (3.42)	<.001
Class 2(Low)	0.65	79.9 (7.14)		16.7 (5.97)		3.4 (1.88)	

Table 5

EMI-index (%) and PTLT (M and SD) for each language class

	EMI-index	PTLT	
	Eyes/[Mouth+Eyes]	Eyes	Mouth
High	29.6%	0.142 (.131)	0.304 (.224)
Low	54.3%	0.284 (.179)	0.224 (.126)

The children's looking behaviour was analysed based on their linguistic class membership at the time of the experiment (Table 5). Children in the two groups differed significantly in the relative attention (EMI) to the eyes or mouth (*Mann Whitney* = 32, *p* = .016). More specifically, children with a high linguistic profile also looked more towards the mouth (mouth/eyes+mouth) (29.6%; an EMI < 50 means that children looked more towards the mouth) than children with a low linguistic profile (54.3%; an EMI > 50 means that children looked more towards the eyes).

A difference also emerged in the PTLT towards the eyes (*Mann Whitney* = 40.0, *p* = .05). Children in the low-group showed a greater PTLT towards the eyes (28.4%) if compared to children in the high-group (14.2%).

Such differences also need to be considered in the light of the results of Study 1, in which more time towards the mouth was reported only in older children. Such a result can also be confirmed when the linguistic profiles (and age) are taken into account.

Correlations (Time 1 – Time 2) and T1 language measures effect. Preliminarily, a correlation analysis was conducted between language measures at T1 (preverbal vocalisations, babbling, words) and the receptive and the expressive vocabulary scores (MB-CDI) at Time 2. Significant correlations emerged between the number of produced words (MB-CDI) at T2 and the babbling (*Spearman's rho* =.645, p <.001) and words (*Spearman's rho* = .657, p < .001) produced at T1.

Finally, Spearman's rho correlations were conducted between the PTLT towards the eyes and the mouth, the EMI-index at Time 1, and the child's receptive and the expressive vocabulary scores (MB-CDI) at Time 2.

A significant positive relationship emerged between the PTLT time towards the mouth (T1) and the expressive vocabulary measured with the MB-CDI (T2) (*Spearman's rho* = .461, p = .018). Moreover, relative attention towards the eyes (EMI-index, eyes/eyes+mouth) is significantly and negatively correlated with the MB-CDI production scores (*Spearman's rho* = -.421, p = .032). With the focus on the mouth, such relationships indicate that the more children look towards the mouth (both as a PTLT and as relative time), the higher their expressive vocabulary scores.

Discussion

There are two main findings of Study 2. First, a significant difference emerged in the looking behaviour of the two language-groups. Specifically, children who are more linguistically advanced at the time of the experiment (i.e., produced significantly more babbling and words) looked more towards the mouth at the time of the experiment (Imafuku et al., 2019; Tenenbaum et al., 2015). A dominant role of the mouth as a visual cue emerged when the language was considered at the time of the experiment (as shown by Hillairet de Boisferon et al., 2018; Morin-Lessard et al., 2019; Pons et al., 2015; Pons et al., 2019; Tsang et al., 2018). Also, this pattern is still significant when looking towards the mouth is regarded as a measure of relative attention. Before the emergence of canonical babbling in our sample, in line with other studies, infants seemed to be less interested in the talker's mouth, probably because they are in an early stage of their language development (Hillairet de Boisferon et al., 2018). However, when they enter the canonical babbling and first words stage, their focus on the mouth increases as a function of their language skills. This pattern could be seen in our more linguistically advanced sample (see also Study 1). Audiovisual information contained in the mouth seems to play a crucial role in language acquisition in the first two years of life. And our evidence shows that this interest depends on the children's linguistic stage. This aspect, however, (i.e., the linguistic stage) had previously been only speculated about and not yet tested (Lewkowitz & Hansen-Tift, 2012; Hillairet De Boisferon et al., 2018). The present study partially goes against some of these speculative ideas, such as the *language expertise hypothesis* (Hillairet De Boisferon et al., 2017; Lewkowitz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowitz, 2015). Unlike many authors, we did not find an important role for the mouth when children are in the lower linguistic stage (i.e., when they are less expert), but we found it when they had already entered a more advanced stage (i.e., babbling and first words acquisition). This finding is in line with more recent studies that show that children older than 12 months still look more towards the mouth, in contrast with the language expertise hypothesis (Hillairet De Boisferon et al., 2018; Morin-Lessard et al., 2019). In our study, it was not the case that once children start producing a large percentage of CV syllables and words, they stopped looking towards the mouth (Lewkowitz & Hansen-Tift, 2012). The contrary was the case, which requires a new interpretation. It could indicate that the mouth is both a facilitative mechanism for learning a language and a cue that supports a child's language development when they have already mastered some skills. In other words, we can ask, is it the child's looking behaviour that shapes language skills or it is their language skills that drive their looking behaviour? More studies are needed to answer this question.

Secondly, the correlational analysis between the children's selective visual attention at T1 and their language outcomes three months later supports the above discussion. The significant positive relationship between the children's expressive vocabulary at T2 and the babbling and word production at T1 indicates that children producing more advanced forms at T1 also produce more words at T2, as measured through the MB-CDI. However, we did not report or control how much vocal advance at T1 could contribute toward explaining or affecting the relationship between looking behaviour and expressive skills (which is one of the limitations of this study). The significant positive relationship between the children's expressive vocabulary and the time they looked towards the mouth is in line with previous studies. Other studies, indeed, confirmed that the rate of expressive vocabulary at the time of the experiment (Morin-Lessard et al., 2019; Tsang et al., 2018) and longitudinally (Tenenbaum et al., 2015) is related to the amount of looking towards the mouth. Thus, looking time towards the mouth could predict later vocabulary size (as shown in an 'atypical population' study by Young, Merin, Rogers, & Ozonoff, 2009). This tendency confirms that looking towards the mouth could provide more opportunities to practice and succeed later in language development in everyday situations (Falck-Ytter, Fernell, Gillberg, & von Hofsten, 2010). This finding supports those studies that suggest an important role for the mouth as a crucial visual cue for language learning.

General Conclusion

The present study could contribute significantly to our understanding of infants' looking behaviour towards adults' faces and its relationship with language development. This study can be considered exploratory or preliminary due to its small sample size, its main limitation. Only 26 infants of different ages were tested; thus, firm conclusions are difficult to draw. A study with more children is needed to confirm the tendencies and the correlations that emerged in the present contribution, by reducing the risk of individual variability and strengthening the significance of the associations we found. Moreover, more studies are required to investigate infants' attentional mechanisms towards faces and their relationship with language acquisition and development. However, the exact role played by looking towards a specific region of the adult's face in children's language development and learning remains to be clarified. In particular, the following question remains unsolved: Is it language knowledge that drives the child's looking behaviour or, on the contrary, is it the child's language knowledge or vocal skills that lead children to look more at one specific region of the talking face than at another?

It could be interesting in the future to study these mechanisms from a naturalistic perspective by observing how the child's looking behaviour changes during interactions (Tsang et al., 2018, p. 105). The study of eye gaze in the first years of a child's life, in spontaneous and non-controlled situations, is also receiving considerable attention (for a recent review of the literature, see Çetinçelik, Rowland, & Snijders, 2021). Again, conclusions are hard to draw because tracking eye-gaze and looking behaviour in spontaneous interactions is challenging. It is possible, however, to provisionally conclude that non-verbal features characterising face-to-face interaction (i.e., facial expression, gestures and eye gaze) do play a crucial role in the child's linguistic development (Lewkowicz & Pons, 2013; Teinonen, Aslin, Alku, & Csibra, 2008).

References

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

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

Berdasco-Muñoz, E., Nazzi, T., & Yeung, H. H. (2019). Visual scanning of a talking face in preterm and full-term infants. *Developmental psychology, 55*(7), 1353. doi: 10.1037/dev0000737

Bremner, A.J., Lewkowicz D.J., & Spence, C. (2012). The multisensory approach to development. In Bremner AJ, Lewkowicz DJ, & Spence C (Eds.), *Multisensory Development*. Oxford, UK: Oxford.

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

Borgi, M., Cogliati-Dezza, I., Brelsford, V., Meints, K., & Cirulli, F. (2014). Baby schema in human and animal faces induces cuteness perception and gaze allocation in children. *Frontiers in psychology, 5,* 411. https://doi.org/10.3389/fpsyg.2014.00411

Caselli, M. C., Bello, A., Rinaldi, P., Stefanini, S., & Pasqualetti, P. (2015). *Il Primo Vocabolario Del Bambino: Gesti, Parole e Frasi. Valori di riferimento fra 8 e 36 mesi delle forme complete e delle forme brevi del questionario MacArthur-Bates CDI*. Milano: Franco Angeli.

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., Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychonomic bulletin & review, 26*(4), 1388–1399. https://doi.org/10.3758/s13423-019-01601-0

DiStefano, C., & Kamphaus, R. W. (2006). Investigating Subtypes of Child Development: A Comparison of Cluster Analysis and Latent Class Cluster Analysis in Typology Creation. *Educational and Psychological Measurement, 66*(5), 778–794. https://doi.org/10.1177/0013164405284033

Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal Of Pediatrics*, *124*(2):199-203. doi: 10.1016/s0022-3476(94)70303-5.

ELAN (Version 6.0) [Computer software]. (2020). *Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive*. Retrieved from https://archive.mpi.nl/tla/elan

Everitt, B (1980). Cluster analysis. *Quality and Quantity*, *14*, 75–100. https://doi.org/10.1007/BF00154794

Fagan, M. K. (2009). Mean Length of Utterance before words and grammar: longitudinal trends and developmental implications of infant vocalizations. *Journal of child language*, *36*(3), 495–527. https://doi.org/10.1017/S0305000908009070.

Falck-Ytter, T., Fernell, E., Gillberg, C., & von Hofsten, C. (2010). Face scanning distinguishes social from communication impairments in autism. *Developmental Science*, *13*(6), 864–875. https://doi.org/10.1111/j.1467-7687.2009.00942.x

Fort, M., Ayneto-Gimeno, A., Escrichs, A., & Sebastian-Galles, N. (2018). Impact of bilingualism on infants' ability to learn from talking and nontalking faces. *Language Learning, 68*(Suppl 1), 31–57. https://doi.org/10.1111/lang.12273

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*, 509–516.

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of Eye Color, Positioning, and Head Movement for Eye-Tracking Data Quality in Infant Research. *Infancy*, *20*(6), 601–633. doi: 10.1111/infa.12093

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. *Developmental Science, 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. *Journal of Experimental Child Psychoclogy*, *172*, 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. (2006). Infant speech perception activates Broca’s area. *NeuroReport, 17*, 957–962. https://doi.org/10.1097/01.wnr.0000223387.51704.89

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

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

Keren-Portnoy, T., Majorano, M., & Vihman, M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language, 36*(2), 235-267. doi:10.1017/S0305000908008933

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

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

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical Babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports*, *6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

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

Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development, 37*(2), 90–94. https://doi.org/10.1177/0165025412467582

MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription format and programs, Volume II: the database. *Comput. Linguist., 26*, 657–657. doi: 10.1162/coli.2000.26.4.657

Majorano, M., & D’Odorico, L. (2011). The transition into ambient language: A longitudinal study of babbling and first word production of Italian children. *First Language, 31*(1), 47–66. https://doi.org/10.1177/0142723709359239

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language and Hearing Research, 44*, 670–684.

Merin, N., Young, G.S., Ozonoff, S., & Rogers, S.J. (2007). Visual Fixation Patterns during Reciprocal Social Interaction Distinguish a Subgroup of 6-Month-Old Infants At-Risk for Autism from Comparison Infants. *J Autism Dev Disord, 37*, 108–121. https://doi.org/10.1007/s10803-006-0342-4

Michon, M., López, V., & Aboitiz, F. (2019). Origin and evolution of human speech: Emergence from a trimodal auditory, visual and vocal network. *Progress in Brain Research, 250*, 345–371. https://doi.org/10.1016/bs.pbr.2019.01.005

Morgan, L., & Wren, Y. E. (2018). A Systematic Review of the Literature on Early Vocalizations and Babbling Patterns in Young Children. *Communication Disorders Quarterly*, *40*(1), 3–14. https://doi.org/10.1177/1525740118760215

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology., 55*(8), 1640–1655. https://doi.org/10.1037/dev0000750

Muthén, L. K., & Muthén, B. O. (2007). *Mplus User's Guide (Sixth Edition)*. Los Angeles, CA: Muthén & Muthén.

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

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders, 32*(4), 223–245. https://doi.org/10.1016/S0021-9924(99)00013-1

Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of speech and hearing research*, *35*(1), 99–107. https://doi.org/10.1044/jshr.3501.99

Pons, F., Bosch, L., & Lewkowicz, D. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *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. *Infant Behavior and Development, 54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Ruff, H., & Rothbart, M. (1996). *Attention in early development: themes and variations*. Oxford University Press.

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

Stoel-Gammon, C. (1985). Phonetic inventories 15-24 months — a longitudinal study. *Journal of Speech and Hearing Research, 28*, 505–512.

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. First *Language, 9*(6), 207–223. https://doi.org/10.1177/014272378900900607

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

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

The jamovi project (2020). *jamovi*. (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org

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

Vermunt, J.K., &Magidson, J. (2000). *Latent GOLD's User's Guide.* Boston: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars, & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge University Press.

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.

Vihman, M.M (2014). *Phonological development: The first two years (2nd ed.)*. Wiley-Blackwell: Oxford, UK.

Vihman, M. M., & McCune, L. (1994). When is a word a word? *Journal of Child Language, 21*(3), 517–542. https://doi.org/10.1017/S0305000900009442

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language (Baltimore), 61*(2), 397–445. https://doi.org/10.2307/414151

Vilain, A., Dole, M., Lœvenbruck, H., Pascalis, O., & Schwartz, J. (2019). The role of production abilities in the perception of consonant category in infants. *Developmental Science, 22*(6), e12830–n/a. https://doi.org/10.1111/desc.12830

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy, 19*(5), 427–460. https://doi.org/10.1111/infa.12055

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

CHAPTER 4

Do Language Skills Shape What An Infant Looks At While Interacting With The Mother? An Exploratory Study

Introduction

There is much evidence in the literature about the longitudinal relationship between selective visual attention towards informative parts of the face (e.g., eyes, mouth; Tenenbaum et al., 2015) or informative areas in the social scene (e.g., attention to the face or objects; Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell, & Tomasello, 1998; Morales et al., 2000; Morales, Mundy, & Rojas, 1998; Mundy & Gomes, 1998; Mundy, Kasari, Sigman, & Ruskin, 1995; Young, Merin, Rogers, & Ozonoff, 2009) and the child's later language development. The idea that attention to faces supports language development in multiple ways is now well-established and well-supported (Brooks & Meltzoff, 2005; Kushnerenko et al., 2013; Morales et al., 1998; Tenenbaum, Amso, Abar, & Sheinkopf, 2014). Moreover, increasing numbers of studies have supported the argument that having access to information from multiple modalities (integration of auditory (voice) and visual (face) stimuli) provides many benefits (Bremner, Lewkowicz, & Spence, 2012) in the perception of speech (Stevenson, Segers, Ferber, Barense, & Wallace, 2014) and in social, communication and language development (Bahrick & Todd, 2012). However, evidence about the nature of the processes involved in the integration of these modalities remains limited, and we still know very little as to how attention to faces changes with age or experience and the maturation of attentional mechanisms or language skills. In particular, there have been few studies of how individual differences may affect such looking patterns. To the best of our knowledge, it is still unclear which aspects of a visual scene contribute to language development and how they do so (Libertus, Landa, & Haworth, 2017). Nor do we know how basic attentional abilities are used to construct higher-level language skills (D'Suoza, D'Suoza, Johnson, & Karmiloff-Smith, 2015).

Several studies have tested the association between the child's visual attention towards specific parts of the human face (i.e., mouth vs eyes), and their language development in experimental settings, by controlling and manipulating a set of variables. For example, exposing children to isolated static or dynamic faces on a screen (Birmingham, Bischof, & Kingstone, 2008a, 2008b, 2009; Fletcher-

Watson, Findlay, Leekam, & Benson, 2008; Kuhn & Land, 2006; Kuhn, Tatler, & Cole, 2009; Lewkowitz & Hansen-Tift, 2012; Walker-Smith, Gale, & Findlay, 1977). However, as reported by Slone and colleagues (2018): "such tasks fail to capture the dynamic nature of natural visual attention and the means by which children's natural visual environments are generated - active exploration" (p. 2). Among these experimental studies, a first attempt to link infants' looking behaviour and their vocal production abilities at the time of the task was made by Imafuku and colleagues (2019). They showed that the amount of imitation done by 6-month-old babies during a passive audiovisual task consisting of a female face reciting some vowels is significantly related to the proportion of time the child looks toward the mouth. The more children looked towards the mouth, the more they imitated. Such a relationship suggests a link between where the child looks and what they can replicate. In this case, the face represents the visual site of the articulatory and motor actions that the child can imitate. Within the same study, in a second experiment, Imafuku and colleagues (2019) found a direct effect of the child's vowel imitation when a speaker (on a screen) directs eye gaze toward the infant rather than away, highlighting an "engagement" effect that elicits the child's vowel reproduction. In support of these findings, it was shown that mouth looking facilitates or supports language development because of the redundant audio and visual stimulus provided by the articulatory action of the mouth. These, together, contribute to the acquisition of speech sounds and language in infancy, which also accords with the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000; Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012). Within this framework, expressive vocabulary is positively related to the time spent looking at the mouth (Morin-Lessard, Poulin-Dubois, Segalowitz, & Byers-Heinlein, 2019; Tsang, Atagi, & Johnson, 2018). Also, 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. However, Pons and colleagues (2019) and Hillairet de Boisferon and colleagues (2018) failed to find any relationships – at 12 and 18 months, respectively – between the child's looking behaviour and their linguistic level (in either comprehension or production), based on a parental questionnaire.

However, the visual exploration of the scene in natural contexts is still an under-investigated research topic, perhaps because it requires a lot of observational work (Fausey, Kayaraman, & Smith, 2016). Even though the relationship between specific patterns of visual attention and language development is now well established in ecological settings (see the studies regarding Joint Attention: Akthar & Gernsbacher, 2007; Beuker, Rommelse, Donders, & Buitelaar, 2013; Brooks & Meltzoff, 2005, 2015; Carpenter et al., 1998; Morales et al., 1998), less is known about the link between face and scene scanning and language abilities at the time of the task. Each child processes the visual scene differently. This may be based on exogenous factors, such as the behaviour of the person with whom they are interacting or the type of stimulus they are receiving (Goldstein & Schwade 2008; Libertus et al., 2017; Oller, Buder, Ramsdell, Warlaumont, Chorna, & Bakeman, 2013; Ramsdell, Oller, Buder, Ethington, & Chorna, 2012). Or it may be based on the infant's developmental abilities (such as their cognitive, socio-emotional or linguistic abilities) (Frank, Simmons, Yurovsky, & Pusiol, 2013; Kretch, Franchak, & Adolph, 2014; Pereira, James, Jones, & Smith, 2010; Raudies & Gilmore, 2014). For example, Fausey and colleagues (2016) suggest that visual experiences in infancy vary with the child's developmental stage. It is now well-established that children can learn the referential use of language by observing the target object or simultaneously listening to speech and attending to other people's movements and gestures, such as their faces (Gogate, Bolzani, & Betancourt, 2006). During the first year of life, the development of both eye movements and attentional control allows infants to attend visual and environmental scenes (bottom-up and top-down processes, Bahrick & Pickens, 1988; Itti & Koch, 2000, 2001; Parkhurst et al., 2002; Torralba, Oliva, Castelhano, & Henderson, 2006).

Some studies, using behavioural and electrophysiological techniques, have analysed how children allocate their attentional resources (i.e., endogenous attention) while exposed to a visual social scene (Helo, Rämä, Pannash, & Meary, 2016). Endogenous attention was found to be fundamental in guiding gaze allocation and exploration of the visual scene in the first years of a baby's life (for reviews, see Colombo, 2001; Johnson, 2002). More recently, Weatherhead, Arredondo, Nácar Garcia and Werker (2021) showed that visual information from talking faces provides additional linguistic information and supports the learning and recognition of novel words. In a previous study, Weatherhead and White (2017) had showed that 11-15-month-old infants exposed to auditory stimuli (familiar and non-familiar words) pronounced by seen human faces raised the level of attention towards unfamiliar words rather than familiar words. In this regard, the dynamic properties of the face (Guellaï et al., 2011, 2014) could affect audiovisual speech perception, in line with the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000). From these studies, two crucial understandings emerged, the first of which is that the attentional system is adaptive and varies with experience and context (Colombo, 2001). The second is that the rapid development of selective attention during the second half of the first year of life is important for later childhood cognitive competencies such as language development (word learning, Yu, Suanda, & Smith, 2018) and socio-emotional skills (e.g., Bhatt and Quinn, 2011), and therefore merits special attention in psychological research.

Recently several studies have tried to find an innovative way of studying this phenomenon in a real-life situation, by using head cameras (Braddik & Atkinson, 2011; Jayaraman, Fausey, & Smith, 2017). The use of a head-mounted camera in child development studies is becoming more and more widespread 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 (Aslin, 2009; Jayaraman, Fausey, & Smith, 2017; Noris, Keller, & Billard, 2011; Pereira, James, Jones, & Smith, 2010; Schmitow, Stenberg, Billard, & von Hofsten, 2013; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013, 2012). These studies confirm that the frequency with which children direct their attention towards faces tends to decline over the first year of life, which has implications for development (Libertus, Landa, & Haworth, 2017; Libertus & Needham, 2011, 2014). The visual scanning of social and interactive scenes depends on a large variety of external and individual factors (Liu et al., 2015; Tham, Bremner, & Hay, 2017) in some way related to the child's language skills at the time of the task. For example, as mentioned in the previous chapter, several studies have tried to account for the attentional shift from the eyes to the mouth region in children between 4 and 8 months, speculating that the shift depends on a specific developmental achievement (e.g., the onset of babbling). However, very few studies have investigated the relationship between the child's selective visual attention towards specific features of the visual scene and their language skills in a natural setting.

The present study

The present study tests the relationship between the child's actual language skills and their selective visual attention to faces in a natural task. Since a strong relationship exists between the child's selective visual attention and individual and external factors, we hypothesised that the child's vocal production abilities, as measured at the time of the task, could explain the directionality of their looking behaviour. Our goal is to investigate how selective visual attention towards faces or objects in social situations is driven by the child's language development and the type of stimulus to which they are exposed (familiar words or non-words).

We will address these research questions:

- Does the time of attention towards specific stimuli (e.g., the object, the adult's face) in an interactive task vary with the child's vocal skills at the time of the task? Is there any correlation between the child's looking behaviour and their vocal skills?

- Is the number of episodes of face fixation of 3+ secs explained by specific features of the child's vocal production (i.e., production of pre-canonical vocalisations, babbling or words)?

The exploration of the child's selective visual attention in an interactive setting will provide us with insight into how selective visual attention towards faces or objects during exposure to words and non-words varies with the children's daily quantity of vocal production and on the level of their vocal production (in terms of preverbal or verbal production).

Method

Participants

Thirty-nine Italian infants participated in the study. Four children were excluded due to problems with the video recordings or because the baby was distress. Six additional children were excluded. For these six children, we did not have the LENA recordings (i.e., quantitative language measure). The final sample is of 29, 12-19-month-old infants (female n = 10). The children had a mean age of 14.9 months (*SD* = 2.34). Their weight at birth was 3.215 kg on average (*SD* = 559 grams). On the day of the observation, the mothers had a mean age of 34 years (*SD* = 4.30) and fathers of 38.8 (*SD* = 6.65). The parents have a medium-high level of education: 17.2 years (*SD* = 2.23) for the mothers and 16.4 years for the fathers (*SD* = 3.88). All the families live in the North of Italy. None of the children wore glasses, none of the mothers had reported any diagnoses of developmental delays or disorders or worries about any aspects of their children's development. All the families participated in the study voluntarily.

Instruments

Parental interview and MacArthur–Bates Communicative Development Inventories

The day before the meeting, families were contacted for a short interview and completed the online MB-CDI questionnaire. When mothers were contacted, they were asked to provide the researcher with a list of sounds or words produced by the children. Mothers were also asked to provide information about the child's comprehension and vocal production of specific lexical items without any preliminary explanation of the task. Families filled out an online Italian short version of the MacArthur-Bates Communicative Development Inventory (containing the target words) – MB-CDI Words and Gestures (children aged 12-

15) – (Primo Vocabolario del Bambino - PVB, Caselli, Bello, Rinaldi, Stefanini, & Pasqualetti, 2015). The short Italian "Words and Gesture" version (100 words) is appropriate for children from 9 to 24 months.

Based on the mother's answers to the interview and the questionnaire, the researcher chooses the best set of stimuli to present to the child the following day.

Single object task – selective visual attention

The main caregiver (the father on just one case; otherwise the mother) was provided with a head-mounted camera to wear during the task. They put it on before starting the experimental session and wore it for five to ten minutes, to give the child the chance to familiarise themselves with it. During these minutes, the experimenter explained to the caregiver how the task works. In addition, a stationary camera placed in front of them recorded the scene from the front. During this familiarisation period, the child was able to gain confidence both with the experimenter and the video cameras. The caregivers were instructed to keep their children in front of them to assure the camera a good view of the child's head and gaze movements.

A set of audiovisual stimuli, words (comprehended by the child) (W) or non-words (NW) (no specific software was used to create these pseudo-words; pseudo-words were build by taking existing words in the Italian language and by changing one letter),were used (Table 1). Two sets of two words and two non-words each were used. Stimuli were chosen based on the parents' answers in the MB-CDI and the parental report, administered by phone the day before the experiment and chosen by the frequency of use. Words were all present in the MB-CDI and they were all objects familiar (comprehended) to the children. Non-words were created ad-hoc. Each set has specific characteristics (one of the two sets contains words with geminates). See Table 2 for more details on the normative data on the child's proportion of understanding/producing for the target words used in our task.

Table 1

Definition of the stimuli used during the vocal imitation task in the two sets (A and B). For words, the table also reports the proportions of the children's production for the target words according to our age range 12-19 months (Wordbank for the Italian language)

Set A

Symbol	Definition	Stimuli	Proportion of children Understanding (12 months-19 months)	Proportion of children Producing (12 months-19 months)
W	Words	cane	0.86-1.00	0.12-0.84
		libro	0.71-0.89	0.03-0.37
NW	Non-words	poda	-	-
		zava	-	-

Set B

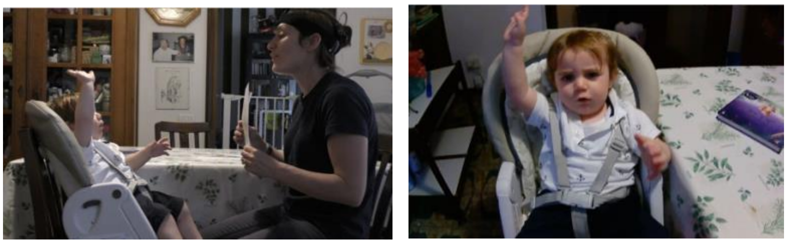
Symbol	Definition	Stimuli	Proportion of children Understanding (12 months-19 months)	Proportion of children Producing (12 months-19 months)
W	Words	palla	0.88-1.00	0.45-0.84
		gatto	0.36-0.95	0.00-0.32
NW	Non-words	banno	-	-
		codda	-	-

All play sessions were double video-recorded both with a stationary video camera (Panasonic, 4K, HC-WX970) (Figure 1A of the view), placed in front of the dyad, and with a head-camera (Polaroid Cube+, 1440p) on the mother's head (Figure 1B of the view). Each target word was presented with a target image to attract the child's attention (one image for each word). We provide the images of the target objects to the mother in plasticised sheets so they could present them to the infants. Each coded session has a median duration of around 3 minutes (*M* = 154.77 sec, *SD* = 53.45). The experimenter explained to the caregiver that they had to interact with their children to facilitate learning of the word reported below the picture in the plasticised sheets. The caregiver was required to use the target word at least six times in the session. They were also asked to engage the child in the task by using some strategies (for example, by telling a short story using that target word/non-word). Sessions in which repetitions of the target word/non-word were less than six were excluded from the analysis. The number of repetitions of each target word in a session was also controlled: The mothers repeated the target word (word or non-word) eight times on average (*SD* = 2.20). The number of vocal productions of the mother for the target word or non-word did not affect the child's looking behaviour in any way (all *ps* > .05). Altogether, 116 conditions (29*4 target stimuli each) were coded and analysed. The presentation of the set A or B did not affect the child's looking behaviour in any way (all *ps* > .05).

Figure 1

Screenshots of the views the experimenter used to code the looking episodes during the vocal imitation task (Figure A is the researcher's view; Figure B is the mother's view)

A. B.



Coding

Each session was coded offline. The videos from the stationary video camera and the head-mounted camera were analysed synchronously at half of their original speed (X 0.5).

The coding starts when the mother presents the picture to the child and ends when the mother presents the next picture to the child. The coding of the looking behaviour involved two steps: Step 1, coding of the child's looking behaviour; Step 2, classification of the behaviour into broad categories representing the attentional indexes.

Step 1. An offline coding scheme was adopted for analysing the child's looking behaviour in the session second by second. The starting coding scheme for the child's looking behaviour comprised five mutually exclusive categories (first column of the Table 2). The looking time, in seconds, that each child spent looking at each category was considered for the present study. Those moments when neither camera caught the child's looking or when the researcher passed the target words to the participants were excluded from the analyses.

Step 2. Firstly, three sub-categories (macro-categories, last column of Table 2) were created from the five categories described in Step 1 (first column of Table

2). The percentage of looking time to each of the three categories was considered. The “0” code was excluded from the count of the percentage. Secondly, since our main focus was on the role of the adult’s face, a more in-depth analysis was conducted on the episodes in which the child stayed focused on the face for a period of at least three seconds (episodes of attention).

Table 2

Categories used to code the child’s looking behaviour second by second (Step 1 — Starting micro-categories; Step 2 —final macro-categories)

Starting Coding categories	Definition	Example	Final categories (sub-categories)
Looking towards the face	The child clearly directs gaze towards adult’s face		FACE (%)
Looking towards the object	The child clearly looks at target object image (on plasticised sheet).		OBJECT (%)
Looking towards mother’s head-camera	The child directs gaze to head-camera		DISTRACTION (%)
Looking elsewhere	The child looks at other parts of room, other objects, or mother’s hands.		DISTRACTION (%)

Looking towards external observer	The child looks towards the stationary camera or toward the researcher/ experimenter.	 	DISTRACTION (%)
“0” code - non-identifiable looking	Within the same second two or more categories occur (e.g., when child looks for half second towards mother’s face and half-second towards experimenter; or when child looks face-obj-face-obj within one sec.). Or when child’s eye movements are not clear enough (e.g., when they bring paper too close to their eyes) or when they look at some parts of paper that do not correspond to image (e.g., the corners).	-	-

Note. Starting categories were the micro-categories used to analyse the child’s visual scene more in-depth. The final categories are the three macro-categories (face, object, and distraction) which were used to analyse the data, based on the specific hypotheses of the study.

Macro-categories (Percentage of attention time)

1) distraction time: sum of time, as a percentage, that children spent looking elsewhere, towards the head and stationary cameras. It is a measure of inattention tallied for the total duration of the task, for words and non-words.

2) looking time towards the target object: the percentage of time children spent looking towards the target object. It is a measure of attention towards the target object tallied for the total duration of the task, for words and non-words.

3) looking time towards the adult's face: the percentage of time children spent looking towards the face region. It is a measure of attention towards the adult's face tallied for the total duration of the task, for words and non-words.

Episodes of attention (number)

The episodes in which children maintain their attention for three consecutive seconds (Ruff & Lawson, 1990; Yu & Smith, 2016) towards the adult's face were also considered. They represent a measure of "face fixation".

Reliability

The same coding method was applied for all children. Two independent coders, the principal investigator and a research assistant, listened to and looked at 30% of the recordings. The reliability score for the categories referred to the child's looking behaviour was excellent (k = .97) (Fleiss, 1981).

Parent-child Interaction (vocal skills – qualitative measure)

After the single object task (see, *Single object task – selective visual attention)*, the caregiver was asked to play with their child as they usually do when alone. Infants were video recorded during around 20 minutes of spontaneous interaction with their caregivers while playing with toys provided by the experimenter (only the first 5 minutes of active play were analysed for all children). In each play session, mothers were asked to interact with their children as they would normally do, to make the situation as natural and spontaneous as possible. The researcher gave some parents a set of toys while others brought their own (an adaptation done in the face of the COVID-19 pandemic). The video observations were conducted at the families' homes, a natural context for sustaining spontaneous production and reducing distractions.

From when the interaction between the mother and the child begins, five minutes of spontaneous speech for each child were transcribed and analysed. Each child's vocalisation was phonetically transcribed by using ELAN (Version 6.0) [Computer software] (2020) and coded by using CHAT of CHILDES (Codes for the Human Analysis of Transcripts, MacWhinney, 2000). Cry, vegetative sounds and shouts were not transcribed. Three classes of production were identified, based on their phonological and semantic properties: (i.e., the percentage of quasi-resonant sounds, cooing sounds, fully-resonant sounds or raspberries not containing phonological structures or CV sounds), 2) babbling (i.e., the percentage of syllables consisting of at least one consonant and one vowel; reduplicated babbling was counted as one production) 3) words (i.e., the percentage of productions that have a sound-meaning correspondence) (Stoel-Gammon, 1989; Vihman, Macken, Miller, Simmons, & Miller, 1985; Vihman & McCune, 1994; see also Lang et al., 2019).

Reliability

A second independent transcriber transcribed 20% of the video recordings. Reliability index based on the child's amount of vocal production was good (80% of agreement).

The Language ENvironment Analysis system (LENA) (vocal skills – quantitative measure)

On the same week of the task and of the direct observation, the participating children wore a LENA device, a small digital recorder that families were requested to switch on at the beginning of the day and leave on for at least 10 hours. The measures of the child’s vocal production (quantity) were retrieved using the LENA system (LENA Colorado, Boulder). The children wore the recorder in the front chest pocket of clothing designed to optimise microphone placement and to reduce noise from clothing friction as much as possible. LENA provided us with two types of data:

- Language measures: Adult Word Count (AWC); Child Vocalisation Count (CVC); Conversational Turns Count (CTC).

- Data on the acoustic environment: meaningful speech, distant speech, noise, silence, TV and electronic sounds.

For the present study, we considered only the CVC. A vocalisation is defined as a "chunk" of speech-related sounds, separated by >=300ms of something else. This measure does not specifically indicate the number of words produced by the child, nor does it distinguish between preverbal and verbal speech. It does not, therefore, provide us with a qualitative measure of the child's language measure, but rather with a raw measure of how much the child vocalizes. For the present study, we retrieved a comparable measure for all children, considering the number of CVC/per minute. To retrieve the data, we subtract from the total recording time (*M* = 737 minutes, around 12 hours of recording, *SD* = 116) the time that LENA recorded as "silence". The amount of CVC was then divided by the final time (omitting all silent moments) (*M* =390 minutes, around 6.5 hours). The result was a measure of CVC/minute for each child.

Measures

To sum up, several measures were considered in the present study:

1) Child's selective attention, considered as the percentage of attention towards the adult's face, as the percentage of attention towards the object, and as the percentage of distraction. The number of episodes of face fixation was also taken into account.

2) Child's talkativeness (quantity) was measured by looking at the number of child vocalisations per minute produced in the 12 hours recorded by the LENA device.

3) Child's vocal skills (qualitative measure) were measured as the percentage of vocalisations, babbling, and words spontaneously produced in five minutes of interaction with their caregiver (video-recordings).

Data analyses

Two sets of analyses were conducted in order to establish how the child's selective visual attention varies with their vocal characteristics at the time of the test.

High vs Low Vocal Skills (CVC/MIN, pre-canonical vocalisations, babbling and words) – group differences. The children's vocal production was taken into account. Children were divided into two groups based on the features of their vocal production: low- vs high-vocal production.

Vocal production and Attention Time (percentage). First, a series of chi-square tests were run to control the variability in the two groups (low vs high vocal production) in terms of gender, age, parental age, parents' level of education. Then, a series of t-tests for independent samples were conducted with each attentional measure as a dependent variable (i.e., looking time towards the object, the face, distraction time) and the two groups of children as the independent variable. Finally, correlations were run between the time of attention towards the areas of the visual scene (object, face, distraction) and each measure of vocal production (pre-canonical vocalisations, babbling, words) for the entire group of children.

Vocal Production and Face Fixation Measures. A series of t-tests or ANOVAs were conducted for the entire group of children, to discover individual variability in language measures in terms of gender, age, parental age, parents' level of education. A series of regression analyses were run to test the extent to which the children's vocal skills in terms of pre-canonical vocalisations, babbling, words (based on 5 minutes of interaction during free play with their mother) predict the child's episodes of face fixation. Episodes of face fixation were considered the dependent variable, and the child's raw number of pre-canonical vocalisations, babbling, words produced in 5 minutes of spontaneous interaction during play was triggered as predictors (by controlling for individual factors, where the ANOVAs showed they were significant).

Results

High vs Low Vocal Skills (CVC/MIN, pre-canonical vocalisations, babbling and words) – group differences. The children were divided into two groups based on their median vocal production per minute as measured by LENA. LENA neither provides a measure of vocal skills nor differentiates between preverbal and verbal speech. The two groups did not differ in terms of gender ($p > .05$), age ($p > .05$), or for other socio-demographic characteristics (i.e., parental age, parental education) (all *ps* > .05). Therefore, in order to establish which are the vocal features of the two groups of children, the percentage of preverbal pre-canonical vocalisations, babbling, words produced in the 5-minutes of interaction with the mother were also taken into account.

High-vocal-production group (15 children). Children producing fewer pre-canonical vocalisations (55.8%), more babbling (12.9%) and more words (31.3%). Children with a median < 3.27 production per minute (LENA).

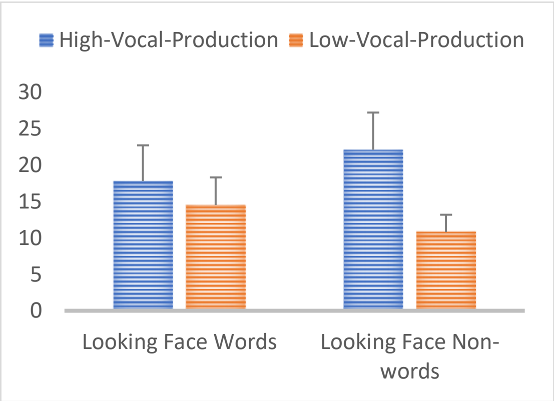
Low-vocal production group (14 children). Children producing more pre-canonical vocalisations (65.2%) but less babbling and fewer word forms (babbling = 10.1%; words = 24.7%). Children with a median > 3.27 production per minute (LENA).

The two groups differed significantly in their vocal production as counted as LENA (raw) ($p < .001$). But they did not significantly differ either in the number of tokens, precanonical vocalisations, babbling and words, as measured during the spontaneous interaction during play with their mother (all *ps* > .05).

Vocal production and Attention Time (percentage). A series of t-tests for independent samples were run for each looking behaviour (percentage of attention towards the object and the face or distraction). No differences emerged between the two groups in amount of time of distraction. The children in the low-production group looked significantly more towards the object (M = 49.1%, SE = 4.63) than children in the high-production group (M = 30.5%, SE = 3.55) when exposed to non-words ($t(27) = 3.167$, $p = .004$). The children in the high-production group looked significantly more towards the adult’s face (M= 22.1%, SE = 3.74) than children in the low-vocal-production group, when exposed to non-words (M = 10.8%, SE = 2.38) ($t(27) = -2.043$, $p = .05$). The same pattern also emerged for words, but it was not statistically significant (Figure 2). Children using more advanced forms look longer in total towards the face and display more episodes of fixation towards the face (4.79 episodes *vs* 1.27 episodes) when exposed to non-words than children with less advanced vocal forms.

Figure 2

Percentage of time looking towards the face during word and non-word exposure in the two groups of children (high vs low vocal production groups)



A correlation analysis (between time of attention and vocal measures) for the entire group of children showed a positive relationship between CV syllables production, i.e., neither 'pre-verbal' nor word production, and the percentage of attention towards the face during the whole task (r =.435, p = .018), and especially when children were exposed to non-words (r = .500, p = .006). This indicates that the more the child babbles, the more they look towards the face when exposed to words they have never encountered before.

Vocal Production and Face Fixation Measures. For the following analyses, the two groups of children were merged. A series of ANOVAs and t-tests were run to test the effect of individual variables (gender, age, parental age, parents' level of education) on the number of pre-canonical vocalisations, babbling, words. Age only significantly affected the number of words produced by the children (F (2,26) = 3.56, p = .043). The parents' level of education does not determine differences in the level of vocalisation in our sample. Regression models were run, controlling for the children's ages.

Regressions were carried out for the episodes of fixation lasting at least 3 seconds for the whole duration of the task, for word and non-word exposure with language measures as predictors, controlling for age. Three significant models emerged (Table 3).

Table 3

Regressions table

Dependent variable	Independent variables	*Adjusted R2*	*df*		*F*	*β*	*t*	*p*
Whole Task		**0.474**	**5**	**23**	**6.05**			**0.001**
	Preverbal vocalisations					0.297	1.91	0.068
	Babbling					**1.611**	**4.32**	**<.001**
	Words					-0.205	-1.529	0.14
	Age							>.05
Words		**0.445**	**5**	**23**	**5.49**			**0.002**
	Preverbal vocalisations					**0.278**	**2.446**	**0.023**
	Babbling					**0.986**	**3.6**	**0.001**
	Words					-0.078	-0.796	0.434
	Age							>.05
Non-words		**0.283**	**5**	**23**	**3.21**			**0.024**
	Preverbal vocalisations					0.018	0.235	0.816
	Babbling					**0.625**	**0.188**	**0.003**
	Words					-0.127	-1.87	0.074
	Age							>.05

Note. df = degrees of freedom.

In the three models, the production of CV syllables significantly explains the episodes of fixation towards the face during the task. To better understand the sign and degree of the predictions that emerged, simple correlations were run between the three language measures retrieved from the recording of the spontaneous interaction (pre-canonical vocalisations, babbling, words) and the episodes of face fixation. The correlation matrix showed that when exposed to words there is a positive relationship between the episodes of face fixation and the child's preverbal productions (both pre-canonical vocalisation, $r = .488$, $p = .007$; and babbling forms, $r = .616$, $p < .001$). And there is a positive relationship between face fixation and babbling production in the non-word condition ($r = .492$, $p = .007$). Overall, these results show that the number of CV syllables significantly predicts the episodes of face fixation. This established a relationship between vocal skills and the child's looking behaviour.

Discussion

In the present exploratory study, the selective visual attention of 29 children, aged 12-19 months, was analysed using a stationary video camera and a camera mounted on the parent's head. The single object task was ad-hoc designed to create a quasi-natural speech situation in which the child receives a series of common words and non-words from their principal caregiver. Unlike previous studies, the children are actively involved: children do not only attend to a screen with an isolated talking face, but are engaged in the task in their homes with their principal caregiver. This situation is similar to a natural setting, allowing us to simulate the child's everyday life. The task seems to be a promising way to observe changes in the child's looking behaviour when exposed to speech in natural settings. The use of words and non-words allows us to at least partially control the variability in the orientation of the children's attention provoked by different conditions (Libertus et al., 2017).

It is important to note that few studies have considered the child's orientation of visual attention, especially in natural contexts, because it requires a lot of observational work (Fausey, Kayaraman, & Smith, 2016). Recently, however, the adoption of different techniques to track eye movements in a natural setting - e.g., the portable eye tracker and the head-mounted camera - are being used more and more (Braddik & Atkinson, 2011; Jayaraman, Fausey, & Smith, 2017). In line with what has emerged from studies adopting these technologies, in our specially designed task the face seems not to be the primary focus of attention for children after the first year of life (Libertus, Landa, & Haworth, 2017; Libertus & Needham, 2011, 2014). Indeed, the children in our sample looked towards the adult's face less than 30% of the time overall (Figure 2).

Even if it is not the primary focus of a child's attention, the adult's face seems to play a crucial role, especially when the child's vocal skills are taken into account. Indeed, the children in the two language groups (low and high production) displayed different looking behaviours, especially in the amount of time they looked towards the face. Children in the low-production group looked longer towards the object than children in the high-production group. Complementarily, the children in the high production group looked longer towards the face than children in the low-production group. The children showing more interest in faces are those with more advanced language skills. These are the same ones who are already showing adult-like speech in their production. Since faces provide visual cues for speech processing, a plausible explanation for this difference in behavior is that children who have entered the babbling phase or who are beginning to produce words have a greater interest in the source of speech because they are more likely to look at something they can understand or even replicate. This is also in line with studies arguing for a link between the child's interest in the internal features of the face (i.e., the mouth) and their language skills. Specifically, this suggests that the looking behaviour reflects the child's active search for linguistically relevant information (see, for example, Tsang et al., 2018).

This result was further confirmed by the correlations and regression analyses that showed that the quantity of CV syllables produced during interaction significantly explains or predicts the episodes of face fixation for the task as a whole, for both words and non-words (in the full group of children). That is, the children who produced more CV syllables looked significantly more towards the adult's face. This supports the idea that the production of specific vocal routines (i.e., babbling) is related to the child's looking behaviour in natural settings. There could be two explanations for this. The first explanation is focused on the child. It concerns both what children know and what they choose to look at in their natural context, and what they can learn from what they see around them. First, it could be that for children producing more CV syllables, the adult's face serves as a visual cue or as a mechanism of visual reinforcement that provides redundancy (Bruderer, Danielson, Kandhadai, & Werker, 2015) to what the children are already able to produce (Vilain et al., 2019). Indeed, looking towards the adult's face could be a sign of mirroring and recognition (i.e., the child's motor knowledge influences their speech perception, Vihman, 1993; Vilain et al., 2019), supporting a link between perception and production (Majorano, Vihman, DePaolis, 2014). On the other hand, looking towards faces can also be seen as an informative mechanism for building phonological knowledge.

One piece of empirical evidence supporting all this concerns the features of early vocal production. Indeed, the phonological repertoire of children beginning to produce adult-like sounds is characterised by what has been called 'visible speech' (that is, the speech that children can get from seeing the adults around them speaking: Caldognetto & Zmarich, 2000). In support of this, Boysson-Bardies and Vihman (1991) analysed the place of articulation (labials, dentals, velars) in the babble and words produced at four different developmental points, based on the number of words produced in a 30-minute recording session (zero, 4, 15 and 25 words), in five infants each from four different language backgrounds (English, French, Japanese, and Swedish). Boysson-Bardies and Vihman found that although French- and English-learning infants produced a proportionally higher number of labials in their babbling and word production than Japanese and Swedish infants, in accord with the proportion of occurrence of those consonants in the adult target words in those languages, in all four groups the proportion of labials was about 5% higher in words than in babble, over the entire period sampled. Vihman and colleagues (1985) also note the greater use (in their American English data) of labials in words as compared with babble (or 'non-words') and comment that 'labials may be preferred, once the children began to acquire adult words in larger numbers ... because of their special perceptual salience: they can be seen as well as heard' (Vihman et al., 1985, p. 435). In other words, these studies provide some evidence that labials, as the most visible place of articulation, support word production and may actually bias infants to select more labial-initial target words to attempt, as they are easy to replicate.

The second explanation is socially based and it relates to what the caregiver does. When children produce CV forms, the caregivers around them recognise such forms (since they are speech-like) and they tend to support this behaviour by repeating or providing a response to what the child produces (Gros-Louis, West, Goldstein, & King, 2006). Although this finding has recently been contested (Athari et al., 2021; Fagan & Doveikis, 2017), the child's attraction to the face could be explained by maternal behaviour (i.e., expressive behaviour, emotional link), but that is beyond the scope of the present study.

To summarise, we believe that the face plays a functional role in the child's language learning process. Indeed, we found that a child's attraction to the adult face is shaped by the child's language skills at the time of the task. On the one hand, the face can be used as a redundant cue due to the articulatory movements that come from the mouth. On the other hand, it provides the child with emotional reinforcement, through emotional expressions that also come from the eyes.

Limits and conclusions

Although this study represents an original contribution from a methodological perspective, it has some limitations. A first limitation is that we do not have a direct measure of word learning during the task. A direct measure of word learning would have been valuable for explaining how visual attention supports novel word learning (Yu & Smith, 2012). Second, we do not have measures related to the caregiver's behaviour (gaze) or their language use (input) during the task. For example, we did not control prosodic aspects of parental language during the task to test if it affected the child's visual orientation. Finally, we should consider the reliability of the LENA measures. LENA cannot provide a realistic measure of vocal skills in children between 12 and 19 months of age. Indeed, basing the evaluation of vocal advance only on quantitative data extrapolated from LENA fails to inform us about qualitative aspects of the child's speech, which would give a better measure of the child's language development. Instead of basing their findings on isolated and automatic language skill scores, language research needs to combine quantitative and qualitative aspects of language, integrating different measures, as we have attempted to do in the present study.

The present study offers insights into what happens in a natural situation when children are exposed to words and non-words. The use of an innovative tool, head-mounted cameras, affords us measures of attention that would be difficult to retrieve in a spontaneous setting, without the use of an eye-tracking device and without interfering with the child's spontaneous behaviour. At this stage, it seems that it would be interesting to explore in greater depth what exactly children look at in a face when listening to new lexicon items or a story and whether the specific patterns of attention within a face could be explained by the same linguistic characteristics.

References

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: official publication of the American Academy of Optometry, 86*(6), 561–565. https://doi.org/10.1097/OPX.0b013e3181a76e96

Athari, P., Dey, R., & Rvachew, S. (2021). Vocal imitation between mothers and infants. *Infant Behavior & Development, 63*, 101531. https://doi.org/10.1016/j.infbeh.2021.101531

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., & 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.

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

Bhatt, R. S., & Quinn, P. C. (2011). Different Approaches to the Study of Early Perceptual Learning. *Infancy: the official journal of the International Society on Infant Studies, 16*(1), 61–68. https://doi.org/10.1111/j.1532-7078.2010.00056.x

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a) Gaze selection in complex social scenes, *Visual Cognition, 16,* M 2-3, 341-355, DOI: 10.1080/13506280701434532

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Social attention and real-world scenes: the roles of action, competition and social content. *Quarterly* *journal of experimental psychology, 61*(7), 986–998. https://doi.org/10.1080/17470210701410375

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition, Special issue, Eye guidance in natural scenes*, 1–21.

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

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision research, 51*(13), 1588–1609. https://doi.org/10.1016/j.visres.2011.02.018

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.). (2012). *Multisensory development*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.001.0001

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental science, 8*(6), 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of child language, 35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

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

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 176. https://doi.org/10.2307/1166214

Caselli, M. C., Bello, A., Rinaldi, P., Stefanini, S., & Pasqualetti, P. (2015). *Il Primo Vocabolario Del Bambino: Gesti, Parole e Frasi. Valori di riferimento fra 8 e 36* *mesi delle forme complete e delle forme brevi del questionario MacArthur-Bates CDI*. Milano: Franco Angeli.

Colombo, J. (2001). The development of visual attention in infancy. *Annual review of psychology*, *52*, 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

D'Souza, D., D'Souza, H., Johnson, M.H., & Karmiloff-Smith, A. (2015) Concurrent Relations between Face Scanning and Language: A Cross-Syndrome Infant Study. *PLoS ONE*, *10*(10): e0139319. https://doi.org/10.1371/journal.pone.0139319

ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fagan, M. K., & Doveikis, K. N. (2017). Ordinary interactions challenge proposals that maternal verbal responses shape infant vocal development. *Journal of Speech, Language, And Hearing Research, 60*(10), 2819–2827. https://doi.org/10.1044/2017_JSLHR-S-16-0005

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

Fleiss, J.L. (1981) *The measurement of interrater agreement. In: Statistical Methods for Rates and Proportions (2nd Edition)*. John Wiley, NewYork, 212-236.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*(4), 571–583. https://doi.org/10.1068/p5705

Frank, MC., Simmons, K., Yurovsky, D., & Pusiol, G. (2013). Developmental and postural changes in children's visual access to faces. In: Knauff, M.Pauen, M.Sebanz, N., Wachsmuth, I., editors. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*; Austin, TX: Cognitive Science Society, p. 454-459.

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

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509–516. https://doi.org/10.1177/0165025406071914

Guellaï, B., Coulon M., & Streri A. (2011). The role of motion and speech in face recognition at birth. *Visual Cognition, 19*,1212–1233 10.1080/13506285.2011.620578

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. https://doi.org/10.3389/fpsyg.2014.00812

Helo, A., Rämä, P., Pannasch, S., & Meary, D. (2016). Eye movement patterns and visual attention during scene viewing in 3- to 12-month-olds. *Visual neuroscience*, *33*, E014. https://doi.org/10.1017/S0952523816000110

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Developmental science, 20*(3), 10.1111/desc.12381. 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. *Journal of Experimental Child Psychology*, *172*, 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental science*, *22*(6), e12825. https://doi.org/10.1111/desc.12825

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*, 194–203. https://doi.org/10.1038/35058500

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

Johnson, M. H. (2002). The development of visual attention: A cognitive neuroscience perspective. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain development and cognition: A reader* (pp. 134–150). Blackwell Publishing.

Kretch, KS, Franchak, JM, & Adolph, KE. (2014). Crawling and walking infants see the world differently. *Child Development, 85*(4), 1503-18. doi: 10.1111/cdev.12206

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

Kuhn, G., & Land, M. F. (2006). There's more to magic than meets the eye. *Current Biology, 16*, R950–R951. doi:10.1016/j.cub.2006.10.012

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition, 17*, 925–944. doi:10.1080/13506280902826775

Kushnerenko, E., Tomalski, P., Ballieux, H., Potton, A., Birtles, D., Frostick, C., & Moore, D. G. (2013). Brain responses and looking behavior 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

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical Babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports, 6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

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

Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental science, 14*(6), 1355–1364. https://doi.org/10.1111/j.1467-7687.2011.01084.x

Libertus, K., & Needham, A. (2014). Face preference in infancy and its relation to motor activity. *International Journal of Behavioral Development, 38*(6), 529–538. https://doi.org/10.1177/0165025414535122

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*, Article 1976. https://doi.org/10.3389/fpsyg.2017.01976

Liu, S., Xiao, W. S., Xiao, N. G., Quinn, P. C., Zhang, Y., Chen, H., Ge, L., Pascalis, O., & Lee, K. (2015). Development of visual preference for own- versus other-race faces in infancy. *Developmental Psychology*, *51*(4), 500–511. https://doi.org/10.1037/a0038835

MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription format and programs, Volume II: the database. *Comput. Linguist.*, *26*, 657–657. doi: 10.1162/coli.2000.26.4.657

Magno 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.

Majorano, M., Vihman, M.M., & DePaolis, 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

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

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*(3), 283–298. https://doi.org/10.1016/S0193-3973(99)00040-

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology, 55*(8), 1640–1655. https://doi.org/10.1037/dev0000750

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior & Development, 21*(3), 469–482. https://doi.org/10.1016/S0163-6383(98)90020-0

Mundy, P., Kasari, C., Sigman, M., & Ruskin, E. (1995). Nonverbal communication and early language acquisition in children with Down syndrome and in normally developing children. *Journal of Speech & Hearing Research, 38*(1), 157–167. https://doi.org/10.1044/jshr.3801.157

Noris, B., Keller, J., & Billard, A. (2011). A wearable gaze tracking system for children in unconstrained environments. *Computer Vision and Image Understanding, 115*, 476-486.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Scinces of the United States of America, 110*, 6318–6323. doi: 10.1073/pnas.1300337110

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research, 42*(1), 107–123. https://doi.org/10.1016/s0042-6989(01)00250-4

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated views. *Journal of Vision, 10*, 1–13.

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. *Infant Behavior and Development*, *54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Ramsdell, H. L., Oller, D. K., Buder, E. H., Ethington, C. A., & Chorna, L. (2012). Identification of prelinguistic phonological categories. *Journal of Speech Language and Hearing Research, 55*, 1626–1639. doi: 10.1044/1092-4388(2012/11-0250)

Raudies, F., & Gilmore, R. O. (2014). Visual motion priors differ for infants and mothers. *Neural Computation*, *26*(11), 2652–2668. https://doi.org/10.1162/NECO_a_00645

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*(1), 85–93. https://doi.org/10.1037/0012-1649.26.1.85

Schmitow, C, Stenberg G, Billard A, & von Hofsten C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development.*, *37*(5), 468–474.

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C. H., Franchak, J. M., Pearcy, D., Suarez-Rivera, C., Xu, T. L., Zhang, Y., Smith, L. B., & Yu, C. (2018). Gaze in Action: Head-mounted Eye Tracking of Children's Dynamic Visual Attention During Naturalistic Behavior. *Journal of visualized experiments, 141*. 10.3791/58496. https://doi.org/10.3791/58496

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother’s view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17.

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in psychology, 5,* 379. https://doi.org/10.3389/fpsyg.2014.00379

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. *First Language, 9*(6), 207–223. https://doi.org/10.1177/014272378900900607

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of child language, 42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Tenenbaum, E., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology, 5,* Article 490. https://doi.org/10.3389/fpsyg.2014.00490

Tham, D.S.Y., Bremner, J.G., & Hay, D. (2017). The other-race effect in children from a multiracial population. A cross-cultural comparison. *Journal of Experimental Child Psychology, 155*, 128-137. https://doi.org/10.1016/j.jecp.2016.11.006

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review, 113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

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

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*(1-2), 61–82.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language (Baltimore), 61*(2), 397–445. https://doi.org/10.2307/414151

Vihman, M. M., & McCune, L. (1994). When is a word a word? *Journal of Child Language, 21*(3), 517–542. https://doi.org/10.1017/S0305000900009442

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. *Developmental Science*, *22,* e12830–n/a. https://doi.org/10.1111/desc.12830

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313–326. https://doi.org/10.1068/p060313

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition, 160,* 103–109. https://doi.org/10.1016/j.cognition.2017.01.002

Weatherhead, D., Arredondo, M. M., Nácar Garcia, L., & Werker, J. F. (2021). The Role of Audiovisual Speech in Fast-Mapping and Novel Word Retention in Monolingual and Bilingual 24-Month-Olds. *Brain sciences, 11*(1), 114. https://doi.org/10.3390/brainsci11010114

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? using a head camera to study visual experience. *Infancy, 13*(3), 229–248.

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

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, L.B. (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

Yu, C., & Smith, L. B. (2016). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current biology, 26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental science, 22*(1), e12735. https://doi.org/10.1111/desc.12735

CHAPTER 5

Conclusion

General Discussion

This dissertation presents significant results about selective visual attention and its relationship with the child's language skills and offers innovative insights in developmental psychology.

The literature review prompted a vital contribution and discussion because until now, to the best of our knowledge, no one has systematically reported the results of studies investigating the selective visual attention phenomenon in infancy. In particular, both endogenous (e.g. age, language background) and exogenous factors (e.g., native vs non-native speech exposure, IDS vs ADS) significantly explain the looking behaviours of children given a passive speech task. Significant looking shifts have emerged about specific ages or developmental periods. What most of these studies have attempted to do in explaining the child's looking shifts was to assign these changes to specific developmental stages, with particular reference to the time at which children start babbling, which represents one of the most important achievements in language acquisition due its relationship with later language skills. But none have included actual measures of babbling production in their studies to empirically support this hypothesis. The most-commonly supported hypothesis to explain these shifts was the language expertise hypothesis, which speculates on the potential link between the child's looking preference for the mouth or the eyes and their linguistic skills. It notes that pre-babbling children look more towards the eyes, babbling children shift to the mouth, and more linguistically advanced children return to the eyes. What has emerged from the more recent literature is a turnaround in children who start producing their first words, i.e., children aged around 12 months or older. Also, when infants become linguistically "independent" or advanced (i.e., children adopting more structured or adult-like vocal forms), they continue to look more towards the mouth. This finding undermines the language expertise hypothesis, which is now difficult to support.

In addition, no one has systematically investigated the relationship between the child's selective visual attention towards a specific region of the human face and their language skills. But, as reported in the second chapter, it has emerged that some studies have linked language skills (expressive or receptive, measured through self-report measures filled out by parents) with children's looking behaviours. Generally, this area of the literature showed an association between the time spent looking towards the mouth and an infant's early expressive language skills. Importantly, this applies even after the end of the first year of life, supporting the idea that language acquisition is part of a long, broader learning process. However, no associations were found between children's receptive skills and their looking behaviour, probably because the areas of interest analysed are more relevant to expressive than to receptive skills. Recently a few studies have advanced hypotheses relating the child's looking preferences to their social development. However, more research is needed to confirm these associations.

The second core chapter (Chapter 3) extended the results of the studies reported in Chapter 2 but with significant differences. Firstly, it was the first study to use this experimental procedure in an Italian context with Italian infants. Second, it was the first study to extrapolate the children's vocal measures from the transcriptions of a direct observation while they were playing with their mother, rather than using an indirect self-report questionnaire.

From this chapter, two main findings emerged. The first is the relationship between the changes in attentional patterns and the child's age: the older the children in our sample, the more they focused on the mouth area, especially when exposed to their native speech. This pattern shows that, in our sample, children older than 12 months still need the redundancy of the visual cue produced by the mouth because at this stage they are not yet productively "independent", even if they are processing language sufficiently well.

The second finding is the relationship between the changes in the child's attentional patterns and their language skills. There are two elements in this relationship: not only older children but also the more linguistically advanced children (i.e., those producing more babbling and words) looked more towards the mouth. This finding is important at this stage since it represents evidence against the language expertise hypothesis. Following this hypothesis we would have expected that children who are entering the babbling stage or who still do produce few identifiable words would have looked more towards the mouth than more linguistically advanced children. Our finding is not what the language expertise hypothesis would predict: here, children did not stop looking towards the mouth in favour of the eyes when listening to their native language. This finding needs more interpretation, but we can attempt several explanations: In the first year of life the mouth represents both a facilitative mechanism for learning a language and a visual cue for children. The mouth also attracts more linguistically advanced children because they recognise the specific vocal sequences coming from it, or because they are more likely to pay attention to the source of the speech because they understand what those movements are about in this specific stage of their development. On the contrary, those children whose phonological repertoire is not yet stable showed no particular attraction to the mouth region, probably because they are still in an exploratory phase. During this phase they are not particularly attracted by the mouth and its articulatory movements and this may be related to their linguistic immaturity.

Finally, this trend, i.e., the more "expert" the children were, the more they looked towards the mouth, was also found in the longitudinal part of Chapter 3. The children who looked more at the mouth at the time of the experiment achieved higher expressive vocabulary scores three months later. In the light of this finding, looking at the mouth provides the child with more opportunities to practise and succeed in vocal development.

The last main chapter (Chapter 4) is the most innovative one methodologically. To the best of our knowledge, nobody has previously tried—in a spontaneous setting in which children are exposed to speech—to link the child's selective visual attention with the child's current language skills. Most of the studies investigating the phenomenon of selective visual attention were conducted in laboratory settings, not taking into account what a child looks at in a spontaneous context or interaction with their principal caregiver.

In recent decades attention has begun to be given to the complexity of reality, and to extending laboratory results to real-life situations. Here, the aim was to understand what a child looked at during a simple speech task in which the caregiver showed the child some pictures of the named target objects. We discovered that children generally looked more towards the object than to the face or other parts of the room during this task. This tendency was shown in previous studies, indicating that younger children (in the first months of life) generally looked significantly more towards the adult face, with a tendency for this pattern to decrease over time. Here, the children are beyond the age threshold identified in the literature. Despite this, we found that only the child's looking towards the face in a natural or spontaneous context is significantly linked to the child's language skills at the time of the task. In particular, we found that looking towards the adult face does not seem to be a common pattern among children who are less vocally competent. As further support, we found that how much children babble significantly predicts the episodes of face fixation for the entire duration of the task. We are therefore led to support the hypothesis that the child's attention to faces may be driven by their particular level of vocal development.

Limitations

This dissertation has several limitations, all suggesting research that will need to be undertaken in future studies. The first is the limited sample size and the age range of the population involved in our empirical studies. Involving more than two age groups, with more children in each, would have improved the statistical power of the analysis. In addition, a longitudinal perspective, with the same children at different age points, would have provided more knowledge of what exactly happens at different developmental stages. This would have allowed us to see how the looking behaviour of each child changes over time, and whether a correspondence exists between the child's looking behaviour and their language skills on test-day and at a range of age points.

A second limitation that diminishes the generalisability of our findings to other populations is the low level of individual variation in our sample. Our results would be wider in scope, and more robust, with the inclusion of children from a different linguistic or cultural background (i.e., bilinguals) or of different socio-economic status. Recent studies have, for example, shown that membership in different cultural traditions may also have an impact on looking behaviour.

The third limitation relates to the adoption of the LENA device in one of the three core chapters. Although this tool has not been validated for Italian, we decided to overcome this limitation by not only including the raw data produced by LENA but also transcribing qualitative data on the language skills of the involved children.

Main findings, future directions and implications for practice

This dissertation helps us to clarify the developmental role of selective visual attention. It also opens new lines of research in the study of children's individual skills in selectively attending to important areas of the face or visual scene. The main findings from this dissertation require in-depth comment since they could have significant repercussions not only for research but also for educational or clinical practice.

The role of the mouth as a visual cue or facilitative cue for language acquisition

The face, and more specifically the mouth region, plays a vital role in human communication and the expression of emotional states. A child who has no access to these channels (mouth, face) will suffer negative consequences in their development, particularly their language development, the focus of the present dissertation. Seeing mouth movements provides information about the temporal and phonetic properties of speech and helps listeners to decode vocal signals (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Grant & Greenberg, 2001; Yehia, Rubin, & Vatikiotis-Bateson, 1998). Additionally, it helps listeners to discriminate words from pseudowords, both in acoustically "normal" and especially in noisy speech conditions (Fort, Spinelli, Savariaux, & Kandel, 2012; Grieco-Calub, & Olson, 2015; Jerger, Tye-Murray, & Abdi, 2009; Lalonde & Holt, 2015). In addition, our findings strongly support a clear relationship between mouth looking and expressive language skills: The more children look towards the mouth, the faster the rate of growth of their expressive vocabulary.

This significant relationship has several implications for practice, not only in children with typical development or in those who are learning a second language (see Hirata & Kelly, 2010), but also in atypical situations (Young, Merin, Rogers & Ozonoff, 2009). For example, in several clinical fields, it is accepted practice to use lip-reading (or speech reading) as a therapy to teach language to children with hearing loss (see Nasim, Fahad, Ahmad, Khan, & Shah, 2017). More recently, it has emerged that approximately the 5% of the world's population suffers for hearing loss (Ong, 2020) and that these are the same who count on visual cues, lip-reading and facial expressions to effectively communicate with other (Fortin, 2020, see also Green, Staff, Bromley, Jones, & Petty, 2021).

In addition, special attention needs to be paid to the effects of the wide spread of SARS-CoV-2 as regards the use of face masks. These days wearing masks is common practice, both in daily life situations 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 mouth of people talking to children play a key role in language acquisition and communication development. Not having access to motor information or having a degraded signal affects the perception of speech. It impedes a child learning a new language because audiovisual speech processing supports language acquisition and development, especially in the first years of life. This is true not only for language per se but also for socio-emotional and interactional development.

All this, again, supports the idea that audio and visual integration of speech signals in natural situations is a basic process that needs to occur for the child to access all the necessary information to process and learn a language.

The role of the mouth movements as a mirroring mechanism for vocal imitation

As reported by Bahrick, Todd, and Soska (2018), “selective attention is the gateway for information pickup and processing and the basis for all we perceive, learn and remember” (p. 2207). Moreover, it is now well established that in the early stages of language acquisition children benefit from selective attentional skills that allow word segmentation, thus reducing the complexity of the speech input (Diego-Balaguer, Martinez-Alvarez, & Pons 2016). The activation of attentional mechanisms influences how children perceive speech. At the same time, attentional mechanisms fashion the production of sounds, motor actions or words. For example, it is well known that selective attention towards specific aspects of the visual scene or specific actions (e.g., motor action) elicit imitative behaviours, both of words and actions (Imafuku et al., 2019). And it is also well-documented that to imitate, children need both auditory and visual (visual-motion processing) information, as supported in the Legerstree study (1990). According to Legerstree, audio and visual information elicits the early imitation of speech sounds in infants at 3-4-months. But, whereas in previous studies (Kuhl and Meltzoff, 1982) infants simultaneously saw the matched faces, the infants in Legerstree’s study were presented with two different conditions (matched/mismatched) to demonstrate the effect of audiovisual integration in language imitation. Legerstree showed how infants use multi-modal cues to build lexical representation and how multimodality is necessary for speech perception and production. Indeed, “children learn by doing, but also learn a great deal by watching others’ actions.” (p.167, Ruff & Rothbart, 1996). This is the main reason why studying these imitative mechanisms contributes to understanding language acquisition processes. Indeed, through looking and gaining information from the adults around them, children were exposed to models that they sometimes imitate. Imitation happens when someone observes particular actions and translates such observed action into an activity (vocal or motor). However, when the children's vocal behaviour is observed in a natural context, it has been shown that the number of vocal imitations is very low (Athari, Dey, & Rvachew, 2021).

The multisensory mechanisms that underpin speech production and imitative behaviours in adults and children are also corroborated by the neurophysiological evidence reported by Rizzolatti and colleagues (1998, 2004). The discovery of mirror neurons helps explaining these mechanisms from another perspective (i.e., neurophysiological, but see Vihman, 2002 for a theoretical account). The discovery of the mirror-neuron system in monkeys and the echo-neuron system in humans prompted Rizzolatti et al. to hypothesise a relationship between motor actions and speech in human beings. No direct evidence has been provided for the presence of the same kind of neurons in humans, but neuroimaging studies have provided evidence of the activation of neurons in the prefrontal cortex during motor imitation tasks. This means that when an infant or an adult is listening to someone talking, the speech-related motor centres are activated. Neuroimaging data has also shown that adult listeners activate motor brain areas during speech perception (in line with the motor theory of speech perception). Similarly, in infants aged 11 and 12 months native speech activates auditory brain areas, while the perception of non-native speech activates motor brain areas (Kuhl, Raminez, Bosseler, Lin, Imada, 2014). These are significant findings in that they support the role of multi-sensoriality in infants' speech perception and production. They also mean that when a child starts to imitate adults' speech behaviour, neural activation of audiovisual and motoric patterns occurs simultaneously (Imafuku et al., 2019).

Our findings suggest that infants who babble more or who produce more words look longer towards the adult face or the mouth. The vocal or motor responses produced by infants may be influenced by, and linked with, what they have heard or seen (i.e., the articulatory filter hypothesis: Vihman, 1991, 1993, 1996), suggesting a dynamic relationship between production and perception. This pattern can be explained in three ways: firstly, the infant recognises what they can already reproduce (DePaolis, Vihman, & Keren-Portnoy, 2010; Majorano, Vihman, & DePaolis, 2014). Secondly, a novelty effect makes them more engaged in some of the vocal patterns coming from the adult face or mouth. Third, children at this stage are also more mature and their attentional system is much more developed. This leads them to look at what they need, based on what they already know or on what they are interested in. A more in-depth analysis needs to be conducted before we can say more on this point.

Concluding remarks on the relationships between selective visual attention and language skills in infancy

This dissertation supports the idea of a link between the child's selective visual attention towards the human face or specific region of the human face (especially the mouth) and the child's current language knowledge. It is now eminently arguable that in a child's development, attention and language are complex, dynamic and complementary processes that influence one another: the growth of one system is a function of that of the other. Indeed, as suggested by Bahrick and colleagues, "*What we perceive, learn, and remember influences what we attend to next, creating a cycle of attention* --> *perception* --> *learning* --> *memory* --> *attention*" (Bahrick, Todd, & Soska, 2018, p. 2207). In other words, language skills drive attention towards specific information sources in the environment or in parts of the face. A child directs their attention towards specific parts of the visual scene according to what they already know and what they are interested in. As further support, attentional shifts emerge in correspondence with specific developmental stages, indicating that children are naturally attracted by stimuli from which they get crucial information to extrapolate novel pieces of information. Also, children's activation of attentional resources enhances their chances of reetaining new lexicon items, thus improving or shaping their language skills or language knowledge. Indeed, attention is a cognitive process that, like other processes, will be affected by individual differences. The attentional abilities that children display depend on two sets of factors, endogenous (including their age, their temperament, their curiosity) and exogenous, both of which help them to learn more and/or different things. The exogenous factors relate to all aspects of the environment around them: the type of environment in which they are growing up, the ambiguity or the novelty of the stimuli they receive. For an in-depth analysis of these factors, see Perone and Spenser (2013).

The evidence taken together strengthens the idea that selective attention is fundamental to language acquisition and learning (see also, Amso & Johnson, 2006; Bhatt & Quinn, 2011; Markant & Amso, 2013; Walther, Rutishauser, Koch, & Perona, 2005). Further research is needed to determine the causes of this relationship (i.e., does selective attention drive language or do the child's language skills affect their selective attention?) and the long-term impact of early selective attentional mechanisms on future learning.

References

Amso, D, & Johnson SP. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*, 1236–1245.

Athari, P., Dey, R., & Rvachew, S. (2021). Vocal imitation between mothers and infants. *Infant behavior & development*, *63*, 101531. https://doi.org/10.1016/j.infbeh.2021.101531

Bahrick, L. E., Todd, J. T., & Soska, K. C. (2018). The Multisensory Attention Assessment Protocol (MAAP): Characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *Developmental Psychology, 54*, 2207–2225. doi:10.1037/dev0000594

Bhatt, R. S., & Quinn, P. C. (2011). How does Learning Impact Development in Infancy? The Case of Perceptual Organization. *Infancy: the official journal of the International Society on Infant Studies, 16*(1), 2–38. https://doi.org/10.1111/j.1532-7078.2010.00048.x

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Diego-Balaguer, R. de, Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7*, Article 44. https://doi.org/10.3389/fpsyg.2016.00044

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants?. *Infant behavior & development*, *34*(4), 590–601. https://doi.org/10.1016/j.infbeh.2011.06.005

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2012). Audiovisual vowel monitoring and the word superiority effect in children. *International Journal of Behavioral Development, 36*(6), 457–467. https://doi.org/10.1177/0165025412447752

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

Grant, K.W., & Greenberg, S. (2001). Speech intelligibility derived from asynchronous processing of auditory-visual information. *Proceedings of the Workshop on Audio-Visual Speech Processing*.

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

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)

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

Jerger, S., Tye-Murray, N., & Abdi, H. (2009). Role of visual speech in phonological processing by children with hearing loss. *Journal of speech, language, and hearing research*, *52*(2), 412–434. https://doi.org/10.1044/1092-4388(2009/08-0021)

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 111*(31), 11238–11245. https://doi.org/10.1073/pnas.1410963111

Lalonde, K., & Holt, R. F. (2015). Preschoolers benefit from visually salient speech cues. *Journal of speech, language, and hearing research*, *58*(1), 135–150. https://doi.org/10.1044/2014_JSLHR-H-13-0343

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior & Development*, *13*(3), 343–354. https://doi.org/10.1016/0163-6383(90)90039-B

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/

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204. https://doi.org/10.1080/15475441.2013.829740

Markant, J., & Amso, D. (2013). Selective memories: Infants' encoding is enhanced in selection via suppression. *Developmental Science, 16*(6), 926–940.

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.

Ong, S. (2020, June 9). How face masks affect our communication. BBC Future. https://www.bbc.com/future/article/20200609-how-face-masks-affect-our-communication

Perone, S., & Spencer, J. P. (2013). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in psychology, 4*, 648. https://doi.org/10.3389/fpsyg.2013.00648

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience, 27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and clinical neurophysiology, 106*(4), 283–296. https://doi.org/10.1016/s0013-4694(98)00022-4

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

Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 69–84). Hillsdale, NJ: Erlbaum.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61–82.

Vihman, M. M. (1996). *Phonological development (1st ed.)*. Oxford, UK: Basil Blackwell.

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.

Walther, D., Rutishauser, U., Koch, C., & Perona, P. (2005). Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Computer Vision and Image Understanding*, *100*(1), 41–63. https://doi.org/10.1016/j.cviu.2004.09.004

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*(1), 23–43. https://doi.org/10.1016/S0167-6393(98)00048-X

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

Ringraziamenti/Acknowledgments

Un ringraziamento speciale alla mia tutor, Prof.ssa Marinella Majorano, per i suoi preziosi consigli e insegnamenti, per avermi sempre sostenuta e supportata in ogni momento di questo percorso. Grazie per il tempo dedicatomi, per l'ammirevole pazienza nel guidarmi e spronarmi ogni giorno di più.

A special thank you to my co-supervisors, Prof Marilyn Vihman and Prof. Tamar Keren-Portnoy, for their precious and helpful advice and all their support. For the time they reserved for me in all these years and especially for the great opportunity they gave me to visit their University and make me join their research group. Thank you for teaching me a lot and for sharing with me your thoughts about my research project.

Un ringraziamento a tutti i miei colleghi, ognuno di loro ha contribuito a farmi sentire un po' meno sola durante questo percorso. Un grazie speciale Beatrice, che con il suo entusiasmo e la sua forza è riuscita persino a rendere meno 'tragici' alcuni miei momenti di sconforto.

Thank you to all the colleagues met during my stay(s) abroad. They have enriched the visiting experience and made it much more magic. Thank you to Rui-Qi and Margherita for our afternoon tea and Thursday cakes sessions (and not only!).

Senza la partecipazione delle famiglie al progetto tutto questo non sarebbe stato possibile. Quindi ci tengo a fare un ringraziamento speciale a tutte le famiglie che hanno aderito a questo studio, soprattutto grazie per averci creduto assieme me (specialmente nel periodo di *lockdown*) e per aver dimostrato un grande interesse nel mio progetto di ricerca. Mi avete insegnato molto.

Un grazie di cuore al mio punto di riferimento, la mia forza più grande, la mia famiglia, che è sempre stata presente anche quando io non lo ero. Grazie per avermi incoraggiata, sostenuta ogni singolo giorno di questo percorso.
Grazie a Matteo, per la sua immensa pazienza, per aver accettato tutte le mie scelte, anche se a volte implicavano immensi sacrifici, e per aver ascoltato i miei quotidiani discorsi (o meglio… monologhi) sul 'tutto' e sul 'nulla'.
Vi sono infinitamente grata.