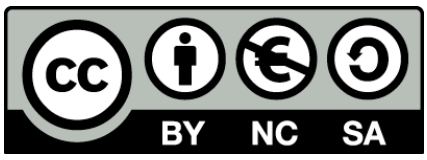
UNIVERSITAT DE BARCELONA

Language Factors Modulate Audiovisual Speech Perception

A Developmental Perspective

Joan Birulés Muntané



Language Factors Modulate Audiovisual Speech Perception

A Developmental Perspective

Acknowledgements

Estic molt agraït a totes les persones que m'han acompanyat durant aquests quatre anys i que han fet possible aquesta tesi. En primer lloc vull donar les gràcies als meus supervisors, en Ferran i la Laura. Gràcies per confiar en mi i oferir-me l'oportunitat d'aprendre i descobrir amb (i de) vosaltres el món del desenvolupament. Gràcies Ferran per les mil hores de feina, discussions i revisions conjuntes – al despatx, camp de futbol, bar o per telèfon –, i també per estar sempre present, involucrar-me en tot i per encoratjar-me a trobar la meva manera d'encaixar en l'estrany mon de la investigació. Gràcies Laura per transmetre'm la teva passió pel coneixement, el teu detallisme i la teva intensitat – transformada en mil hores més de discussions que, tot i deixar-nos sense dinar, acostumen a ser molt enriquidores.

Moltes gràcies també a la Jessica Sánchez, l'assistent de recerca del Babylab de l'APAL, per totes les hores invertides i per aguantar-nos a tots, i a tots els nadons, nens, famílies i escoles que han col·laborat amb nosaltres, voluntàriament, i sense els quals aquests estudis no podrien realitzar-se.

I am also very grateful to our main collaborator, David Lewkowicz, who from the beginning has been like a third supervisor, and who through many discussions, re-writings and corrections has helped me improve in my writing and my understanding of both our studies and the field of developmental science.

Likewise, I want to thank the people from Grenoble's baby lab; Mathilde Fort in particular but also Julien Diard, David Meary, Olivier Pascalis, Hélène Lœvenbruck and Anne Vilain who welcomed me with open arms and from whom I learned a great deal during my PhD stay in their lab.

Vull donar gràcies al Salva, que em va introduir al món de la ciència cognitiva – quan no sabia gairebé què volia dir – i em va oferir l'oportunitat de conèixer el seu grup i de començar a fer la meva recerca, que finalment em va dur a començar aquesta tesi.

I also want to thank the masters' students Nari Seo, Ricarda Brieke, and Crystal Tarragó who have helped me develop, improve and rethink each of the smaller projects that are here combined to form this larger piece of work.

A més a més vull donar mil gràcies als companys, amics i familiars que m'han acompanyat – i aguantat – aquests anys. A l'Anna per transmetre'm la seva passió per la ciència i pel feminisme, a la Carlota per aportar-me moltíssima energia renovadora i molt ben acompanyada de Tonis, Marcs del Valle, belchitenses o neu mastegada. Al Marc Colomer, amb qui ha estat un plaer compartir dies de ciència i música – i estranys open mics en conferències o intensius de tesi acompanyats de banys gelats – i també als seus companys de grup Jesús, Camille, Chiara, Gonzalo i Konstantina que m'han acollit en cursos, inundacions, congressos i banys turcs i que ho han tornat tot plegat molt més interessant i divertit.

Per acabar, agraeixo als meus pares, germana, al Sisu, a la Paula – ara experta en cognició i gràfics de puntets –, al meu cosí – i dissenyador d'aquest llibre –, al Genís, Mariona i Toni, a la Lídia, i també a la nova Marieta, mil gràcies pel vostre suport i estima d'aquests anys.

Table of Contents

List of Figures

List of Tables

Abstract

In most natural situations, adults look at the eyes of faces in seek of social information (Yarbus, 1967). However, when the auditory information becomes unclear (e.g. speech-in-noise) they switch their attention towards the mouth of a talking face and rely on the audiovisual redundant cues to help them process the speech signal (Barenholtz, Mavica, & Lewkowicz, 2016; Buchan, Paré, & Munhall, 2007; Lansing & McConkie, 2003; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). Likewise, young infants are sensitive to the correspondence between acoustic and visual speech (Bahrick & Lickliter, 2012), and they also rely on the talker's mouth during the second half of the first year of life, putatively to help them acquire language by the time they start babbling (Lewkowicz & Hansen-Tift, 2012), and also to aid language differentiation in the case of bilingual infants (Pons, Bosch & Lewkowicz, 2015).

The current set of studies provides a detailed examination of the audiovisual (AV) speech cues contribution to speech processing at different language development stages, through the analysis of selective attention patterns when processing speech from talking faces. To do so, I compared different linguistic experience factors (i.e. types of bilingualism – distance between bilinguals' two languages –, language familiarity and language proficiency) that modulate audiovisual speech perception in first language acquisition during infancy (Studies 1 and 2), early childhood (Studies 3 and 4), and in second language (L2) learning during adulthood (Studies 5, 6 and 7).

The findings of the present work demonstrate that (1) perceiving speech audiovisually hampers close bilingual infants' ability to discriminate their languages, that (2) 15-month-old and 5 year-old close language bilinguals rely more on the mouth cues of a talking face than do their distant bilingual peers, that (3) children's attention to the mouth follows a clear temporal pattern: it is maximal in the beginning of the presentation and it diminishes gradually as speech continues, and that (4) adults also rely more on the mouth

speech cues when they perceive fluent non-native vs. native speech, regardless of their L2 expertise.

All in all, these studies shed new light into the field of audiovisual speech perception and language processing by showing that selective attention to a talker's eyes and mouth is a dynamic, information-seeking process, which is largely modulated by perceivers' early linguistic experience and the tasks' demands. These results suggest that selectively attending the redundant speech cues of a talker's mouth at the adequate moment enhances speech perception and is crucial for normal language development and speech processing, not only in infancy – during first language acquisition – but also in more advanced language stages in childhood, as well as in L2 learning during adulthood. Ultimately, they confirm that mouth reliance is greater in close bilingual environments, where the presence of two related languages increases the necessity for disambiguation and keeping separate language systems.

Resum

Generalment, els adults mirem als ulls quan ens parlen, a la recerca d'informació social (Yarbus, 1967). Tot i això, quan el senyal auditiu es torna confús (per exemple, quan hi ha soroll) movem l'atenció visual a la zona de la boca i així ens beneficiem de la informació audiovisual que ens ajuda a processar millor el senyal de la parla (Barenholtz, Mavica, i Lewkowicz, 2016; Buchan, Paré, i Munhall, 2007; Lansing i McConkie, 2003; Vatikiotis-Bateson, Eigsti, Yano, i Munhall, 1998). Paral·lelament, els infants – que són sensibles a la correspondència entre el senyal acústic i visual de la parla (Bahrick i Lickliter, 2012) – també atenen a la boca d'un parlant durant la segona meitat del primer any de vida. Suposadament, aquest comportament els ajuda en el procés d'adquisició del llenguatge, just en el moment en què comencen a produir sons de balboteig (Lewkowicz i Hansen-Tift., 2012), i també, en el cas dels infants d'entorns bilingües, per ajudar-los a discriminar entre les seves dues llengües (Pons, Bosch i Lewkowicz, 2015). El següent conjunt d'estudis proporciona un examen detallat sobre la contribució dels senyals audiovisuals al processament de la parla en diferents etapes del desenvolupament del llenguatge, a través de les anàlisis dels patrons d'atenció selectiva a una cara parlant. Així, compararé diferents factors lingüístics (*i.e.* tipologies de bilingüisme – la distància entre les llengües d'un bilingüe –, la familiaritat i la competència amb l'idioma) que modulen la percepció audiovisual de la parla en l'adquisició del llenguatge durant la primera infància (Estudis 1 i 2), en nens d'edat escolar (Estudis 3 i 4) i també en l'aprenentatge d'una segona llengua (L2) durant l'edat adulta (Estudis 5, 6 i 7). Els resultats d'aquests estudis demostren que (1) la percepció audiovisual de la parla dificulta la capacitat dels infants bilingües de discriminar les seves llengües properes, que (2) els bilingües de llengües properes de 15 mesos i de 5 anys d'edat posen més atenció a les pistes audiovisuals de la boca d'un parlant que els seus companys bilingües de llengües distants, que (3) l'atenció dels nens a la boca del parlant segueix un patró temporal regular: és màxima al començament de la presentació i disminueix gradualment a mesura que continua la parla,

i que (4) els adults també es recolzen més en els senyals audiovisuals de la boca d'un parlant quan perceben un discurs en una llengua no nativa (L2), independentment de la seva competència en aquesta. En resum, aquests estudis aporten nova evidència al camp de la percepció de la parla audiovisual i el processament del llenguatge, demostrant que l'atenció selectiva als ulls i a la boca d'un parlant és un procés dinàmic i de cerca d'informació, i que aquest és, en gran mesura, modulat per l'experiència lingüística primerenca i les exigències que comporten les diferents situacions comunicatives. Aquests resultats suggereixen que atendre de forma selectiva a les pistes audiovisuals i redundants de la boca d'un parlant en els moments adequats millora la percepció de la parla i és crucial per al desenvolupament normal del llenguatge, no només durant la primera infància sinó també en les etapes més avançades del llenguatge, així com en l'aprenentatge de segones llengües durant l'edat adulta. En última instància, aquests resultats confirmen que l'estratègia de recolzar-se en les pistes audiovisuals de la boca d'un parlant s'utilitza en major mesura en entorns bilingües propers, on la presència de dues llengües relacionades augmenta la necessitat de desambiguació i de mantenir els dos sistemes lingüístics separats.

Chapter 1

General Introduction

Overview

A newborn is suddenly faced with a world of novel stimuli, most of which involve more than one developing sensory system. During the first years of postnatal life, infants must learn to make sense of the complex multisensory experiences that are found in this new and rich environment. Indeed, infants develop an attentional system that is capable of selecting the stimuli that are relevant to them and deploying their attention resources selectively on those, whilst filtering the rest (Amso & Scerif, 2015). Through studying infants' selective attention scientists have found a window to reveal the underlying cognitive processes during infants' development.

One of the most remarkable cognitive challenges that infants face is language acquisition. In the last decades, the study of infants' selective attention has shed new light onto the field of language learning through studying the perception of audiovisual speech (for reviews see: Kuhl, 2004; Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012).

In a seminal study by Lewkowicz and Hansen-Tift (2012), it was shown that infants' selective attention to a talker's face follows a developmental pattern that starts at the eyes of a talker and shifts to the mouth at the second half of the first year of life. From the perspective of infants' language acquisition process, it is worth noting that this shift to the mouth coincides with the emergence of endogenous attention (Colombo, 2001), the onset of canonical babbling (Oller, 2000) and is associated to language growth in the second year of life (Tenenbaum et al., 2015; Young, Merin, Rogers, & Ozonoff, 2009). Thus, the shift to the mouth at this stage was interpreted as infants' intentional reliance on the source of the audiovisual speech cues, that is, the talker's mouth.

Remarkably, a recent study with bilingual infants by Pons, Bosch, and Lewkowicz (2015) has found that language background influences infants' pattern of selective attention to a talking face. Pons et al. (2015) study has shown that bilingual infants perform the shift to the mouth earlier in development and show a greater preference for the mouth at the end of the first year of life, as compared to their monolingual peers. This has been interpreted as bilingual infants' additional resourcing to the audiovisual speech cues, in face of their extra challenge of learning two languages whilst keeping them separate. In turn, the fact that infants' different language background modifies selective attention patterns to a talking face supports the linguistic explanation of the developmental changes in selective attention, first described in Lewkowicz and Hansen-Tift (2012).

In this dissertation I explore the influence and use of the audiovisual cues in speech perception. To do so, I first investigate the effect of presenting a talking face in a language discrimination task at an early stage of language acquisition. Second, I explore the

developmental trajectory of the selective attention patterns to a talking face, from infants to children and adults. Last, I evaluate the possible modulatory effect of various linguistic factors such as bilingualism (and types of bilingualism), language familiarity and language proficiency onto selective attention to a talking face.

In the introduction I first review the fundamental aspects of multisensory perception, and more specifically of adults' audiovisual speech perception. After going through the general perceptual advantages of perceiving talking faces, I explain the selective attention strategies that are employed in the perception of talking faces, together with their cognitive significance and implications for audiovisual speech perception. Third, I describe the development of acoustic, visual and audiovisual speech perception and its associated selective attention mechanisms, from very early stages of language acquisition in infancy to the age of 6 years, when basic phonology of the language has already been acquired (Dodd, Holm, Hua, & Crosbie, 2003). In the last two sections, I review two linguistic factors that have been seen to modulate attention mechanisms in general and the selective attention patterns to talking faces more specifically: bilingualism and the perception of second languages.

Audiovisual speech perception

The world is inherently multisensory: we perceive information from our surroundings with our senses, and we code it through the different sensory modalities. This presents a cognitive challenge to our brain; it must decipher what information originates from the same object or event and is to be integrated and what information does not, and hence should be segregated. This process, which has been named sensory binding[1], must be rapidly solved in order to form coherent perceptual representations and produce adequate and adaptive behaviors. Indeed, there are specialized neural processing mechanisms that enable the combination and integration of multisensory information (Meredith & Stein, 1986; Stein & Stanford, 2008), which are located both in higher order cortical regions, such as the superior temporal sulcus or the intraparietal complex, and also in the lower primary somatosensory cortex (Ghazanfar & Schroeder, 2006 for a review). Importantly, being able to integrate multisensory information does not only solve the problem of sensory binding but it also provides us with a more salient signal, that results in behavioral and perceptual benefits over the perception of each modality separately. Generally, these include faster and more accurate detection, discrimination and localization of stimuli (for a review, see Mark Murray & Wallace, 2011). For example, previous research in the speech domain showed that, in the presence of noise, seeing the speaker's mouth increases the intelligibility of the auditory signal (Sumby & Pollack, 1954). Moreover, when the auditory and visual stimuli are incongruent, the brain fuses them to form a new audiovisual percept, which corresponds to the combination of the two (McGurk & MacDonald, 1976)[2]. Last, exploring the neural correlates of these findings, a study showed that in fact, learning to lipread – i.e. watch mouth movements to understand speech –, results in extensive cross-modal plasticity within cortical regions (Calvert et al., 1997; for a review see Bavelier & Neville, 2002). These three examples illustrate the multisensory nature of perception and its cross-modal characteristics, both at the behavior and at the brain level. In turn, they all focus on the subject of multisensory integration in the audiovisual domain, which is closely linked to speech perception and spoken communication and has received a great deal of attention in the last decades (for

[1] Sensory Binding is defined as the processes whereby the different information perceived is coded as originating from the same object or event (i.e. bound). Thus, the "audiovisual binding" refers to the fusion of the auditory and visual speech token, as occurs in the McGurk effect (see below).

[2] When presenting the syllable /ga/ visually and the syllable /ba/ auditorily adults perceive the syllable /da/, which corresponds to the combination of the auditory and visual information presented, in an intermediate place of articulation (McGurk & MacDonald, 1976).

reviews see: Bailly, Perrier, & Vatikiotis-Bateson, 2012; Campbell, 1998; Munhall & Johnson, 2012; Soto-Faraco et al., 2012; Summerfield, 1992).

Indeed, most natural interactions with other people are audiovisual, that is, we can both hear and see our interlocutor. The talker's face gives us access to a great deal of information. Firstly, the eyes help us identify the talker and through its movements we can infer the talker's state of mind, attitudes, and potential intentions (for a review see: Birmingham & Kingstone, 2009). Moreover, the talker's mouth provides spatiotemporally and acoustically congruent auditory and visual speech cues (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Yehia, Rubin, & Vatikiotis-Bateson, 1998). These dynamic audiovisual speech cues are especially interesting in light of the multisensory studies above presented; our brain is able to process the two streams of information together and perceive them as a coherent multisensory speech signal. Importantly, research has shown that this fused multisensory speech signal is perceptually more salient than the auditory-only signal and can improve the comprehension of speech. This audiovisual performance increase was first illustrated in a classic study by Cotton (1935). In that study, participants listened to the live auditory signal of a speaker talking through a microphone in a non-illuminated booth, and then the experimenter distorted the auditory signal until participants could not understand the speech. Thereafter, when they turned the booth lights on – thus the speaker's face could be seen –, participants showed a nearly full recovery of the speech's comprehension. A great body of studies have replicated and extended Cotton's seminal work in the last 80 years, by better describing the audiovisual gain in various situations. For example, studies have shown that 1) the audiovisual gain increases as a function of the amount of noise: the lower the speech-to-noise ratio (SNR) the higher the audiovisual gain (Sumby & Pollack, 1954), that 2) it is dependent on the visual information of the lips, as the performance improvement disappears when only visual information of syllabic timing is presented (Summerfield, 1979), and that 3) the amount of audiovisual speech comprehension improvement varies as a function of vowels' audiovisual intelligibility (Benoît, Mohammadi, & Kandel, 1994) and of sentences' visual readability (Macleod & Summerfield, 1987). Last, studies have also replicated the audiovisual gain effect by using filtered speech instead of adding noise to the speech (Risberg & Lubker, 1978; Sanders & Goodrich, 1971), or by adding a second and irrelevant passage on top of the relevant passage to process (Reisberg, 1978). Figure 1 shows Risberg and Lubker's (1978) findings, here presented as a clear example of the non-linear audiovisual improvement (here, the percentage of correctly perceived words as a function of the filtered-audio, visual or audiovisual condition).

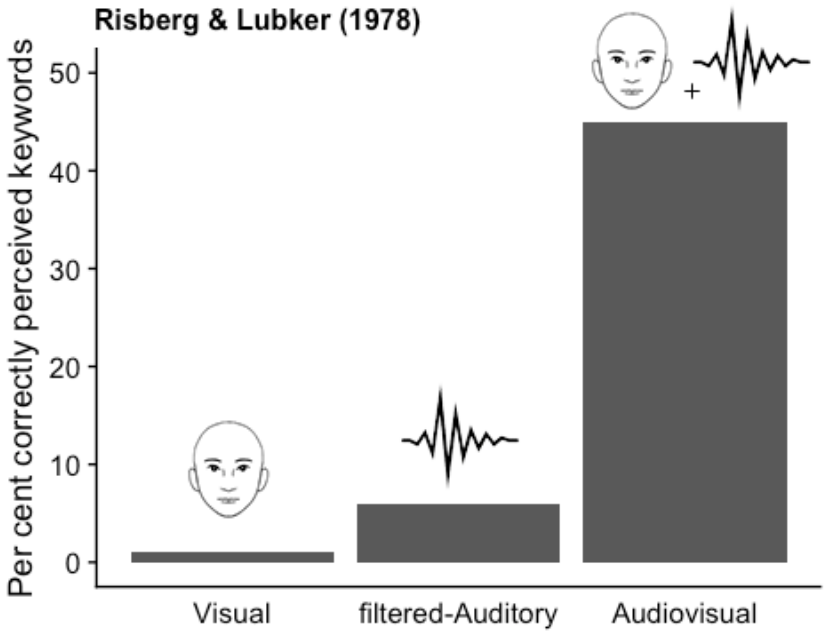


Figure 1. Using a 180Hz low-pass filter, the auditory performance decreases to 6%, visual-only performance was 1%, but the combination of the two reached 45% of correctly perceived keywords. Adapted from Risberg & Lubker (1978).

It is worth noting that all these studies decreased the auditory-only performance from ceiling by deteriorating the auditory information, and as a consequence they could find the audiovisual gain. However, other studies have demonstrated that the perceptual advantage offered by the audiovisual speech over auditory-only speech is not limited to offering a "back-up system" in face of environmental noise (Johnstone, 1996), but that audiovisual speech can also enhance processing in situations where the auditory signal is clear. Indeed, studies have found a performance increase in the audiovisual condition as compared to auditory-only when presenting clear but highly complex speech (syntactically and semantically, a fragment of Kant's *Critique of Pure Reason*) or with speech uttered in an unfamiliar accent or language (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987).

Overall, this evidence demonstrates a clear beneficial effect of perceiving audiovisual speech over perceiving auditory-only speech, which is larger in the perception of deteriorated auditory signal but also present in the processing of intact auditory signal. In the latter case, the difficulty for comprehension may come from complex, accented or non-native language speech.

Above and beyond the performance improvement of audiovisual speech, other studies have focused on the way adults visually explore talking faces. In this way, researchers aim to uncover the attentional strategies underlying this audiovisual improvement, and in turn, explore the extent to which these selective attention patterns can in fact modulate the perception and processing of speech.

Selective attention to talking faces

As earlier noted, it is known that whenever adults interact with one another they tend to look at their social partners' eyes (Yarbus, 1967), where they can gain access to deictic social cues (for a review see: Birmingham & Kingstone, 2009). When, however, the audiovisual speech becomes ambiguous or difficult to comprehend, e.g. speech-in-noise, the redundant audiovisual speech cues become especially relevant and thus adults deploy more of their attention to the talker's mouth (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998). Specifically, Vatikiotis-Bateson and colleagues (1998) showed that although participants deployed most of their attention to the talker's eyes, when the levels of noise increased, they shifted more to mouth (see Figure 2).

Interestingly, the authors also noted that even at the highest noise levels, perceivers only explored the mouth about half the time, and that eye motions did not correlate with comprehension scores. It was suggested that extrafoveal fixations may be sufficient for the extraction of the dynamic phonetic information from the talker's mouth, and thus perceivers could spend half the time attending to other more distributed speech cues whilst maintaining equal comprehension. Still, the fact that the increasing levels of noise accentuated the mouth-fixations suggested that fixating directly on the mouth facilitated the speech processing task.

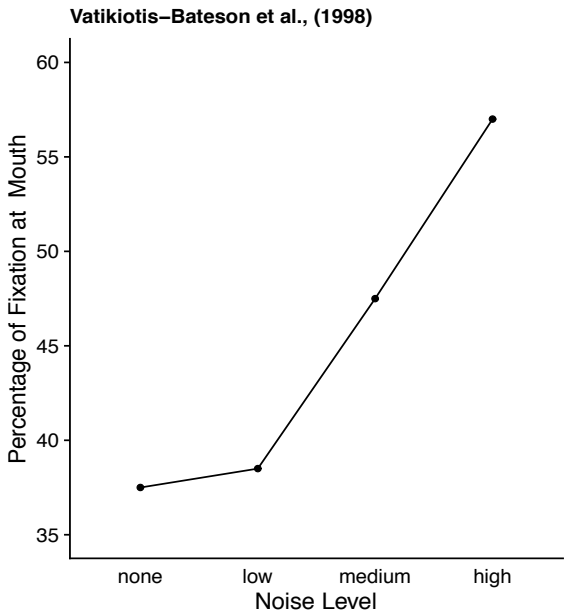


Figure 2. When the speech-to-noise ratio (SNR) decreased, the PTLTmouth increased by ≈ 20%, adapted from Vatikiotis-Bateson et al. (1998).

Instead of adding noise to the speech, Lansing and McConkie (2003) manipulated the intensity of the auditory signal (i.e. volume) and evaluated perceivers' comprehension to four different speakers whilst recording their eye gaze. In this way, the authors showed again that participants looked preferably at the speaker's eyes during silence periods (i.e. prior to and after the speech) and that they shifted to the mouth of the talker during low intensity speech periods, as they did in the presence of noise (Vatikiotis-Bateson et al., 1998). Moreover, Lansing and McConkie (2003) also found a parametric increase of mouth-looking when the auditory intensity lowered, albeit at the maximal levels of speech comprehension difficulty (i.e. very weak or missing auditory signal) the mouth fixations increased up to 60-85%, different from the 50-65% in the former study. Last, increased attention to the mouth did not correlate with speech comprehension either. In this case, comprehension performance only correlated with individual lipreading ability and sentences' difficulty. Together, these two studies confirmed that perceivers deploy greater attention to the mouth under auditory-compromised situations, although they also suggest that this modulation of eye motions is rather macroscopic, as comprehension does not seem to be impeded when attention is distributed to other parts of the face.

In a similar vein, a study by Thompson and Malloy (2004) explored intact auditory signal but in older adults (between 65 and 75 years of age), who exhibit slower speed of language processing and hearing loss. The results showed that indeed, older adults deployed more attention to the mouth than younger adults did (20 to 40 years of age). They interpreted these differences as older adults' greater reliance on the mouth audiovisual speech cues, as a compensation strategy in face of comprehension difficulties (Thompson, 1995; Thompson & Malloy, 2004).

Later studies have extended these findings by showing that in fact, participants actively adjust their selective attention depending on the task they are performing. For example, when participants have to judge the emotion of faces they fixate more on the eyes than when they have to identify words, and when noise is added to the speech signal participants' fixations are longer and more centralized on the talker's face, that is, on the nose and mouth of the speaker (Buchan, Pare, & Munhall, 2008; Buchan et al., 2007). Additionally, when participants have to perform a specific speech processing task – in the absence of noise –, they also resource more to the mouth's speech cues (Barenholtz et al., 2016; Lusk & Mitchel, 2016). In Lusk and Mitchel's (2016) study, participants were asked to segment words of an artificial language presented audiovisually. Interestingly, participants started fixating more on the mouth of the speaker, but as familiarization progressed – and they learned the new words – attention to the mouth decreased (Lusk & Mitchel, 2016). On the other hand, in Barenholtz, Mavika and Lewkowicz's (2016) study, participants had to

decide whether a short auditory sentence coincided with one of two audiovisual sentences they heard, in both their native and a non-native language. The results showed that overall attention was deployed preferably to the mouth of the talker, and that the non-native language elicited even greater mouth looking than did the native one (Barenholtz et al., 2016). Thus, jointly, these studies reveal that adults' allocation of attention to a talking face is a rather dynamic, information-seeking process, which is largely dependent on the ongoing processing task.

Last, a recent study has revealed that when the stimuli consist of more naturalistic and dynamic scenes of talking faces, the previously reported general preference for the eyes decreases, and participants allocate their gaze more dynamically to the eyes, nose and mouth of a speaker in response to the currently depicted event; that is, on the eyes when a face makes eye contact, on the mouth when it starts speaking and on the nose when it moves quickly (Võ, Smith, Mital, & Henderson, 2012).

In sum, the reviewed studies suggest that indeed, selective attention to talking faces mediates the audiovisual gain in speech perception. Specifically, these studies demonstrate that attention allocation on a talker's face is highly dependent on the demands of the particular cognitive process (or task) that participants are undergoing, as they actively adjust their selective attention to improve processing performance.

Then, if greater attention to the redundant audiovisual speech cues enhances speech processing in adults, is it possible that infants might also attend more to the audiovisual cues during speech and language acquisition? In the following two sections, I review the development of audiovisual speech perception together with the development of selective attention to a talking face, from infancy to late childhood.

The development of audiovisual speech perception

To explore infants' audiovisual speech perception, we must first understand the manner in which infants perceive the acoustic and visual information of speech and its changes across development.

Previous research shows that after the in-womb low-pass filtered speech experience, newborns have already built some basic acoustic representations of speech. This is supported by studies using the high amplitude sucking (HAS) paradigm, where the novelty of a stimulus is indexed by the infant's higher sucking rate after having habituated to a previous stimulus. In this way, studies have demonstrated an increase in newborns' sucking rate when listening to speech over listening to complex non-speech stimuli (Vouloumanos & Werker, 2007), and also newborns' ability to acoustically discriminate languages from different rhythmic classes (Byers-Heinlein, Burns, & Werker, 2010; Mehler et al., 1988), even when both languages are unfamiliar to them (Nazzi, Bertoncini, & Mehler, 1998). Remarkably, newborn infants are also able to discriminate most phonetic contrasts of both native and unfamiliar languages, which suggests that their phonetic discrimination skills are initially quite broad (Werker & Tees, 1999). Thereafter, during the first months after birth, infants become gradually specialized in the contrasts that are present in their native language speech (Kuhl et al., 2006; Kuhl, Tsao, & Liu, 2003). In turn, this specialization in their native language also entails a concomitant decrease in the perceptual sensitivity to those phonetic contrasts that are not relevant in their native language, which has been called perceptual narrowing or attunement (for a review see Kuhl, 2004). Noteworthy however, later studies have extended this idea by showing that in fact, there are a few exceptions where phonetic contrasts are not yet discriminated at birth and depend on this specialization. For example, Narayan and colleagues found that some nasal place speech sounds require months of experience with infant's native language to be discriminated (Narayan, Werker, & Beddor, 2010). Indeed, a recent review on perceptual narrowing argues for a more flexible system of contrasts discriminability as a function of experience and concluded that, although perceptual narrowing may be a useful label for describing a general phenomenon observed across various domains, each specific stimulus might present different trajectories and timings of discriminability (Maurer & Werker, 2014).

Different from the auditory system which is fully functional at birth, the visual system is not completely functional until 3 to 4 postnatal months (Boothe, Dobson, & Teller, 1985). Although newborns are able to direct their gaze toward relevant visual stimuli such as motion

and strong contrasts (e.g. they can segregate figure from ground), and already show a preference for face-like patterns (Johnson, Dziurawiec, Ellis, & Morton, 1991), their perception of coherent and stable objects requires learning and maturation over the first months after birth (Johnson, 2013).

Nonetheless, in day to day situations, infants do not perceive the acoustic or visual information of speech in isolation, but they usually perceive it together through talking faces, as a redundant audiovisual speech signal. Moreover, research indicates that from a very early age, infants are sensitive to the percept that results from the simultaneous perception of both the auditory and visual information. A study by Sai (2005) provides an illustrative example of an early audiovisual cross-modal effect, by showing that newborns recognized their mother's face only when they had had previous exposure to the mother's voice-face combination (Sai, 2005). But, to what extent are young infants sensitive to the match between the heard and seen speech information of talking faces?

Ground-breaking work by Kuhl and Meltzoff (1982) first showed infants' sensitivity to the visual correspondence of acoustic speech. Specifically, they demonstrated that four-month-old infants attended more to the face that articulated the sound they heard – i.e. the "ee" and "ouu" vocalic sounds –, which suggests that infants already possess some knowledge of the relationship between audition and articulation. This correspondence was later shown in earlier ages; in two-month-old infants (Patterson & Werker, 2003), and even in neonates, using both human and primate faces (Aldridge, Braga, Walton, & Bower, 1999; Lewkowicz, Leo, & Simion, 2010 respectively). Furthermore, neural evidence for integration of heard and seen speech has been found from as early as 2.5 months (Bristow et al., 2008; Kushnerenko, Teinonen, Volein, & Csibra, 2008), which is consistent with the fact that at 5 months of age, infants can integrate incongruent acoustic and visual information, in a manner consistent with the previously reported McGurk effect (McGurk & MacDonald, 1976; Rosenblum, Schmuckler, & Johnson, 1997). Last, a study by MacKain and colleagues (1983) showed that at that age, the audiovisual matching of consonant sounds is more robust when the matching face is on the right side, suggesting an involvement of the language areas (left hemisphere) for performing the audiovisual match.

Later in development, as previously reported in the auditory and visual domains, the sensitivity to audiovisual congruency also undergoes perceptual specialization and narrowing. In the second half of the first year of life, infants become sensitive to only those faces and speech sound contrasts they have had continuous exposure with and are now specialized in (Lewkowicz et al., 2010; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; Vouloumanos, Hauser, Werker, & Martin, 2010; for reviews see Lewkowicz, 2014;

Maurer & Werker, 2014; Murray, Lewkowicz, Amedi, & Wallace, 2016; Werker & Hensch, 2015).

Generally, the studies mentioned above explore sensitivity to audiovisual congruency by contrasting matching vs. non-matching audiovisual stimuli; that is, audiovisual matching is considered successful when infants attend more to the matching condition. However, it is likely that these studies miss to reveal the more subtle developmental differences in the way that infants perform these audiovisual matching tasks.

To investigate this question in a finer-grain fashion, other studies have explored infants' ability to detect the temporal synchrony of talking faces. As expected, these studies have shown that although infants from 10 to 16 weeks of age are able to detect the temporal congruency of lip movements and speech sounds (Dodd, 1979), their temporal integration window is rather wide (i.e. ± ~650ms, compared to ~100ms in adults; Lewkowicz, 2000, 2010), and that it does not depend on specific linguistic experience or familiarity with the language (Pons & Lewkowicz, 2014). After these results, the authors concluded that infants' perception of audiovisual speech synchrony is driven by a low-level, domain-general mechanism, and that it is consistent with the broad perceptual tuning reported previously.

The studies above described demonstrate that infants are capable of matching auditory and visual speech from a talking face, and that they seem to do so according to the specific correspondences between the two senses. Then, it is likely that infants benefit from the redundant information of the speech to facilitate its processing, as multimodal information has been found to enhance infants' general perception and learning in other non-speech stimuli (for a review see: Bahrick & Lickliter, 2012). To our knowledge, only two studies have specifically explored whether perceiving and integrating the visual information of speech together with the auditory signal is in fact advantageous and improves infants' perception of speech sounds and thus, language learning.

Teinonen and colleagues (2008) were the first to test such question. In their study, they presented 6-month-old infants with a continuum of /ba/ and /da/ speech sounds that followed a unimodal distribution centered at the average adult category boundary[3]. The results showed that infants could only succeed in a post-test discrimination of a /ba3/ and /da6/speech sounds when they had been previously familiarized with a face providing the articulatory information for the two phonemes (Teinonen, Aslin, Alku, & Csibra, 2008). These results provided the first direct evidence of a visual enhancement in a speech

[3] In a continuum of speech sounds from /ba/ to /da/, the authors divided the distribution in 8: ba1, ba2, ba3, ba4 and da5, da6, da7, da8, where ba4 and da5 composed the center of the distribution (i.e. adults' category boundary).

perception task. More recently, Ter Schure, Junge, & Boersma (2016) have extended these results by showing that visual information is also helpful to 8-month-old infants when learning to discriminate a native language contrast, and that additionally, it facilitates the discrimination of unfamiliar language contrasts. Together, as had been noted earlier in adults, these two studies demonstrate that infants also exhibit an enhancement when perceiving speech sounds in the presence of the concurrent articulatory visual information.

It becomes clear from the findings above reviewed that the development of audiovisual speech processing undergoes important changes during the first year of life, and it would seem reasonable that perceptual narrowing or attunement also gave closure to these perceptual learning processes. Nevertheless, research shows that children's audiovisual perceptual system is still underdeveloped and far from adult-like levels. For example, a body of studies have shown that children are less influenced by the visual information of speech than adults are. The classic McGurk effect provided the first illustration of this phenomenon, by showing that 3- to 8-year-olds show less fusion than adults do (McGurk & MacDonald, 1976). Later studies have replicated these findings and additionally reported that English speaking children's audiovisual fusion – i.e. visual influence on the audiovisual percept – is still increasing between 7 and 11 years of age (Sekiyama & Burnham, 2008), and that children are poorer lip-readers than adults, which correlates with the amount of visual influence they experience (i.e. the better lip-reading abilities the more visual information influences audiovisual perception) (Desjardins, Rogers, & Werker, 1997), and also with children's speech production abilities (Massaro, Thompson, Barron, & Laren, 1986).

Above and beyond the McGurk effect, other studies have also suggested children's audiovisual perception is still under development. These have revealed that children's temporal judgment of audiovisual simultaneity is immature (Kaganovich, 2016; Lewkowicz & Flom, 2014), that they benefit less from visual articulations and display considerably less audiovisual enhancement (Ross et al., 2011), that they do not show the congruence-independent attenuation of amplitude typically found in adults' ERPs when perceiving audiovisual speech (Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014) and that their processing of audiovisual speech is less efficient, as shown by a differential activation of the brain regions involved (Dick, Solodkin, & Small, 2010). Finally, Baart and colleagues recently showed that only at 6.5 years of age did children benefit from having previous phonetic knowledge of the speech sounds perceived, thus indicating that audiovisual matching based on phonetic cues – and not only on lower-level temporal cues as in infants – develops quite late in development (Baart, Bortfeld, & Vroomen, 2015).

In sum, the above reviewed research show that 1) the audiovisual perceptual system is already functional early in life and that 2) it develops slowly and is not fully established

until adulthood. Albeit in an immature form, the fact that the audiovisual integration mechanisms are already present in both infants and children allows them to perceive multisensory information and, more importantly, to take advantage of the audiovisual information when processing speech. As a consequence, is it possible that infants deploy their attention preferably to these redundant multisensory cues of a talker's face in situations of social interactions?

Infants' selective attention patterns are known to be established early on, in order to attend and process what they perceive as most relevant in their environment (Colombo, 2001). During their development, they continuously update these patterns of attention with experience so that they become increasingly efficient in solving the cognitive tasks they are faced with (see for a review Bahrick & Lickliter, 2012). Bahrick and Lickliter suggest that infants' early attention allocation is highly influenced by the salience of multimodal stimuli, and that selectively attending to the multimodal redundant stimuli is key to infants' perceptual processing, learning and memory during the first months of life. Moreover, they also note that attending to some properties of stimulations can come at the expense of others, particularly when attentional resources are most limited and less efficient (Bahrick & Lickliter, 2012, 2015).

These studies highlight the crucial role of studying infants' selective attention and suggest that doing so can give us valuable insight into infants' ongoing cognitive processes. In the following section I will review the findings to date on infants' attention to talking faces and the developmental trajectory of these selective attention patterns.

Infants' selective attention to talking faces

Infants show an early capacity to orient to faces. This bias has been seen to be present in newborns as longer looking times to face-like patterns (Goren, Sarty, & Wu, 1975; M. H. Johnson et al., 1991) and more strongly when the faces engage them in mutual gaze (Farroni, Csibra, Simion, & Johnson, 2002; Farroni et al., 2005). Also, recent evidence suggests that even third semester fetuses engage preferentially with upright face-like stimuli than with inverted face-like stimuli (Young et al., 2017). Later, at 4- to 12- months of age, infants show a preference for faces amongst multiple competing objects (Frank, Vul, & Johnson, 2009; Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016).

Thus, it is clear that infants are indeed attracted and tend to orient to faces. Noteworthy, this early sensitivity to faces and to mutual gaze has been regarded as a foundational component for developing their future social skills (Farroni et al., 2005; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; for a review see Klin, Shultz, & Jones, 2015). However, the specific cues within a talker's face that attract infants' attention remain to be known.

One of the first studies to address selective attention within a talker's face in early infancy was performed by Haith, Bergman and Moore in 1977. By using a mirror and infrared video cameras, the experimenters recorded very young infants' eyes fixations on real faces, and in this manner, they were able to reconstruct their scanning patterns. The results showed that 1- to 3-month-old infants fixate mostly on the talker's eyes, and that attention to the face and eyes increased when the face started to speak (Haith et al., 1977). Only twenty-five years later, Lewkowicz and Hansen-Tift (2012) performed an influential and comprehensive eye-tracking study which described in more detail the evolution of selective attention to a talker's face (i.e., to the talker's eyes and mouth) across the first year of life.

In their seminal work, Lewkowicz and Hansen-Tift (2012) presented videos of a talker's face producing fluent-speech monologues in English (native) or in Spanish (non-native), to 4-, 6-, 8-, 10-, and 12-month-old monolingual, English-learning infants and to a group of English-speaking adults. Results indicated that when the talker spoke in the native language, 4-month-olds attended more to the eyes, 6-month-olds looked equally to the eyes and mouth, 8- and 10-month-olds attended more to the mouth, 12-month-olds attended equally to the eyes and mouth, and adults attended more to the eyes. The results were identical in response to a talker speaking in the non-native language except that this time 12-month-olds continued to attend more to the talker's mouth.

The authors related the attentional shift from the talker's eyes to the talker's mouth between four and eight months of age with two coinciding important developmental events: the emergence of canonical babbling (Oller, 2000), and the emergence of endogenous

attention (Colombo, 2001). Specifically in the perception of faces, two studies demonstrate that physical salience can still explain infants' selective attention patters at 3 and 4 months of age, but that between 4 and 8 months they attend increasingly to faces, regardless of their salience (Frank et al., 2009; Kwon et al., 2016). Therefore, the attentional shift from the eyes to the mouth was interpreted as a reflection of infants' emerging interest in audiovisual speech. Later findings supported this conclusion by showing that greater attention to the mouth during the first year of life is associated with concurrent expressive language skills (Tsang, Atagi, & Johnson, 2018) and with greater rates of expressive language growth in the second year (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013; Tenenbaum et al., 2015; Young et al., 2009).

As well as noting infants' new interest in audiovisual speech, Lewkowicz and Hansen-Tift (2012) also concluded that the greater attention to a talker's mouth by eight months of age reflects the greater salience of redundantly specified audiovisual speech, in a developmental stage where infants' are starting to rely on audiovisual speech for their acquisition of native speech forms. This idea is also supported by studies showing that around 6 months of age, greater looking to the mouth is associated with infants' tendency to vocally imitate them (Imafuku, Kanakogi, Butler, & Myowa, 2019), and that at that age, if infants' tongue is blocked, infants fail to learn new speech sound contrasts properly (Bruderer, Danielson, Kandhadai, & Werker, 2015), which supports again a linguistic interpretation of infants' shift to the mouth. Last, infants' greater mouth looking in the unfamiliar speech condition was related to the fact that 12-month-old infants have already acquired their initial phonological expertise for native speech, and because of experience-based perceptual narrowing (Maurer & Werker, 2014), the responsiveness to non-native speech has become more difficult and therefore they still need to rely on the audiovisual speech cues of the talker's mouth.

Thereafter, during the second year of life, Hillairet de Boisferon and colleagues (2018) revealed that 14 and 18 month old infants continue to rely on the audiovisual speech cues of the talker's mouth, with a stronger mouth preference at 18 months of age, which the authors related to infants' vocabulary-explosion during the latter part of the second year of life (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2018).

It is not until most crucial aspects of language phonology are already acquired, at around 5 to 6 years of age (Bosch Galceran, 2004; Dodd et al., 2003) that the attention to the mouth starts to decrease and children's attention becomes equally distributed between

the eyes and mouth (Byers-Heinlein, Morin-Lessard, Poulin-Dubois, & Segalowitz, 2014[4]; Król, 2018; Nakano et al., 2010). Interestingly, in the study by Król (2018), she found that children with higher word recognition proficiency and higher average pupil response had an increased likelihood of fixating the mouth, indicating a stronger motivation to decode speech.

These results fit well with previous evidence claiming that children's audiovisual perceptual system is still under development (Desjardins et al., 1997; Knowland et al., 2014; Lewkowicz & Flom, 2014; Ross et al., 2011), less influenced by the visual information (McGurk & MacDonald, 1976; Ross et al., 2011) and neuronally less efficient (Dick et al., 2010), and hence it is likely that their increased attention to the mouth serves as a compensating mechanism.

Taken together, these studies associate the selective attention patterns to a talking face with language development and speech processing effort, from early infancy until adulthood.

In the following section I review two linguistic factors that may trigger an increased use of the audiovisual speech cues and may also modulate the exploratory patterns to a talking face; 1) the case of infants growing up to be bilingual and 2) the case of non-native speech perception.

[4] Although statistics for the eyes-mouth preference are not provided, Byers-Heinlein et al., (2014) show a significant decrease of attention to the mouth throughout development.

Modulatory language factors

Bilingualism

Although most of the developmental studies have been performed in monolingual populations, many children grow up in bilingual environments and acquire two first languages instead of one. Previous research has claimed that bilingual – or multilingual – infants' language acquisition processes are similar to those of monolingual infants and that they pass the developmental milestones at approximately the same age (Kuhl, 2004; Werker, 2012; Werker & Byers-Heinlein, 2008). However, bilingual-to-be infants also exhibit some relevant differences, which include not only aspects related to language acquisition but also related to more general cognitive processes (for reviews see: Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2009; Costa & Sebastián-Gallés, 2014).

In this section, I will first review bilingual infants' acoustic, visual and audiovisual language discrimination abilities and thereafter the influence of bilingualism onto selective attention to a talking face.

Language discrimination in bilinguals

One of the initial challenges that bilingual infants must face is the necessity to discriminate between the languages spoken around them (Genesee, Nicoladis, & Paradis, 1995; Mehler & Christophe, 1995; Werker & Byers-Heinlein, 2008). Evidence from monolingual infants shows that already at birth, they can discriminate rhythmically distant languages (i.e. Russian and French, Mehler et al., 1988; Ramus et al., 2000), even when both languages are unfamiliar to them (Nazzi et al., 1998). By two months of age however, they only perform successfully when one of the languages tested is native to them (Christophe & Morton, 1998). Importantly, Christophe and Morton (1998) also showed that English-learning 2-month-old infants could not discriminate English from Dutch, which exemplifies that the ability to separate utterances from rhythmically close languages – such as Dutch and English – has not yet been developed at 2 months of age.

Crucially, these studies were conducted with monolingual participants and hence they could discriminate the languages by telling apart a familiar vs. an unfamiliar language. In the scenario of infants growing up in bilingual environments, infants face the more difficult task of telling apart their two familiar languages. Remarkably, a study by Byers-Heinlein and collaborators (2010) demonstrated that bilingual newborn infants are also able to discriminate between their languages. In that study, newborns who had been prenatally

exposed to two rhythmically distant languages (i.e. Tagalog and English) showed equal preference for each of the two languages, whilst being able to differentiate between the two – as opposed to monolingual infants who showed a preference for their native language. Byers-Heinlein et al. (2010) suggested that the fact that bilingual neonates could recognize their two languages and pay selective attention to them ensured further learning from each of the two languages.

In the case of close bilingual infants – i.e. those infants learning a pair of rhythmically close languages –, the discrimination between their languages requires at least 4 months of linguistic experience (Bosch & Sebastián-Gallés, 1997, 2001; Molnar et al., 2014; Nazzi et al., 2000 for Catalan-Spanish, Basque-Spanish and English-Dutch, respectively), and neural maturation (Peña, Pittaluga, & Mehler, 2010), and is only successful when one of their native languages is present (Nazzi, Jusczyk, & Johnson, 2000). Worthy of mention, some bilingual infants face the even more difficult challenge of acquiring two very closely related languages such as Catalan and Spanish. On top of being rhythmically close languages – both syllable-timed (Ramus, Nespor, & Mehler, 1999) – Catalan and Spanish share a great number of features such as phonetic-phonological categories, phonotactic structures, morphological complexity, lexical stress patterns and a high number of cognate words (Bosch, 2018; Bosch & Ramon-Casas, 2014).

Bilinguals' language proximity is worthy of mention, since it has been seen to modulate various processes related to the acquisition of auditory-only speech. For example, infants' acquiring a pair of distantly related languages (English-Spanish) can establish some of their vowel categories earlier than infants acquiring a pair of closely related languages (Catalan-Spanish, Bosch & Sebastián-Gallés, 2003; Sundara & Scutellaro, 2011). These studies suggest that both similarity in global rhythm and the degree of overlap in phonetic categories reduce the perceptual distance for some language pairs and, as a result, preclude their early differentiation. On the other hand, studies have also found that language proximity facilitates vocabulary building and word learning due to the phonological similarity between the words in the two languages being learned (Bosch & Ramon-Casas, 2014; Havy, Bouchon, & Nazzi, 2016). Together, these data suggest that infants learning two close languages face different cognitive challenges than those learning one or two distant languages, and that, in regard to language discrimination, close bilingual infants face a much harder task. However, Bosch and Sebastián-Gallés (2001) showed that Catalan-Spanish bilingual infants could discriminate their languages at 4.5 months of age just as the monolingual group did. Moreover, the authors noted that although both groups succeeded in the task, the bilingual group presented longer orientation times to their native language (Bosch & Sebastián-Gallés, 1997). This was interpreted as bilinguals' slower detection of

their native language, due to the additional challenge of having two familiar and closely related languages. In the same line, in a language discrimination task, Nacar Garcia and colleagues (2018) have shown that Catalan-Spanish bilingual 4.5-month-olds present a later processing of the novel language, as compared to an earlier detection by monolingual infants. These data support the interpretation that monolingual and bilingual infants may employ different strategies for language discrimination, the earlier based on familiarity and the latter on an increased attention to the speech signal.

On the other hand, other studies have also explored monolingual and bilingual infants' ability to discriminate languages presenting only the visual information. Weikum and colleagues (2007) showed that 4- and 6-month-old monolingual infants were able to discriminate their native language (English) from a rhythmically distant one (French) in a visual-only presentation. Interestingly however, whilst at 8 months of age the monolingual infants no longer succeeded in the task, a group of English-French bilingual infants could visually discriminate their two languages at 4-, 6- and 8 months of age. A subsequent extension of this study with Catalan-Spanish bilingual infants (hence unfamiliar to the presented languages; English and French) confirmed that bilingual infants' ability of visually discriminating French from English was not caused by familiarity with the languages, but it was rather caused by the fact of being upbrought in a bilingual environment (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012). These two studies demonstrate infants' ability to use visual speech cues – in isolation – for detecting a distant language switch. In turn, they also show an advantage in bilingual infants' visual discrimination of languages at 8 months of age. It has been suggested that this bilingual advantage may result in heightened perceptual vigilance for linguistics cues, which might contribute to the emergence of broader cognitive advantages seen in bilingual infants and adults (Werker, 2012). However, whether this capacity can help young bilingual infants discriminate their languages in the presence of the redundant audiovisual signal – as infants most frequently perceive it – remains fairly unexplored.

The only study to our knowledge that has specifically explored audiovisual language discrimination is a study by Bahrick and Pickens (1988). In this study, the researchers found that 5-month-old monolingual and bilingual infants succeeded in the discrimination of Spanish and English audiovisual passages from a bilingual speaker and thus, they confirmed that 5-month-old infants could also discriminate the two languages when presented audiovisually. Crucially, this study used distant language pairs that can be discriminated auditorily from birth (Ramus et al., 2000), and hence it is not surprising that 5-month-old infants would also discriminate them audiovisually. To date, the early discrimination of close languages presented audiovisually remains an open question.

Noteworthy, it has been argued that in fact, having to separate their two native languages – and keep them separate – is likely to be the cause of other cognitive advantages associated to bilingualism such as enhanced attention to faces (Mercure et al., 2018), faster search, habituation and encoding of visual stimuli (Chabal, Schroeder, & Marian, 2015; Friesen, Latman, Calvo, & Bialystok, 2014; Singh et al., 2015), facility for simultaneous segmentation of two artificial languages (Antovich & Graf Estes, 2018) or even individual sounds (Sebastián-Gallés & Bosch, 2009), as well as executive functioning and enhanced cognitive control (Comishen, Bialystok, & Adler, 2019; Kovács & Mehler, 2009; Mehler & Kovács, 2009). Although I will not discuss the bilingual advantage further, it is worth noting that there also seems to be an advantage in perceiving some visual aspects of language, as exemplified by Weikum and colleagues (2007).

Selective attention to a talking face in bilinguals

As already mentioned, infants show an early capacity to orient to faces (Farroni et al., 2005; Goren et al., 1975; Johnson et al., 1991; Young et al., 2017), and they attend to talking faces more than to any other competing object (Frank et al., 2009; Kwon et al., 2016). Crucially, the cognitive effects that derive from being exposed to two languages not only include a general visual processing enhancement – as previously described –, but it also includes an enhanced attention to faces. Indeed, Mercure and colleagues (2018) have recently demonstrated that bilingual infants orient faster and attend longer at faces than monolingual or bimodal bilingual infants (i.e. those acquiring an oral language together with sign language), when these faces are presented within arrays of images that contain faces among other competing objects (Mercure et al., 2018).

Other studies have also explored whether bilingualism modulates attention within a talker's face. In the first study to explore bilingual infants' attention to a talking face and how these attention patterns might change across development, Pons, Bosch and Lewkowicz (2015) compared 4-, 8-, and 12-month-old monolingual and bilingual infants learning either Catalan, Spanish or both languages simultaneously, and they tracked infants' eye gaze while presenting them with audiovisual faces talking in the infants' dominant language and in English. Results from the monolingual group exhibited the same developmental pattern of shifting attention as did the monolingual infants in the Lewkowicz and Hansen-Tift (2012) study: 4-month-olds attended more to the eyes, 8-month-olds to the mouth and 12-month-olds attended equally to both areas, except in the non-native language condition where they attended preferably to the mouth. Interestingly, the results from the bilingual group indicated

that 1) they started shifting their attention to the talker's mouth earlier in development than their monolingual counterparts, and 2) that they continued to focus on the talker's mouth in response to audiovisual speech in their dominant language, at an age (12 months) when their monolingual counterparts no longer do so in response to native speech. Finally, the results showed that bilingual 12-month-old infants attended more to the talker's mouth when she spoke in a non-native language than did their monolingual counterparts, suggesting that bilingual infants rely on redundant audiovisual speech cues to help them with their greater speech-processing needs. Other studies have given support to this idea by showing that 8-month-old Catalan-Spanish bilinguals attend more to the mouth of a person expressing different affective expressions than do monolingual infants (Ayneto & Sebastián-Gallés, 2016) and that 15-month-old Catalan-Spanish bilingual infants do not learn to anticipate a movement that originates in a person's eyes because they attend more to the person's mouth (Fort, Ayneto-Gimeno, Escrichs, & Sebastián-Gallés, 2017).

Concerning older children, only two studies to date have investigated bilinguals' selective attention to talking faces in children. In one of these studies, Pons et al., (2018) employed the same free viewing method used by Lewkowicz and Hansen-Tift (2012) to examine selective attention to a talker's eyes and mouth in 7-year-old close bilingual children (i.e. Catalan and Spanish) with Specific Language Impairment (SLI) and in their typically developing (TD) peers. Findings indicated that, similar to infants, close language bilingual TD children deployed more attention to the talker's mouth. In a second study, Byers-Heinlein and colleagues (2014) also employed a free-viewing method with monolingual and distant bilingual infants (i.e. English and French) and children ranging in age from five months of age to six years of age. This study revealed greater attention to the mouth throughout this developmental period, regardless of children's language background (Byers-Heinlein et al., 2014). Overall, attention to the mouth was most accentuated at 20 months of age and then it slowly transitioned to equal attention to the eyes and mouth by 5 years of age, as previously reported in monolingual infants (Król, 2018; Nakano et al., 2010).

From the combination of these studies it emerges that the linguistic distance (or in this case, proximity) of bilingual infants' and children's two languages may modulate their attentional looking pattern to a talker's eyes and mouth. Yet, no study to date has directly tested this subject.

Together, evidence suggests that bilingual infants and children develop a greater reliance on the redundant audiovisual cues of a talker's mouth, which is reflected not only in speech perception (Pons et al., 2015, 2018), but also when perceiving non-linguistic dynamic faces (Ayneto & Sebastián-Gallés, 2016; Fort et al., 2017), which fits well with previously reported enhanced visual perception in bilingual infants (Chabal et al., 2015; Friesen et al.,

2014; Sebastián-Gallés et al., 2012; Singh et al., 2015; Weikum et al., 2007). Furthermore, it has been argued that bilinguals' general increased attention to faces (Mercure et al., 2018) and more specifically to the mouth of talking faces, reflects infants' attempt to disambiguate the two languages they are learning (Pons et al., 2015).

Non-native speech perception

It is generally accepted that infants' language environment modifies their initial perception of speech sounds. As already noted earlier in this chapter, during the second half of the first year of life infants' perception of language becomes gradually attuned to their native language contrasts, whilst the perception of non-native contrasts declines (Kuhl et al., 2006, 2003; but see Narayan et al., 2010). Although this process of perceptual narrowing or attunement is typically described in the auditory modality, it is also present in the visual and audiovisual modalities (Maurer & Werker, 2014; Pons et al., 2009). Indeed, the process of perceptual narrowing substantially changes the way infants perceive unfamiliar speech; once infants have established their native categories, processing a novel contrast becomes much harder since it has now become a non-native contrast. Also, infants' neural activity reflects the perceptual narrowing process by showing different patterns of cortical activity in response to native and non-native speech throughout the first year of life (Fava, Hull, & Bortfeld, 2014).

Crucially, recent work by Ter Schure and collegues (2016) showed that when infants are faced with the challenging situation of perceiving non-native speech, they resource to additional information such as visual articulations (Ter Schure et al., 2016). This goes in line with Lewkowicz & Hansen-Tift's (2012) previous findings that 12-month-old infants increase their attention to the mouth when perceiving non-native speech, in the same way that bilinguals rely more on the mouth, likely for language separation (Pons et al., 2015).

Thereafter, in a globalized world, the perception of non-native languages – or in the case of adults learners, second languages (L2) – becomes increasingly relevant: we are often faced with situations where we must understand people talking in L2 languages and, as a consequence, we have to cope with both the imperfect naturalistic signals of spoken language and with our imperfect knowledge of such language (for a review, see: Lecumberri, Cooke, & Cutler, 2010). For comprehension to succeed in such situations, non-native listeners must do an extra "listening" effort, as is reflected in an increased pupil size (Borghini & Hazan, 2018), and by the fact that they require between 1 and 7 dB increased speech-to-noise ratio to perform equally than native listeners (van Wijngaarden, Steeneken, & Houtgast, 2002). It would follow logic that then, in the same way that infants resource to the talker's mouth when they are faced with non-native language speech (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015; Ter Schure et al., 2016), adults would also resource to the mouth when they process speech in a second language.

As noted earlier, a body of studies illustrates that certainly, adults do resort to the audiovisual speech cues when they have difficulties for understanding speech. This was first demonstrated by the classic speech-in-noise studies that showed a non-linear performance

increase when speech is presented audiovisually, as compared to the audio- or visual-only presentation (Risberg & Lubker, 1978; Sumby & Pollack, 1954). In the same vein, other studies have shown that listening comprehension of non-native speech (Arnold & Hill, 2001; Reisberg, 1978; Sueyoshi & Hardison, 2005) and the perception of specific L2 speech contrasts (Navarra & Soto-Faraco, 2007) can also be facilitated by the presence of the talker's facial cues. Although in this case the auditory information may be intact, the difficulty for comprehending the speech originates in the listeners' limited knowledge of the non-native language. It has been proposed that the visible articulatory movements of the talker's mouth reduce the set of potential targets the speaker will likely produce and hence primes word identification and facilitates speech comprehension (Skipper, Van Wassenhove, Nusbaum, & Small, 2007).

Together, these studies indicate that perceiving the talker's face is helpful in non-native speech perception. However, whether (and how) adults selectively attend to the talker's mouth to take advantage of such audiovisual gain is still unclear. In the same way that when noise levels increase, adults deploy their attention gradually more to the talker's mouth (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998) and that 8- and 12-month-old infants also deploy more attention to the mouth when perceiving non-native speech (Lewkowicz & Hansen-Tift, 2012; Ter Schure et al., 2016), it would be expected that adults resourced more to the talker's mouth in the perception of non-native speech.

The only study to our knowledge that has explored selective attention in non-native speech perception is Barenholtz et al. (2016). Different from previous studies that had used long (45 s) speech fragments, in this study adults were tracked their eye gaze while preforming an audiovisual classic ABX task in both a native and a non-native language. Specifically, a talker's face uttered short sentences (3 s) in pairs (AB), and after each pair an audio-only presentation of one of the two audiovisual sentences was played (X). Participants had to report whether the audio-only sentence coincided with the first or the second sentence they had just been presented. Results from Barenholtz et al. (2016) study showed that 1) participants deployed overall attention preferably to the mouth of the talking face, and 2) participants significantly increased their attention to the mouth in the non-native speech condition. However, when participants freely-watched the same materials (i.e. without performing the sentence identification task), although overall, they still exhibited a preference for the talker's mouth, the greater mouth-attention in the non-native speech condition was no longer significant. The authors interpreted these results as adults' greater reliance on the mouth speech cues in response to the more challenging situations of having to specifically process the speech – to perform the task – and even more so when the language used in the task was unfamiliar to the participants. Remarkably, participants in this

study were naïve (inexperienced) in the non-native language, hence it was both non-native and unfamiliar. It remains to be known whether this finding would extend to the perception of more naturalistic and longer non-native speech segments, and also whether participants' knowledge of the non-native language (i.e. second language proficiency) would modulate selective attention towards the talking face.

Chapter 2

Research Aims

Research Aims

A clear message to extract from the evidence reviewed in the introduction is that speech is inherently audiovisual, and that both monolingual and bilingual infants and adults are sensitive to and take advantage of the redundant audiovisual information when processing speech. However, the specific manner in which the audiovisual cues influence speech perception and the extent to which infants and adults attend and use these cues is yet to be well described. This thesis therefore addresses the following issues:

1. Perceiving speech audiovisually in the first stages of language acquisition: Does the addition of the visual information of speech modulate infants' language discrimination abilities?

2. Language factors modulate AV speech perception:

    a) Does the distance between bilinguals' two languages influence infants' and children's selective attention patterns to a talking face?

    b) Does the familiarity and/or proficiency with a language influence infants', children's and adults' selective attention patterns to a talking face?

Study 1

In order to address these research questions, I will evaluate monolingual and bilingual infants, children and adults in different perception tasks of audiovisual talking faces. Specifically, in Study 1 I will explore the first question here described, by studying monolingual and bilingual infants' capacity for discriminating languages in the audiovisual modality; that is, their ability to detect when a talking face switches between two languages. I will do so at the age of 4 months, when infants have been found to be able to discriminate their two close languages acoustically. Based on the reviewed literature, it is expected that both groups will easily detect a distant language switch. However, the evidence is not as clear regarding the detection of a close language switch; the prediction is that although both groups may succeed, bilingual infants may show a delayed detection of the close language switch.

Studies 2, 3 & 4

In Studies 2, 3 and 4 I will explore the influence of bilinguals' language background onto selective attention to talking faces. I will do so in a population that is in the midst of acquiring their two languages, at 15 months of age (Study 2), and thereafter in a population where most aspects of language have already been acquired, at 5 to 6 years of age (Studies 3 and 4). More specifically, I will compare the effect of acquiring a pair of close languages versus acquiring a pair of distant languages (Studies 2 and 3), and also the temporal dynamics of monolingual and bilingual children's attention to a talking face (Study 4). The prediction is that those infants learning a pair of close languages may rely more on the audiovisual speech cues of a talker's mouth, putatively for aiding language differentiation. Then, although to a lesser extent, the prediction is that this effect may also be observed in children, although caused by different cognitive processes than in infants. Last, studying the temporal dynamics (Study 4) may give us some insight into children's selective attention strategies to a talking face.

Studies 5, 6 & 7

In Studies 5, 6 & 7 I will explore the influence of language factors such as language familiarity and proficiency on adults' selective attention to talking faces. In Study 5 I aim to replicate previous work by Barenholtz et al. (2016) – i.e. adults' greater attention to the mouth when performing a speech processing ABX task with short sentences – with our different set of stimuli and languages. Subsequently, in Study 6, I will explore adults' attention to a talker's face in more naturalistic and long speech segments in a non-native language. Last, in Study 7, I will evaluate the influence of different levels of non-native language proficiency onto the selective attention patterns to a talking face. Based on the reviewed literature, my expectation is that adults will deploy more of their attention to a talker's mouth when perceiving a non-native language, and that this mouth-preference will decrease with language proficiency (i.e., highly proficient non-native listeners should attend less to a talker's mouth than low level learners).

Chapter 3

Audiovisual Language Discrimination

Study 1

Detection of a language switch from a talking face: Evidence from monolingual and bilingual 4-month-old infants

Introduction

Previous studies show that by 4 months of age, both monolingual and bilingual infants can discriminate auditorily any language pair. However, whether this ability extends to the more naturalistic situation of perceiving talking faces is yet to be known. This issue is particularly relevant for infants growing up in bilingual environments, because on top of learning and keeping apart their languages, some bilingual infants are exposed to speakers switching between these languages (i.e. code-switching). Thus, bilingual infants growing up in such environments should learn to detect when one single speaker switches between the two languages in order to process the two languages adequately.

The only study to our knowledge that has specifically addressed audiovisual language discrimination is Bahrick and Pickens (1988). The authors found that 5-month-old monolingual and bilingual infants could discriminate Spanish and English audiovisual passages from a bilingual speaker. However, whether 4-months-old infants can discriminate close languages presented audiovisually remains an open question. Moreover, infants' exposure to code-switching was not considered in Bahrick and Pickens (1988) study, nor in any language discrimination studies reported previously, and it could well be modulating infants' performance.

The main goal of the current study was to evaluate whether 4-months-old bilingual infants could detect when a talking face switches between their two native and close languages. Specifically, we assessed 4-month-old monolingual and bilingual infants' ability to

detect when a talking face switches from speaking in their native or dominant language (Catalan or Spanish) to the other close language (Catalan or Spanish). Since we hypothesized that the close language switch would be difficult to detect for 4-month-old infants, we also evaluated a second switch to a distant language (English), to ensure that both groups would show a positive result (as in Bahrick & Pickens, 1988). Moreover, based on previous evidence (Bosch & Sebastián-Gallés, 1997; Nacar Garcia et al., 2018), we also predicted that bilingual infants might show a delayed close language switch detection. Thus, we assessed the language switch detection in two test trials, allowing infants more time to explore the before-mentioned delayed detection. Last, we expected that those bilingual infants with higher exposure to code-switching might also show a faster detection as compared to those bilingual infants with no or very low exposure to code-switching.

Method

Participants. We recruited seventy-seven 4- to 5-month-old infants from a maternity hospital for this experiment. All were healthy, full-term infants with no history of vision or hearing problems according to parental report. The Language Exposure Assessment Tool (LEAT) (DeAnda, Bosch, Poulin-Dubois, Zesiger, & Frienda, 2016) was used to establish an estimate of daily exposure to the language(s) being learned by the infants. Participants were divided in two groups based on their linguistic environment: Spanish or Catalan monolinguals and Spanish-Catalan bilinguals. As in previous bilingual studies (Bosch & Sebastian-Galles, 2001; see also Byers-Heinlein, 2015), we required that an infant's daily exposure to each of the input languages range between 50%-50% and 25%-75% of exposure time to each input language. The final sample was composed of 55 infants (Mean age = 4 months, 9 days, Range = 3 months, 12 days - 5 months, 7 days). This included 29 monolingual infants (8 Catalan, 21 Spanish; Mean age = 4 months, 9 days, Range = 3 months, 17 days – 4 months, 27 days, 16 boys) and 26 bilingual infants (Mean age = 4 months, 10 days; Range = 3 months, 12 days – 5 months, 7 days; 13 boys). The remaining 22 infants were tested but excluded from the final data analysis due to the following reasons: crying or fussiness (8), parental interference (n=2), failure to reach the habituation criterion (n=7), failure to complete a minimum of 6 trials of habituation (n=5).

Stimuli. The stimuli consisted of thirty-six 10s-long audiovisual video clips, extracted from three popular children's stories and recorded by a female actor in Catalan, Spanish and English in a child-directed manner (twelve videos per language). The recording took place

in a soundproof booth. The speaker was a 21-year-old Catalan, Spanish and English simultaneous trilingual woman (i.e. English father and Catalan-Spanish bilingual mother). The selected videos segments were checked by independent experimenters for consistency in lighting, positioning of the actress on screen and audio pitch and clarity. Moreover, in order to remove any abrupt entrances all videos began and ended with a 1-second fade-in/out.

Apparatus. The experiment was conducted in a dimly lit and sound-attenuated test booth. The infant sat on an infant seat approximately 2.5ft (75cm) away from a 65" LG TV screen while the parents sat behind them. The face of the speaker presented in the screen measured 1.15x1.24ft (35x38cm). The audio played through two loudspeakers (Sony SS-125 E) situated below the screen and covered by black fabric. Infants were recorded using a digital video camera (Canon MV750i), which was connected to a display and recording device for online and offline coding of the infant's response. The experiment was controlled by the Habit program (Habit X v.1.0; Cohen, Atkinson, & Chaput, 2000) run on a Mac computer (OS X; v.10.4.11). The experimenter monitored the infant's eye gaze direction from an adjacent room, unaware of the trial status, by pressing a key on the computer's keyboard.

Procedure. We used the habituation-switch discrimination design (Stager & Werker, 1998; Werker, Cohen, Lloyd, Casasola, & Stager, 1998) with the addition of a second switch trial. The experiment started and ended with an animated rotating wheel. Each trial was 10s long and they were presented pseudo-randomly. All trials were preceded by the blue flower attention getter until the experimenter ensured the infant was oriented to the screen and triggered the trial. Depending on their native – or dominant – language infants were divided into two groups; 19 infants were assigned the Catalan habituation, and 36 were assigned to the Spanish habituation. The habituation criterion was set to 60% of the total looking time with a moving average of three trials – i.e. the software ended the habituation and triggered the test phase when the average looking time across a three-trial block decreased to the criterion of 60% of that infant's maximum looking time. If the criterion was not reached by the end of 24 trials the phase also ended. Infants who habituated in 6 or fewer trials and who failed to habituate within 24 were excluded from analysis.

The test phase consisted of one same trial and four switch trials, with the same inter-trial attention getter as in the habituation phase. First a video clip belonging to the same language as in the habituation phase was presented, the "same" trial. Thereafter the "switch" trials were presented. The "switch" trials consisted of two blocks, the "close language

switch" and the "distant language switch", and each block was composed of two 10 s trials[5]. Hence, the test phase sequence was the following: "same", "close switch 1", "close switch 2", "distant switch 1", "distant switch 2".

We expected that if infants detected the language switch, they would recover their attention (i.e. dishabituate) and hence look longer during that switch trial compared to the same trial. We also anticipated that once infants had detected the first language switch and dishabituated, the second language switch would not differ from the first switch, regardless of the language presented.

After the experiment, a trained coder who was blinded to the stimuli calculated infants' looking times on a frame-by-frame observation of the video records. Looking times obtained from this offline coding were used in the analyses.

Last, parents completed the Language Mixing Questionnaire (Byers-Heinlein, 2012) to measure the amount of code-switching in the bilingual families. This questionnaire contains 5 questions (1 to 7), and it evaluates code-switching at the word level, sentence level, and also the direction of switch. Both parents completed the questionnaire and hence we collected two scores (1-35) per infant.

Results

To ensure that overall habituation levels did not vary across groups, we first analyzed the number of trials and the habituation rate with a one-way ANOVA with Linguistic Group (Bilingual, Monolingual) as the between-subjects factor (see Figure 3). This analysis revealed that the two groups did not differ in number of habituation trials [bilinguals M = 17.85, SD = 2.85, monolinguals M = 15.66, SD = 5.13; $F(1,54) = 3.70$, p = n.s.] nor in habituation rate bilinguals M = 49.28%, SD = 9.13, monolinguals M = 50.67%, SD = 7.59, $F(1,54) = .38$, p = n.s.].

To answer the first question of whether infants detected the first (close) or second (distant) language switch, we averaged the close and distant language switch blocks (averaged the two 10 s trials) and computed a mixed analysis of variance (mixed-ANOVA) with the Total Looking Times (TLT) and the following factors: Test Trial (same, average close switch, average distant switch) as a within subject's factor, and Group (Bilingual, Monolingual) and

[5] Since an earlier study shows longer orientation latencies towards their native language in bilingual infants (Bosch & Sebastián-Gallés, 1997), we considered the possibility that bilinguals' switch detection may require more time, and thus we composed each switch trial of two 10 s trials of that language (i.e. 20 s).

Habituation Language (Catalan, Spanish) as between subject's variables. Results revealed a main effect of Test Trial [$F(1,51) = 20.79$, $p < .001$, $\eta_p^2 = .290$], and a significant interaction between Test Trial and Group [$F(1,51) = 4.35$, $p = .042$, $\eta_p^2 = .079$]. The lack of interaction between habituation language and test trial indicates there is no asymmetry of switch detection between Catalan and Spanish. The significant interaction between Test Trials and Group was further explored using paired t-tests to compare the same trial to the two switch trials separately in each group (see Figure 4). The monolingual group showed a significant difference between same and average close switch [$t(28) = 3.06$, $p = .005$, $d = .63$] and also between same and average distant switch [$t(28) = 3.63$, $p = .001$, $d = .74$]. In the bilingual group, the difference between same and average close switch did not reach significance [$t(25) = .293$, $p >. 1$] but the difference between same and average distant switch did [$t(25) = 3.42$, $p = .002$, $d = .65$].

These results indicate that whilst the monolingual group dishabituated in the first language switch (Spanish to Catalan or vice-versa) and then continued to attend in the second language switch (English), the bilingual group did not dishabituate until they reached the distant language switch. In other words, the monolingual group detected the close language switch whereas the bilingual group only detected the distant language switch.

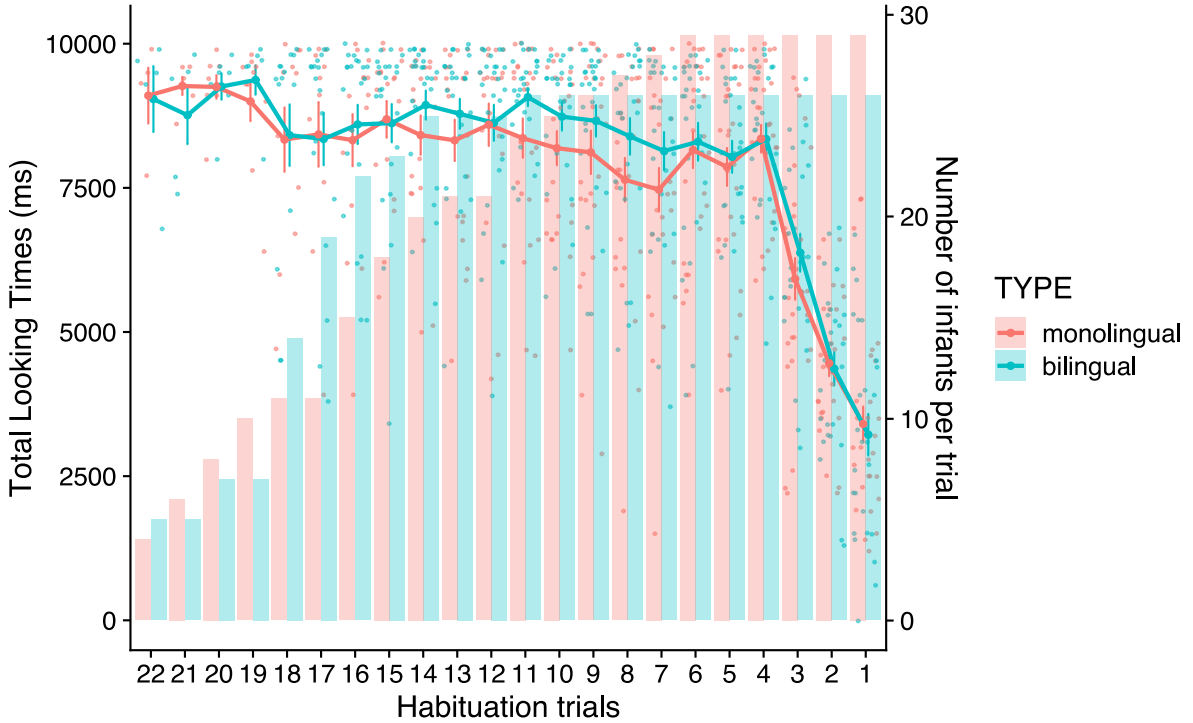


Figure 3. Attention curve throughout the habituation process. The x axis depicts the inverted habituation trials; trial 0 being the last habituation trial before the test trials start. On the left y axis, the total-looking-time (TLT) is shown in milliseconds, and on the right y axis the number of infants included in each trial. Dots represent the individual TLT means. Error bars represent the standard error of the means (SE), and bars the number of infants per trial.

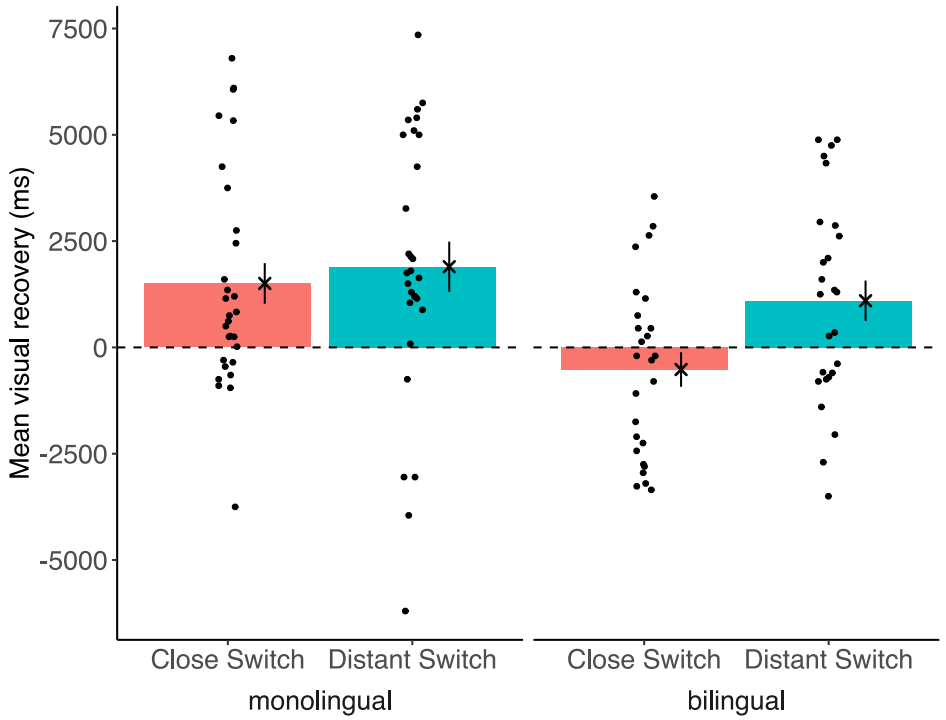


Figure 4. Mean visual recovery times for the close and distant switch trials of monolingual and bilingual 4-month-old infants. Bars and crosses represent group means, and the dots the individual mean times. Vertical lines depict standard errors (SE) of the means.

Based on Bosch & Sebastián-Gallés (1997) results, we also explored the possibility of a delayed language switch detection in bilingual infants. To do so, we performed the same ANOVA but without averaging the test trials. Thus, we computed a new mixed ANOVA with the Total Looking Times (TLT), the Test Trial (same, close switch 1, close switch 2, distant switch 1, distant switch 2) as a within subject's, and Group (Bilingual, Monolingual) and Habituation Language (Catalan, Spanish) as between subject's variables. As in the first analysis, the analysis yielded a main effect of Test Trial [$F(1,51) = 24.65$, $p < .001$, $\eta p2 = .326$] and an interaction between Test Trial and Group [$F(1,51)= 3.77$, $p = .058$, $\eta p2 = .203$]. Even though the interaction here was only marginal, we felt that based on our theoretical predictions there was an a priori justification for examining the data separately for each linguistic group. Tests of these a priori hypotheses were conducted using Bonferroni adjusted alpha levels of 0.012 per test (0.05 / 4 comparisons). Paired t-tests for the monolingual group indicated that the same trial was significantly different to the second trial of the close language switch (same vs. close switch 2 [$t(28) = 3.06$, $p = .005$, $d = .79$], and to the two distant language switches (same vs. distant switch 1 [$t(28) = 2.75$, $p = .01$, $d = .62$], same vs. distant switch 2 [$t(28) = 3.98$, $p < .001$, $d = .88$]). The first trial of the close language switch did not reach significance (same vs. close switch 1 [$t(28) = 2.04$, $p = .051$]). In the bilingual group, the Bonferroni-adjusted t-tests only yielded significant results for the second trial of the distant language switch (same vs. distant switch 2 [$t(25) = 3.26$, $p = .002$, $d = .71$], all other ps > .1).

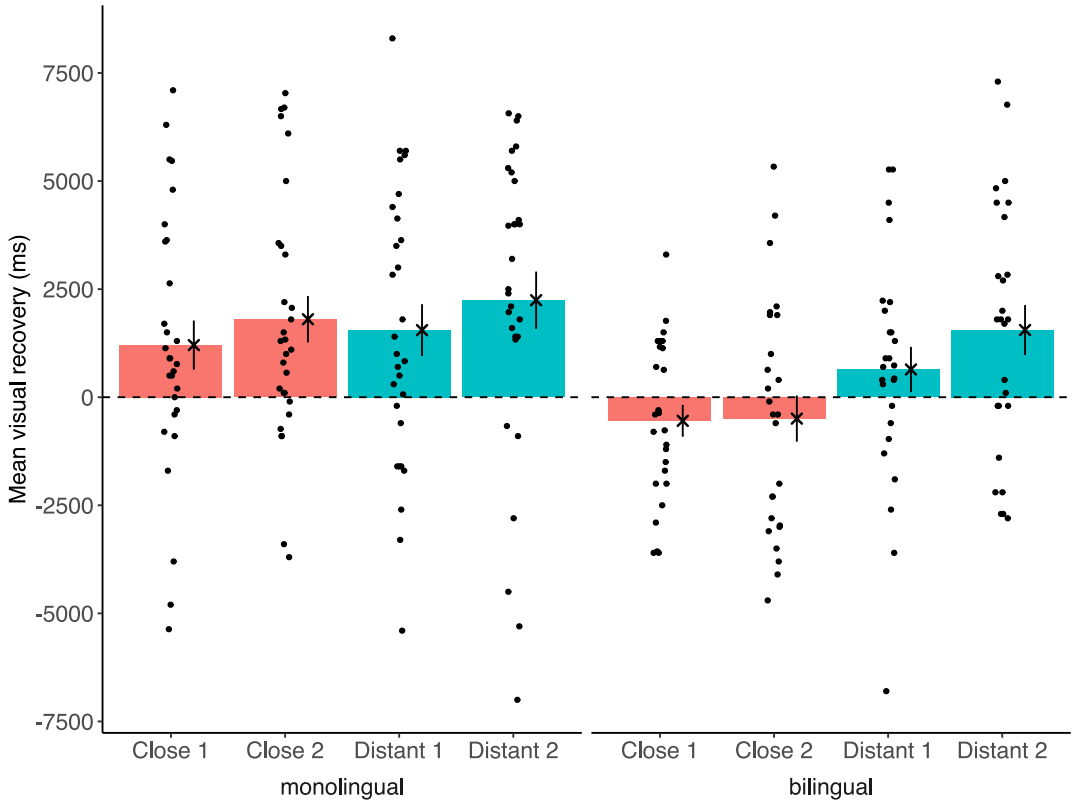


Figure 5. Mean visual recovery times for the close 1, close 2, distant 1 and distant 2 switch trials of monolingual and bilingual 4-month-old infants. Bars represent group means, and the dots the individual mean times. Vertical lines depict standard errors (SE) of the means.

The last goal of the current study was to explore whether bilinguals' language switch detection ability would be influenced by their previous experience of their parents code-switching languages. To explore this hypothesis, we used the obtained scores of the language Mixing questionnaire of each infant's main caretaker. The results showed a distributed pattern with a mean score of 12.58 out of 35 and a high standard deviation of 8.14. To analyze these results, we divided the sample in high switchers and low to none-switchers (high CS Score >14 items; low to none-switchers Score <12 items) on basis of a median split at 13. Following the logic from the previous two analysis, we performed two mixed ANOVAs: one with the averaged Test Trial (same, average close switch, average distant switch) as a within subject's factor and Code-switching (high CS, low CS) as between subject's, and the other with Test Trial (same, close switch 1, close switch 2, distant switch 1, distant switch 2) and Code-switching (high CS, low CS) as between subject's.

The results of the two ANOVAs did not yield any significant results (all ps > .5). However, to ensure that there was no modulation of the code-switching experience, we also analyzed the individual scores of the language mixing questionnaire by way of Pearson Correlations with the amount of looking recovery from the same trial to the four different switch trials, first averaged (same to average close switch and same to average distant switch) and then with the full trials (same to close switch 1, same to close switch 2, same to distant switch 1, same to distant switch 2). Once again, all correlations were non-significant (All ps > .1). These null results indicate that bilinguals' experience of code-switching in their environment does not seem to influence their capacity for detecting a language switch from a talking face.

Discussion

The main goal of the current study was to determine whether 4-months-old monolingual and bilingual infants would detect a close and a distant language switch from a talking face. To accomplish this, we habituated 4-month-old monolingual and bilingual infants to a talking face speaking in their native language, and then we measured attention recovery when the speaker switched to a close language (close switch, Catalan or Spanish) and then to a distant language (distant switch, English). As expected, we found that the monolingual group detected when the speaker switched from Catalan to Spanish – or vice-versa. In contrast, however, the bilingual group did not recover their attention when the speaker switched between their two languages. Regarding the distant language switch to English, both monolingual and bilingual groups showed an attention increase. Interestingly, when splitting the switch trials into two 10 s trials, both groups showed that in fact, it was not until the

second 10 s trial that they detected the language switch. Last, and different from our expectations, we also found that bilinguals' amount of code-switching exposure did not modulate their detection of the language switches.

The findings obtained here provide the first evidence of close language audiovisual discrimination in 4-month-old infants. Our results that monolingual infants detected the close language switch fit well with previous evidence showing close languages auditory discrimination at this age (Bosch & Sebastián-Gallés, 2001; Nazzi et al., 2000), and add to this evidence that discrimination is also possible when perceiving the same single talking face switching between the languages.

Importantly, the most relevant finding is that the bilingual group did not detect the close language switch (Catalan to Spanish or vice-versa) from a talking face, that is, the switch between their two familiar languages. This is surprising in light of previous evidence showing auditory discrimination (Bosch & Sebastián-Gallés, 2001). Although studies exploring audiovisual speech perception in close bilingual infants – specifically Catalan-Spanish bilingual infants – would predict an enhanced sensitivity for visual speech cues (Sebastián-Gallés et al., 2012), and more attention to the mouth than their monolingual peers (Pons et al., 2015), our results show that in the task of language discrimination, presenting the talking face seems to hamper bilinguals' performance.

Reasonably, bilinguals' task of differentiating their two familiar languages – with half the amount of input – involves quite different and likely more complex cognitive mechanisms than monolinguals' task of separating their native from a non-native one. Indeed, a recent study supports this idea by proposing that monolingual infants base their early discrimination on familiarity, whilst bilingual infants perform a later processing, compatible with an increased attention to the speech signal (Nacar Garcia et al., 2018). If seeing the face of the talker – constant across the language switch – distracts infants' attention from speech, it would be reasonable that it hampers switch detection the most when it is based on attention mechanisms, as is proposed to be the case of bilingual infants. Moreover, the fact that code-switching scores did not modulate the detection of the switch suggests that infants' ability is not directly related to their previous experience with talking faces changing languages, but rather is related with the familiarity of the languages perceived. Further audiovisual language discrimination studies that specifically assess familiarity and bilingualism independently are needed to disentangle these two factors and confirm or shape this interpretation.

Noteworthy, the habituation-switch procedure is not without limitations; although bilingual infants did not show an attention recovery in the close language switch, some may argue that they could have detected the switch, but that due to the familiarity of the two

languages and the constant face, the detection of the switch did not lead to an attention increase. Indeed, studies using techniques such as EEG could detect these more subtle effects, but the fact that in the present study both groups showed a positive effect (i.e. an attention recovery) with the same task and conditions suggests that at least, if anything, the detection would have been much weaker than in the other groups and languages, and hence it would lead to similar conclusions.

Our results that both monolingual and bilingual infants detected when the speaker switched between two rhythmically distant languages (i.e. Spanish or Catalan to English) are consistent with previous studies showing that distant languages are discriminated auditorily at birth (Byers-Heinlein et al., 2010; Nazzi et al., 1998), they can be discriminated visually-only at 4 months of age (Weikum et al., 2007), and audiovisually at 5 months of age (Bahrick & Pickens, 1988). In the same line, a recent study by Berdasco-Muñoz, Nazzi, and Yeung (2019) has shown that 6-month-old infants change their visual exploration pattern of a talking face when the speaker switches between two distant languages (French and English). Also, concerning the bilingual group, it is relevant to mention that having positive results in the second language switch validates their null attention recovery in the first, close language switch.

Last, the results of our second analysis – the two switch trials separately – did not reflect bilinguals' slower detection of the switch, as expected from Bosch and Sebastián-Gallés (1997). Instead, both groups detected the language switch at the second time window, that is, between 10 and 20 s of exposure. These results support the idea that the addition of the visual information reduced infants' overall capacity for detecting the switch, putatively due to the fact that it adds a visual factor that remains unchanged when the switch occurs. However, these temporal results must be interpreted with caution because neither group showed an early detection of a switch. Therefore, future studies that explore specifically the timing of a language switch detection are needed to strengthen and extend these conclusions.

In sum, this study is the first to show that monolingual 4-month-old infants notice when a talking face switches from their native language to a close one, but that bilingual 4-month-old infants can only detect the switch when it involves a distant and/or unfamiliar language. The fact that bilingual 4-month-old infants did not detect the switch between their two close languages and that their experience with code-switching did not modulate this detection has further implications in the underlying processes of bilingual language acquisition, and in the understanding of the manner in which bilingual infants perceive their linguistic environment. Further studies looking into bilinguals' specific attentional mechanisms involved in the audiovisual discrimination of their languages will help us elucidate a more complete picture of this phenomenon.

Chapter 4

Language Factors Modulate Selective Attention to a Talking Face

Evidence from Infancy

Study 2

The influence of language distance on bilingual infants' selective attention

Introduction

The combination of the reviewed literature and the results from Study 1 demonstrate that discriminating close languages – such as Catalan and Spanish – is a difficult task for infants and that there are cognitive consequences that derive from this process. If learning two close languages entails a greater linguistic challenge for bilingual infants, is it possible that learning such close languages also modifies the way that infants selectively attend to the speech cues of a talking face?

As already noted, Pons, Bosch, and Lewkowicz (2015) showed that Catalan-Spanish bilingual infants (a) show equal looking times to the eyes and mouth at 4 months of age, different to their monolingual peers that attend to the eyes at this age and (b) they attend preferably more to the mouth of the talker's face during the second part of the first year of life, in response to both their dominant and non-native language. However, what is not clear from these findings is whether the rhythmic and phonological distance of the two languages being learned by bilingual infants might differentially affect their deployment of attention to a talker's mouth. Specifically, it is possible that the bilingual Catalan- and Spanish-learning infants in the Pons et al. (2015) study deployed more attention to the mouth because Catalan and Spanish are rhythmically and phonologically close languages (Bosch & Sebastián-Gallés, 2001; Ramus et al., 1999).

The evidence suggests that infants learning two close languages might find it more difficult to separate the sounds of each of their two languages. If that is the case, then it is

also possible that close-language bilingual infants take greater advantage of audiovisual redundancy than do their more distant-language bilingual counterparts in situations where they have access to a talking face.

The goal of this study was to test this language-distance hypothesis and, thus, examine whether bilingualism is a heterogeneous phenomenon in which the specific properties of the languages being acquired play an important role in infants' audiovisual speech perception. To test this hypothesis, we investigated whether the patterns of selective attention found in bilingual infants' response to a talker's eyes and mouth differ as a function of the proximity of the two languages that they are learning. Thus, in Study 2 we examined selective attention to a talker's eyes and mouth in 15-month-old bilingual infants who were either learning a pair of close or a pair of distant languages.

Method

Participants. We recruited forty-seven 15-month-old infants from a maternity hospital for this experiment. All were healthy, full-term infants with no history of hearing problems according to parental report. The Language Exposure Assessment Tool (LEAT) (DeAnda et al., 2016) was used to establish an estimate of daily exposure to the language(s) being learned by the infants. To adequately represent the infants' bilingual environments, parents were instructed to indicate word productions in any of the infants' two languages.

Participants were divided into two groups based on their linguistic environment: Close bilinguals (Catalan-Spanish) and distant bilinguals (Catalan or Spanish and a rhythmically and/or phonetically distant language, described below). As in previous bilingual studies (Bosch & Sebastian-Galles, 2001; see also Byers-Heinlein, 2015), we required that an infant's daily exposure to each of the input languages range between 50%-50% and 25%-75% of exposure time to each input language. Nine additional infants were tested but not included in the final data analyses due to: crying (4), failure to complete the calibration phase of the procedure (1), or failure to obtain a minimum of 9s of data during a 45s test trial (this equals to 20% of the test trial) (4).

The final sample was composed of 38 infants (Mean age = 15 months, 5 days, Range = 14 months, 8 days - 15 months, 13 days). This included 20 close bilingual infants (Spanish-Catalan; Mean age = 14 months, 29 days, Range = 14 months, 20 days – 15 months, 7 days, 11 boys) and 18 distant bilingual infants (Spanish-Other, where Other refers to 1 Swedish, 6

German, 4 Russian, 3 Arabic, 3 French and 1 Rumanian infant; Mean age = 15 months, 12 days; Range = 14 months, 26 days – 15 months, 27 days; 5 boys)[6].

Stimuli. The stimuli were identical to those used by Pons et al. (2015) and consisted of 45 s audiovisual video clips in which one of two female actors recited a prepared monologue. One of the actors (a highly proficient Catalan-Spanish bilingual) recited a Spanish or a Catalan version of the monologue, whereas the other actor (a native speaker of American English) recited an English version of the monologue. To elicit maximal attention, the actors recited the monologues in an infant-directed manner (Fernald, 1985).

Apparatus and procedure. Infants were seated on an infant seat while the parents sat behind them. Testing took place in a dimly lit and sound attenuated room and the stimuli were presented on a 17 in computer monitor using Tobii Studio software (Tobii Technology AB, Danderyd, Sweden). Eye gaze was recorded with a Tobii X120 stand-alone eye tracker at a sampling rate of 60 Hz. We used the Tobii eye tracker's five-point calibration routine to calibrate each participant's gaze. The experiment started with the calibration routine. Once calibration was successfully completed, we presented two videos, one in the infants' dominant native language (Catalan or Spanish) and one in a non-native language (English). The order of the videos was counterbalanced across infants. While the infants watched the videos, the eye-tracker monitored their gaze at two areas of interest (AOI), the eyes and the mouth. The AOIs used here were identical to those used by Pons et al. (2015).

Results

To compare the proportion of time deployed to the eye and mouth AOIs, we computed proportion-of-total-looking-time (PTLT) scores for each participant by dividing the amount of time they looked at each AOI, respectively, by the total amount of time they looked at the face. We then analyzed these scores by way of a mixed ANOVA, with AOI (eyes, mouth) and Test Language (native, non-native) as within-subjects factors and Linguistic Distance (close bilingual, distant bilingual) as the between-subjects factor. Results revealed a main effect of Test Language [$F(1,36) = 4.97$, $p = .032$, $\eta_p^2 = .12$], a main effect of AOI [$F(1,36) = 62.63$, $p < .01$, $\eta_p^2 = .63$], a significant AOI x Test Language interaction [$F(1,36) = 7.39$, p

[6] Although French and Rumanian are also Romance languages, each of them is nevertheless substantially different from Spanish, either at the phonological, the morphological or the lexical stress levels.

= .01, η_p^2 = .17], and a significant AOI x Linguistic Distance interaction [$F(1,36) = 4.70$, p = .037, η_p^2 = .12]. The Test Language main effect reflects infants' greater total PTLT (i.e., PTLT to the eyes AOI + PTLT to the mouth AOI) when exposed to the face talking in the non-native language than to the face talking in the native language. The AOI main effect was due to the fact that, overall, infants looked more at the mouth than the eyes. The Test Language x AOI interaction was mainly due to the fact that infants looked longer at the mouth when they were exposed to the non-native than to the native language. Finally, the most interesting result was the AOI x Linguistic Distance interaction. Figure 6 displays the mean PTLT scores for each AOI, collapsed across the two languages, as a function of linguistic distance. As can be seen, even though both the close and the distant bilingual groups looked more to the mouth than eyes [t(19) = 7.99, p < .01, d = 3.27; t(17) = 3.65, p < .01, d = 1.42, respectively], the close bilingual group looked more to the mouth than the eyes than did the distant bilingual group. distance. As can be seen, even though both the close and the distant bilingual groups looked more to the mouth than eyes [$t(19) = 7.99$, p < .01, d = 3.27; $t(17) = 3.65$, p < .01, d = 1.42, respectively], the close bilingual group looked more to the mouth than the eyes than did the distant bilingual group.

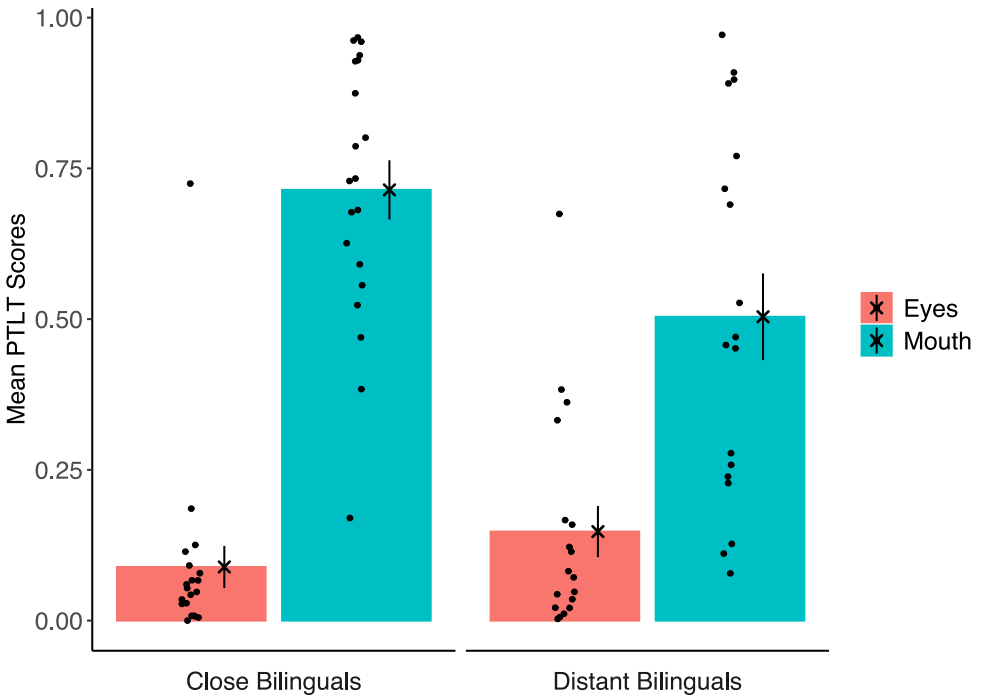


Figure 6. Distribution of mean proportion-of-looking-time (PTLT) scores to the eyes and mouth as a function of Linguistic Distance (Close and Distant Bilinguals), collapsed across languages. Dots represent each infant's Mean PTLT score; Bars and crosses with error bars represent Mean PTLT score and standard error of the mean (SE) for each group.

Discussion

The current experiment yielded three principal findings. First, we found that, overall, 15-month-old bilingual infants attended more to the mouth than the eyes when they were exposed to a talking face. Second, we found that the infants attended more to the mouth than the eyes when the talker spoke in a non-native than native language and that they did so regardless of the linguistic distance between English and the bilinguals' second language. Finally, as predicted, we found evidence consistent with our language distance hypothesis. Close bilingual infants attended longer to a talker's mouth than distant bilingual infants did.

Previous studies had demonstrated that close bilingual infants (learning Catalan and Spanish) deploy more attention to a talker's mouth than their monolingual counterparts (Ayneto & Sebastián-Gallés, 2016; Fort et al., 2017; Pons et al., 2015). These results are now extended by the findings from the current study demonstrating that in fact, greater attention to a talker's mouth is not a characteristic of bilingualism *per se*. Rather, it appears that greater attention to a talker's mouth is deployed by bilingual infants exposed to close languages.

Critically, the fact that close bilingual infants attended more to a talker's mouth than eyes indicates that linguistic distance plays an important role in selective attention to talking faces at 15 months of age. Consistent with findings from adult studies showing that audiovisual redundancy cues facilitate adults' speech processing under challenging conditions as well (Barenholtz et al., 2016; Lansing & McConkie, 2003; Reisberg et al., 1987; Vatikiotis-Bateson et al., 1998), the present results suggest that access to redundant audiovisual cues may help infants learning close languages to disambiguate them.

Chapter 5

Language Factors Modulate Selective Attention to a Talking Face

Evidence from Childhood

Study 3

The influence of language distance on bilingual children's selective attention

Introduction

The previous study confirmed the hypothesis that linguistic distance is indeed modulating bilingual infants' selective attention to a talking face, putatively to help them face their greater challenge of learning and separating two close languages. The purpose of the current study was to investigate whether the greater amount of selective attention deployed to a talker's mouth by close as opposed to distant bilingual infants extends into early childhood.

As earlier noted, children's phonological system is already established and its production is nearly error-free (Bosch Galceran, 2004; Dodd et al., 2003). If phonological expertise alone mediates selective attention to a talker's mouth, then children might exhibit a reduction in selective attention to the mouth. If, however, a bilingual context contributes to the mouth bias, then they may also focus their attention on a talker's mouth presumably because this may facilitate their comprehension and/or disambiguation of the two languages. Thus, the specific question addressed here was whether selective attention to the talker's face is modulated by the proximity of the input languages in early childhood in the same way that it is in infancy.

We expected that young bilingual children learning two close languages might take greater advantage of redundant audiovisual cues than bilingual children learning two distant languages. This prediction would not only be consistent with our infant findings but also with findings in children showing that monolingual and distant bilingual children attend equally to the eyes and mouth (Byers-Heinlein et al., 2014; Król, 2018). Moreover, it would also be consistent with adults' increased attention to the mouth in face of speech processing

difficulties (Barenholtz et al., 2016; Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998). Therefore, it is reasonable to expect that those children learning close languages may take greater advantage of the audiovisual speech cues of a talker's mouth, given that they have a more demanding task (i.e. language disambiguation) than children learning two distant languages do. No studies to date have specifically investigated whether the linguistic proximity of young bilingual children's two languages modulates their attentional deployment to a talker's eyes and mouth. Therefore, we tested this possibility in 4-6-year-old bilingual children learning pairs of either close or distant languages.

Method

Participants. We recruited 46 children from a school located in Barcelona (Catalonia, Spain). None of the children had a history of hearing problems according to parental report. Parents completed an online language questionnaire to establish the language background of the participants. All participants came from Spanish, Catalan or Russian homes and they had been exposed to their second language early in life (i.e., before entering school). As a result, the participants were early sequential bilinguals, either close bilinguals (Catalan and Spanish) or distant bilinguals (Russian and Spanish). After entering school at 3 years of age, participants also had exposure to English as a third language.

Seven children were tested but not included in the final data analysis because another language was spoken at home that was not Catalan, Spanish or Russian (4) or because we failed to obtain a minimum of 2 s of looking per each 10 s trial (3) (Frank, Vul, & Saxe, 2012). Thus, the final sample consisted of 32 children (Mean age = 5 years, 8 months, Range = 4 years, 2 months - 6 years, 9 months) of whom 17 were close bilinguals (Spanish-Catalan, Mean age = 5 years, 8 months, Range = 4 years, 2 months - 6 years, 8 months, 11 boys) and 15 were distant bilinguals (Spanish-Russian, Mean age = 5 years, 9 months, Range = 4 years, 4 months - 6 years, 9 months, 9 boys).

Stimuli. We first conducted a pilot test with the stimuli presented in Study 2 (two videos in infant-directed speech, 45 s-long each) and found that children at this age did not sustain sufficient attention for videos this long. Therefore, we made similar but more appropriate stimuli for 4- to 6-year-old children. The new stimuli consisted of shorter video clips (10 s). In each video, one of two female actors uttered a short monologue (part of "the Snowman" story by R. Briggs, 1978) in their native language (one in Spanish and the other in American English) in a child-directed manner. Note that English was not an unfamiliar language for the participants: it was non-native but familiar (L3).

Procedure. Children were seated on an adjustable chair, 60 cm in front of the computer monitor, in a small and dimly lit room of the school. Similar to Study 2, the stimuli were presented with Tobii Studio software (Tobii Technology AB, Danderyd, Sweden) and eye gaze was recorded using a Tobii T120 eye-tracker integrated into a 17-inch TFT monitor, at a sampling rate of 60 Hz. We used the Tobii eye tracker's nine-point calibration routine to calibrate each participant's gaze. After the calibration was completed, each participant watched the two video clips, one in Spanish and the other in English. The order of the videos was counterbalanced across children. While the children watched the videos, the eye-tracker monitored their gaze to the same two AOIs, namely the eyes and the mouth as in Study 2.

Results

First, considering the linguistic background of these children, and to ensure that they had comparable competence in Spanish (L1 /L2) and English (L3), participants' competence in these two languages was obtained from the school (teachers' formal assessment) and compared with a Mann–Whitney U test. The test showed that the children's competence level in the two languages used in the experiment was equivalent across the two groups (for Spanish U= 76.5, n.s.; for English U= 161, n.s.).

Then, identical to Study 2, we used a mixed ANOVA, with AOI (eyes, mouth) and Test Language (Spanish, English) as within-subjects factors and Linguistic Distance (close, distant) as the between-subjects factor to analyze the PTLT scores. Results yielded a main effect of Test Language [$F(1,30) = 7.87$, $p < .01$, $\eta_p^2 = .21$] and an AOI x Linguistic Distance interaction [$F(1,30) = 5.13$, $p = .03$, $\eta_p^2 = .15$].

The Test Language main effect reflects children's greater total PTLT (i.e., PTLT to the eyes AOI + PTLT to the mouth AOI) when exposed to the face talking in the non-native language than to the face talking in the native language. The AOI x Linguistic Distance interaction indicates that the differential amount of attention deployed to the eyes and mouth depended on children's linguistic background. Figure 7 displays the mean PTLT scores for each AOI, collapsed across the two languages, as a function of linguistic distance.

Follow-up paired t-tests showed that the close language group looked more at the mouth than the eyes [$t(16) = 2.51$, $p = .023$, $d = 1.18$], whereas the distant language bilingual group looked equally at the eyes and mouth [$t(14) = .78$, $p = .44$, $d = .92$]. The absence of a Test Language x AOI interaction could be due to the fact that the non-native language was familiar (L3) to the participants.

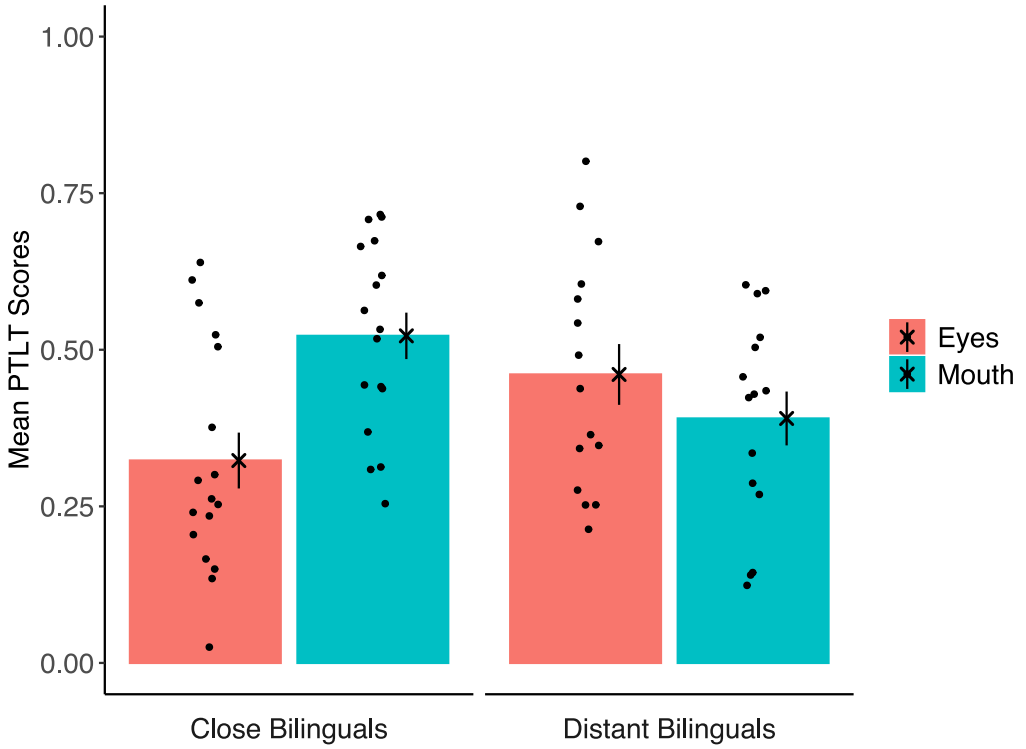


Figure 7. Distribution of mean proportion-of-looking-time (PTLT) scores to the eyes and mouth as a function of Linguistic Distance (Close and Distant Bilinguals), collapsed across languages. Dots represent each infant's Mean PTLT score; Bars and crosses with error bars represent Mean PTLT score and standard error of the mean (SE) for each group.

Discussion

Overall, the results from this experiment show that, as a group (i.e., regardless of the proximity of their two languages), 4- to 6-year-old bilingual children looked equally at a talker's eyes and mouth. Nonetheless, when language proximity was taken into account, the findings showed that the distant language bilingual children looked equally at the eyes and mouth but that the close language bilingual children looked longer to the mouth than eyes. The different patterns of attention found in the two groups indicate that language proximity continues to play a role in attentional responsiveness to talking faces into early childhood.

Our results from the bilingual children provide new insights into the role of audiovisual redundancy in speech processing during development. Whereas bilingual infants in Study 2 attended more to a talker's mouth regardless of language, and regardless whether they were learning close or distant languages, bilingual children only did so when they were learning close language pairs. One reason why bilingual children did not exhibit a preference for the talker's mouth may be due to the fact that phonological development is largely established by 6 years of age (Bosch Galceran, 2004) and that children at this age may also be interested in overall facial expressions rather than just the social information located in a talker's eyes. In addition, it should be noted that children deploy equal attention to a talker's eyes and mouth spontaneously (i.e., in the absence of any specific task) (Byers-Heinlein et al., 2014; Nakano et al., 2010) but adults only show such pattern when their task requires them to explicitly process linguistic input (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998, or even a mouth preference in Barenholtz et al., 2016). This suggests that young children's attentional strategy is more like that of adults in that they only deploy greater attention to a talker's mouth when language processing is challenging (i.e., when they are learning close-language pairs). Indeed, the fact that our close-bilingual children group deployed more attention to the mouth than eyes is consistent with previous findings from Pons et al. (2018), where a typically developing control group – i.e. Catalan-Spanish bilingual children as well – also showed a preferential attention for the mouth. Interestingly, it is also consistent with a recent study that found that children with higher word recognition proficiency and higher average pupil response have an increased likelihood of fixating the mouth, indicating a stronger motivation to decode speech (Król, 2018). Together, these results support the idea that multisensory processing and integration are very much task-dependent processes (Murray et al., 2016).

In conclusion, the results from the present study show that linguistic distance plays an important role in mediating selective attention to talking faces in 4- to 6-year-old children. If Study 2 showed that close-language bilingual infants begin relying on the greater salience of redundant audiovisual cues to more easily disambiguate and separate the languages they are learning, the present results confirm that this pattern of attention is still present during childhood.

Study 4

Temporal dynamics of children's attention to a talking face

Introduction

Previous studies have shown that children deploy an equal amount of attention between a talker' eyes and mouth (Byers-Heinlein et al., 2014; Król, 2018; Nakano et al., 2010). The results of the previous study extend this evidence by showing that Catalan and Spanish bilingual children exhibit a greater reliance on the mouth audiovisual cues than distant bilingual children do. An important question that emerges after these results is: what is the main factor causing close bilingual children to still deploy most of their attention towards the mouth of a talker, when their monolingual counterparts no longer do so? As suggested in Study 3, one possibility is that albeit their high proficiency in both languages at this age, children still resource to the mouth's audiovisual speech cues to help them separate their two languages. On the other hand, it is also possible that close bilingual children exhibit such mouth-preference as a result of their early experience with learning and disambiguating their two close languages, which may modify or bias their later exploratory looking pattern of a talking face.

Crucially, previous studies have only used the average of looking time towards the face's different areas of interest (AOIs). Nevertheless, studying the temporal dynamics of children's selective attention to a talking face may help confirm one of these possibilities.

We expect that a temporal pattern consistent with language disambiguation should eventually show a decrease of attention to the mouth, similar to adults' progressive decrease in mouth-looking as they become familiarized with new artificial words (Lusk & Mitchel, 2016), and consistent with a perceptual adaptation process to speech, as shown in accented

or non-native speech perception (A. Bradlow & Bent, 2008; Clarke & Garrett, 2004). Alternatively, a rather stable pattern of selective attention over time would be more consistent with the hypothesis of an earlier-shaped face exploration strategy or mouth-bias.

Therefore, in the current study we used longer (60 s) monologues of a speaker talking in children's native and non-native language, and we explored the temporal dynamics of children's selective attention, with the aim of revealing their different processing strategies and/or use of the audiovisual speech cues. Moreover, to assess the modulatory effects of language background, we did so in monolingual and close bilingual children.

Method

Participants. A total of sixty-six 4- to 6-year-old children were tested. Participants were recruited from three different schools, two located in a Catalan-Spanish bilingual environment in Barcelona, and one located in a Spanish-monolingual environment in Madrid, Spain. None of the children had a history of hearing problems according to parental report. Parents completed an online language questionnaire to establish the language background of the participants. Participants were classified accordingly in two groups; Spanish monolingual children and Catalan-Spanish early sequential bilingual children. We defined early sequential bilinguals as those bilingual children that came from Catalan or Spanish monolingual homes and had been exposed to their second language early in life (i.e., before entering school). After entering school at 3 years of age, participants also had some exposure to English as a third language, according to the Spanish study program.

Ten children were tested but not included in the final data analysis because they had had exposure to another language at home that was not Catalan or Spanish (2) or they failed to properly calibrate the nine fixation points (8). None were excluded in base of their minimum looking times – i.e. 20% minimum per trial (Frank et al., 2012) –, since all children had above 30% of eye tracking signal. Thus, the final sample consisted of 56 children (Mean age = 5 years, 8 months, Range = 4 years, 2 months - 6 years, 9 months) of whom 28 were early sequential bilinguals (Spanish-Catalan, Mean age = 5 years, 11 months, Range = 5 years, 5 months - 6 years, 6 months, 11 boys) and 28 were monolinguals (Spanish, Mean age = 5 years, 8 months, Range = 5 years, 5 months - 6 years, 6 months, 15 boys).

Stimuli. We used the same recorded material from Study 1, only that this time we used the full-length videos (60 s) in order to evaluate gaze evolution across time. Moreover, this allowed to have the same speaker across the native and non-native language conditions. As earlier mentioned, the video clips consisted of a 21-year-old Catalan-Spanish-English

trilingual female actor (i.e. English father and Catalan-Spanish bilingual mother) who was filmed from her shoulders up and who spoke in a natural voice while she kept her head still. The recording took place in a soundproof booth, where the actor was recorded speaking a set of three short popular children's stories in Catalan, Spanish and English, respectively. The average length of each story was 57.3 s.

Procedure. The procedure, software and hardware used were identical to Study 3, with the exception that in the current study each participant watched three video clips, one in Catalan, one in Spanish and one in English. The order of the videos was counterbalanced across children. While the children watched the videos, the eye-tracker monitored their gaze to the same two AOIs, namely the eyes and the mouth as in Study 2 and 3.

Results

Average PTLT Scores from the initial 10s

First, in order to compare the results of the current experiment to those of Study 3 – where video trials were 10 s long – we analyzed the first 10 s of the native (or dominant) and the non-native language trials. Identical to Study 3, we analyzed the averaged PTLT to the eyes and mouth by way of a mixed ANOVA, with AOI (eyes, mouth) and Test Language (native, non-native) as within-subjects factors and Group (monolingual, sequential bilingual) as the between-subjects factor. The results of the ANOVA revealed an effect of AOI [$F(1,54) = 13.12$, $p < .001$, $\eta_p^2 = .20$], an interaction of Group x AOI [$F(1,54) = 4.78$, $p = .033$, $\eta_p^2 = .08$] and a triple interaction between Group, AOI and Test Language [$F(1,54) = 4.03$, $p = .050$, $\eta_p^2 = .07$]. The AOI main effect reflects an overall preference for the mouth. Then, the paired t-tests after the AOI x Group interaction showed that the monolingual group looked equally at both AOI [$t(27) = 0.86$, $p = .395$, $d = .16$], whilst the bilingual group preferred the mouth over the eyes [$t(27) = 5.21$, $p < .001$, $d = .98$]. To understand the triple interaction between Group, AOI and Test Language (illustrated in Figure 8) we conducted two mixed ANOVAs, one per language group. The monolingual group ANOVA showed a significant effect of Test Language x AOI [$F(1,27) = 6.31$, $p = .018$, $\eta_p^2 = .19$] whilst the bilingual group only showed the AOI main effect [$F(1,27) = 27.15$, $p < .001$, $\eta_p^2 = .50$]. Together, the results of the two ANOVAs indicate that the significant triple interaction was due to the fact that only the monolingual children exhibited a "non-native effect" – i.e. greater attention to the mouth in the non-native language. However, further paired t-tests inside the monolingual

group comparing eyes *vs.* mouth in the native and non-native conditions revealed that in fact, they looked equally at both AOIs in the two conditions [native: $t(27) = .13$, $p = .899$, $d = .02$; non-native: $t(27) = 1.48$, $p = .151$, $d = .28$], and hence the interaction was due to the fact that they looked more to the mouth in the non-native condition as compared to native one.

These results were equivalent to those previously observed in Study 3, demonstrating again a greater reliance on the audiovisual speech cues in close bilingual children. Moreover, the monolingual group attended equally to the eyes and mouth, analogous to the distant bilingual children results from Study 3. Subsequently, in order to unveil the underlying factors behind the different patterns of attention we proceeded to explore the temporal dynamics of selective attention to the talker's eyes and mouth.

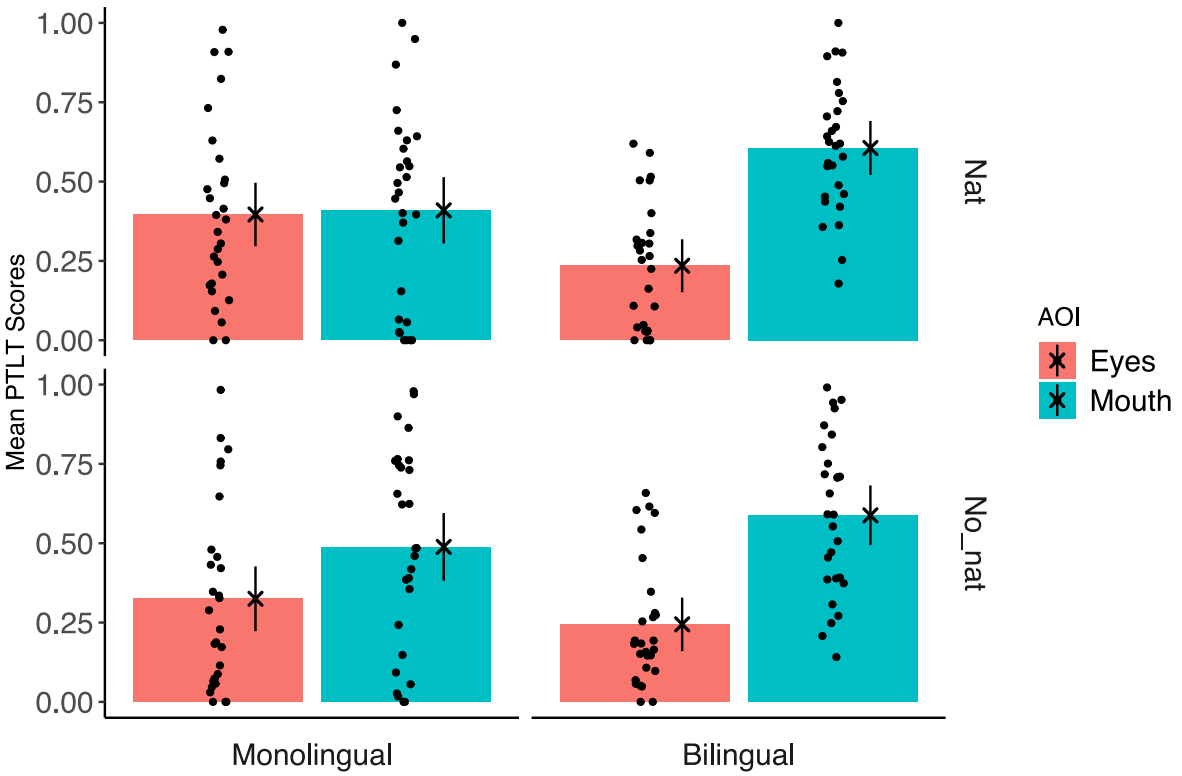


Figure 8. Distribution of mean proportion-of-looking-time (PTLT) scores to the eyes and mouth AOIs of the first 10 seconds of the video, as a function of Linguistic Background (Bilingual and Monolingual) and Test Language (Native, Non-native). Dots represent each child's Mean PTLT Score; bars and crosses with error bars represent Mean PTLT Score and Standard Error of the Mean (SE) for each group.

Time Course analysis from the total of 60s

After confirming the expected differences between monolingual and close bilingual children, we proceeded to the main analysis of interest; the time course of selective attention. We used a growth curve analysis (Mirman, 2014) and a binomial logistic mixed-effects model (Generalized linear mixed model (GLMM) with the R package lme4 version 1.1-21, function glmer) to analyze children's fixation data during the 60s of the trials. Instead of the previously used PTLT scores to the eyes and mouth, here we chose to use the difference score values (PTLTeyes - PTLTmouth) for model simplicity and to avoid auto-correlation (when PTLTeyes increases PTLTmouth decreases and vice versa). The overall time course to the two AOIs was modeled with a third-order (cubic) orthogonal polynomial, with the same fixed effects as in previous analysis: Test Language (native, non-native) and Group (monolingual, bilingual) on all time terms (intercept, linear, quadratic, and cubic). The model also included participant random effects on all time terms except the cubic[7]. Native language and Monolingual children were used as the baseline in the model, and relative parameters were estimated for the Bilingual group and Non-native conditions.

Table 1 shows the forward building of the model; a comparison of each model fit with the same model plus one other variable. In this way, each variable is added one by one, and only if it significantly improves the model. The results showed that the full model – containing linear, quadratic and cubic time polynomials and the two fixed effects plus all interactions – fitted the best, without compromising its convergence[8]. The model was coded in R as: [full model <- PTLT ~ (time + time^2 + time^3) * Group * Test Language + (time + time^2 | Participant)].

[7] Estimating random effects is "expensive" in terms of the number of observations required, so this cubic term was excluded because it tends to capture less-relevant effects in the tails.

[8] By the principle of marginality, a factor must be kept if the interaction is significant, regardless of the main effect. Moreover, changing the order of the comparisons – maximal and drop1 or forward (Barr, Levy, Scheepers, & Tily, 2014)- yielded the same results (i.e. keeping the maximal model).

Table 1. GLM models' forward comparisons statistics.

term	df	AIC	BIC	logLik	deviance	statistic	Chi.Df	p	p<.05
PTLT base mod	2	89,798	89,817	-44,897	89,794.28	NA	NA	NA	NA
+Test Language	3	89,536	89,564	-44,765	89,529.59	265	1	<0.0001	***
+Group	4	89,536	89,573	-44,764	89,527.62	2	1	0.16	
+Test Lang x Group	5	89,365	89,411	-44,677	89,354.53	173	1	<0.0001	***
+Time	11	87,591	87,694	-43,785	87,569.37	1,785	6	<0.0001	***
+Time^2	18	85,492	85,659	-42,728	85,455.69	2,114	7	<0.0001	***
+Time^3	22	85,215	85,420	-42,586	85,171.16	285	4	<0.0001	***

Note. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2 and Figure 9 summarize the results of the statistical model. The significant main effect of Test Language reflects an overall increased mouth looking in the non-native condition [Estimate = - 0.47 (0.03), z = - 18.01, p < .001]. However, the lack of Group main effect reflects the fact that, when analyzing the full 60 s of the trial, the bilingual children greater mouth looking is no longer significant [Estimate = - 0.96 (0.52), z = - 1.83, p = .07].

When considering the time polynomials, the results showed a main effect of the cubic term, reflecting a general steep initial decline of mouth-attention, together with a rise of eyes-attention, across the two groups and conditions [Estimate = - 0.15 (0.07), z = - 2.14, p= .03]. Last, the triple interaction between the three time terms, Group and Test Language [Estimate = - 0.92 (0.12), z = - 7.58, p < .001; Estimate = - 0.56 (0.12), z = - 4.53, p < .001; Estimate = - 1.58 (0.12), z = - 13, p < .001, respectively] indicate that the characteristics of the curve are significantly different between the two groups and language conditions.

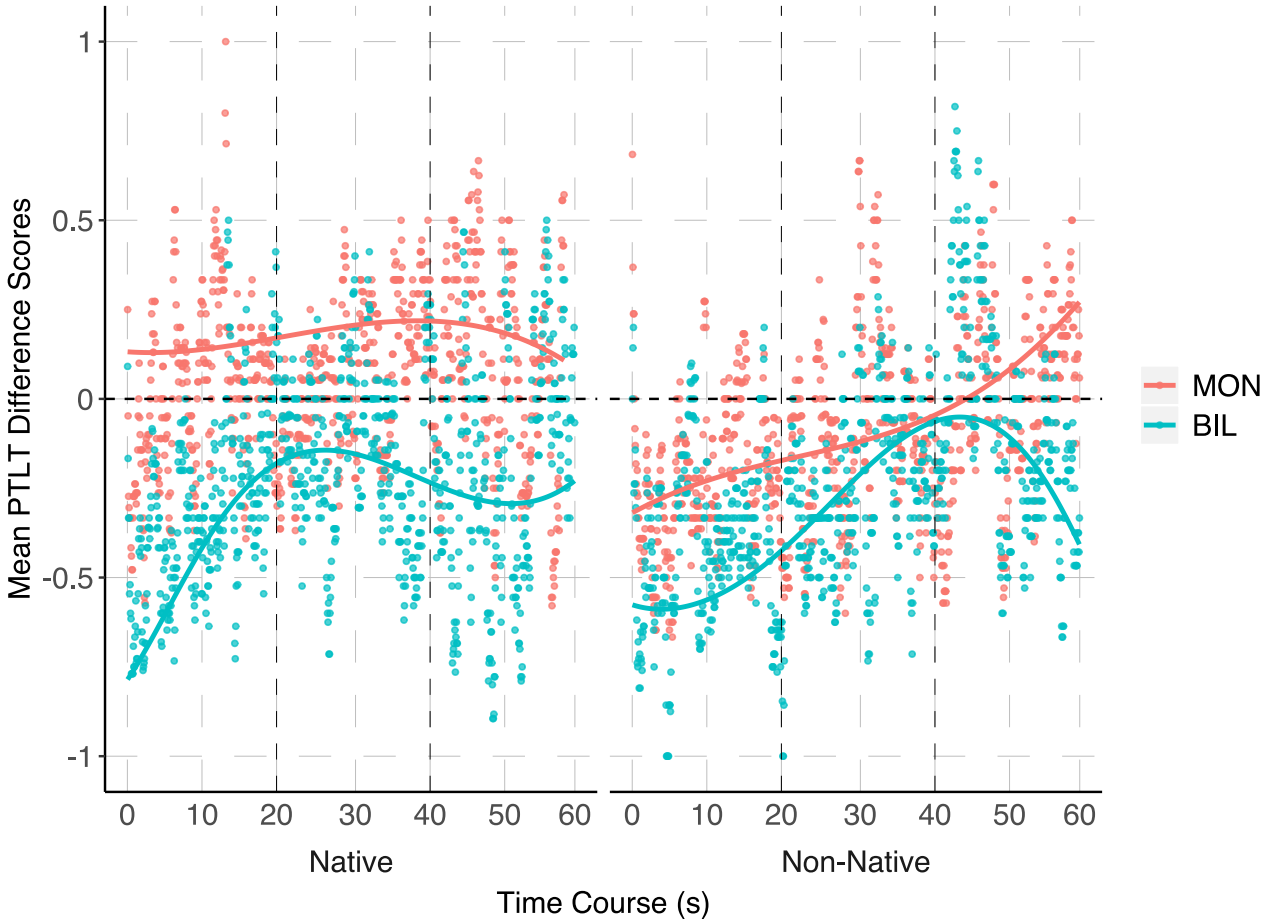


Figure 9. Time course graphs for each group of participants' Mean Difference Score (i.e. PTLTeyes-PTLTmouth) during the 60 sec length of the Native and Non-native conditions. The Monolingual group is represented in red, the Bilingual group in blue. The lines represent the fitted GLMM including time up to cubic term for each condition. Dots represent each group Mean Difference Score for each timepoint (i.e. every 60ms).

Visual inspection of the data (Figure 9) suggests that within the native language condition, bilingual children initially fixate more on the mouth of the talker and slowly shift to equally looking to the two AOIs, whereas the monolingual group present a rather flat time course of attention. Differently, in the non-native language condition, both monolingual and bilingual children seem to perform a similar pattern of initial attention to the mouth and later decrease to more distributed attention, although with overall greater attention to the mouth in the bilingual group [Estimate = 0.41 (0.04), $z = 11.69$, $p < .001$].

Table 2. Summary of fixed effects of the full GLM model.

	Estimate	Std. Error	z value	p	p<.05
(intercept)	0.32	0.37	0.84	0.4000	
time	0.00	0.26	-0.01	0.9954	
time^2	-0.26	0.37	-0.71	0.4801	
time^3	-0.15	0.07	-2.14	0.0325	*
Group (Bilingual)	-0.96	0.52	-1.83	0.0679	.
Test Lang (Non-native)	-0.47	0.03	-18.01	<0.0001	***
time x Group (Bilingual)	0.72	0.35	2.06	0.0395	*
time^2 x Group (Bilingual)	-0.59	0.51	-1.16	0.2465	
time^3 x Group (Bilingual)	0.81	0.09	9.20	<0.0001	***
time x Test Lang (Non-native)	1.10	0.09	11.93	<0.0001	***
time^2 x Test Lang (Non-native)	0.51	0.09	5.40	<0.0001	***
time^3 x Test Lang (Non-native)	0.29	0.09	3.15	0.0016	**
Group (Bil) x Test Lang (No-nat)	0.41	0.04	11.69	<0.0001	***
time x Group x Test Lang	-0.92	0.12	-7.58	<0.0001	***
time^2 x Group x Test Lang	-0.56	0.12	-4.53	<0.0001	***
time^3 x Group x Test Lang	-1.58	0.12	-13.00	<0.0001	***

Note. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

PTLT Scores averaged in three temporal windows (0-20s, 20-40, 40-60s)

Last, to further explore these differences over time and to allow for an easier comparison with the previous studies – that have used ANOVAs onto averaged looking time – we divided the timeline into three 20 s time bins and analyzed the PTLT data by way of a mixed ANOVA, including AOI (eyes, mouth), Test Language (native, non-native) and Time bins

(1,2,3) as within-subjects factors and Group (monolingual, sequential bilingual) as the between-subjects factor.

The results yielded a significant Test Language main effect [$F(1,51) = 7.14$, $p = .010$, $\eta_p^2 = .12$] a Test Language x AOI interaction [$F(1,51) = 4.22$, $p = .045$, $\eta_p^2 = .08$], and a Time bin x AOI interaction [$F(2,102) = 13.87$, $p < .001$, $\eta_p^2 = .21$]. The Test Language main effect reflects a greater overall attention in the native language condition. Next, the interaction of AOI with Test Language is due to the fact that children overall attended more to the mouth of the talker in the non-native language condition (see Figure 9). The interaction of AOI with Time bins shows that children's attention deployment differs across the time windows. Further t-tests first showed that overall, children attended equally at both AOIs in the native language condition [$t(52) = .92$, $p = .364$, $d = .13$], whereas they deployed more of their attention to the mouth in the non-native condition [$t(52) = 2.13$, $p = .038$, $d = .29$]. Last, paired t-tests after the AOI x Time bins interaction showed a mouth-preference in the first 20 s [$t(52) = 2.86$, $p = .006$, $d = .39$] and equal looking at eyes and mouth at the second and third time bins [2nd: $t(52) = 1.17$, $p = .249$, $d = .16$; 3rd: $t(52) = .31$, $p = .760$, $d = .04$].

Discussion

The results from the current study show that children's allocation of attention to a talking face is a highly dynamic process that changes over time, and that this change is dependent on both the language of the speaker and children's language background.

First, when analyzing the average of the initial 10 seconds of the trial, the results showed that as expected, close bilingual children exhibited greater mouth looking than monolingual children did. This is consistent with our results from Study 3 – i.e. increased attention to the mouth in close than in distant bilingual children – and with the idea that learning two close languages increases attention to the mouth of a talker, not only in infancy (Study 2) but also during the linguistically advanced age of 5 to 6 years.

Then, instead of analyzing averaged PTLT values as in previous studies, we analyzed the full 60 s of the trial and introduced time as a continuous variable, which was modelled with a cubic-term polynomial. Crucially, the results of this temporal analysis provided us with some new insight into the possible causes behind children's differential pattern of selective attention to a talking face.

First, the analysis revealed that bilinguals' greater attention to the talker's mouth is not present throughout the 60 s length of the trial, but instead is rather restricted to the initial phase, between the first 20 and 30 seconds. This pattern of attention fits well with the interpretation that close bilingual children initially rely more on the mouth redundancy for

aiding in the task of language disambiguation, and that as time goes by and language ambiguity decreases, their attention pattern becomes more similar to that of a monolingual child, that is, equally distributed between the eyes and mouth.

Furthermore, the fact that monolingual children also performed such attention temporal pattern when perceiving a non-native language – i.e. initial mouth preference followed by the decrease to a balanced pattern between the two AOIs – suggests that this pattern is not limited to close bilingual children, but that it may be a general pattern of selective attention to perceptually adapt to a new language, accent or speaker. In this manner, the mouth redundant cues would augment speech perception in the initial phase, where it contains higher ambiguity or uncertainty. Then, as the perceiver becomes more adapted to the language and speaker characteristics, the preference for the mouth withdraws and attention is again more distributed. This idea is also consistent with previous evidence showing a similar decrease of attention to the mouth when adults become familiarized with an audiovisual task (Lusk & Mitchel, 2016) or with a perceptual adaptation to foreign speech (A. Bradlow & Bent, 2008; Clarke & Garrett, 2004). It may be that bilingual children always perform such "adaptation pattern" since they are often unsure of the incoming language, whereas monolingual children only exhibit such pattern when they detect speech in a different language from their native one.

Last, the fact that bilingual children's curve of attention from the mouth to the eyes peaked earlier in the native language than in the non-native (at around 20 s in the native and 40 s in the non-native language condition, see Figure 9) suggests that the velocity of the decrease is proportional to the difficulty of the processing task, that is, the more difficult the adaptation, the longer the mouth preference period. Remarkably, this study allows us to report in detail the comparison "native - non-native" speech, since we used a trilingual speaker and therefore the speaker is constant across the test languages (unlike previous studies that used different speakers).

In sum, the results from this study are the first evidence to support the idea that children's temporal dynamics of selective attention to a talking face reflects a general strategy for perceptually adapting to the speech they perceive. In the case of close bilingual children, this includes disambiguating their two languages and hence this adaptive pattern is shown in both their native and the non-native language, whilst in the monolingual children it only shows when perceiving a non-native language.

Chapter 6

Language Factors Modulate Selective Attention to a Talking Face

Evidence from Adult Participants

Overview

The previous studies 1 and 2 added further insight to the earlier reviewed evidence on infants' perception of audiovisual speech by showing that language disambiguation conveys a greater challenge for those infants learning two closely related languages, as compared to monolingual and distant bilingual infants. This is illustrated by the fact that 1) they do not detect a switch between their two languages presented audiovisually at 4 months of age, when their monolingual peers do (Study 1), and 2) they exhibit greater attention to the mouth of a talker than their bilingual peers learning two distant languages at the age of 15 months (Study 2).

In studies 3 and 4, these findings were extended to children by showing that 1) at the age of 5 to 6 years close bilingual children still show a stronger reliance on the mouth than their distant bilingual peers do (Study 3), and that 2) attention to the mouth in children is initially accentuated and slowly decreases over time until equal attention to the eyes and mouth is reached (Study 4). This temporal pattern of attention to the mouth of a talker supports the interpretation that children attend to the talker's mouth to help them disambiguate language and perceptually adapt to the characteristics of the speech they are hearing. Remarkably, this decrease of attention to the mouth is slower when perceiving non-native speech, which suggests that the greater challenge of perceiving non-native speech sounds requires more time before disengaging from the talker's mouth and distributing their attention over other parts of the face, such as the eyes.

If speech processing elicits greater attention to a talker's mouth in children, and it does so for a longer period of time when it is in a non-native language, then this raises an interesting question. Is it possible that adults might rely more on the audiovisual cues located in a talker's mouth when trying to comprehend non-native speech?

Indeed, as earlier reviewed in the introduction, adults' processing and comprehension of speech is enhanced by the audiovisual signal – *vs.* auditory only –, not only in the presence of noise (Sumby & Pollack, 1954) but also when perceiving non-native speech (Arnold & Hill, 2001; Reisberg et al., 1987). As a consequence, although adults usually look at their social partners' eyes (Yarbus, 1967), when they need to specifically process and/or disambiguate audiovisual speech, they also attend more to the talker's mouth to augment speech processing (Buchan et al., 2007; Lansing & McConkie, 2003; Lusk & Mitchel, 2016; Vatikiotis-Bateson et al., 1998; Võ et al., 2012), which includes performing a speech processing task in a non-native language (Barenholtz et al., 2016). However, Barenholtz et al. (2016) study explored non-native speech perception in inexperienced perceivers (i.e. naïve in the non-native language), and hence it remained to be known whether

participants' knowledge of the non-native language (i.e. second language proficiency) would modulate selective attention towards the talking face.

In chapter 6, we investigated whether the degree of second language proficiency modulates selective attention to the mouth of a talker speaking in that language. To do so, we conducted 3 experiments: first, Study 5 aimed to replicate Barenholtz et al.'s (2016) study with our set of languages and stimuli. Then, in Study 6 we investigated adults' selective attention to relatively longer and more natural fluent speech in a native and non-native language. Last, in Study 7 we directly examined the hypothesis that second language proficiency modulates selective attention to a talking face, by exploring participants with varying degrees of second language proficiency.

Study 5

Selective attention to a talking face whilst performing an *ABX* task using short sentences in native and non-native language

Introduction

As earlier described, Barenholtz et al. (2016) is the only study to date that has compared selective attention in a native vs. a non-native language, showing that participants deployed more attention to the mouth when performing a native-language speech processing task, and even more when performing such task in the non-native language. In sum, when faced with the greater difficulty of having to process unfamiliar audiovisual speech sounds, adults resorted to greater lipreading as they do when speech is presented in noise.

The purpose of this experiment was to extend the Barenholtz et al. (2016) study but with different stimuli and languages than those used in that study. Moreover, we used a crossed-languages design, that is, we presented individuals from a Spanish-speaking and an English-speaking community with a talker speaking in Spanish and English.

Method

Participants. A total of 40 subjects participated in this study. Of these, 20 subjects were native Spanish speakers who were students at the University of Barcelona and 20 were native English speakers who were students at Northeastern University in Boston. The students

participated in the study for course credit. Two more subjects were run but excluded due to technical issues with the eye-tracking equipment (n=2). All subjects self-described as having no or very little knowledge (max. A2 level, *Common European Framework of Reference for Languages*) of the non-native language.

Stimuli. As in the Barenholtz et al.'s (2016) study, the stimulus materials consisted of short video clips of a Spanish-English bilingual female actor who was filmed from her shoulders up and who spoke in a natural voice while she kept her head still. The actor was recorded speaking 20 different Spanish and 20 different English sentences. The average length of each sentence was 2.5 s.

Apparatus and procedure. Participants were tested in a quiet laboratory either at the University of Barcelona or at Northeastern University. Selective attention was measured with a REDn SensoMotoric Instruments (SMI, Teltow, Germany) eye tracker running at a sampling rate of 60 Hz. The participants sat at a table with a Dell Precision m4800 laptop computer in front of them at a distance of 60 cm from their eyes. The eye tracker camera was attached to the bottom of the computer screen and SMI's iViewRed software controlled the camera and processed eye gaze data. SMI's Experiment Center software controlled the stimulus presentation and data acquisition. The video clips were presented on the computer's 11 x 13 in screen and the soundtrack corresponding to the videos was presented through a pair of Sony headphones which participants wore throughout the experiment. We used a 9-point calibration routine to calibrate eye gaze by presenting a small yellow star in the center of the screen as well as in the 4 corners of the screen and the 4 midpoints between the corners and the center of the screen.

The experiment began with a training trial where two videos showing the actor uttering different sentences were presented in turn followed by an audio-only clip of one of the two previous sentences. Participants had to choose which of the two sentences that they saw and heard corresponded to the audio-only clip by pressing a key on the keyboard. Once they understood the procedure, we proceeded to the test phase, which was divided in two blocks of ten pairs of sentences. The order of language presentation (i.e. familiar or unfamiliar first) was counterbalanced across participants. That is, half the participants were presented with 10 familiar-language sentence pairs first followed by 10 unfamiliar-language sentence pairs while the other half were presented with the same sentences but in reverse order. To control for any possible language-specific effects, the same sentences were familiar for one group of participants and unfamiliar for the other group.

Results

Consistent with Barenholtz et al. (2016) we defined three areas of interest (AOIs): the mouth, the eyes, and the face. Then, we calculated the proportion of total looking time (PTLT) deployed to the eyes and mouth, respectively, by dividing the amount of time spent fixating the eyes and the mouth, respectively, by the total amount of fixation of the face. We then used a mixed, repeated-measures analysis of variance (ANOVA) to analyze the PTLT scores, with Language Group (Spanish, English) as a between-subjects factor and Language (native and non-native) and area of interest (AOI; eyes and mouth) as within-subjects factors. The results revealed a main effect of AOI [$F(1,38) = 6.30$, $p = .016$, $\eta_p^2 = .142$], reflecting an overall preference for the eyes. The results also yielded an AOI x Language interaction [$F(1,38) = 8.97$, $p = .005$, $\eta_p^2 = .191$, reflecting the fact that the amount of attention deployed to the eyes and mouth, respectively, differed for the two Language Groups (Figure 10).

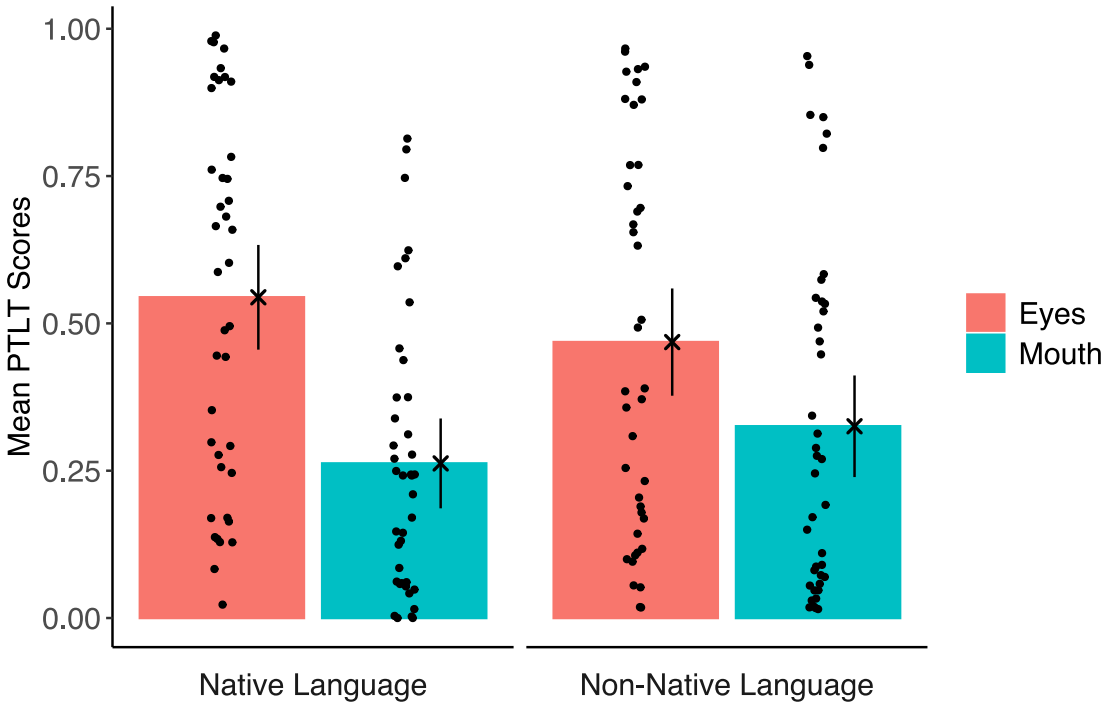


Figure 10. Distribution of mean proportion-of-total-looking-time (PTLT) scores to the eyes and mouth as a function of Test Language (Native and Non-native) collapsed across Spanish and American participants. Dots represent individual Mean PTLT Scores; Bars and crosses with error bars represent Mean PTLT scores and standard error of the mean (SE) for each group.

Follow-up paired t-tests compared the eyes and the mouth PTLT scores from the native condition against the non-native one. As expected, results revealed that attention to the eyes decreased [$t(39) = 2.767$, $p < .01$, d = .439] and attention to the mouth increased [$t(39) = 2.649$, $p = .012$, d = .460] in the non-native language condition compared to the native one. Finally, paired t-tests comparing PTLT to the eyes against PTLT to the mouth in each of the languages separately showed an eyes preference in the native language condition [$t(39) = 3.481$, $p < .01$, $d = 1.43$], and equivalent looking to eyes and mouth in the non-native language condition [$t(39) = 1.529$, $p = .134$, $d = .656$].

Discussion

These results replicate the main finding from Barenholtz et al. (2016) demonstrating again that participants deployed greater attention to the mouth when exposed to the non-native language video than when exposed to the native one. Consistent with Barenholtz et al.'s (2016) interpretation, the current results can be interpreted as reflecting adults' greater reliance on audiovisual cues emanating from a talker's mouth cues when the speech processing task is more difficult because the speaker is using a non-native language to produce the speech utterance. Crucially, it should be noted that this effect is independent of the type of language spoken or the specific person speaking because these two factors were counterbalanced across the two language groups tested here.

Finally, it is interesting to note that, overall, we obtained greater attention to the talker's eyes whereas Barenholtz et al. (2016) obtained greater attention to the talker's mouth in a task that is similar across the two experiments. The most likely reason for this difference is the fact that we used different stimuli. Despite this difference, however, the findings of principal interest, namely those reflecting differential processing of native vs. non-native audiovisual speech, were consistent across the two studies: participants deployed more attention to the talker's mouth when they were exposed to a talker speaking in a non-native language than in their native language.

Study 6

Selective attention to a talking face uttering passages in a native and a non-native language

Introduction

Both in Study 5 and in the Barenholtz et al. (2016) study, participants were required to encode the short audiovisual speech sentences (3 s long) presented during the first phase of the experiment and subsequently asked to perform a simple match-to-sample task. Hence, the combination of the speech processing task and the short length of the stimuli make it reasonable for participants to rely more on the mouth audiovisual cues when the presented language is non-native. However, whether adults will also rely on the mouth redundancy cues when perceiving longer, fluent speech without having to perform any speech processing task remains an open question.

To test this hypothesis, we presented adults with relatively extended, fluent speech utterances (60s long) in the participants' native and non-native languages. Moreover, we counterbalanced subjects' native language by conducting the experiment in Spain and in the US. This enabled us to explore the effect of a non-native language on the deployment of selective attention to a talker's eyes and mouth, independent of the specific language in which the speech was uttered. Finally, even though our participants were not given a specific speech-processing task, they were told that they would first see and hear some audiovisual speech utterances and that they would then be given some questions related to these utterances at the end of the experiment.

Method

Participants. A total of 45 adults participated in this study. Of these, 22 were native Spanish and Catalan bilingual speakers who were students at the University of Barcelona and 23 were native, monolingual, English speakers who were students at Northeastern University in Boston. The students participated in the study for course credit. All participants self-described as having no or very little knowledge (max. A2 Level, *Common European Framework of Reference for Languages*) of the non-native language.

Stimuli. Identical to Study 4, the stimulus materials consisted of a Catalan-Spanish-English trilingual female actor who was filmed from her shoulders up and who spoke in a natural voice while she kept her head still. The actor was recorded speaking a set of 3 short children's stories in Catalan, Spanish and English, respectively. The average length of each story was 57.3 s. It should be noted that the population in Barcelona is bilingual, meaning that people are native speakers of both Catalan and Spanish. Consequently, these two languages were presented in the experiment as native for the Spanish group and non-native for the English group.

Apparatus and procedure. We used the same hardware and software as described in Study 5. Once the eye tracker calibration was completed, we presented three videos to each participant. These consisted of videos in which the actor could be seen and heard speaking in Catalan, in Spanish, and in English, respectively. Participants were given the following instructions: "You are going to watch a woman telling you three different short stories, in three different languages. Please listen carefully because I will ask you some questions about the stories you heard". These instructions were only given to ensure that participants were fully engaged in the experiment. The order of the videos and the specific stories were assigned randomly and counterbalanced across participants. Additionally, using a crossed design between the Spanish and the American participants eliminated any possible language-specific effect and constrained the effects to language familiarity per se.

Results

First, to ensure that the Spanish and American participants did not respond differently to the Catalan and Spanish videos, first we used a repeated-measures analysis of variance (ANOVA), with Language Condition (Catalan and Spanish) and area of interest (AOI; eyes and mouth) as within-subjects' factors, to analyze the PTLT scores in each participant group,

respectively. The ANOVA of the Spanish participants' data yielded an AOI main effect [$F(1,21) = 5.98$, $p = .023$, $\eta_p^2 = .222$], indicating greater overall looking at the eyes. Crucially, the Language Condition x AOI interaction was not significant [$F(1,21) = 1.75$, $p = .200$, $\eta_p^2 = .077$], indicating that the Spanish participants looked more at the eyes in both language conditions. The ANOVA of the American participants' data did not yield a significant AOI effect [$F(1,22) = .78$, $p = .386$, $\eta_p^2 = .034$], indicating that the American participants looked equally to the two AOIs. Also, like the Spanish participants, the American participants exhibited the same pattern of selective attention to the eyes and mouth across the two language conditions (Language Condition x AOI interaction [$F(1,22) = 2.18$, $p = .154$, $\eta_p^2 = .090$]). Given that responsiveness to the Spanish and Catalan videos did not differ in either group, we only used the data from the Spanish video condition for the main analysis (a supplementary analysis of responsiveness in the Catalan video condition yielded results that were identical to those from the Spanish video condition). Overall, the native-language condition was Spanish for the Spanish participants and English for the American participants while the non-native language condition was English for the Spanish participants and Spanish for the American participants. This enabled us to both simplify the design to one native and one non-native language condition–similar to the design in the two previous studies (Barenholtz et al., 2016; Lewkowicz & Hansen-Tift, 2012) – and to then make a balanced comparison between the Spanish and American participants.

Next, we analyzed the data from the native and non-native language conditions for both groups of participants as defined above. To do so, we used a mixed, repeated-measures ANOVA, with Language Group (Spanish, English) as a between-subjects factor and Language Condition (native and non-native) and AOI (eyes, mouth) as within-subject's factors. Results revealed a main effect of AOI [$F(1,43) = 9.27$, $p < .001$, $\eta_p^2 = .177$] and an AOI x Language Condition interaction [$F(1,43) = 46.29$, $p < .001$, $\eta_p^2 = .518$]. Figure 11 shows these two statistically significant findings. As can be seen, even though participants exhibited an overall preference for the eyes, they deployed their selective attention to the eyes and mouth differently depending on whether the actor spoke in a native or non-native language. Follow-up t-tests, comparing the PTLT to the eyes and mouth, respectively, across the native and non-native language conditions revealed that participants attended less to the eyes in the non-native language condition [$t(44) = 6.35$, $p < .01$, $d = .95$] and that they attended more to the mouth in the non-native condition [$t(44) = 6.41$, $p < .01$, $d = 1.07$]. Paired t-tests comparing PTLT to the eyes and mouth within each of the language conditions, respectively, indicated a preference for the eyes in the native condition [$t(44) = 5.63$, $p < .01$, $d = 2.00$] and equal attention to the eyes and mouth in the non-native condition [$t(44) = .70$, $p = .49$ $d = .277$].

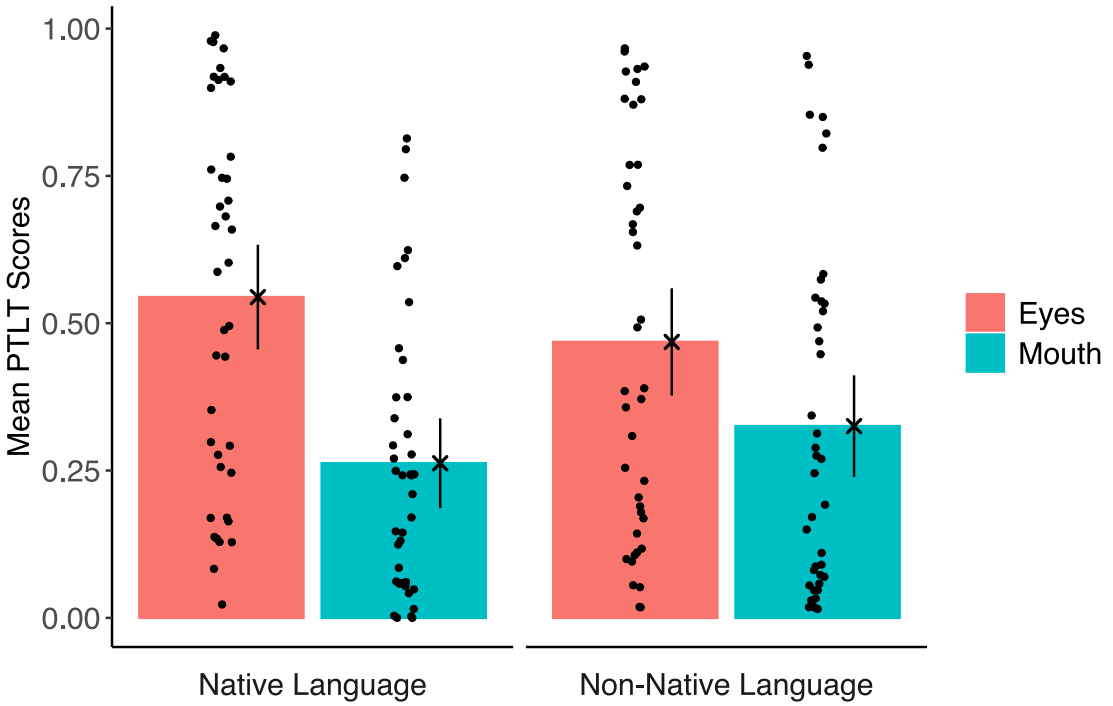


Figure 11. Distribution of mean proportion-of-looking-time (PTLT) scores to the eyes and mouth as a function of Test Language (Native and Non-native) collapsed across Spanish and American participants. Dots represent individual Mean PTLT Scores; Bars and crosses with error bars represent Mean PTLT scores and standard error of the mean (SE) for each group.

Discussion

The results from this experiment indicate that when adults are trying to comprehend 60 s-long, fluent audiovisual speech they exhibit differential patterns of selective attention to the talker's eyes and mouth as a function of whether the speech is in their native or non-native language. Specifically, adults attend more to the talker's eyes than mouth in the native language condition, whereas they deploy more attention to the mouth when the speech is not in their native language, resulting in equal attention to the eyes and mouth. This pattern of findings is consistent with evidence from speech-in-noise experiments showing that adults usually attend more to a talker's eyes except in the context of noise when they attend equally to the talker's eyes and mouth (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998). The current findings add to this evidence by showing that adults' strategy of deploying greater attention to a talker's mouth under challenging conditions includes the processing of long and fluent speech in an unfamiliar language. Specifically, our findings indicate that selective attention to different parts of a talker's face is modulated by adults' familiarity and, thus, prior experience with a specific language. When the speech is in a familiar language, adults direct most of their attention to the talker's eyes. This is presumably because their familiarity with their native language enables them to engage in relatively 'automatic' speech processing. In contrast, when the speech is in an unfamiliar language, adults deploy more of their selective attention to the talker's mouth. This enables them to take advantage of the greater perceptual salience of audiovisual speech, to help them overcome the greater challenge of trying to comprehend the message inherent in an utterance spoken in an unfamiliar language.

Importantly, the fact that the American and the Spanish participants exhibited the same pattern of attention in response to native and non-native speech suggests that these effects are not specific to English or Spanish but rather that they reflect a general feature of responsiveness to an unfamiliar language. Moreover, the lack of differences also indicates that participants' language background (i.e. bilingual vs. monolingual) did not affect their relative deployment of selective attention to a talker's eyes and mouth.

Study 7

Language proficiency modulation of selective attention to a talking face uttering passages in an L2

Introduction

The combination Barenholtz et al. (2016) and Studies 5 and 6 demonstrate that there is a robust difference in the pattern of attention to a talking face when adults perceive a native and a non-native language, that is, adults rely more on the mouth audiovisual cues when they are presented with non-native speech. Crucially however, participants in these studies had no or very little knowledge of the non-native language (under the A2 level), and hence they did not comprehend the non-native speech monologues. Consequently, the increased mouth-looking in face of non-native speech may reflect their intention to extract and understand some words, but it is not comparable to a more usual second-language social interaction. In such situations, the interlocutors normally have an intermediate or high level of the non-native language, and they must use all of their attention resources to try to understand as much content of the speech as possible. If indeed, adults allocate their attention dynamically towards the eyes and mouth of a talker depending on their present processing task difficulty, it is likely that their need to resource towards the talker's mouth will vary as a function of their level of proficiency in that second language.

Thus, the goal of Study 7 was to explore whether the proficiency level of the non-native language would modulate the reliance on the audiovisual speech cues of the mouth of a talking face. If language proficiency affects selective attention to a talker's eyes and mouth, then one plausible prediction is that highly proficient adult speakers of a second language might spend most of their time attending to a talker's eyes when the talker speaks in a native

language, as expected for native speaker. A second and equally plausible prediction is that adults who possess low or intermediate proficiency in a non-native language are likely to attend more to a talker's mouth, as has been shown in the two previous studies. However, if we consider that L2 learners only exceptionally attain native-like levels of speech perception performance (Lecumberri et al., 2010), a third prediction would be that the highly proficient L2 learners may still rely more on the mouth than native speakers do. To examine these predictions, in the present experiment we presented a video of a talker speaking in English to Spanish-Catalan bilinguals differing in the degree of language proficiency in a non-native language (i.e., English) and to monolingual native speakers of English and recorded their selective attention to the talker's eyes and mouth.

Method

Participants. We tested a total of 76 participants. The majority of the participants (n=57) were undergraduate students at the University of Barcelona. All of these students were native Catalan and Spanish bilingual speakers. The remainder of the participants were 19 undergraduate students from Northeastern University in Boston who were native English speakers. The Spanish participants were subsequently classified into three groups: 19 who were highly proficient in English (high B2 to a C2 levels of the *Common European Framework of Reference for Languages*), 19 who had an intermediate-level of proficiency (high A2 to a B1 levels), and 19 who had a low level of English proficiency (A1 to A2 levels [9]). When we first recruited the participants, we asked them to self-report their level of English, based on their previous official exams (i.e. Cambridge English tests, TOEFL, IELTS etc.). Once the participants completed the experiment, their English proficiency level was re-evaluated by administering the "Cambridge General English Placement Test". Three participants were excluded from the sample because their self-reported proficiency level and the level obtained with the English test did not match.

[9] As a reference of the English level of the students, the CEFRL B1 (Intermediate) level is defined as someone who can understand the main points of clear standard input on familiar matters, can deal with most travelling situations in that language, and can produce simple connected text on familiar topics and briefly give reasons and explanations for opinions and plans. The CEFRL C2 (highly proficient) level is defined as someone who can understand with ease virtually everything heard or read, can summarize information from different sources in a coherent presentation, and can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

Stimuli. We recorded three new videos that consisted of an American female speaker reciting 20s English monologues of everyday-life situations (including anecdotes and opinion pieces on social topics, 60s in total as in Study 6). The video characteristics were comparable to those presented in Study 6. That is, the actor was recorded from her shoulders up, her eyes and mouth size and position were similar to that in the videos presented in Studies 5 and 6, and she spoke in a natural voice while she held her head still.

Apparatus and procedure. The apparatus and procedure were identical to that in Study 6. The current experiment was conducted at the University of Barcelona and at Northeastern University. The laboratories in both locations were dimly lit and sound-attenuated.

Results

First, to ensure that the three pre-selected non-native English groups actually comprehended the stories according to their English level, we conducted an ANOVA on the post-test questionnaire scores to determine if they differed as a function of English proficiency level (low, intermediate, high). As expected, the results showed that the three groups differed in their performance [Low: M = .20, SD = .14; Intermediate: M = .54, SD = .15; High: M = .80, SD = .08, $F(56) = 98.92$, $p < .001$].

We then conducted the principal analysis whose purpose was to determine whether the three English proficiency groups differed in terms of their selective attention to the talker's eyes and mouth. We used a mixed, repeated-measures ANOVA, with Proficiency (low, intermediate, high) as a between-subjects factor and AOI (eyes and mouth) as a within-subjects factor to analyze the data. Contrary to expectations, the ANOVA yielded no significant effects, indicating that the three proficiency groups distributed their selective attention to the talker's eyes and mouth in similar ways. Due to the fact that visual exploration of the data (Figure 12) seems to display a mild reduction of attention to the mouth in the higher proficiency groups, we extracted each participant's 1) English Test Scores and 2) Post-viewing Comprehension Scores and tested the correlation of these scores and their PTLT difference scores. The Pearson Product Moment correlation yielded null results [r = .068, n = 57, $p = .615$, r = .10, n = 57, $p = .444$] and, thus, confirmed the results of the ANOVA (see Figure 13).

Finally, we collapsed the data for the three proficiency Spanish groups and compared their data to the data from the American group of participants for whom the talker spoke in their native language. For this comparison, we used a mixed, repeated-measures ANOVA, with Group (Spanish, American) as a between-subjects factor and AOI (eyes and mouth) as a within-subjects factor. Results yielded a significant AOI main effect [$F(1,74) = 11.21$, $p = .001$, $\eta_p^2 = .132$] and a significant AOI x Group interaction [$F(1,74) = 20.00$, $p < .001$, $\eta_p^2 = .213$]. The AOI main effect reflects an overall preference for the eyes while the significant interaction (see Figure 12) indicates that the distribution of selective attention depended on whether the language spoken was the participants' native language or a non-native one. To identify the source of the AOI x Group interaction, we first used paired t-tests to compare the PTLT eye versus mouth scores in each of the groups, respectively. Results revealed that the Spanish group looked equivalently to the two AOIs [$t(57) = 1.02$, $p = .31$] but that the American group looked more to the eyes than to the mouth [$t(18) = 7.93$, $p < .001$]. Finally, we used independent t-tests to compare attention to the mouth and eyes, respectively, across the two groups. Results confirmed that the non-native group looked less to the eyes [$t(74) = 4.46$, $p < .001$] and more to the mouth [$t(74) = 3.96$, $p < .001$] than the native group.

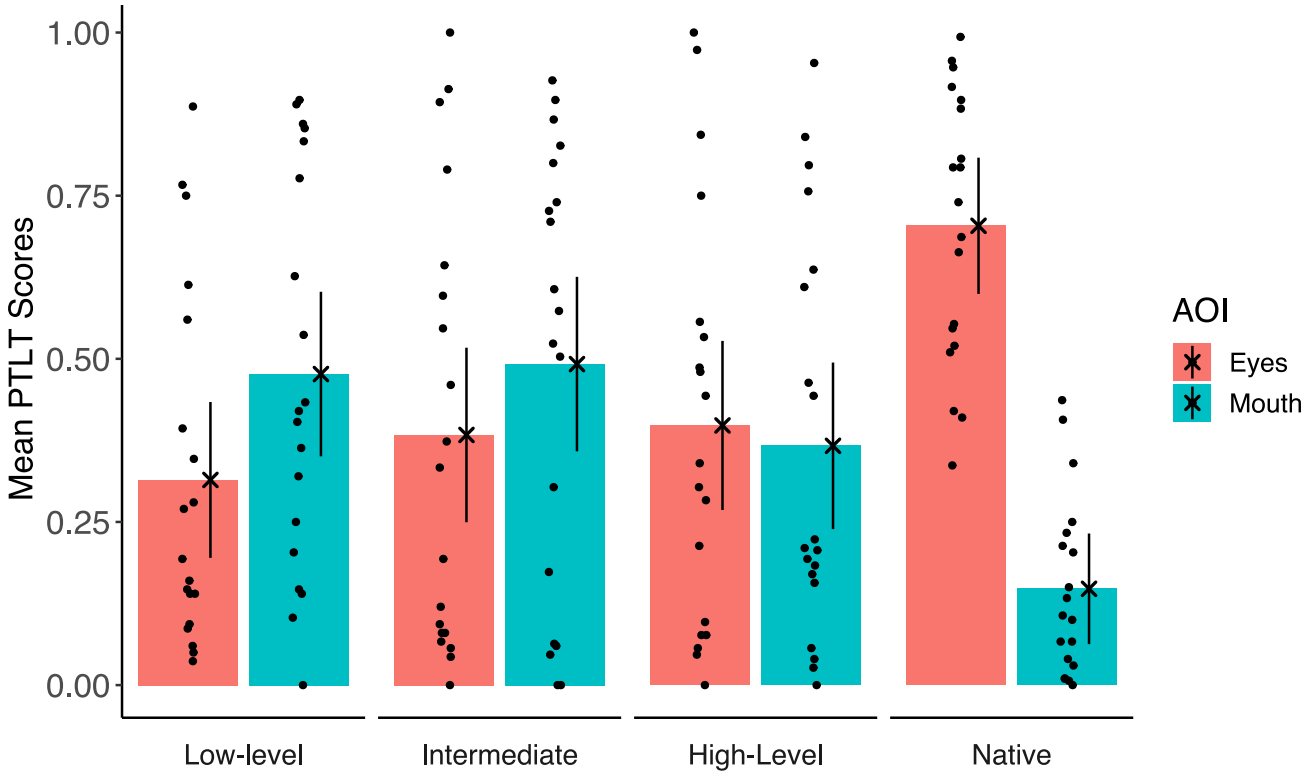


Figure 12. Distribution of mean proportion-of-looking-time (PTLT) scores to the eyes and mouth as a function of English Proficiency (Low-, Intermediate, High and Native levels). Dots represent individual Mean PTLT Scores; Bars and crosses with error bars represent Mean PTLT scores and standard error of the mean (SE) for each group.

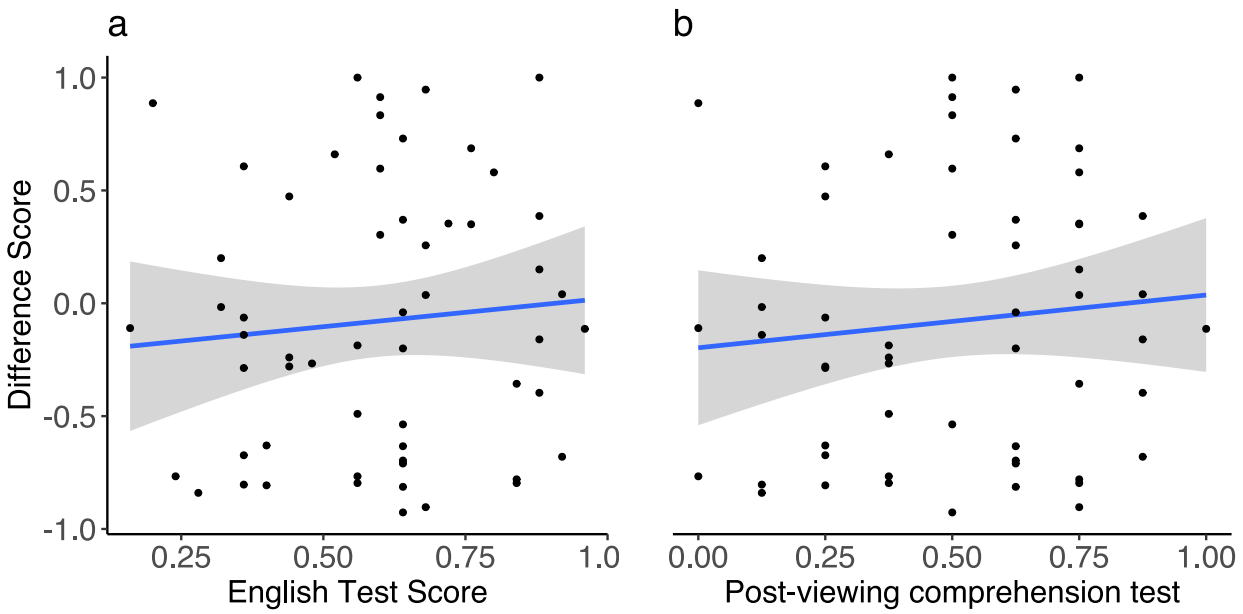


Figure 13. Correlation between the Difference Score (PTLTeyes - PTLTmouth) and (a) the English Test Scores, and (b) the Post-viewing comprehension test of non-native participants. Dots represent individual means, the line a fitted linear model and the shaded area represent standard errors of the mean.

Discussion

The results from Study 7 give support to our alternative prediction, that is, they indicated that the degree of non-native language proficiency did not affect the relative deployment of selective attention to a talker's eyes versus mouth in Spanish bilingual speakers tested with English audiovisual utterances. Interestingly, however, and consistent with the findings from Study 6, whereas native English speakers attended more to the talker's eyes than mouth, Spanish speakers attended equally to the talker's eyes and mouth regardless of their proficiency in English. Follow-up comparisons showed that the Spanish speakers attended less to the talker's eyes than the English speakers and that they attended more to the talker's mouth than did the English speakers.

If adults deploy greater attention to the mouth under challenging processing conditions, including the processing of non-native speech, it follows that the difficulty of the listening task might modulate the amount of attention directed to the mouth. Indeed, Vatikiotis-Bateson et al. (1998) found that adults' attention to the mouth increased continuously with the amount of noise (i.e. none, low, medium and high). Similarly, in an audiovisual speech segmentation task, Lusk & Mitchel (2016) found that attention to the mouth decreased as familiarization progressed and adults learned the new artificial words' boundaries. Based on such findings, we expected that participants' level of non-native language proficiency would modulate the amount of attention directed to the mouth. In other words, we expected that highly-proficient L2 learners of English would not need to rely on the audiovisual speech cues to the same extent as a speaker with lower proficiency. Accordingly, we expected highly proficient L2 speakers to exhibit a selective attention pattern similar to that found in native speakers. On the other hand, however, we also noted earlier that even highly proficient speakers differ from native ones in some crucial aspects of language perception such as phonology (McClelland, Fiez, & McCandliss, 2002), and hence we considered the possibility that the highly proficient group may still need to rely more on audiovisual redundancy.

Remarkably, the results of Study 7 are more consistent with the alternative prediction; that is, they show that albeit their significantly different levels of English competence and of speech's comprehension, the three non-native groups of participants exhibited similar patterns of selective attention and that, together, they differed in terms of their attention to the mouth with respect to the native-language group. As in Study 6, the non-native group exhibited equal attention to the eyes and mouth whereas the native-language group exhibited a clear preference for the eyes.

Although our results are also in line with the fact that increases in processing difficulty correspond with increases in selective attention to a talker's mouth, they also

suggest that this relationship is a non-linear one. That is, at least in the case of speakers with different levels of non-native language expertise, increasing expertise does not correspond with decreasing levels of selective attention to a talker's mouth. On the one hand, such results are consistent with previous evidence showing that adults' selective attention patterns to a talking face cannot be attributed to single attentional shifts to the mouth to disambiguate an ambiguous phoneme or a word that is difficult to understand (Vatikiotis-Bateson et al., 1998; Võ et al., 2012). Given this, it may be that participants' selective attention patterns are indeed rather macroscopic, and not sufficiently sensitive to the more subtle differences in processing capacity that speakers with different levels of non-native language proficiency may exhibit.

On the other hand, the fact that the highly proficient group clearly differed from the native group is consistent with second-language learning literature showing that the production and perception of L2 phonology is quite an arduous task for L2 learners. These studies show that learners' plasticity is limited, and that highly proficient L2 speakers rarely attain the ultimate phonological competence of native speakers (McClelland et al., 2002; Pallier, Bosch, & Sebastián-Gallés, 1997). Even when their speech recognition performance may be native-like, the addition of noise makes highly competent non-native listeners become less accurate than native speakers (Cutler, Garcia Lecumberri, & Cooke, 2008). Cutler et al (2008) study also illustrates the fact that L2 perception requires more cognitive effort (Borghini & Hazan, 2018), because the strategies used tend to be less efficient than those of native speakers. For example, in phoneme's discrimination, highly proficient L2 speakers sometimes focus on different – and less informative – formants than native speakers do (Iverson et al., 2003). Moreover, they rely less on contextual plausibility (Mattys, Carroll, Li, & Chan, 2010) due to the fact that their lexical and semantic knowledge is not as easily accessed (A. R. Bradlow & Alexander, 2007).

All in all, the combination of such findings with those of Study 7 suggests that even highly proficient participants find second language speech perception challenging, and hence they cannot engage in the earlier noted "relatively automatic speech processing" as do native speakers. Instead, they still need to rely more on the mouth speech cues – when they are available – for reassuring comprehension of the message in their second language, by means of audiovisual integration.

The current results provide the first evidence concerning the differences between adults' selective attention patterns to a talking face, as a function of their language background (i.e. native speakers vs. L2 learners), but not dependent from L2 language expertise. Once this result with averaged PTLT scores has been set, future studies should then explore the time course of L2 learners' attention when perceiving L2 speech. Taking into account the differences in L2 comprehension and also based on the results of Study 4,

it would follow theoretical logic that the lower-level participants exhibited a slower mouth-decrease than the higher-level participants, albeit not showing in the overall averaged scores.

Chapter 7

General Discussion

Overview

The collection of studies presented in this thesis were designed to shed new light into the topic of audiovisual speech perception and language processing by exploring the manner in which the audiovisual cues of a talker's face influence speech perception, and more specifically, the extent to which infants, children and adults attend and use these redundant cues under various linguistically diverse conditions. Specifically, the current studies tested the hypothesis of whether (1) the addition of the visual information of speech would modulate infants' language discrimination abilities, and whether (2) language factors such as the distance between bilinguals' two languages and the familiarity or proficiency with a language would influence perceivers' selective attention patterns to a talking face.

To do so, I performed seven experiments in which different aspects of attention to audiovisual speech were explored, in infants (Studies 1 and 2), children (Studies 3 and 4) and adult participants (Studies 5, 6, 7) summarized in Table 3. This last chapter comprises a summary of the main findings of the current work and a general discussion of the results, considering their contribution to the field of audiovisual speech perception. Finally, the limitations of the present work and future directions are discussed.

Table 3. Summary of the studies as a function of task and age group

Task	Factors	**4mo**	**15mo**	**4 - 6yo**	**Adults**
Av language discrimination		S. 1			
Selective attention to a talker's face	Language Distance		S. 2	S. 3 & 4	
	Language Familiarity		S. 2	S. 3 & 4	S. 5, 6 & 7

Summary of results

Study 1 explored monolingual and bilingual 4-month-old infants' capacity for discriminating languages in the audiovisual modality. As expected, we found that both groups detected when the talker switched from their native language to a distant language (English). However, when the switch involved a close language (from Catalan to Spanish or vice versa) only the monolingual group showed successful detection. Bilingual infants showed no attention recovery in the close language switch – between their two languages –, which was not correlated to their previous code-switching exposure. *It was concluded that seeing the constant face of the talker would hamper switch detection. Moreover, differential strategies for detecting a language switch between monolingual and bilingual infants are discussed.*

Study 2 investigated the influence of learning two close vs. distant languages onto bilingual infants' selective attention patterns to a talking face, speaking in their native in a non-native language. The results revealed that 1) overall, 15-month-old infants show a preference for the talker's mouth, 2) infants attended more to the mouth than eyes when the talker spoke in a non-native than native language, and 3) although both groups showed a preference for the mouth, close bilingual infants attended to it longer than distant bilingual infants did. *Close bilinguals' harder task of disambiguating between their languages is discussed, as a possible explanatory factor for their greater reliance on the mouth speech cues.*

Study 3 explored the same question in 4- to 6-year-old children. The results revealed that 1) overall, bilingual children looked equally at the talker's eyes and mouth, 2) there was no difference between perceiving native and non-native speech, and 3) distant bilingual children looked equally at the eyes and mouth whereas the close bilingual children looked longer to the talker's mouth. *The fact that bilinguals' language distance still influences selective of attention to a talker's face in children at this advanced age is discussed.*

Study 4 aimed at gaining insight into the underlying processes driving children's – distinct – selective attention patterns to a talking face. To do so, Study 4 analyzed the temporal dynamics of selective attention to a talking face in monolingual and bilingual 5- to 6-year-old children. The results showed that in the native language condition, monolingual children exhibited a balanced pattern of attention between the talker's eyes and mouth, which was constant across the trial length. Differently, close bilingual children exhibited an initial mouth preference that decreased with time until reaching equal attention to the eyes and mouth (~ 20 s until the video's completion). Concerning the non-native language condition, the results

showed that both monolingual and bilingual children exhibited a similar pattern – i.e. initial mouth attention followed by equal looking to the talker's eyes and mouth –, only that the initial mouth preference lasted longer (from ~ 40 s). *These results support the idea that children rely on the mouth speech cues as a general strategy for perceptually adapting to the speech they perceive.*

Study 5 investigated the influence of language familiarity onto adults' selective attention to talking faces. In this study, adults performed a speech processing task – identical to Barenholtz et al. (2016) – where they had to auditorily identify one of two 3s-long audiovisual utterances (ABX task), spoken in their native and in a non-native language. The results showed that 1) participants deployed greater attention to the mouth when exposed to the non-native language video than when exposed to the native one, and that 2) overall, participants deployed more attention to the talker's eyes than mouth. *It was concluded that participants rely more on the talker's mouth under the more challenging situation of identifying short snippets of speech in a non-native language.*

Study 6 explored again the modulation of language familiarity onto adults' selective attention to talking faces, but this time in longer (60s-long) more naturalistic audiovisual speech, without asking participants to perform any specific processing task other than attending to the speech. The results of this study demonstrated that 1) adults attended more to the eyes than mouth in the native language condition, and that 2) they deployed more attention to the mouth in the non-native language condition, resulting in equal attention to both areas. *This study demonstrated that the earlier conclusion – i.e. increased attention to the mouth may help perceive non-native speech – also applies to longer, more naturalistic unfamiliar speech perception.*

Study 7 evaluated the influence of second language proficiency onto the selective attention patterns to a talking face. In this study, low, intermediate and high-level learners of English as well as native speakers of English were tested with English audiovisual utterances. The results showed L2 learners attended equally to the talker's eyes and mouth regardless of their proficiency in English, whereas the native English group attended more to the talker's eyes than mouth. *This study indicated that the degree of non-native language proficiency does not modulate relative deployment of attention to the talker's face. The role of the mouth speech cues for enhancing speech processing together with the different demands that L2 speech processing involves are discussed.*

Integration of results

Language Discrimination

One of the first challenges that infants must face to successfully acquire language is the construction of a representation of the sound properties of their native language/s, which is specially challenging for those infants exposed to more than one language. Albeit the high complexity of the task, previous research shows that already at birth, both monolingual and distant bilingual newborns can recognize their language/s and discriminate them from other rhythmically distant languages (Byers-Heinlein et al., 2010; Nazzi et al., 1998). Differently however, in the case of bilingual infants learning a pair of rhythmically close languages, they require at least 4 months of experience and neural maturation to auditorily detect the switch between the two languages (Bosch & Sebastián-Gallés, 2001; Molnar et al., 2014; Nazzi et al., 2000; Peña et al., 2010).

Additionally, as earlier introduced, the present pair of languages (Catalan and Spanish) is quite particular in that on top of being rhythmically close languages (both syllable-timed), they also overlap in a great number of non-rhythmic features such as phonetic-phonological categories, phonotactic structures, high number of cognate words, etc. (see Bosch, 2018). This is relevant to the present work because high linguistic proximity has been found to influence language acquisition: it can delay the establishment of some vowel categories, as compared to learning two more distant languages (Bosch & Sebastián-Gallés, 2003; Sundara & Scutellaro, 2011), but at the same time it can also accelerate vocabulary building and word learning (Bosch & Ramon-Casas, 2014; Havy et al., 2016). In other words, the evidence shows that language proximity can reduce the perceptual distance for some language pairs, which increases the difficulty of their differentiation.

Such is the case of Catalan and Spanish bilingual infants. It is not until 4 months of age that there is proof for acoustic discrimination of their native languages, and such discrimination has only been found when using a variation of the head-turn procedure (Bosch & Sebastián-Gallés, 2001), but not when analyzing orientation times to their native languages (Bosch & Sebastián-Gallés, 1997). Additionally, the latter study showed that bilingual infants needed more time to orient to their native languages than monolingual infants did, which suggested different mechanisms of switch detection by the two groups (Bosch & Sebastián-Gallés, 1997). These studies demonstrate that although achievable, the discrimination of Catalan and Spanish at 4 months of age is still a hard task for bilingual infants.

Study 1 did address the issue of infants' Catalan-Spanish discrimination but in this case in the audiovisual domain, that is, presenting a talking face that switched between the languages. The results of the experiment added to the previous evidence from acoustic-only studies by showing that the bilingual group did not detect the close language switch, whilst the monolingual group did. Moreover, the fact that both groups detected a subsequent distant language switch (to English) verifies that non-detection of the close language switch is indeed due to the language and not to experimental settings. Taken together with previous studies showing that bilinguals can discriminate their languages acoustically (Bosch & Sebastián-Gallés, 2001) and that, in fact, they show an enhanced sensitivity to visual speech cues (Sebastián-Gallés et al., 2012) suggests that in this case, rather than taking advantage of the visual information, seeing the talker's face is precluding bilinguals' ability to discriminate their two languages. This is highly interesting since it is indeed audiovisually that infants most usually perceive languages, and thus it suggests that their task may be even harder than is presently assumed.

These results are also interesting in light of previous studies suggesting different language discrimination strategies between monolingual and bilingual infants; whilst monolinguals can more readily detect familiar *vs.* unfamiliar speech, bilingual infants show a slower processing of a language switch, which involves an increased attention to the speech signal (Ferjan Ramírez, Ramírez, Clarke, Taulu, & Kuhl, 2017; Kuipers & Thierry, 2015; Nacar Garcia et al., 2018; Singh et al., 2015) and renders slower orientation times to their native languages (Bosch & Sebastián-Gallés, 1997). Our results support this interpretation by showing that when the talking face is presented – and therefore attention may be more distributed and not only focused on the acoustic speech signal – only those infants with a linguistic experience restricted to a single language exhibit an attention recovery to the language switch. Concerning the bilingual group, it is possible that such detection may have appeared with longer exposure to the switch. Last, the fact that code-switching exposure did not modulate switch detection strengthens the idea that it is in fact familiarity with the two languages and not the experience of observing people switch that weakens bilingual infants' detection of the switch between their native languages.

In sum, Study 1 results give support to the idea that the discrimination of two closely related and familiar languages is quite an arduous task for infants and that, at the age of 4 months, they only succeed when the auditory information is presented in isolation, reflecting again the fact that increased attention to the speech signal is needed to succeed. In turn, this suggests that under more naturalistic settings, bilingual infants may need more time, experience and/or maturation time to be able to tell apart their two native languages when spoken by the same talking face. Further studies that explore language discrimination in more

naturalistic (audiovisual) settings and assess bilingualism, language proximity and language familiarity independently will help better understand this process and the influence of each factor onto the development of language discrimination.

Development of selective attention to a talking face

Taking on a different but complementary approach, other investigations have explored infants' selective attention patterns to their environment and how they change throughout development. Interestingly, the study of infants' selective attention gives us insight into what infants perceive as most relevant and hence pay most attention to at a given moment (Amso & Scerif, 2015). Following this idea, a great body of work – including Study 2 to Study 7 – has explored the topic of audiovisual speech perception in monolingual and bilingual infants, children and adults by investigating their selective attention patterns to a talking face.

It is known that already in their first months of life, infants attend preferably to face-like patterns than to other types of stimuli (Johnson et al., 1991), and that, within a face, they attend more to the talker's eyes (Haith et al., 1977). Thereafter, between 6 and 8 months of age infants shift towards attending more to the talker's mouth (Lewkowicz & Hansen-Tift, 2012), which has been interpreted as infants' intentionally focusing on and relying on the source of audiovisual speech cues – i.e. the talker's mouth –, to aid them acquire their language. Later studies revealed that, in fact, bilingual infants performed the shift to the mouth earlier than their monolingual peers, at 4 months of age (Pons et al., 2015), and that they showed increased attention to the mouth at 8 months (Ayneto & Sebastián-Gallés, 2016), 12 months (Pons et al., 2015) and at 15 months of age (Fort et al., 2017). These results have been interpreted as bilingual infants' additional resourcing to the audiovisual speech cues, in face of their extra challenge of learning two languages whilst keeping them separate.

However, it is worth noting that these studies were performed in Catalan-Spanish bilingual infants which, as earlier mentioned in Study 1, is a pair of two very closely related languages that are harder to differentiate than other more distant language pairs. Therefore, in Study 2 we tested the hypothesis that bilinguals' language proximity may be modulating selective attention to the talker's face. The results of Study 2 confirmed this hypothesis and thus suggested that the previously reported increased attention to the mouth was not a consequence of bilingualism *per se,* but rather derived from the greater cognitive challenge of being exposed to a pair of close languages.

As earlier discussed in Study 1 (see page 149-150), previous studies had already raised the idea that bilingual infants may need to deploy more attention to the speech signal to help them deal with their dual-language input. For example, studies have found that bilingual

infants tend to habituate faster and fixate longer on new stimuli (Singh et al., 2015), and that they present increased attention to speech (Kuipers & Thierry, 2015; Shafer, Yu, & Garrido-Nag, 2012). Similarly, previous language discrimination studies have reported that bilingual infants exhibit slower detection times (Bosch & Sebastián-Gallés, 1997; Ferjan Ramírez et al., 2017; Nacar Garcia et al., 2018) and that they may not even detect a language switch from the same talking face (Study 1). The combination of Study 2 results with the above-mentioned studies gives support to the interpretation that, in order to help in their harder task of disambiguating two close languages, close bilingual infants increase their attention to speech in general and to the talker's mouth in particular.

Interestingly, the results of Study 3 extended those of Study 2 by showing that, in fact, the reported greater attention to the talker's mouth in close than in distant bilingual infants is also present in 5- to 6-year-old children. One may argue that at the linguistically advanced age of 5 years, children are not likely to need the additional audiovisual cues of a talker's mouth for processing speech any longer. If that were the case, then their increased attention to the mouth may reflect an earlier-shaped mouth bias, a maintained exploratory behavior due their early experience of relying on the mouth cues when learning their languages. On the other hand, it is also feasible that the dual language input is still a source of ambiguity for 5-year-old children, and therefore that they are still relying on and purposely attending to the mouth audiovisual cues. Yet, regardless of the underlying motivations for such behavior, the results of Study 3 showed that the distance between bilinguals' two languages still plays an important role in selective attention to a talking face during childhood.

Consistent with these results, a recent study by Morin-Lessard and colleagues (2019) has demonstrated that monolingual and distant bilingual (French and English) infants and children exhibit comparable patterns of selective attention to a talking face. In this study, the researchers explored 5-, 9-, 12- and 14-month-old infants and 2-, 3- and 4- to 5-year-old children and showed that language background (i.e. bilingualism) had no significant effect in any age group. Again, these results together with Studies 2 and 3 reinforce the idea that it is not the mere fact of learning any two languages that is linked to the previously reported greater attention to the mouth, but rather the fact of learning two languages that are linguistically close and difficult to disambiguate.

Noteworthy, it has been argued that, in fact, the constant practice of having to separate their two native languages – and keep them separate – is likely to be one of the main causes of other cognitive advantages associated to bilingualism such as enhanced attention to faces (Mercure et al., 2018), enhanced visual-only discrimination of languages (Weikum et al., 2007) faster search, habituation and encoding of visual stimuli (Chabal et al., 2015; Friesen et al., 2014; Singh et al., 2015), facility for simultaneous segmentation of two artificial

languages (Antovich & Graf Estes, 2018) or even individual sounds (Sebastián-Gallés & Bosch, 2009), as well as executive functioning and enhanced cognitive control (Comishen et al., 2019; Kovács & Mehler, 2009; Mehler & Kovács, 2009). If that is the case, then, the present results suggest that the proximity between bilinguals' two languages is a highly relevant factor that will modulate their cognitive and linguistic abilities. Therefore, future studies that explore bilinguals' language learning process and their associated cognitive benefits should go beyond the monolingual and bilingual comparison and embrace more systematic comparisons of different groups of bilingual language learners.

Beyond the monolingual-bilingual comparison, it is relevant to discuss the general developmental trajectory of selective attention to the talker's face. Morin-Lessard et al. (2019) have revealed that 5-month-old infants pay equal attention to both eyes and mouth, 9-month-olds show a preference for the mouth, 12-month-olds show again a balanced distribution between the two areas and then 14-month-old infants to 5 year-old children show a preference for the mouth. In their study, attention to the mouth peaks at around 2 years of age and then slowly declines, albeit still showing a preference for the mouth at 5 years of age.

Taken these results together with the previously reviewed evidence and Studies 2 and 3 shows that, overall, following the shift towards the mouth at around 8 months of age (Lewkowicz & Hansen-Tift, 2012) the mouth-preference is maintained during infancy as a general pattern of selective attention to a talking face – i.e. regardless of language background or test language – until at least early childhood (i.e. at 15 months: Study 1 & Fort et al., 2017; at 14 and 18 months: Hillairet de Boisferon et al., 2018; at 14 and until 5 years of age Morin-Lessard et al., 2019), despite the exception of 12-month-old infants that attend equally to both AOIs. This decrease of mouth-looking in 12-month-olds has been related to the fact that perceptual narrowing is at its endings (Maurer & Werker, 2014) and hence it was argued that infants may not need the AV redundant cues to the same extent as before (Lewkowicz & Hansen-Tift, 2012). However, the fact that only 2 months later infants show again a clear preference for the mouth indicates that the mouth decrease found in 12-month-olds is a quite specific effect (restricted to this age group), and therefore, beyond the possible interpretations behind it, the overall evidence argues in favor of the earlier-mentioned general mouth preference in infancy and early childhood.

Thereafter, during childhood, a common finding amongst the different studies of selective attention is that mouth-looking peaks at around the age of 2, and then starts decreasing (Jones & Klin, 2013; Morin-Lessard et al., 2019) until reaching equal looking to the eyes and mouth at around 5 years of age (Król, 2018; Nakano et al., 2010) – with the exception of Morin-Lessard et al.'s (2019) findings still showing a mouth preference in the

5-year-old group. Children's attention pattern seems to stand as a middle point between infants' mouth-dominance and adults' eyes-dominance. However, as earlier discussed, the causes that motivate children with such a cognitively advanced age and a large linguistic experience to still attend to the talker's mouth at least half the time, and close bilingual children to pay an even greater amount of attention to the mouth remained to be well understood.

In Study 4, we hypothesized that exploring children's dynamic allocation of selective attention over time would help understand the causes of these attention patterns and the underlying strategies they may be performing. Thus, as opposed to averaging across the whole duration of the videos – as previous studies had done, including Studies 2 and 3 –, in Study 4 we explored the time course of selective attention to a talking face in 5- to 6-year-old monolingual and close bilingual children by fitting a growth curve model to the full 60 s of data.

First, although the results of Study 4 do replicate the predicted increased mouth-looking in close bilingual (than in monolingual) children, it is crucial to mention that they also restrain these differences to the initial phase of the speech. This initial (until about 20s) attention to the mouth supports the idea that, albeit being already highly proficient in their native languages, close bilingual children do not yet identify the language that is spoken with the same automaticity than adults or distant bilingual infants do, and that as a consequence, they initially rely more on the mouth audiovisual cues to help them in this task and gradually cease to do so as the trial advances.

Secondly, the fact that monolingual infants exhibited a similar pattern of attention in the non-native language condition – initial mouth preference and decrease to equal looking to eyes and mouth –, suggests that in fact, this may be a general pattern of selective attention to perceptually adapt to a new language, accent or speaker. Similar to previous studies showing perceptual adaptation processes to foreign or artificial speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Lusk & Mitchel, 2016), children might initially rely on the mouth cues to augment speech perception when the levels of uncertainty are high, and thereafter, as they become more adapted to the speech characteristics, they distribute again their attention equally to both areas.

Last, the fact that monolingual children perform such an "adaptation pattern" only when they perceive a non-native language whereas close bilingual children do it in both languages suggests again that bilingual children may have higher levels of uncertainty when being spoken to. Consequently, close bilingual children would perform the "non-native attention pattern" when perceiving their native languages as well. Future studies that explore

specifically speech's perceptual adaption together with selective attention patterns to a talker's face in children are needed to further interpret these results.

Altogether, these studies suggest that the audiovisual redundant speech cues play a key role in speech perception throughout infants' (language) development, and that the amount of attention to the mouth generally reflects the moments in which infants and children favor the processing of speech cues over other cues also present in a talker's face, such as the eyes. Previous studies exploring interindividual variability have shown that indeed, what infants prioritize and therefore deploy their attention to at a given moment is closely associated to their ongoing cognitive processes, as illustrated by the fact that greater attention to the mouth in the first year of life correlates with higher language skills (Tenenbaum et al., 2013, 2015; Tsang et al., 2018; Young et al., 2009), and likewise, greater attention to the eyes correlates with communication and social skills (Pons, Bosch, & Lewkowicz, 2019).

In the case of bilingual infants and children, the results from the present study demonstrate that linguistic distance (or proximity in this case) plays an important role in mediating selective attention to a talking face throughout early childhood. In other words, close bilinguals' increased attention to speech in general and to the mouth cues in particular seems to be a consequence of their harder task of disambiguating the languages perceived. However, regardless of bilinguals' greater use of the AV cues, attending to the mouth cues seems to be a general pattern of attention to a talking face, as we have seen that the talker's mouth is the strongest attractor of attention during infancy and early childhood.

Last, it is important to place these findings under the bigger picture of natural language learning. Previous studies suggest that faces are in infants' visual field between 25 and 40% of the time (Kretch, Franchak, & Adolph, 2014; Sugden, Mohamed-Ali, & Moulson, 2014), and that within this time, they actively fixate on faces about 50 to 80% of the time (Frank, Amso, & Johnson, 2014; Frank et al., 2012). These numbers indicate that the talker's face is not always accessible or attended to, and that communication and language acquisition occur in both its presence and its absence. The clearest example of this is that congenitally blind children are capable of acquiring their native language in a largely typical fashion (Bohannon, Landau, & Gleitman, 1986). In sum, these studies indicate that vision (and the audiovisual speech cues) is not indispensable for learning a language. Rather, they suggest that infants are cognitively flexible and that the different constraints and linguistic situations modify the used strategies and learning styles. Then, our results argue for a strong audiovisual support in sighted children's acquisition and processing of language, and especially so when the speech perception situations become more challenging, due to active learning of speech articulation (e.g. babbling stage; {Formatting Citation} close bilingual

environments (Study 2, 3 and 4) to noisy situations (Król, 2018; Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998) or to non-native language speech (Kubicek et al., 2013; Lewkowicz & Hansen-Tift, 2012; Ter Schure et al., 2016).

Our collection of studies – together with the reviewed literature – sets a valuable baseline for the comprehension of infants' and children's selective attention patterns to a talking face, together with the language factors that modulate it. Further developmental studies exploring AV speech perception by combining selective attention measures with other approaches such as NIRS, ERPs or heart-rate and specific speech perception tasks will help reduce the variability present in these studies and confirm or redefine its interpretations, and eventually they will give new insight into the cognitive consequences of the described attention patterns.

Selective attention to a talking face in a non-native language

As described in the introduction, infants' perception of speech sound contrasts is modulated by their language experience. In the second half of the first year, perception of their native contrasts remains – or improves – while perception of non-native contrasts declines (e.g. Kuhl et al., 2006). Relevant to the selective attention studies before presented, there is evidence that at 6 to 8 months of age infants start allocating more of their attention resources on the talker's mouth when they perceive non-native speech as compared to when they perceive native-language speech (Berdasco-Muñoz et al., 2019; Ter Schure et al., 2016), and that infants continue to do so at 12 months of age (Kubicek et al., 2013; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). The present findings add to this evidence that this "increased mouth-looking in non-native speech" is also present at 15 months of age (Study 2), and at the age of 5 years (Study 4, although see Morin-Lessard et al., 2019).

Morin-Lessard et al., (2019) report no selective attention differences between test languages in any of the age groups (i.e. 5 months to 5 years). Importantly however, the researchers used different speakers for the native and non-native language videos which makes it impossible to separate language from speaker effects. The other studies here reported used one speaker only (Study 4, Berdasco-Muñoz et al., 2019; Ter Schure et al., 2016; Kubicek et al. 2013) or have replicated the findings by crossing infants' native languages (i.e. the English and Spanish videos shown to English-learning infants in Lewkowicz and Hansen-Tift (2012) were later presented to Spanish-learning infants in the Pons et al. (2015) study and in Study 2, finding the same pattern of results).

Altogether, these studies suggest that once infants' native categories begin to be established, both infants and children increase their attention to a talker's mouth when they detect unfamiliar speech sounds, putatively to aid in their processing. Ter Shure and colleagues' (2016) findings give support to this interpretation by demonstrating that not only do infants deploy their attention to the mouth but they also use the mouth AV cues to learn novel contrasts (Ter Schure et al., 2016).

The studies examined so far have concerned with infants and children perception of non-native speech. Following, I discuss whether the same attentional response to non-native speech applies to adult participants, and whether their proficiency in the non-native language (i.e. L2 proficiency) modulates their selective attention patterns to L2 speech.

We know from previous studies that adults' comprehension of non-native speech improves when presented audiovisually (Arnold & Hill, 2001; Hardison, 2005; Reisberg et al., 1987), similar to the classic studies showing that speech-in-noise is better understood when seeing the talker's face (Cotton, 1935; Sumby & Pollack, 1954). However, to our

knowledge, only Barenholtz et al.'s (2016) study had explored specifically adults' allocation of selective attention to a talker speaking in a non-native language. As earlier described, the results of Barenholtz et al. (2016) showed increased mouth-looking in the non-native speech condition, again supporting the idea that adults do rely more on the mouth to help them decode non-native speech. However, the participants in this study performed a short-sentence (3 s-long) identification task, and hence it remained to be demonstrated whether these results would extend to the more naturalistic situation of a talker producing longer, non-native speech monologues.

The combination of Studies 5 and 6 demonstrated that indeed, adults increased their attention to the mouth in response to non-native speech, not only when performing a replica of Barenholtz et al. (2016) short-sentence identification task with different languages and materials (Study 5), but also when participants perceived longer monologues (60 s) and their only task was to try to comprehend its content, mimicking a more naturalistic situation of non-native language speech perception (Study 6).

Last, we noted that the participants from all these studies were naïve (inexperienced) in the non-native language, and hence although they may rely on the mouth cues to extract some words or help disambiguate speech sounds, their task is likely to differ from that of an adult who is learning an L2 and therefore has some previous knowledge of the non-native language.

Study 7 explored the latter situation by evaluating whether participants' different levels of the non-native language (i.e. L2 proficiency) would modulate selective attention to the talker's face. Interestingly, the results showed only two different patterns of attention; that of a native perceiver – i.e. preference for the eyes – and that of a non-native perceiver – i.e. equal attention to eyes and mouth. Non-native participants' proficiency level in their L2 did not affect selective attention patterns. Different from the gradual increase of attention to the mouth with increasing levels of noise (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998), these results suggest that the attention to a talker's face follows a more dichotomous pattern, dictated by participants' early linguistic experience, where even the highly proficiency L2 learners behave as intermediate- or low-level learners and only native speakers exhibit the classic preference for the eyes (Yarbus, 1967).

In turn, they also suggest that language proficiency only does not explain the variability found within L2 perceivers, but that other variables – such as for example listening effort, listening confidence or L2 learning strategies – may play an important role in selective attention to a talker's eyes and mouth. As earlier mentioned, previous studies show that L2 learners very rarely reach native-like levels of speech perception performance (Lecumberri et al., 2010) and that they exhibit an increased cognitive effort when processing their L2

(Borghini & Hazan, 2018). As a consequence, it is likely that their level of confidence in L2 speech perception will be lower than that of a native speaker, and this in turn may motivate attending more to the talker's mouth – when it is available – for reassuring the comprehension of the content.

In the same vein, these differences in efficiency and confidence between native and non-native listeners may also affect the processing of the AV speech cues *per se*. In other words, it is possible that the amount of attention necessary to process the AV speech cues can vary between groups. It is known that the AV integration of speech requires attention (Alsius, Navarra, Campbell, & Soto-Faraco, 2005), and that the addition of distractors (i.e. noise or multiple talkers) reduces the maximal distance from the talker's mouth from which speech intelligibility is maintained (Yi, Wong, & Eizenman, 2013). Based on these studies, one may argue that native speakers may present a more efficient processing of the AV speech cues which allows them to focus more on the talker's eyes, without compromising the perception of the peripheral mouth's AV cues. On the other hand, non-native speakers may need more direct attention to the mouth in order to obtain a similar AV signal augmentation.

An alternative explanation for the lack of differences between the different levels of L2 proficiency would be that L2 learners do not easily change their speech processing attentional strategies while becoming more proficient in their L2. In other words, as discussed in bilingual children's results of Study 4, it may be that the earlier-acquired exploratory behavior – when learners were actually lower-level and thus, in need of the audiovisual signal enhancement – is later maintained, regardless of the fact that they may no longer need the audiovisual support.

In conclusion, the combination of Studies 5, 6 and 7 corroborate findings from other studies by demonstrating that greater speech-processing difficulty elicits greater reliance on the audiovisual perceptual cues available in a talker's mouth. In addition, our findings show for the first time that this general principle extends to people with differing levels of non-native language proficiency but with an important caveat: the degree of selective attention to a talker's mouth is not affected by the level of non-native language expertise. Overall, findings to date suggest that (1) perceivers resort to the greater saliency of the audiovisual speech cues located in a talker's mouth to enhance their speech comprehension and that (2) they rely on such cues even if they are expert L2 speakers. Finally, our findings have practical implications; they support the idea that second-language learning can be maximized by audiovisual training with audiovisual (rather than auditory-only) L2 materials (Bernstein, Auer, Eberhardt, & Jiang, 2013; Heikkilä et al., 2018). Future studies that incorporate other more fine-grained measures of L2 perception and processing will contribute to gain a better understanding of the current results and their implications.

The present set of studies support the idea that the audiovisual information of speech plays a key role in both first language/s acquisition during development and second language acquisition in adulthood. Moreover, the current work sheds some new light into the understanding of *when* and *why* a perceiver attends to the AV speech cues and relies on such cues for processing speech. Analyzing the selective attention patterns to a talking face has provided us with highly valuable insight on infants', children's and adults' attentional priorities at various linguistically different situations. The results of these studies reveal that the attentional patterns to a talker's face are highly dependent on internal factors such as age, cognitive capacities and language background (i.e. which specific language or languages are spoken or being learned), and also on external factors such as the specific task at hand, the language perceived (i.e. native, non-native) and the quality of the input signal.

Last, it is worth mentioning that there are relevant clinical implications that derive from the study of selective attention to audiovisual speech. Previous studies have demonstrated that variations from the selective attention patterns to a talking face here described (i.e. from typically developing participants) can be related to cognitive disorders such as autism spectrum disorder (ASD) (Chawarska, MacAri, & Shic, 2012; Jones & Klin, 2013; Nakano et al., 2010), specific language impairment (SLI) (Pons et al., 2018) or developmental risks associated to preterm birth (Berdasco-Muñoz et al., 2019). These studies suggest that measures of selective attention could be used for the detection or diagnosis of certain developmental disorders, and also to help develop possible clinical interventions (Irwin & DiBlasi, 2017).

Limitations and future directions

I have reported in this dissertation my attempts to further develop the understanding of audiovisual speech perception, by investigating selective attention to talker's faces in a series of experiments with infants, children and adults. Nonetheless, these conclusions also come with some limitations that need be considered.

First, in all the studies here presented, the stimuli consisted of an isolated and centrally presented talking face, with the only movement of the mouth and minor facial expressions. We chose this type of stimuli to be able to compare our findings with previous research, which has been very useful for interpreting our results and has allowed to grow more accumulative knowledge on the topic. Nevertheless, the extent to which these results and conclusions translate to real-life communication situations – which generally involve more movement and richer visual scenes – remains to be well described.

A few studies have approached the matter by using more ecologic stimuli; for example, Võ and colleagues (2012) explored adults' selective attention to videos of more naturalistic and dynamic scenes of talking faces (i.e. real interviews of casual people in the street), and they showed that participants perceiving speech in their native language did not exhibit the previously reported preference for the eyes, but rather allocated their attention more dynamically by focusing on the eyes when a face made eye contact, on the mouth when it started speaking and on the nose when it moved quickly (Võ et al., 2012). Additionally, Yi et al. (2013) found that the amount of attention to the mouth area of the talker increased in the presence of a second, distractor face (Yi et al., 2013). However informative, further studies that explore selective attention in more naturalistic settings – for example by using head-mounted eye-trackers and allowing participants to move freely inside a setting (Hernik & Broesch, 2019; Kretch et al., 2014; Suarez-Rivera, Smith, & Yu, 2019) – are still needed in order to better understand the ecological translatability of the present results and to eventually build a complete picture of the role of the AV speech cues in real-life speech perception situations.

The second issue worth discussing is the free-viewing eye-tracking paradigm, used in Studies 2 to 7 as well as in many studies here reviewed. Although it is used to reflect participants' unconstrained exploratory behavior, one may argue that there is a theoretical leap of interpretation between placing once visual attention (foveal fixation, i.e. overt attention) to an area of a presented video, and actually processing that information. Indeed, when selective attention is the only measure obtained, the possibility that an individual is covertly attending to another area or that s/he is generally uninterested and not engaged with the material cannot be fully disregarded. For this reason, we tried to use different measures

of attention such as post-viewing comprehension tests and memory tests to ensure the reliability of the data (except in pre-verbal infants). In the case of children, we failed to adequately quantify their rather inventive answers and did not find a systematic way to introduce the data into the analysis. In the adults' studies the tests ensured overall attention, but they were not sensitive to more fine-grained effects such as listening effort or more subtle comprehension differences.

Future selective attention studies that wish to use free-viewing paradigms would benefit from incorporating additional measures of processing (physiological measures such as pupillometry, heart rate or electrophysiology, on top of the behavioral tests), not only to ensure participants' processing but also to be able to reveal further more fine-grained cognitive mechanisms underlying the deployment of selective attention.

Last, the analysis of eye-tracking data is a topic that is also worth examining. As in most previous studies, here (except in Study 4) we counted the time participants spent looking to a specific area of the screen (sum of hits to a-priori defined AOI, such as the eyes), then we divided these by the time spent in the face, and finally we analyzed these relative scores to the different AOIs by way of mixed-ANOVAs and t-tests. This method for analyzing eye-tracking data spatially has been broadly used due to its straight-forward interpretation and it has unveiled a great body of knowledge regarding selective attention allocation to different stimuli. Also, as earlier commented in regard to the typology of stimuli used, using the same analyses as previous studies facilitates the direct comparison of the results conclusion and strengthens their validity. On the other hand, however, this method comes with a few limitations as well. The first one is that working with relative measures of attention forces developmental researchers to use cut-off values (i.e. minimum time to include a trial or an infant), which is a hard decision to make since there is no clear consensus across previous studies, and hence it adds variability to the data processing[10]. The second limitation concerns the pre-definition of AOIs, which, similar to the cut-off values, depends on the stimuli used and experimenters' decisions, and hence is susceptible to inconsistencies – for example, the definition of an AOI's size and borders. The third and last remark is about the temporal information. The averaging across time may be useful to interpret a general preference of attention, but it can sometimes occlude highly valuable information that may

[10] In the present case, we chose to use a 20% cut-off value, based on Frank, Vul, & Saxe (2012). Additional analysis validated that small changes of this cut-off value (i.e. ± 10 %) did not change the results obtained.

help interpret the data. One way to help better characterize perceivers' visual exploration strategies is indeed to examine the data temporally as well as spatially.

As performed in Study 4, incorporating the temporal information into the analysis can be done by using linear mixed-effects models (LMM) together with growth curve analysis (GCA). The combination of these statistical models allows for analyzing the time-course of the data, as well as to control for trial and participant effects. Alternatively, another interesting approach to finer-grained analysis is the use of Hidden Markov Models (HMM), which includes the temporal information (in the form of transition matrixes) and data-driven states identification (AOIs). Although not included in the present work, we are currently exploring the use of HMMs to analyze the data from selective attention studies (Birules, Fort, Diard, Bosch, & Pons, 2019; Birules, Lewkowicz, Pons, & Bosch, 2019). Future studies that analyze patterns of selective attention to a talking face by using both temporal and spatial analysis methods may help us take the current findings one step further and provide new insights into the cognitive mechanisms at play in language acquisition and face processing.

Chapter 8

Conclusions

Conclusions

This doctoral thesis provides new insights into the way that infants, children, and adults attend to and process the audiovisual speech cues originated in a talker's face, and how this behavior affects their perception of speech and language. Specifically, two broad questions were raised: first, whether perceiving speech audiovisually would modulate infants' language discrimination ability, and second, whether different aspects of language background would modulate perceivers' selective attention patterns to a talking face.

In regard to the first question, the findings of the present work suggest that in the case of close bilingual infants, perceiving speech audiovisually does in fact hamper their ability to discriminate languages. It is here demonstrated by the fact that Catalan-Spanish bilingual 4-month-old infants could not detect when a talker switched between their two languages, whereas monolingual infants showed detection of the switch from their native to a close language. These results support the idea that Catalan-Spanish bilingual infants' strategy for detecting a switch between their languages requires more time and attention to the speech signal, as compared to monolinguals' faster switch detection, based on familiarity.

Concerning the second question, the current results indicate that overall, the tested language factors modulate perceivers' attentional patterns, reinforcing the understanding that there is a linguistic motivation behind perceivers' selective attention patterns to a talking face. Specifically, here we found that first, the distance between bilingual infants' and children's two languages (and not bilingualism *per se)* affects their deployment of selective attention; we demonstrated that those learning a pair of close languages rely more on the speech cues of a talker – as compared to distant bilinguals and monolinguals –, likely for aiding in the differentiation of their two close languages.

Above and beyond the influence of close bilingualism, the present results also indicate that children's pattern of attention – i.e. initial increased attention to the mouth and later more balanced exploration between the talker's eyes and mouth – reflects a general strategy for perceptually adapting to the characteristics of the perceived speech.

Last, in regard to the perception of non-native languages, the findings from this work showed that perceivers (i.e. infants, children and adults) generally attend more to the talker's mouth under the challenging situation of perceiving non-native speech, which suggests that they use the visual information of the mouth to augment the processing of non-native speech. Remarkably, the results also showed that the degree of proficiency in the non-native language (i.e. second language) did not modulate the amount of attention deployed onto the talker's mouth, which suggests that even expert L2 learners reinforce their L2 processing by attending to the talker's mouth speech cues.

In conclusion, the present work adds a new piece of evidence to the field of audiovisual speech perception and language learning by showing that selective attention to the eyes and mouth of a talker is a highly dynamic process, which is largely modulated by the perceivers' early linguistic experience and their ongoing processing task. Ultimately, this work suggests that accessing to the redundant audiovisual speech cues at the adequate moment enhances speech perception and is crucial for normal language development and speech processing.

Chapter 9

References

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, *2*(1), 42–46. https://doi.org/10.1111/1467-7687.00052

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Amso, D., & Scerif, G. (2015). The attentive brain : insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606–619. https://doi.org/10.1038/nrn4025

Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: bilingual experience supports dual language statistical word segmentation. *Developmental Science*, *21*(2), 1–11. https://doi.org/10.1111/desc.12548

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339–355. https://doi.org/10.1348/000712601162220

Ayneto, A., & Sebastián-Gallés, N. (2016). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science*, 1–11. https://doi.org/10.1111/desc.12446

Baart, M., Bortfeld, H., & Vroomen, J. (2015). Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech. *Journal of Experimental Child Psychology*, *129*(4), 157–164. https://doi.org/10.1016/j.jecp.2014.08.002

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory Development* (pp. 183–206). New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bahrick, L. E., & Lickliter, R. (2015). 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.Learning

Bahrick, L. E., & Pickens, J. N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant Behavior and Development*, *11*(3), 277–296.

https://doi.org/10.1016/0163-6383(88)90014-8

Bailly, G., Perrier, P., & Vatikiotis-Bateson, E. (2012). *Audiovisual Speech Processing*. (G. Bailly, P. Perrier, & E. Vatikiotis-Bateson, Eds.). Cambridge, UK: Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511843891

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, *147*, 100–105. https://doi.org/10.1016/j.cognition.2015.11.013

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2014). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 1–43. https://doi.org/10.1016/j.jml.2012.11.001.Random

Benoît, C., Mohammadi, T., & Kandel, S. (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. *Journal of Speech and Hearing Research*, 1195–1203.

Berdasco-Muñoz, E., Nazzi, T., & Yeung, H. H. (2019). Visual scanning of a talking face in preterm and full-term infants. *Developmental Psychology*. https://doi.org/10.1007/s00216-013-7487-8

Bernstein, L. E., Auer, E. T., Eberhardt, S. P., & Jiang, J. (2013). Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Frontiers in Neuroscience*, *7*(7 MAR), 1–16. https://doi.org/10.3389/fnins.2013.00034

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3. https://doi.org/10.1017/S1366728908003477

Birmingham, E., & Kingstone, A. (2009). Human social attention: A new look at past, present, and future investigations. *Annals of the New York Academy of Sciences*, *1156*, 118–140. https://doi.org/10.1111/j.1749-6632.2009.04468.x

Birules, J., Fort, M., Diard, J., Bosch, L., & Pons, F. (2019). Using Hidden Markov Models to understand infants ' developmental pattern of visual attention to talking faces ; evidence from monolingual and bilingual infants. In *Workshop on Infant Language Development*. Potsdam, Germany.

Birules, J., Lewkowicz, D., Pons, F., & Bosch, L. (2019). The role of language proficiency in visual attention to a talking face: Evidence from both Looking Times to pre-defined AOIs and Hidden Markov Model (HMM) analysis. In *The XIV International Symposium of Psycholinguistics (ISP)*. Tarragona, Spain.

Bohannon, J. N., Landau, B., & Gleitman, L. (1986). *Language and Experience: Evidence from the Blind Child. Language* (Vol. 62). https://doi.org/10.2307/414686

Boothe, R. G., Dobson, V., & Teller, D. Y. (1985). Postnatal development of vision in human and nonhuman primates. *Annual Review of Neuroscience*, *8*, 495–545.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased

for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*(MAR), 1–13. https://doi.org/10.3389/fnins.2018.00152

Bosch Galceran, L. (2004). *Evaluación fonológica del habla infantil*. Barcelona, Spain: Masson.

Bosch, L. (2018). Language proximity and speech perception in young bilinguals: revisiting the trajectory of infants from Spanish–Catalan contexts. In M. Gibson & J. Gil (Eds.), *Romance Phonetics and Phonology*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oso/9780198739401.003.0017

Bosch, L. (2019). Language proximity and speech perception in young bilinguals. In *Romance Phonetics and Phonology* (pp. 353–366). Oxford University Press. https://doi.org/10.1093/oso/9780198739401.003.0017

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322. https://doi.org/10.1177/0165025414532559

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*(1), 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/S15327078IN0201_3

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life *. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 1–22. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010027707001126

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Briggs, R. (1978). *The Snowman*. London, UK: Hamish Hamilton.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2008). Hearing Faces: How the Infant Brain Matches the Face It Sees with the Speech It Hears. *Journal of Cognitive Neuroscience*, *21*(5), 905–921. https://doi.org/10.1162/jocn.2009.21076

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*, *112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Buchan, J. N., Pare, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162–171. https://doi.org/10.1016/j.brainres.2008.06.083

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, *2*(1), 1–13. https://doi.org/10.1080/17470910601043644

Byers-Heinlein, K. (2012). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, *16*(01), 32–48. https://doi.org/10.1017/S1366728912000120

Byers-Heinlein, K. (2015). Methods for Studying Infant Bilingualism. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing*. Cambridge, UK: Cambridge University Press.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343–348. https://doi.org/10.1177/0956797609360758

Byers-Heinlein, K., Morin-Lessard, E., Poulin-Dubois, D., & Segalowitz, N. (2014). Monolingual and bilingual infants ' attention to talking faces from 5-57 months. In *Poster session presented at: Boston University Conference on Child Language Development, Boston, MA*. (p. 2014).

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596. https://doi.org/10.1126/science.276.5312.593

Campbell, R. (1998). *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory–Visual Speech*. (R. Campbell, B. Dodd, & D. Burnham, Eds.), *Trends in Cognitive Sciences* (Vol. 3). East Sussex, UK: Taylor & Francis Routledge. https://doi.org/10.1016/S1364-6613(99)01368-6

Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics*, *77*(8), 2684–2693. https://doi.org/10.3758/s13414-015-0973-7

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

Chawarska, K., MacAri, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(8), 903–913. https://doi.org/10.1111/j.1469-7610.2012.02538.x

Christophe, A., & Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-

month-olds. *Developmental Science*, *1*(2), 215–219. https://doi.org/10.1111/1467-7687.00033

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cohen, L., Atkinson, D., & Chaput, H. (2000). Habit 2000: A new program for testing infant perception and cognition. Retrieved from https://scholar.google.es/scholar?cluster=5640921205697487167&hl=ca&as_sdt=2005&sciodt=0,5

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*(June), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Comishen, K. J., Bialystok, E., & Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Developmental Science*, (July 2018), 1–11. https://doi.org/10.1111/desc.12797

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews. Neuroscience*, *15*(5), 336–345. https://doi.org/10.1038/nrn3709

Cotton, J. C. (1935). Normal "Visual Hearing." *Science*, *82*(2138), 592–593.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268. https://doi.org/10.1121/1.2946707

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Frienda, M. (2016). A Tutorial on Expository Discourse: Structure, Development, and Disorders in Children and Adolescents. *Journal of Speech, Language, and Hearing Research : JSLHR*, *59*(December), 1–15. https://doi.org/10.1044/2016

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*(1), 85–110. https://doi.org/10.1006/jecp.1997.2379

Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, *114*(2), 101–114. https://doi.org/10.1371/journal.pone.0178059

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*(4), 478–484. https://doi.org/10.1016/0010-0285(79)90021-5

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics*, *17*(8), 617–643. https://doi.org/10.1080/0269920031000111348

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

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

Fava, E., Hull, R., & Bortfeld, H. (2014). Dissociating Cortical Activity during Processing of Native and Non-Native Audiovisual Speech from Early to Late Infancy. *Brain Sciences*, *4*(3), 471–487. https://doi.org/10.3390/brainsci4030471

Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2017). Speech discrimination in 11-month-old bilingual and monolingual infants: a magnetoencephalography study. *Developmental Science*, *20*(1). https://doi.org/10.1111/desc.12427

Fort, M., Ayneto-Gimeno, A., Escrichs, A., & Sebastián-Gallés, N. (2017). Impact of Bilingualism on Infants' Ability to Learn From Talking and Nontalking Faces. *Language Learning*, 1–27. https://doi.org/10.1111/lang.12273

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces in early infancy. *Journal of Experimental Child Psychology*, *31*(9), 1713–1723. https://doi.org/10.1109/TMI.2012.2196707.Separate

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

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

Friesen, D. C., Latman, V., Calvo, A., & Bialystok, E. (2014). Attention during visual search: The benefit of bilingualism. *International Journal of Bilingualism*, *19*(6), 693–702. https://doi.org/10.1177/1367006914534331

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, *22*(3), 611–631. https://doi.org/10.1017/S0305000900009971

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

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. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/1165958

Haith, M. M., BERGMAN, T., & Moore, M. J. (1977). Eye Contact and Face Scanning in Early Infancy. *Science*, *198*(November), 853–855. https://doi.org/10.1126/science.918670

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training , visual cues , and phonetic environment. *Applied Psycholinguistics*, *26*, 579–596.

Havy, M., Bouchon, C., & Nazzi, T. (2016). Phonetic processing when learning words: The case of bilingual infants. *International Journal of Behavioral Development*, *40*(1), 41–52. https://doi.org/10.1177/0165025415570646

Heikkilä, J., Lonka, E., Meronen, A., Tuovinen, S., Eronen, R., Leppänen, P. H., … Tiippana, K. (2018). The effect of audiovisual speech training on the phonological skills of children with specific language impairment (SLI). *Child Language Teaching and Therapy*, (September), 026565901879369. https://doi.org/10.1177/0265659018793697

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *22*(4), 1–8. https://doi.org/10.1111/desc.12779

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*(April), 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*, (March), e12825. https://doi.org/10.1111/desc.12825

Irwin, J., & DiBlasi, L. (2017). Audiovisual speech perception: A new approach and implications for clinical populations. *Linguistics and Language Compass*, *11*(3), 77–91. https://doi.org/10.1111/lnc3.12237

Iverson, P., Kuhl, P. K., Akahane-Yamadac, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *38*, 361–363. https://doi.org/10.1016/S0

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

Johnson, S. P. (2013). Development of the Visual System. In *Neural Circuit Development and Function in the Brain* (Vol. 3, pp. 249–269). https://doi.org/10.1016/B978-0-12-397267-5.00033-9

Johnstone, R. A. (1996). Multiple displays in animal communication: "backup signals" and

"muliple messages." *Proceedings of the Royal Society of London Series B-Biological Sciences*, *351*(Real 1990), 329–338.

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

Kaganovich, N. (2016). Development of Sensitivity to Audiovisual Temporal Asynchrony during Mid-Childhood. *Developmental Psychology*, *52(2)*(4), 232–241. https://doi.org/10.1038/bmt.2013.152.Hematopoietic

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12215080

Klin, A., Shultz, S., & Jones, W. (2015). Social visual engagement in infants and toddlers with autism: Early developmental transitions and a model of pathogenesis. *Neuroscience and Biobehavioral Reviews*, *50*, 189–203. https://doi.org/10.1016/j.neubiorev.2014.10.006

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*(1), 110–124. https://doi.org/10.1111/desc.12098

Kovács, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science (New York, N.Y.)*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518. https://doi.org/10.1111/cdev.12206

Król, M. E. (2018). Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price: An eye-tracking study. *PLoS ONE*, *13*(3), 1–14. https://doi.org/10.1371/journal.pone.0194491

Kubicek, C., Boisferon, A. H. de, Dupierrix, E., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants : The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, *37*(2), 106–110.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

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

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants

show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2). https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Kuipers, J. R., & Thierry, G. (2015). Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain and Language*, *149*, 27–32. https://doi.org/10.1016/j.bandl.2015.07.004

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11442–11445. https://doi.org/10.1073/pnas.0804275105

Kwon, M. K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537–555. https://doi.org/10.1037/dev0000080

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, *65*(4), 536–552. https://doi.org/10.3758/BF03194581

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Lewkowicz, D. J. (2000). The Development of Intersensory Temporal Perception: An Epigenetic Systems/Limitations View. *Psychological Bulletin*, *126*(2), 281–308. https://doi.org/10.1037/0033-2909.126.2.281

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*(1), 66–77. https://doi.org/10.1016/j.actpsy.2013.12.013

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

Lewkowicz, D. J., & Flom, R. (2014). The Audiovisual Temporal Binding Window Narrows in Early Childhood. *Child Development*, *85*(2), 685–694. https://doi.org/10.1111/cdev.12142

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., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy*, *15*(1), 46–60. https://doi.org/10.1111/j.1532-7078.2009.00005.x

Lusk, L. G., & Mitchel, A. D. (2016). Differential Gaze Patterns on Eyes and Mouth During Audiovisual Speech Segmentation. *Frontiers in Psychology*, *7*(February), 52. https://doi.org/10.3389/fpsyg.2016.00052

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant Intermodal Speech Perception is a Left-Hemisphere Function. *Science*, *2019*, 1347–1349.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, *41*(1), 93–113. https://doi.org/10.1016/0022-0965(86)90053-6

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11–12), 887–899. https://doi.org/10.1016/j.specom.2010.01.005

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., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology and Behavior*, *77*(4–5), 657–662. https://doi.org/10.1016/S0031-9384(02)00916-2

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 764. https://doi.org/10.1038/260170a0

Mehler, J., & Christophe, A. (1995). Maturation and learning of language in the first year of life. In *The cognitive neurosciences* (M. S. Gazz, pp. 943–954). Cambridge, MA, US: The MIT Press. Retrieved from https://psycnet.apa.org/record/1994-98810-061

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Mehler, J., & Kovács, Á. M. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*.

Mercure, E., Quiroz, I., Goldberg, L., Bowden-Howl, H., Coulson, K., Gliga, T., … MacSweeney, M. (2018). Impact of language experience on attention to faces in infancy:

Evidence from unimodal and bimodal bilingual infants. *Frontiers in Psychology*, *9*(OCT), 1–10. https://doi.org/10.3389/fpsyg.2018.01943

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Mirman, D. (2014). *Growth curve analysis and visualization using R*. (J. M. Chambers, T. Hothorn, D. Temple Lang, & H. Wickham, Eds.), *The R Series* (Vol. 26). Boca Raton (FL): CRC Press/Taylor & Francis. https://doi.org/10.1177/0962280215570173

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm class native language discrimination abilities of basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, *19*(3), 326–337. https://doi.org/10.1111/infa.12041

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Heinlein, K. B.-. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, 1–60.

Munhall, K. G., & Johnson, E. K. (2012). Speech perception: When to put your money where the mouth is. *Current Biology*, *22*(6), R190–R192. https://doi.org/10.1016/j.cub.2012.02.026

Murray, Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences*, *39*(8), 567–579. https://doi.org/10.1016/j.tins.2016.05.003

Murray, M., & Wallace, M. (2011). *The Neural Bases of Multisensory Processes*. (M. M. Murray & M. T. Wallace, Eds.) (1st Editio, Vol. 20115459). Boca Raton, FL, US: CRC Press/Taylor & Francis. https://doi.org/10.1201/b11092

Nacar Garcia, L., Guerrero-Mosquera, C., Colomer, M., & Sebastián-Gallés, N. (2018). Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. *Scientific Reports*, *8*(1), 2770. https://doi.org/10.1038/s41598-018-20824-0

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., … Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2935–2943. https://doi.org/10.1098/rspb.2010.0587

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory

information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12. https://doi.org/10.1007/s00426-005-0031-5

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by english-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19. https://doi.org/10.1006/jmla.2000.2698

Oller, D. K. (2000). The emergence of the speech capacity. *Journal of Child Language*, *30*(3), 731–734. https://doi.org/10.1121/1.1388001.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*, B9–B17. https://doi.org/10.1016/S0010-0277(97)00030-9

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences*, *107*(8), 3823–3828. https://doi.org/10.1073/pnas.0914326107

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 and Development*, *54*(December 2018), 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

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

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10598–10602. https://doi.org/10.1073/pnas.0904134106

Pons, F., Sanz-Torrent, M., Ferinu, L., Birules, J., & Andreu, L. (2018). Children With SLI Can Exhibit Reduced Attention to a Talker's Mouth. *Language Learning*, (68), 180–192. https://doi.org/10.1111/lang.12276

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language

Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, *288*. https://doi.org/10.1126/science.288.5464.349

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Reisberg, D. (1978). Looking where you listen: visual cues and auditory attention. *Acta Psychologica*, *42*(4), 331–341. https://doi.org/10.1016/0001-6918(78)90007-0

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). New Jersey, US: Lawrence Erlbaum Associates, Inc.

Risberg, A., & Lubker, J. (1978). Prosody and speechreading. *Quarterly Progress and Status Report*, *4*, 1–16. Retrieved from http://www.speech.kth.se/prod/publications/files/qpsr/1978/1978_19_4_001-016.pdf

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

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329–2337. https://doi.org/10.1111/j.1460-9568.2011.07685.x

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*(1), 29–50. https://doi.org/DOI: 10.1002/icd.376 The

Sanders, D. A., & Goodrich, S. J. (1971). The Relative Contribution of Visual and Auditory Components of Speech to Speech Intelligibility under Varying Conditions of Frequency Distortion. *Journal of Speech Language and Hearing Research*, *14*(1), 154–159. https://doi.org/10.1121/1.2143572

Schure, S. Ter, Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*(APR), 1–11. https://doi.org/10.3389/fpsyg.2016.00525

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A Bilingual Advantage in Visual Language Discrimination in Infancy. *Psychological Science*, *23*(9), 994–999. https://doi.org/10.1177/0956797612436817

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants : is the distributional account all there is to it ? *Developmental*

Science, *6*, 874–887. https://doi.org/10.1111/j.1467-7687.2009.00829.x

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11*(2), 306–320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

Shafer, V. L., Yu, Y. H., & Garrido-Nag, K. (2012). Neural mismatch indices of vowel discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neuroscience Letters*, *526*(1), 10–14. https://doi.org/10.1016/j.neulet.2012.07.064

Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., … Rifkin-Graboi, A. (2015). Back to Basics: A Bilingual Advantage in Infant Visual Habituation. *Child Development*, *86*(1), 294–302. https://doi.org/10.1111/cdev.12271

Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*(10), 2387–2399. https://doi.org/10.1093/cercor/bhl147

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J. F., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In *Multisensory Development* (pp. 207–228). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0009

Stager, C. L., & Werker, J. F. (1998). Methodological Issues in Studying the Link Between Speech-Perception and Word Learning. *Advances in Infancy Research*, 237–256. Retrieved from https://scholar.google.es/scholar?hl=ca&as_sdt=0%2C5&q=Methodological+issues+in+studying+the+link+between+speech-perception+and+word+learning.+In+C.+Rovee-+Collier%2C+L.+Lipsitt%2C+%26+H.+Hayne+%28Eds.%29+%26+E.+Bavin+%26+D.+Burnham+%28Vol.+Eds.%29%2C+Adv

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. https://doi.org/10.1038/nrn2331

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109. https://doi.org/10.1037/dev0000628

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

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

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, *36*(4–5), 314–331. https://doi.org/10.1159/000259969

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 71–78. https://doi.org/10.1098/rstb.1992.0009

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505–513. https://doi.org/10.1016/j.wocn.2010.08.006

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., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased Focus on the Mouth Among Infants in the First Year of Life: A Longitudinal Eye-Tracking Study. *Infancy*, *18*(4), 534–553. https://doi.org/10.1111/j.1532-7078.2012.00135.x

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

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, *10*(2), 215–228. https://doi.org/10.1037/0882-7974.10.2.215

Thompson, L. A., & Malloy, D. (2004). Attention resources and visible speech encoding in older and younger adults. *Experimental Aging Research*, *30*(3), 241–252. https://doi.org/10.1080/03610730490447877

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 Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *The Journal of the Acoustical Society of America*, *112*(6), 3004–3013. https://doi.org/10.1121/1.1512289

Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*(6), 926–940. https://doi.org/10.3758/BF03211929

Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13), 3–3. https://doi.org/10.1167/12.13.3

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, *81*(2), 517–527. https://doi.org/10.1111/j.1467-8624.2009.01412.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science (New York, N.Y.)*, *316*(5828), 1159. https://doi.org/10.1126/science.1137686

Werker, J. F. (2012). Perceptual foundations of bilingual acquisition in infancy. *Annals of the New York Academy of Sciences*, *1251*(1), 50–61. https://doi.org/10.1111/j.1749-6632.2012.06484.x

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Sciences*, *12*(4), 144–151. https://doi.org/10.1016/j.tics.2008.01.008

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309. https://doi.org/10.1037/0012-1649.34.6.1289

Werker, J. F., & Hensch, T. K. (2015). Critical Periods in Speech Perception: New Directions. *Annual Review of Psychology*, *66*(1). https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (1999). INFLUENCES ON INFANT SPEECH PROCESSING: Toward a New Synthesis. *Annual Review of Psychology*, *50*(1), 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

Yarbus, A. L. (1967). *Eye movements and vision* (Translated). New York, New York, USA: Plenum Press. https://doi.org/10.1016/0028-3932(68)90012-2

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, *26*(1–2), 23–43. https://doi.org/10.1016/S0167-6393(98)00048-X

Yi, A., Wong, W., & Eizenman, M. (2013). Gaze Patterns and Audiovisual Speech

Enhancement. *Journal of Speech Language and Hearing Research*, *56*(2), 471. https://doi.org/10.1044/1092-4388(2012/10-0288)

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

Young, R. J., Amu, J., Reissland, N., Reid, V. M., Donovan, T., & Dunn, K. (2017). The Human Fetus Preferentially Engages with Face-like Visual Stimuli. *Current Biology*. https://doi.org/10.1016/j.cub.2017.05.044