The development of audiovisual vowel processing in monolingual and bilingual infants: A cross-sectional and longitudinal study

PhD candidate: Jovana Pejovic

Departamento de Lengua Vasca y Comunicación
Euskal Herriko Unibertsitatea

Supervisor: Dr. Monika Molnar
Co-supervisor: Dr. Eiling Yee

Donostia/San Sebastián, 2019

BCBL Basque Center on Cognition, Brain and Language

Paseo Mikeletegi, 69,

Donostia-San Sebastián, Spain

April 2019

Acknowledgments

When I took my first step on the PhD road I could not imagine how much unconditional support I would receive and how much I would be thankful for it.

First and foremost, my eternal gratitude goes to my supervisor **Dr. Monika Molnar** for all her the help she provided during my PhD training. Monika kindled my interest in infant language development with her brilliant lectures and discussions. She generously shared her knowledge on infant research methods, theory of language development, and on being an academic in general. She helped me understand the literature on monolingual and bilingual language acquisition and to develop my own research questions within that field. I feel privileged to learn from such an eminent expert and be part of her research team. Furthermore, Monika was always there to listen to my every doubt, every question or dilemma that I encountered (and I have encountered many). She encouraged every step I took towards this point, providing advice and support. I will always be grateful for the motivational words she shared when the PhD road got too bumpy. Thank you Monika, for helping me overcome all obstacles on this long journey.

My great appreciation also goes to my co-supervisor **Dr. Eiling Yee**. I am grateful for every fruitful discussion we had, for her guidance in looking beyond my own findings and grasping the bigger picture, for all the amazing suggestions regarding eye-tracking data, and for her patience. I am also thankful for her support and encouragement when I felt overwhelmed by the data. Thank you Eiling, for reminding me that every step is a progress even when it did not seem like it in the moment.

Both Monika and Eiling were excellent exemplars of what a PhD supervisor should be. Even though part of their supervision was long-distance, I felt that they were always following my steps. For this and much more—Thank you!

The thesis would not have been complete without support from the amazing Baby Lab at BCBL. Firstly, I would like to thank **Elena** for her tremendous help during participant recruitment and in the organization of the lab. Elena, thank you! **Borja**, my BabyLab buddy, helped me so much during data collection and I will always be grateful! Also, I am thankful for the help I received from research assistants, especially **Manex** and **Sara**. Thank you for being so great and professional. The greatest thank you goes to all **families** that took part in this study and invested their time to help me collect the data.

I am very thankful for the many fruitful discussions we had in the **SLIG** group! Thank you for all the suggestions during the earlier stages of this work. I am also very thankful for the help I received from the administrative team. **Eider**, thank you for being there to help me face all matters of bureaucracy and for making the thesis submission process less painful. Thank you **Ana,** for all the help I received during grant applications, and thank you **Ainhoa**, for your help in translating those applications to Spanish.

All these years I have received abundant encouragement and support from the amazing BCBL **predoc** community. Thank you for all prolific (and not so prolific ☺) meetings, for our R/mixed-models discussions, for listening to the many practice talks and rehearsals, and for the academic and non-academic advice. Thank you **Myriam**, for bringing positive energy when I needed it the most. Thanks to **Patricia**, who was always there to share my struggles

and bring hope. Thank you **Irene** for all the stats guidance and for bringing sanity to all us PhD students. Thank you **Sophie** for taking care of me (and for all those cakes). Thank you **Saioa** for the great help I received in my first days at the BCBL and for the days we shared in Paris. And thank you **Ioanna** for bringing a fresh optimistic perspective that we all need.

A very special thanks goes out to my PhD cohort, the *Postmodern Antisocials.* I honestly do not know how I would have finished my thesis without you. Thank you for being there for all professional doubts, for all your help whenever an error message appeared on the screen, and for any and all questions regarding the results. But more importantly, thank you for your friendship. **Saúl**, thank you for reminding us what is important in life and for making our PhD days fun. Thank you **Lelo** for all your help with editing and for being a true friend. It meant a lot knowing that you were always there for me. **Paula**, thank you… Oh thank you for so many things… For being of great help in a professional capacity, for supporting your *pack*, for your advice, and for the many shared laughs. **Eneko**, thank you for your significant ($p < .001;$ pun intended) support when it was direly needed; I will always be grateful for all your kind words. **Peter**, thank you for all the days you filled with joy and for being there for me during the good and the not so good days. My dear **Alexia**, thank you for your support from day one. You inspired me during a difficult time when I needed words of encouragement; you don't even know how much your help meant to me. The greatest support both at work and at home I received from **Bojana**. There are not enough words in any language to express how I feel fortunate to have someone like you in my life. Thank you for being my friend, my sister, my constant encouragement. You brought sunshine on rainy days (and there were many in San Sebastian) and made our place *home.* Volim te mnogo.

Thank you *Antisocials,* for cheering me up and making my PhD days memorable.

My last thank you goes to my family and friends in Serbia. Hvala mojim prijateljicama koje su male strpljenja za sve moje teške dane, za podršku i za ohrabrenje. Hvala **Ivana, Jasmina, Zorana, Dijana i Stefani**. Ne bih uspela bez vas. Hvala mojim **roditeljima** što su pružali podršku da uvek radim ono u čemu uživam. Hvala što ste verovali u mene i onda kada ja nisam. Hvala mojoj **baki** za sve lepe želje ka kojima sam stremila. Hvala **Biljani** i **Andreju** za njihovu podršku. **Biljana**, ti si najodgovornija što sam pratila svoje snove i što nisam odustajala. Ti zaista jesi moja najveća podrška. *To the moon and back.* Moju najveću inspiraciju dugujem malim bićima **Tari** i **Vanji** koje su mi najviše nedostajale tokom studija. Svaki dan sa vama je radost i mogu samo da vam kažem da vas jako volim i da ćemo konačno imati više vremena za filmove, za igranje, za nas. Thank you all for your encouragement and for your constant love. Hvala.

Resumen amplio en castellano

El presente documento resume la tesis titulada *The development of audiovisual vowel processing in monolingual and bilingual infants: A cross-sectional and longitudinal study* (*El desarrollo del procesamiento audiovisual de vocales en bebés bilingües: un estudio transversal y longitudinal*). El objetivo principal de esta tesis fue el de investigar la trayectoria evolutiva del procesamiento de la información lingüística audiovisual en condiciones de congruencia e incongruencia. Para tal fin, examinamos las habilidades de correspondencia audiovisual en bebés monolingües y bilingües de euskera y castellano de 4,5 y 8 meses de edad. En concreto, utilizamos la técnica de *eye-tracking* para medir dos aspectos: i) las diferencias de atención hacia los estímulos congruentes frente a los incongruentes; y ii) las diferencias de atención hacia la zona de los ojos frente a la zona de la boca del interlocutor. La tesis consta de cinco capítulos que se resumen en las secciones siguientes.

Capítulo 1: Introducción

Los bebés que adquieren dos lenguas en la infancia (bebés bilingües) conviven en un entorno lingüístico distinto a los bebés que adquieren una única lengua (bebés monolingües). A diferencia de los monolingües, la exposición lingüística de los bilingües se divide entre sus lenguas. Además, los bilingües están expuestos a más tipos de regularidades lingüísticas en sus dos idiomas, incluidas las regularidades fonéticas, léxicas y sintácticas. A pesar de que, dadas estas diferencias entre los entornos de aprendizaje lingüístico de monolingües y bilingües, podrían esperarse diferencias en la velocidad a la que estos dos grupos de bebés adquieren el lenguaje, el hecho es que los bebés monolingües y bilingües desarrollan el lenguaje a un ritmo similar (ej., Sundara, Polka, y Molnar, 2008; Fennell y Lew-Williams, 2017; Barreña, Ezeizabarrena, y García, 2011). En cualquier caso, todavía se desconocen los mecanismos a través de los cuales los bebés bilingües consiguen desarrollar el lenguaje a la

misma velocidad que sus homólogos monolingües. Una posibilidad sería que los bebés bilingües hicieran uso de estrategias de aprendizaje distintas durante la adquisición del lenguaje. De hecho, estudios anteriores han revelado que, ya desde los 4 meses, los bebés bilingües miran más a la boca del interlocutor que los monolingües (Pons, Bosch, & Lewkowicz, 2015). Aunque este resultado ha dado pie a la propuesta de que la experiencia lingüística bilingüe incrementa la sensibilidad de los bebés hacia los aspectos visuales del lenguaje, entendidos estos últimos como la información que se obtiene al observar los movimientos articulatorios de los hablantes, la hipótesis está todavía sujeta a debate (Mercure et al., 2018; Tsang, Atagi y Johnson, 2018).

En este contexto, la presente tesis se centró en el estudio del procesamiento de los aspectos visuales del lenguaje en bebés, y más concretamente en el examen de la atención de los bebés hacia los movimientos articulatorios del interlocutor. Con el fin de estudiar este aspecto, medimos la sensibilidad al lenguaje audiovisual de bebés monolingües y bilingües. A este respecto, cabe destacar que, en comparación con estudios anteriores, utilizamos el mismo set de estímulos para la totalidad de participantes, asegurando así un nivel óptimo de comparabilidad entre el grupo de bilingües y monolingües. Además, abordamos el análisis de los datos desde tres perspectivas complementarias. En concreto, en el Capítulo 2 analizamos la habilidad de los bebés para asociar correctamente las realizaciones visual y auditiva de las vocales. A continuación, el Capítulo 3 analiza la atención que prestan los bebés a la boca durante la tarea de congruencia audiovisual. Por último, en el Capítulo 4 analizamos los datos longitudinales sobre la sensibilidad de monolingües y bilingües a los aspectos visuales del lenguaje. Dado que utilizamos la misma tarea audiovisual en todos los capítulos, ésta solo se describe en el Capítulo 2. En las siguientes secciones se detallan los hallazgos obtenidos con dichos análisis.

Capítulo 2: El desarrollo de la habilidad de correspondencia audiovisual de vocales en bebés monolingües y bilingües: un estudio transversal.

El objetivo principal del experimento que se describe en este capítulo fue el de examinar si los bebés bilingües muestran mayor sensibilidad a los aspectos visuales del lenguaje que los bebés monolingües. Con tal fin, los bebés realizaron una tarea en la que un interlocutor pronunciaba una vocal que podía ser *congruente* (vocalizaba /i/ y se oía /i/; vocalizaba /u/ y se oía /u/) o *incongruente* (i.e., vocalizaba /i/ y se oía /u/; vocalizaba /u/ y se oía /i/) con lo que oían los bebés. Esta tarea tiene como objetivo la medición de la *habilidad de correspondencia audiovisual* (*audiovisual matching ability*), entendida ésta como la capacidad de los bebés para asociar correctamente la realización oral de los sonidos con los movimientos articulatorios del interlocutor. En un estudio previo con un diseño similar, se observaron diferencias en los tiempos de fijación de la mirada entre las condiciones de congruencia e incongruencia, de manera que los bebés miraban más tiempo en la condición de congruencia que en la de incongruencia (Altvater-Mackensen et al., 2015). Nuestra hipótesis para este experimento era la de que, si el bilingüismo incrementa la sensibilidad a los aspectos visuales del lenguaje, el grupo bilingüe debería mostrar mayor habilidad de correspondencia audiovisual en comparación con el grupo monolingüe.

Realizamos el experimento con 84 bebés monolingües y bilingües (42 en cada grupo) de 4,5 meses de edad y con 75 bebés de 8 meses de edad (39 monolingües y 36 bilingües). El análisis se basó en la medición de la duración total de la mirada en la condición de congruencia audiovisual en comparación con la condición de incongruencia. Los resultados revelaron que tanto los monolingües como los bilingües mostraban una habilidad similar en la tarea de correspondencia audiovisual. En concreto, a los 4,5 meses de edad ambos grupos miraron más tiempo en la condición de incongruencia audiovisual que en la de congruencia, mientras que no se observaron diferencias entre condiciones a los 8 meses de edad. Además,

los resultados revelaron que los bebés más jóvenes mostraban habilidad de correspondencia para uno de los interlocutores, pero no para el otro. Estas diferencias entre interlocutores están en consonancia con hallazgos anteriores sobre cómo la prominencia visual/acústica modela la habilidad de correspondencia audiovisual (Altvater-Mackensen et al., 2015).

Capítulo 3: El desarrollo de la atención selectiva hacia la boca en bebés monolingües y bilingües durante el procesamiento audiovisual de las vocales: un estudio transversal.

En este capítulo analizamos a qué zona miraban los bebés (a la zona de los ojos frente a la zona de la boca) durante la tarea de correspondencia audiovisual presentada en el Capítulo 2. En este contexto, las investigaciones previas sobre la atención de los bebés hacia la boca han arrojado conclusiones contradictorias; aunque se ha observado una mayor atención hacia la boca en bebés bilingües (Pons et al., 2015), el efecto no se ha replicado de forma consistente (Mercure et al., 2018; Tsang et al., 2018). Nuestra hipótesis a este respecto era que, si efectivamente el bilingüismo incrementa la atención hacia la boca, los bilingües deberían prestar más atención a la boca durante la tarea de correspondencia audiovisual.

Con el fin de poner a prueba esta hipótesis, analizamos el tiempo relativo que invertían los bebés en mirar a la zona de los ojos frente a la zona de la boca del interlocutor, y observamos cuatro hallazgos relevantes. En primer lugar, no observamos diferencias entre la atención de los monolingües y bilingües hacia la boca en ninguno de los grupos de edad. En segundo lugar, observamos un incremento de la atención hacia la boca a lo largo del desarrollo, de manera que los bebés de 4,5 meses prestaban más atención a los ojos que a la boca, mientras que los de 8 meses miraban por igual a los ojos y a la boca. En tercer lugar, observamos que la dificultad de la tarea (condición de incongruencia) no incrementaba la atención de los bebés hacia la boca. Por último, descubrimos que las diferencias entre atención a los ojos y atención a la boca estaban moduladas por diferencias entre interlocutores. Cabe destacar que la preferencia de los bebés por los ojos frente a la boca

variaba considerablemente dentro del grupo de los bebés de 8 meses; el origen de esta variación se aborda en el Capítulo 4.

Capítulo 4: Variación individual en el desarrollo de la habilidad de correspondencia audiovisual y la atención selectiva hacia la boca en bebés monolingües y bilingües; un estudio longitudinal.

El objetivo del experimento presentado en este capítulo era entender si las diferencias individuales observadas dentro del grupo de bebés de 8 meses estaban relacionadas con el comportamiento de los bebés a los 4,5 meses de edad. Para tal fin, evaluamos de forma longitudinal el desarrollo de la habilidad de correspondencia audiovisual y la atención hacia la boca del interlocutor en 18 bebés monolingües y bilingües a los 4,5 y a los 8 meses de edad. Los resultados demostraron que la habilidad de correspondencia audiovisual a los 4,5 meses no guardaba relación con dicha habilidad a los 8 meses de edad. Sin embargo, la atención que los bebés prestaban a los ojos y a la boca mostraba una correlación positiva entre los 4,5 y los 8 meses de edad. En otras palabras, los bebés que miraban más a los ojos que a la boca a los 4,5 meses de edad continuaban haciéndolo a los 8 meses. El patrón observado no varió entre bebés monolingües y bilingües, por lo que no hallamos ninguna evidencia de que el bilingüismo incrementara la atención de los niños hacia la zona de la boca del interlocutor.

Capítulo 5: Conclusiones

Los estudios expuestos en la presente tesis no revelaron diferencias claras de sensibilidad a los aspectos visuales del lenguaje entre bebés monolingües y bilingües. Cabe destacar que tales resultados se obtuvieron con un paradigma de investigación en el que se usó el mismo set de estímulos lingüísticos para ambos grupos. Además, los resultados mostraron que la sensibilidad a los aspectos visuales del lenguaje estaba modulada por la edad, la relevancia visual de los estímulos y la dificultad de la tarea, así como por el valor de los estímulos. En términos generales, esta tesis realiza importantes contribuciones y plantea

numerosas preguntas de investigación cruciales para entender los procesos subyacentes a la

adquisición del lenguaje en poblaciones con características lingüísticas distintas.

Abstract

This thesis was designed to evaluate the role of monolingual vs. bilingual experience on audiovisual speech processing. A growing body of recent research has demonstrated that infants are sensitive to the visual attributes of speech (i.e., speakers' articulatory movements; e.g., Kuhl & Meltzoff, 1982; Lewkowicz & Hansen-Tift, 2012). Moreover, whether/how bilingual vs. monolingual experience shapes audiovisual speech processing during the first year of life, as early as 4 months of age, is currently under heated debate (Mercure et al., 2018; Pons, Bosch, & Lewkowicz, 2015; Tsang, Atagi, & Johnson, 2018).

In the current series of experiments, Spanish and Basque monolingual and bilingual infants' processing of audiovisual speech was assessed at 4 and at 8 months of age. In an audiovisual vowel matching task, the infants were presented with the vowels /i/ and /u/ in congruent (i.e., seeing and hearing a speaker producing /i/ and /u/) and incongruent condition (i.e., seeing a speaker mouthing /i/ but hearing /u/; seeing a speaker mouthing /u/ but hearing /i/). By testing vowels that are shared between Spanish and Basque, we tested whether Spanish and Basque monolingual and bilingual infants will differ in their strategies when processing stimuli that are linguistically equally meaningful to both populations. Using an eye-tracker, we measured the infants' eye vs. mouth looking preferences during the congruent and incongruent trials in a cross-sectional and a longitudinal design. In Chapter 2, we present our analyses based on infants' audiovisual matching ability. In Chapter 3, we present analyses based on infants' attention to the mouth and the eyes during the audiovisual matching task. Finally, in Chapter 4, the results of the longitudinal findings are discussed.

Our results revealed several important findings. First, in a testing paradigm where the linguistic stimuli are shared across the monolingual and bilingual infants, similar audiovisual matching abilities across the two language groups were observed (Chapter 2). More importantly, we found that both monolingual and bilingual infants shift their attention from

the eyes to the mouth of a speaker between 4 and 8 months of age, and we found no evidence that bilingualism affects infants' looking behavior at 4 months of age (Chapter 3). Further, we observed that infants' audiovisual vowel matching ability depends on the visual salience of the linguistic stimuli (within and between speakers; Chapter 2). Finally, our longitudinal results revealed that the eye-mouth preference bias at 4 months of age is likely to predict the preference bias at 8 months of age (Chapter 4). Overall, the results presented in this dissertation fundamentally contribute to our understating of how audiovisual speech processing emerges during the first year of life across linguistically diverse populations (monolingual vs. bilingual).

Table of contents

Abbreviations and acronyms list

ANCOVA   analysis of covariance

ANOVA   analysis of variance

AOI   area of interest

AV   audiovisual

CVC   Consonant-Vocal-Consonant

EM   eyes-mouth

LT   looking time

MP   matching preference

PLT   proportion of looking time

1 Chapter 1: Introduction

1.1 General introduction

The aim of the current dissertation is to investigate to what extent infants acquiring one language (monolinguals) and infants acquiring two languages from birth (bilinguals) share their audiovisual speech processing strategies. Bilingual infants encounter a different linguistic environment than monolingual infants. In contrast to monolinguals', bilinguals' language exposure is divided between their languages. In addition, bilinguals are exposed to more types of linguistic regularities (including phonetic, lexical, and syntactic regularities) across their two languages. Despite different language learning environments, bilingual infants reach language milestones at a similar pace as their monolingual peers (e.g., Barreña et al., 2011; Fennell & Lew-Williams, 2017; Molnar, Gervain, & Carreiras, 2014; Paradis & Genesee, 1996).

Considering the differences between monolingual and bilingual linguistic environments, it is puzzling how bilingual infants develop language at the same pace as monolinguals. One possibility is that bilingual infants employ different learning strategies than their monolingual peers during language acquisition. One strategy that has been recently proposed to serve bilingual acquisition is bilinguals' increased sensitivity to visual aspects of speech (i.e., attending to interlocutors' articulatory movements; Pons et al.,2015). Whether this strategy is specific to bilingualism is currently under debate (Mercure et al., 2018; Tsang et al., 2018). The current dissertation addressed crucial issues that emerged from this literature, to gain a more fine-grained understanding of monolingual vs. bilingual audio-visual speech processing.

Prior to the description of the current research (Chapter 2-4), this Chapter provides a theoretical background on monolingual and bilingual audiovisual speech processing. The

background is divided into three main sections. First, Section 1.2 describes the multimodal nature of speech, with particular interest in auditory and visual modality. The next two sections compare monolingual versus bilingual audiovisual speech processing, focusing on studies investigating sensitivity to visual speech. Specifically, Section 1.3 reviews the literature on monolinguals' sensitivity to visual speech, whereas Section 1.4 reviews the currently limited literature on bilingual visual speech processing. Section 1.4 is divided into four parts. First, we provided an overview of bilingual language acquisition (Section 1.4.1) to outline the similarities/differences with monolingual language acquisition that could lead to their differences in sensitivity to visual speech. Next, we present findings on bilinguals' greater sensitivity to visual speech in adults (Section 1.4.2) and compare it to findings in infants in the following section (Section 1.4.3). Importantly, Section 1.4.3 represents the nucleus of the literature review, describing mixed findings on bilingualism and sensitivity to visual speech in preverbal infants, which are directly related to the goals of the dissertation. Finally, we identified the possible methodological issues in those studies that revealed mixed findings (Section 1.4.4) and described how those issues were addressed in the current thesis (Section 1.5).

1.2   Speech as multimodal signal

In everyday communication we have the opportunity to perceive speech both auditorily (i.e., hearing speech), and also visually by seeing the interlocutor's articulatory, facial and body movements. A large body of research demonstrates the importance of visual, sensorimotor, and tactile information in speech processing (for a review see e.g., Keough, Derrick, & Gick, 2019). Therefore, speech is not a unimodal, but a *multimodal* signal, and its visual and auditory features are perceptually dominant (for a review see Navarra, Yeung, Werker, & Soto-Faraco, 2012; van Wassenhove, 2013).

2

The current study focused on the two most perceptually dominant modalities in speech perception: auditory and visual. Speech that is simultaneously perceived through the auditory and visual modality is referred to as audiovisual (AV) speech. By visual aspects of speech, we refer to the visual information we perceive by seeing the interlocutors' lips, tongue, and teeth during speakers' articulation. By auditory aspects of speech, we refer to auditory information we perceive by hearing the interlocutors' speech. Auditory and visual speech are temporally aligned. Mouth opening and closing signals the onset/offset of the auditory speech, and other articulatory movements reveal visual information that is important for speech sound identification. For instance, the place of articulation[1] (e.g., bilabial, dental, etc.) can aid in phoneme distinction: seeing articulatory movements helps in distinguishing two similar words, such as *boy (bilabial onset consonant)* versus *toy (alveolar onset consonant)*. Furthermore, mouth movements transmit information about the amplitude of the speech signal—a more open mouth corresponds to higher amplitudes. Overall, the auditory and visual speech signals co-occur and complement one another both providing cues to decode speech.

In over four decades of research on the development of infant speech perception (for a review see Gervain & Mehler, 2010; Maurer & Werker, 2014; Werker, 2018), most of the studies have investigated the auditory domain of speech perception, while the visual (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999), and sensorimotor (Bruderer, Danielson, Kandhadai, & Werker, 2015; Yeung & Werker, 2013) domains have received relatively little attention. To illustrate this point, we estimated the number of publications on auditory vs. visual or audiovisual speech processing with infants. We searched the Pubmed database (from 1972-2018) using the following keywords; "infant auditory speech perception"; "infant visual

---

[1] In phonetics, the place of articulation refers to the place of the vocal tract where air obstruction occurs during producing a certain consonant. For instance, *bilabial* consonants (i.e., /b/, /p/, /m/) occur when the both lips are closed and constrict airflow; *dental (/θ/* as in **th**ick) when the tongue is placed behind upper teeth, *alveolar* (e.g., /t/, /n/) when the tongue reaches the alveolar ridge to block airflow.

speech perception"; "infant audiovisual speech perception"; and "infant sensorimotor speech perception". Studies tagged as "infant *visual* speech perception" are combined with "infant audiovisual speech perception" studies because the database merges those two terms. The results that appear in **Figure 1-1** clearly demonstrates that infant research has been dominantly focused on the auditory domain, whereas visual and sensorimotor studies are much less frequent.
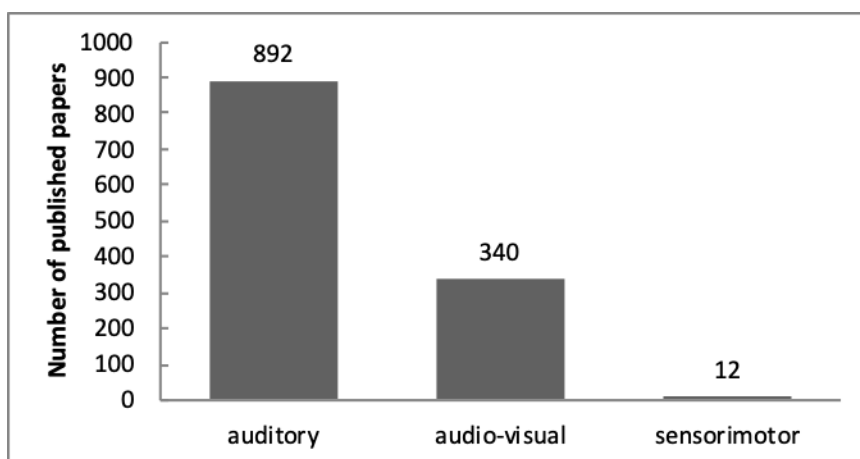
Figure 1-1. Number of studies published and indexed in the PubMed database on infant speech perception investigated in the auditory, audiovisual, and sensorimotor speech domain. Note that this Figure is not based on a systematic literature search but is an estimation of the relative amount of research on each modality in infant speech perception. The data is indexed from 1972 to the 2018 year and retrieved from https://www.ncbi.nlm.nih.gov/pubmed.

1.3 Monolingual (audio)visual speech processing

While infants' experience of auditory speech starts in the womb, around three months prior to birth (for a review on prenatal speech experience see e.g., Gervain, 2015), visual speech cues are not available to infants before birth. Still, newborns demonstrate sensitivity to visual speech. For instance, newborns were capable of imitating AV presented vowels by opening their mouth more to vowel /a/ and spreading the mouth more to /i/ (Coulon, Hemimou, & Streri, 2013).

A growing body of research demonstrated the importance of visual speech in infants'
language processing. For instance, infants, similarly to adults (Soto-Faraco et al., 2007) relied
on visual speech to discriminate languages (Dorn, Weinert, & Falck-Ytter, 2018; Weikum et
al., 2007). In one of these studies, infants were presented with two identical faces in silence
(i.e., muted), one articulating utterances in the infants' native language (English) and another
in a non-native/not-familiar language (French). The study revealed that 4- and 6-month-old
English-learning infants discriminated between the two visually presented languages, but at 8
months of age this ability was not present (Weikum et al., 2007). In addition, learning
phonemic contrasts in 6-month-old infants was enhanced by visual speech input (Teinonen,
Aslin, Alku, & Csibra, 2008), and recognition of familiar word forms was enhanced in the
presence of the visual speech in 12-13-month olds (Weatherhead & White, 2017).

Infants' ability to process visual speech is often assessed by their sensitivity to
matching vs. mismatching audiovisual speech (e.g., Kuhl & Meltzoff, 1982). In the current
study we also used the audiovisual matching task. In the next section, those previous infant
studies that relied on this methodology are described.

1.3.1   Monolingual infants' sensitivity to audiovisual congruence

Numerous studies have assessed infants' AV congruence/matching abilities. Table 1-1
provides an overview of AV matching studies in infants. In a typical AV matching task,
infants are presented with two side-by-side dynamic *silent* videos of the same speaker who
synchronously articulate two different vowels (e.g., /i/ and /a/; see **Figure 1-2** for an
illustration of an AV matching task). In synchrony with the two videos, infants hear one of the
seen vowels (either /i/ or /a/). In such experimental design, 4.5-month-old infants looked
longer to the face that matched the vowel they heard, demonstrating infants' ability to match
the auditory and the visual speech signals (Kuhl & Meltzoff, 1982). The same AV matching
ability has been replicated in 4.5-month-old-infants (Patterson & Werker, 1999), regardless of
the gender of the voice (Patterson & Werker, 2002), and with other vowel pairs such as /i/ and

/u/ (Yeung & Werker, 2013). Infants as young as two months of age demonstrated the same matching ability with vowel pairs (i.e., /i/-/u/, /i/-/a/) and vowel-syllable pairs (/i/-/wi/; (Baier, Idsardi, & Lidz, 2007; Patterson & Werker, 2003).
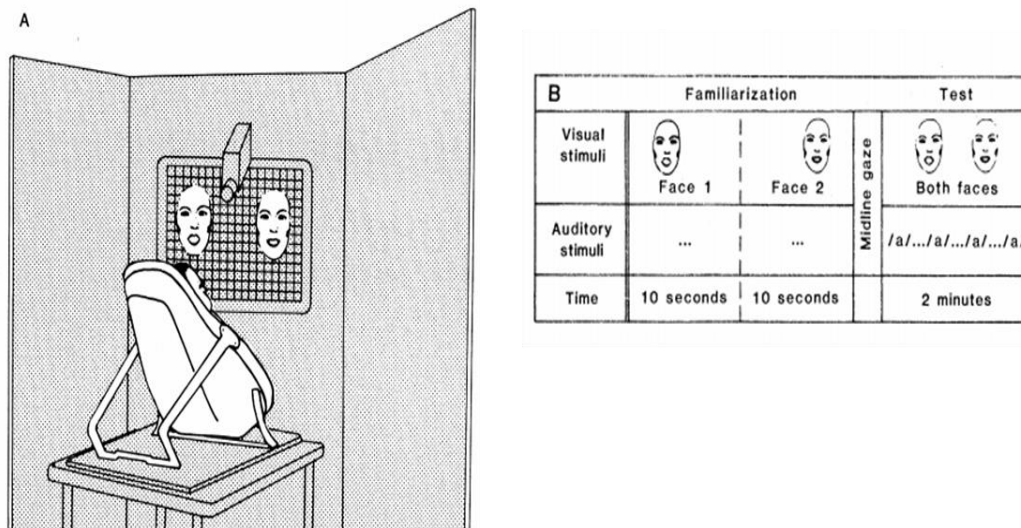
Figure 1-2 Typical experimental design for the infant audiovisual matching task. Panel A illustrates the AV matching test trial with an infant presented with two side-by-side faces articulating /i/ and /a/ vowels while hearing one of them. Panel B illustrates the full experimental procedure. From ″The bimodal perception of speech in infancy″ by P. Kuhl and A. Meltzoff, 1982, *Science, 218*, p. 1139. Copyright 1982 by the American Association for the Advancement of Science. Reprinted with permission.

In addition, the AV matching ability has been demonstrated with syllables (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983; but only for the visual stimuli presented on the right of a display). Later in development, infants start to match non-linguistic AV stimuli. For instance, 8-month-old infants matched non-linguistic visual stimuli, such as schematic faces (i.e., dynamic point-lines), to the corresponding infant-directed continuous speech (Kitamura, Guellaï, & Kim, 2014; Kitamura & Kim, 2010), as well non-linguistic auditory stimuli, such as bilabial trills, to the corresponding articulating face (Mugitani, Kobayashi, & Hiraki, 2008). Therefore, infants' AV matching ability is not limited to linguistic stimuli. Even

spontaneous AV matching of unfamiliar labels to different types of shapes has been observed (e.g., Pejovic & Molnar, 2017).

Infants' AV matching ability is present even when the auditory and visual stimuli are sequentially, rather than simultaneously, presented (Kubicek et al., 2014; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). For instance, when 6- and 11-month-old English-learning infants were auditorily presented with one syllable (i.e., /ba/ or /va/) after which they were visually presented with two side-by-side faces mouthing the two syllables (i.e., /ba/ and /va/), they looked longer to the face mouthing the previously heard syllable (Pons et al., 2009). Similar results are observed in 4.5- and 6-month-old infants tested with fluent speech in the sequential AV paradigm (Kubicek et al., 2014; however in the simultaneous paradigm the AV matching ability of fluent speech is not observed before 12 months of age Lewkowicz, Minar, Tift, & Brandon, 2015).

Overall, ample research suggests that from their second month of life, infants can match auditory speech to the corresponding speakers' articulatory movements (e.g., Kubicek et al., 2014; Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 2003; Pons et al., 2009; Yeung & Werker, 2013). In all these studies, infants were always presented with two faces articulating speech side-by-side (i.e., a two-face display), where one of the faces matched what infants heard. Relevant to the current thesis work is that in this experimental design, infants tend to look longer at the face that matches the auditory speech (see Table 1-1). Although it is clear that infants can perform AV speech matching from an early age, the exact mechanism through which infants detect the AV match or mismatch event is not understood. For instance, it has been demonstrated that temporal (i.e., temporal synchronization between auditory and visual speech) and spectral cues are not essential for AV matching task (Kuhl & Meltzoff, 1984), but more research is needed to understand the perceptual cues on the basis of which infants detect the AV matched event.

In addition to studies that used the side-by-side AV design, infants' AV matching ability has been observed in single-face paradigms too. In the latter paradigm, infants see only one face on the screen and the auditory signal either matches or mismatches the visual information. Hence, in these paradigms, AV match and AV mismatch is presented across trials). The ability of German-learning 5.5- to 6-month-old infants to match the auditory and visual components of the vowels /a/, /e/, and /o/ was assessed (Altvater-Mackensen, Mani & Grossmann, 2015). Using single-face displays, matching and mismatching trials were created for the /a/-/e/ and the /a/-/o/ vowel contrasts separately. Infants exhibited sensitivity to matching AV cues (i.e., they attended longer to match over mismatch trials) for the /a/-/o/ pair, but not for the /a/-/e/ pair, suggesting that infants' AV matching ability is affected by the visually presented vowel. Moreover, other studies using the single-face display paradigm have shown that 7- to 11-month-old infants attended longer to the mouth of the speaker in response to AV mismatch trials (e.g., visual /ba/ and auditory/ga/), than in response to AV match trials (Danielson, Bruderer, Kandhadai, Vatikiotis-Bateson, & Werker, 2017; Tomalski et al., 2013). In the current thesis, we opted for the single-face display, because it represents a more ecologically valid task than presenting infants with videos of two side-by-side faces. In addition, the single-face experimental set up is advantageous when infants' face scanning behavior (as measured by eye-tracking) during AV processing is also under investigation (Chapter 3 & 4 in the current thesis).

Table 1-1

Overview of studies on infant audiovisual matching ability[2]

Study	Auditory Stimuli	Visual presentation	Audiovisual presentation	Preceding familiarization/habituation phase	Age (months)	Findings
Kuhl & Meltzoff, 1982;1984	Vowels /i/ and /a/	Two-face display	Simultaneous	Familiarization	4	Matching Preference[3]
Patterson & Werker, 1999	Vowels /i/ and /a/	Two-face display	Simultaneous	Familiarization	4.5	Matching Preference
Yeung & Werker, 2013	Vowels /i/ and /u/	Two-face display (Exp1)	Simultaneous	Familiarization	4.5	Matching Preference
Patterson & Werker, 2003	Vowels /i/ and /a/	Two-face display	Simultaneous	Familiarization	2	Matching Preference
Altvater-Mackensen, Mani, & Grossmann,2015	Vowels /a/, /e/, and /o/	Single-face display	Simultaneous	Familiarization	5-6	Matching Preference only for /a-o/, but not for /a-e/ pair
Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009	Syllables /ba/ and /va/	Two-face display	Sequential	Baseline + Familiarization	6 and 11	Matching Preference
MacKain, Studdert-Kennedy, Spieker, Stern, 1983	Disyllables as *zuzu, vava*	Two-face display	Simultaneous	NA	5-6	Matching Preference only for right-side presented stimuli
Kubicek et al., 2014	Fluent speech	Two-face display	Sequential	Baseline + Familiarization	4.5 and 6	Matching Preference
Kubicek et al., 2014	Fluent speech	Two-face display	Simultaneous	Baseline + Familiarization	6	Matching Preference

To summarize, this section described infants' sensitivity to visual speech as reflected by their AV matching ability, with a focus on monolingual infants. In addition, we presented evidence that visual speech cues are relevant for monolingual infants' language development (e.g., in visual language discrimination, phoneme discrimination, etc.). Earlier in this chapter we mentioned that recent studies have suggested that acquiring two languages in early infancy increases sensitivity to visual speech cues. To understand these findings and the rationale behind them, the following sections describe the characteristics of bilingual language

---

[2] Note that in Table 1 we only present studies testing AV matching ability in infants, while we are not presenting studies on AV integration ability, or studies using a cross-model transfer paradigm.

[3] By preference we refer to infants' longer looking times to matched versus mismatched condition.

acquisition (in comparison to monolingual) and how bilingualism might affect (audio)visual speech processing.

1.4   Bilingual (audio)visual speech processing

1.4.1   Bilingual language acquisition

Here, we use the term *bilingualism in infancy* or *simultanous bilingual exposure* to refer to those infants who received regular exposure to two languages since their birth or even in their prenatal life. The monolingual and bilingual linguistic environment differs in several ways. First, a bilingual environment, in contrast to a monolingual one, contains two communication systems. This has both qualitative and quantitative consequences on the linguistic environment. The quantity of exposure to languages differs across the bilingual and monolingual linguistic environment. Even though overall monolinguals and bilinguals hear a similar amount of speech presumably, the bilingual speech input is distributed between the two native languages. Therefore, the exposure to *each* language is reduced in comparison to monolingual exposure.

Further, the bilingual environment is linguistically more varied and less consistent than the monolingual environment. While monolingual infants acquire a phonological system of one language, bilingual infants acquire phonemes[4] of both languages. For instance, a Spanish-learning infant will acquire five vowels (/a, e, i, o, u/), while a Spanish-Catalan will in addition to these five vowels acquire vowels specific to Catalan (/ɛ, ɔ, (ə)/). Bilingual infants' word learning takes place in a different context as well, since bilingual infants need to acquire two labels for the same concept. Bilingual language acquisition also requires the computation of two syntactic systems. Sometimes the syntactic rules can be very distinct

---

[4] The phoneme is the smallest unit of speech that distinguishes one word from another, like the phonemes /p/ and /b/ in the words *pin* and *bin*. Languages differ in their phonemes.

across the two languages (e.g., Spanish vs. Basque). For instance, the canonical word order in Spanish requires the subject to be followed by the verb, while this is reversed in Basque.

Regardless of the differences between the bilingual and monolingual linguistic environment, research has demonstrated that bilingual exposure does not hinder language or cognitive development. For instance, bilingual infants are sensitive to the rhythmic and prosodic information in their native languages to the same degree and at the same developmental stages as their monolingual peers (as demonstrated for German-French bilingual infants: Bijeljac-Babic, Höhle, & Nazzi, 2016; and for Spanish-Basque bilingual infants: Molnar, Carreiras, & Gervain, 2016; Molnar, Gervain, et al., 2014; Molnar, Lallier, & Carreiras, 2014). Moreover, monolingual and bilingual infants develop phonetic representations of their native language(s) at the same pace (e.g., Burns, Yoshida, Hill, & Werker, 2007; Larraza, Molnar, & Samuel, under review; Sundara, Polka, & Molnar, 2008). The overall vocabulary size of monolingual and bilingual infants is also comparable. While bilingual infants' vocabulary size within each of their languages is reduced compared to a monolingual vocabulary size (e.g., Bialystok, Luk, Peets, & Yang, 2010), bilinguals total vocabulary (in both languages) is similar to that of monolinguals (Hoff et al., 2012; Pearson, 1998).

Studies that focused on receptive lexical development (i.e., word comprehension) further supported similar development across monolingual and bilingual infants (e.g., Legacy, Zesiger, Friend, & Poulin-Dubois, 2016; Umbel, Pearson, Fernandez, & Oller, 1992). Interestingly, some studies demonstrated that expressive lexical development (i.e., speech production) might differ across monolingual and bilingual infants (for further description see Byers-Heinlein, 2018). For instance, 21- to 23-month-old bilinguals who acquired two similar phonetic systems, such as Spanish and Basque, outperformed their monolingual peers in producing Consonant-Vocal-Consonant (CVC) structures (Ezeizabarrena & Barreña, 2013). In this study, the observed bilingual advantage was not a long-lasting effect, since by 24

months of age Spanish monolinguals and Spanish-Basque bilinguals produced a similar number of CVC structures. Therefore, access to two similar phonological systems can accelerate, at least for a short period, bilingual production development.

Furthermore, recent findings suggest that overall language development (measured using the MacArthur-Bates Communicative Development Inventories — MCDI) in one of the bilingual infant's languages is similar to that of a monolingual infant, if the amount of exposure to that language is high. In particular, 16- to 30-month-old Basque monolinguals and Spanish-Basque bilinguals with more than 60% of exposure to Basque demonstrated similar MCDI scores in Basque, whereas Spanish-dominant bilinguals demonstrated lower MCDI scores in Basque (Barreña et al., 2011). Similar results were obtained in 25- to 50-month olds while measuring the mean length of utterance (Ezeizabarrena & Fernandez, 2018). Also, recent findings suggested that bilingual infants, despite their reduced exposure to morphosyntactic regularities in one language, keep up with their monolingual peers when processing grammatically correct or incorrect noun phrases (Molnar, Alemán Bañón, Mancini, & Caffarra, under review).

Considering the differences across monolingual and bilingual linguistic environments, it is unclear how bilingual infants develop language at the same pace as monolinguals. Research has proposed that bilingual infants may rely on different strategies during language acquisition (e.g., Werker & Byers-Heinlein, 2008). For instance, when AV speech processing is considered, recent studies have suggested that bilingual infants and children pay more attention to the mouth than their monolingual peers, which is presumably a strategy for dual language acquisition (e.g., Birules, Bosch, Brieke, Pons, & Lewkowicz, 2018; Pons, Bosch, & Lewkowicz, 2015). In addition, studies on adults provided further support on bilingual sensitivity to visual speech (e.g., Navarra & Soto-Faraco, 2007). Here, we will discuss in detail studies on bilingual audiovisual speech processing in both adults and infants.

1.4.2   Bilingual (audio)visual speech processing in adults

Recent research has suggested that (audio)visual processing is also modulated by linguistic experience (i.e., monolingual vs. bilingual experience) in adults. For instance, in a study on second language (L2) phoneme perception, two groups of Spanish-Catalan speakers were compared—Spanish-dominant and Catalan-dominant (Navarra & Soto-Faraco, 2007). The two groups were compared on the distinction of /ε/–/e/ vowels, a contrast that exists in Catalan, but not in Spanish. When the phonemes were presented auditorily, Spanish-dominant, but not Catalan-dominant bilinguals failed to discriminate the contrast. Interestingly, when the contrast was presented audiovisually (i.e., with a face pronouncing the vowels) both Spanish- and Catalan-dominant bilinguals discriminated the phonemes, suggesting that visual information enhances perceptual ability when processing the non-dominant language.

Furthermore, recent studies demonstrated that exposure to one versus to two languages shaped AV speech processing. Specifically, in a study by Soto et al. (2007) Spanish monolinguals and three groups of Spanish-Catalan bilinguals (i.e., Spanish-dominant, Catalan-dominant, and Spanish-Catalan balanced bilinguals) were compared on their ability to *visually* discriminate (without any auditory input) between Spanish and Catalan spoken utterances. Both Spanish monolinguals and Spanish-Catalan bilinguals could discriminate visually presented Spanish and Catalan utterances, but the bilingual groups tended to discriminate better than the monolingual group. When the monolingual group was directly compared with each bilingual group separately, the Catalan-dominant group discriminated with higher accuracy than the monolingual group.

Next, exposure to two languages modulated adults' sensitivity to the McGurk effect (Marian, Hayakawa, Lam, & Schroeder, 2018). Korean-English bilinguals perceived the AV incongruent trials more often as McGurk-like stimuli than English monolinguals. Hence, visual speech seems to alter AV speech integration more in bilinguals than in monolinguals.

Findings on selective attention to the mouth further corroborate that bilingualism shapes visual speech processing. Specifically, when participants were explicitly instructed to comprehend speech, monolingual adults attended more to the mouth than to the eyes while comprehending familiar and unfamiliar speech, whereas bilingual adults attended equally to the mouth and eyes when processing familiar speech (Barenholtz, Mavica, & Lewkowicz, 2016).

Overall, the above studies demonstrated that linguistic experience (bilingual vs. monolingual) interacts with audiovisual speech processing in adults (Barenholtz et al., 2016; Marian et al., 2018; Soto-Faraco et al., 2007). At what stage of development visual speech processing is shaped by bilingual experience is addressed in studies conducted with infants. These studies are discussed below.

1.4.3   Bilingual audiovisual speech processing in infancy

Several recent studies have proposed that exposure to two languages in infancy increases sensitivity to visual speech (Birules et al., 2018; Pons et al.,2015; Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; but see Tsang et al., 2018). For instance, in one of the studies described earlier (Section 1.2), English monolingual 4- and 6-month-olds visually discriminated between native (English) and non-native (French) languages (i.e., discriminated between two videos of a speaker silently mouthing in English and French), whereas at 8 months of age only French-English bilinguals, but not English monolinguals, demonstrated discrimination ability (Weikum et al., 2007). However, in Weikum et al.'s study, both languages were native to the bilinguals, whereas French was a non-native language to the monolinguals. Hence it was not clear whether bilingual infants relied more on visual speech cues because they were tested with two native languages. Another study addressed whether this bilingual advantage in visual language discrimination would be present if the two languages were not native to the bilinguals. In that study, Spanish-Catalan

monolingual and bilingual 8-month-olds were tested on visually discriminating English and French (Sebastián-Gallés et al., 2012). Similarly to Weikum et al.'s findings, the study demonstrated that unlike monolinguals, bilinguals discriminated between two languages. Therefore, to visually discriminate languages, bilingual infants did not have to be familiar with any of them. This further supports the claim that bilingual infants, as compared to monolinguals, might be more sensitive to the visual cues when it comes to speech processing.

Bilinguals' increased sensitivity to visual speech cues is also supported by eye-tracking studies revealing that bilinguals spent more time looking at the mouth (i.e., visual speech cue) than monolinguals. For instance, in a recent study, Spanish-Catalan 4-, 8-, and 12-month-old monolingual and bilingual infants were presented with AV continuous speech (i.e., AV presentation of a speaker fluently speaking an utterance). Monolingual infants were presented with AV speech in their *native* language (i.e., Spanish or Catalan), while bilingual infants were presented with AV speech in their *dominant* language. Spanish-Catalan bilingual infants attended to the mouth of speaker *earlier* (at 4 months of age) than Spanish/Catalan monolinguals (at 8 months of age). In addition, at 12 months of age, bilinguals attended to the mouth *more* than their monolingual peers (Pons, Bosch, & Lewkowicz, 2015, but see Tsang et al., 2018). Further, 8-month-old Spanish-Catalan bilingual infants attended more to the mouth of the speaker even with non-linguistic stimuli, i.e., videos of a person laughing or crying (Ayneto & Sebastian-Galles, 2017). A detailed description of the development of attention to the mouth in monolingual and bilingual infants will be provided in Chapter 3.

Pons et al. (2015) proposed that bilinguals' increased sensitivity to visual speech cues, such as to speakers' articulatory movements, aids bilingual infants in learning and separating their two languages. Therefore, one strategy that bilinguals might employ during their dual language acquisition is to rely more on visual speech cues. However, findings from two recent studies challenged the view that bilingualism in infancy increases sensitivity to speakers' articulatory movements. First, in a study by Tsang et al. (2018) 6- to 12-month-old

monolingual (mostly English monolingual) and a heterogeneous group of bilingual infants (exposed to English and various other languages) were presented with AV continuous speech in English. The authors observed that monolingual and bilingual infants did not differ in their attention to the mouth. Next, a recent study by Mercure and colleagues (2018) assessed 4- to 8-month-old monolingual and bilingual infants with AV congruent (i.e. seeing and hearing a speaker producing a syllable /ba/ or /ga/) and AV incongruent stimuli (i.e., seeing /ga/ and hearing /ba/; seeing /ba/ and hearing /ga/) while their eye-movements were recorded. The results revealed that older (6.6- to 8-month-olds) monolingual, but not bilingual infants attended more to the mouth in the AV incongruent condition.

1.4.4   Questions/issues that remain open about bilingual acquisition and audiovisual speech processing

As the pervious section demonstrates, findings on monolingual vs. bilingual sensitivity to visual speech are mixed. Some studies report that early exposure to a bilingual environment increases sensitivity to visual speech cues (Ayneto & Sebastian-Galles, 2017; Pons et al., 2015; Sebastián-Gallés et al., 2012; Weikum et al., 2007). However, such findings are challenged by other studies reporting no difference between monolingual and bilingual infants in their AV speech processing (Tsang et al., 2018), or reporting that monolinguals detected AV discrepancies, but bilinguals did not (Mercure et al., 2018).

By addressing the potential limitations of prior research, the current thesis investigated whether and how bilingual exposure affects AV speech processing in early infancy. Focusing on the studies that assessed AV speech processing, the issues that likely influenced prior findings can be categorized into three points—(1) sample and stimuli, (2) task difficulty, and (3) individual variability.

1.   Sample and stimuli

The heterogeneous linguistic sample

In the above studies, differences in the linguistic background of the bilingual samples could potentially be linked to the discrepancies in their findings. On the one hand, the study reporting increased sensitivity of bilingual infants to visual speech cues tested a homogeneous sample of bilingual infants—exposed only to Spanish and Catalan (Pons et al., 2015). In contrast, the studies showing no evidence of bilinguals increased sensitivity to visual speech tested a heterogeneous bilingual sample—those infants were exposed to English, but they varied in the other language they were exposed to (Mercure et al., 2018; Tsang et al., 2018). Hence, one explanation for the contradictory findings could be the difference between testing a homogeneous versus a heterogeneous sample of bilinguals. In the present study we addressed this issue by testing a *homogeneous group* of Spanish and Basque monolingual and bilingual infants. This allowed a more direct comparison with the previous study that tested a homogeneous Spanish and Catalan monolingual and bilingual group (Pons et al., 2015). An additional advantage of our target population is that both the bilingual populations in the current study and in the Pons et al. study came from a culturally similar bilingual environmental context, being regularly exposed to both of their languages at home and outside their home.

Stimuli

Relevant to the issue of heterogeneous sample, prior AV studies did not ensure that selected stimuli had the same linguistic value for monolingual and bilingual infants. Here, by "the same linguistic value" we refer to stimuli that are linguistically shared between groups. For instance, a phoneme that is native to both monolingual and bilingual infants (i.e., phonological representation of a phoneme overlaps across the groups) is a stimulus with the same linguistic value for both groups. However, previous studies did not control for the

linguistic value of stimuli. For instance, in Mercure et al. (2018) study, the bilingual sample

was exposed to English in addition to one or more other languages (the authors reported 18

non-English languages that infants were exposed to). Such a diverse sample might have

affected the processing of the selected AV stimuli— syllables /ba / and /ga/ produced by an

English speaker. Specifically, from the description of the sample, it is unclear whether the

bilinguals' other languages had similar phonetic categories particularly with regards to the /b/

and /g/ phonemes. For instance, stop consonants (e.g., /b/ and /g/) are present in both English

and Spanish, but differ in voice-onset-time values (i.e., the time that passes from the release of

a stop consonant until its voicing). Spanish voiced stop consonants (e.g., /b/ and /g/) are

characterized by the onset of voicing *prior* to the release, while in English the voice onset

occurs *after* the consonant release. As a consequence, the phonetic categories for stop

consonants are different between English and Spanish (e.g., Flege & Eefting, 1987; Muru &

Lee, 2017; Williams, 1977). Thus, if some bilingual infants in Mercure et al.'s study were

Spanish-learning, their phonetic categories for the English phonemes /b/ and /g/ would be

different than those of English-learning infants[5]. This could have in turn affected AV

discrepancy detection by the Spanish-learning subset of Mercury et al.'s sample.

To account for this issue of the different phonetic value/representations of stimuli for

monolinguals and bilinguals, in this study we presented infants with vowels that are shared

across our monolingual and bilingual sample. Spanish and Basque have a system of five

vowels /a, e, i, o, u/[6] (Hualde, 1991; Hualde et al., 2010). Infants in the current study were

---

[5] A recent finding suggests that AV processing of phonemes is indeed shaped by linguistic experience (Pons et al., 2009). In particular, Spanish- and English-learning infants were tested on AV matching of the phonemes /b/ and /v/. Crucially, these phonemes are distinct in English, but not in Spanish. Six- and 11-month-old English-learning infants were able to AV match previously heard /ba/ or /va/ syllables with the corresponding face mouthing /ba/ or /va/. In contrast, in Spanish-learning infants this AV ability was observed at 6, but not at 11 months. The decrease of AV matching ability in Spanish-learning infants is explained by attunement/narrowing to Spanish that does not distinguish between the /b/ and /v/ phonemes.

[6] Note that here we are referring to the vowel system in Castilian Spanish and most Basque dialects because subjects are recruited from the region where those dialects are spoken. Other dialects might differ in their vowel systems. For instance, the Soule dialect of Basque has an additional vowel /ü/ (Hualde, 1991).

recruited from regions where the Spanish and Basque vowel systems overlap, and Spanish and Basque monolingual and bilingual infants are exposed to the same vowel inventory. Therefore, the linguistic value of the stimuli is comparable across our monolingual and bilingual groups.

2. <u>Task difficulty</u>

An additional factor that could have influenced previous findings was that of task difficulty. Concretely, previous studies assessed infants in an AV speech processing task that may have been more challenging for bilinguals than for monolinguals (Pons et al., 2015; Tsang et al., 2018). Focusing first on studies on AV fluent speech (i.e., AV continuous speech), Pons et al. assessed Spanish-Catalan bilinguals in their dominant language (i.e., Spanish or Catalan). In a similar way, Tsang and colleagues assessed monolinguals (English) and bilinguals (English + other language) only in English, regardless of whether English was their dominant language. Therefore, the two studies also (i.e., in addition to differences in their sample as presented above) differed regarding whether they assessed infants in their dominant language or not. More importantly, in both studies bilinguals, but not monolinguals, were implicitly required to do an additional task. Specifically, monolingual infants were presented with a language that was familiar (i.e., native) or not to them. Thus, monolinguals' *implicit* task was to recognize whether the presented language was their native language or not. In contrast, bilinguals, in addition to recognizing whether the presented language was native to them, were also performing another implicit task: identifying which of the two languages they are familiar with was being presented (i.e., language discrimination). One can argue that bilinguals are always presented with the task of language discrimination when processing either of their native languages. Still, this is not the case for monolinguals. Overall, placing bilinguals in an implicit language discrimination task between their two languages means that the nature and difficulty of the task differed across the monolingual and bilingual groups. Specifically, AV speech processing of fluent speech might be a more difficult task for

bilingual than for monolingual infants. Task difficulty issue was also present in the Mercure et al. (2018) study where bilinguals' processing of possibly non-native phonemes might be a more difficult task than monolinguals' processing of their native phonemes.

In the current study, by using linguistic stimuli (i.e., vowels) that are shared across our Spanish-Basque monolingual and bilingual groups, we avoided implicitly engaging the bilingual group in a language recognition task. Therefore, we controlled that the task demands were similar for both groups. In addition to equalizing the task demands across the groups, we also varied the task difficulty independently. Specifically, we modulated the difficulty of the AV task by presenting infants with AV congruent and incongruent linguistic stimuli. Both monolingual and bilingual infants were tested in the AV processing of auditory and visual congruent stimuli (i.e., AV match condition), and in an overall more difficult task (but equally difficult for both groups), in which auditory and visual stimuli were incongruent (i.e., AV mismatch condition). That means that with our design we could disentangle *the role of bilingualism* per se in AV speech processing from setting bilinguals in a more difficult/different task. Details on how the match and mismatch AV stimuli were created are given in the method section (page 31).

3. Individual variability

Infant research data is often characterized as data with high individual variability (e.g., Colombo & Fagen, 1990) especially in infant looking data (e.g., Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016; Piantadosi, Kidd, & Aslin, 2014). Specifically, a recent study on the developmental changes in 6- to 12-month-old infants' attention to the mouth during AV speech processing revealed significant individual variability between subjects (Tenenbaum et al., 2015). The authors demonstrated that in response to AV speech, infants' tendency to look longer to the eyes than to the mouth (i.e., an eyes preference) or longer to the mouth than to the eyes (i.e., a mouth preference) was stable over development. For instance, if infants

showed an eye preference at 6 months of age, it is more likely that they will show an eye preference at 12 months too. However, previous studies did not account for infants' individual eyes-mouth preference (Mercure et al., 2018; Pons et al., 2015; Tsang et al., 2018). Therefore, it is unclear from studies on monolingual and bilingual attention to the mouth how much variation in eyes-mouth preference was accounted for by individual differences and how much was accounted for by being exposed to a different linguistic environment. To consider individual variability, the current study assessed data both longitudinally and cross-sectional. We opted for both assessments so we could compare the data collected longitudinally and cross-sectional, and also to compare these data with previous studies.

Because of the longitudinal nature of the study we employed two stimuli sets produced by two different speakers (as described in 2.3 section), one presented to infants at the first testing session and the other presented at the second session, to ensure that the findings from the second session were not confounded by previous exposure to the stimuli (i.e., a learning effect). Moreover, testing with two stimuli sets ensured the findings are not specific to one stimulus set, or some specific features of the stimuli (data described in Chapter 2 will underscore the importance of this point).

1.5   The current study

The overall aim of the dissertation is to study whether monolingual and bilingual infants follow similar developmental patterns when it comes to audiovisual speech processing. We measured infants' sensitivity to visual speech in an AV vowel matching task. Similarly to Altvater-Mackensen et al. (2015) we presented infants with AV match and mismatch conditions, across trials, using a single-face display. However, we used a perceptually more distinct vowel pair (/i/ and /u/) than Altvater-Mackensen et al., because the majority of previous infant AV matching studies demonstrated an AV matching ability in 4.5-month old infants using more visually/acoustically distinct vowel pairs (see Table 1-1).

Infants' looking behavior during AV matching task was collected using an eye-tracker. Infants' sensitivity to visual speech has been measured with two variables; (1) infants' AV matching ability, and (2) infants' attention to the eyes and to the mouth during the AV matching task. These two measures were analyzed across cross-sectional and longitudinal data.

Regarding the age groups tested in the current study, we focused on age groups for which bilinguals' increased sensitivity to visual speech has been reported. Specifically, previous studies demonstrated greater bilingual as compared to monolingual sensitivity to visual speech cues at 4 (Pons et al.,2015) and at 8 months of age (Sebastián-Gallés et al., 2012), hence we tested processing of vowels /i/ and /u/ in AV congruent and incongruent conditions in 4.5- and 8-month-old infants.

The remainder of this thesis is divided into three main sections:

1. The development of audiovisual matching ability of vowels in monolingual and bilingual infants- A cross-sectional study (Chapter 2)

2. The development of selective attention to the mouth in monolingual and bilingual infants during audiovisual matching task-A cross-sectional study (Chapter 3)

3. Individual variation in the development of audiovisual matching ability of vowels and selective attention to the mouth in monolingual and bilingual infants- A longitudinal study (Chapter 4)

Because we tested infants in their AV matching ability of vowels and their selective attention to the mouth during the same task, the methods used to collect the data presented in these three chapters are the same. Therefore, we will present the methods only in Chapter 2. Chapter 4 differs from Chapters 2 and 3 only with regards to the

participant section. Within each of the chapters, specific hypotheses will be presented,

followed by the results and discussion.

2   Chapter 2

The development of audiovisual matching ability of vowels in monolingual and bilingual infants. A cross-sectional study

2.1   Introduction

In the previous chapter we presented evidence that infants are sensitive to the visual modality of language. Infants as young as two months of age are able to match an auditorily presented vowel to a face mouthing that vowel (Patterson & Werker, 2003; see Table 1–1 for overview of studies on AV matching ability). It has also been proposed that bilingual infants develop greater sensitivity to visual speech cues (Pons et al., 2015; Sebastián-Gallés et al., 2012; Weikum et al., 2007). Specifically, they spend more time looking at the speaker's mouth compared to their monolingual peers (Pons et al., 2015).

Why would bilingual infants develop such greater sensitivity to visual speech? As discussed in section 1.4.1, bilinguals are exposed to diverse linguistic input due to developing two language systems. As a consequence of being exposed to two language systems, it has been proposed that bilingual infants might rely more on visual speech cues in order to keep their languages separated (Pons et al., 2015). However, recent evidence has challenged this idea. One study found that 6- to 12-month-old monolinguals and bilinguals did not differ in their attention to the mouth of a speaker (Tsang et al., 2018). Another study showed that 6.6- to 8-month-old monolingual, but not bilingual infants detected AV incongruence by attending longer to the mouth in response to AV incongruent syllables /ba/ and /ga/, in comparison to AV congruent syllables (Mercure et al., 2018). For a detailed description of the studies see section 1.4.3.

Yet, it remains unclear whether and under what conditions bilingual infants exhibit a higher sensitivity to visual speech cues than monolinguals. This question is relevant to various

theories of speech perception development (e.g., Murray, Lewkowicz, Amedi, & Wallace, 2016; Werker & Curtin, 2005) and to theories of bilingual acquisition (e.g., Bialystok & Werker, 2017). To systematically assess monolingual and bilingual infants' sensitivity to visual speech, we employed an AV matching task. We presented Basque and Spanish monolingual and bilingual infants with a speaker mouthing a vowel that either matched or mismatched what the infants heard. If infants demonstrate AV matching ability, their looking times to AV match (congruent) and AV mismatch (incongruent) conditions should differ (as observed in Altvater et al., 2015). Moreover, if bilingualism increases sensitivity to visual speech, then we expected bilingual infants to behave differently than the monolinguals in this task (e.g., longer looking times in response to the mismatch trials). Further details regarding the predictions are given in the next section.

A previous AV matching study that employed a similar design to our study demonstrated that 5.5- to 6-month-old (German-learning) infants looked longer to AV match than AV mismatch trials, but only for certain vowel pairs – in particular, only when the /a/-/o/ pair was presented, but not the /a/-/e/ pair (Altvater et al., 2015). As mentioned in the previous chapter, we selected the vowel pairs that are visually/acoustically more distinct (i.e., /i/-/u/), similarly to the most prior AV vowel matching studies (see Table 1-1). Additionally, by testing vowels in Spanish and Basque monolingual and Spanish-Basque bilingual infants, we tested stimuli that are shared between groups. This further ensured that the task demands were similar across the groups (see section 1.4.4 where we described research limitations if stimuli/task differed for the bilingual and monolingual group). Further details on the task are given in the Methods section.

As described, in this study, we tested two age groups: 4.5- and 8-month old infants. The majority of studies demonstrated AV vowel matching ability in 4.5-month-old monolingual infants (see Table 1-1), but, it remains unclear whether bilingualism would

modulate the ability at this age. Recent work suggests that it might – in one recent study, 6.6- to 8-month-old monolinguals, but not bilinguals detected the AV incongruence (Mercure et al., 2018). However, this finding is inconsistent with a previous study that observed greater bilingual sensitivity to visual speech (i.e., to the mouth of a speaker; Pons et al., 2015). We tested 8-month-olds to understand the relationship between bilingualism and sensitivity to visual speech at this age.

2.2   Predictions

<u>4.5-month-old group</u>

Considering previous robust findings on AV matching ability in monolingual 4.5-month-old infants (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Yeung & Werker, 2013), we predicted that the monolingual group would exhbit sensitivity to AV congruency in the current study. The direction of the effect (i.e., longer looking times to AV match or AV mismatch condition) was difficult to predict, given that a recent study demonstrated that the direction has varied in 3- to 6-month-old infants. For instance, for vowels /i/ and /u/ 3-month-olds attended longer to AV match, whereas 6-month-olds attended longer to AV mismatch conditions (Streri, Coulon, Marie, & Yeung, 2016).

More importantly, we predicted that if bilingualism at 4 months of age increases sensitivity to visual speech cues (Pons et al., 2015), then our bilingual group should be more sensitive to visual speech than their monolingual peers. Such greater sensitivity to visual speech could be observed by showing a greater AV matching ability (i.e., more bilingual than monolingual infants showing AV match ability) in the bilingual compared to the monolingual group.

The majority of AV matching studies have tested infants from 4 to 6 months of age and there is little evidence on AV *vowel* matching ability in 8-month-old infants. However, recent studies have shown that around 8 months of age monolingual infants detected AV mismatch of syllables by looking longer to the mouth in response to incongruent AV syllables (Danielson et al., 2017; Mercure et al., 2018; Tomalski et al., 2013). Therefore, we expected that 8-month-old infants would demonstrate AV vowel matching ability. Regarding the direction of the effect, following the previously reported AV incongruence preference (Danielson et al., 2017; Mercure et al., 2018; Tomalski et al., 2013), we predicted that monolinguals would attend longer to the AV incongruent (mismatch) condition.

If bilingualims increases the sensitivity to visual speech at this age (Sebastián-Gallés et al., 2012; Weikum et al., 2007), then our 8-month-old bilinguals should be more sensitive in their AV mismatch detection than monolinguals (i.e., more bilingual than monolingual infants showing AV match ability).

2.3   Method

2.3.1   Participants

The sample size was determined based on a previous study (Pons et al., 2015) that tested 20 monolingual and 21 bilingual infants from each age group (4-, 8-, and 12-month olds). We aimed to test the same sample size. As previously described, a subset of infants was tested longitudinally, as well as cross-sectionally. For this reason, we employed two stimuli sets (i.e., two speakers) to avoid any potential learning effect in infants that were tested longitudinally. Therefore, we aimed to have a sample size comparable to Pons' study, for each stimuli set (i.e., each speaker), leading to ~40 monolingual and 40 bilingual infants in each age group. The participants are described separately for the two age groups below.

4.5-month-old monolingual and bilingual infants

A total of 84 monolingual and bilingual infants were included in the final analysis. The monolingual group consisted of 42 infants (average age 4.5 months, range 123-146 days, 17 female infants). The bilingual group consisted of 42 infants (average age 4.5 months, range 122-146 days, 19 female infants). The full description of participants across the two speakers (i.e., two stimuli sets) is given in Table 2-1. Participants were recruited from the Spanish-Basque bilingual region of San Sebastian, Spain. Infants' exposure to Spanish and/or Basque was carefully evaluated via a language exposure questionnaire (see Appendix A. 2) previously used in Molnar et al. (2014). The questionnaire estimated the overall exposure to Spanish and Basque that infants received via interactions with various caregivers at home and at the daycare. Only infants exposed to at least 95% to one language were included as monolingual infants. Average exposure to one of the languages in the monolingual group was 99.4% (range 95.3-100, 17 exposed to Basque and 25 to Spanish). For the bilingual group, only infants exposed at least 15% to their other language were included. Average exposure to the other language in the bilingual group was 32.5% (range 15-47.6; 15 infants dominantly exposed to Basque, 17 to Spanish, and 10 balanced in their exposure to Spanish and Basque).

Fifty additional infants were tested but their data was not included in the final sample, resulting in an attrition rate of 37%, similar to the attrition rates reported in other eye-tracking studies at this age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). Data was excluded if infants did not finish the experiment due to crying or fussiness (16), failure to calibrate or a very poor calibration (18), equipment failure (1), infants' inattentiveness-looking away from the screen immediately when a trial starts (4), extreme movements (1), not looking at the face of a speaker or not looking at the face for more than one block (3), and if it was hard to determine whether an infant was monolingual or bilingual based on their language background since their exposure to the other language (L2) was 85-95% (7). All the recruited infants were healthy and born at full-term with no reported vision or hearing issues.

8-month-old monolingual and bilingual group

A total of 75 monolingual and bilingual 8-month old infants were included in the final analysis. Thirty-nine monolinguals were assessed (average age 8 months, range 240-255 days, 20 female infants). Thirty-six bilingual infants were assessed (average age 8 months, range 237-259 days, 18 female infants). The full description of participants across the two speakers is given in Table 2-1. The average exposure to one language in the monolingual group was 99% (range 95-100; 20 infants exposed to Basque, and 19 to Spanish). Only infants exposed at least 15% to the other language were included as bilingual infants. The average exposure to the other (less dominant) language was 33 % (range 15-50; 13 infants dominantly exposed to Basque, 14 to Spanish, and 9 balanced in their exposure to Spanish and Basque).

An additional thirty-seven infants were tested but their data were excluded from the final sample, resulting in an attrition rate 33%, common to the eye-tracking studies at this age (Ayneto & Sebastian-Galles, 2017; Pons et al., 2015). Data was excluded if infants did not finish the experiment due to crying or fussiness (13), failure to calibrate or a very poor calibration (10), equipment failure (4), infants' inattentiveness-looking away from the screen immediately when a trial starts (1), extreme movement (3), and because it was hard to determine whether an infant was monolingual or bilingual based on their language background since their exposure to other language (L2) was 85-95% (6). All the recruited infants were healthy, full-term infants, with no reported vision or hearing issues.

Table 2-1

Demographic description of participants across the two speakers (i.e., stimuli sets).

Age group	Speaker	The age range in days	Nº of participants	Nº of female participants	Language group	Exposure to L1/L2 (%)
4.5 months	Speaker 1	123-144	20	9 f	monolingual	99.3 (96.2-100)
	Speaker 2	128-146	22	8	monolingual	99.4 (95.3-100)
	Speaker 1	123-142	20	9	bilingual	29.8 (15-44.1)
	Speaker 2	128-146	22	10	bilingual	33.3 (16.7-47.6)
8 months	Speaker 1	241-252	20	10	monolingual	99.2 (95-100)
	Speaker 2	240-255	19	10	monolingual	98.5 (95.3-100)
	Speaker 1	241-259	18	10	bilingual	30.3 (14.5-50)
	Speaker 2	237-253	18	8	bilingual	34.9 (15.4-50)

Note. Nº refers to a number of participants (sample size). The range of exposure to L1/L2 is given in the brackets. L1 refers to infants' dominant language, whereas L2 refers to infants' less dominant language. For monolinguals, exposure to L1 is presented.

2.3.2 Stimuli

In this study, two vowels were used; /i/ and /u/. Both vowels are part of the Basque and the Spanish vowel inventories (the vowel categories across the languages are highly overlapping, if not virtually identical).

Video recording of the vowels

We recorded videos of two female Spanish-Basque native speakers while they were producing /i/ and /u/. The speakers were recorded separately with the Canon LEGRIA HF G10 camera, with a white background behind the speaker, in a 1920 x 1080 resolution. The frame rate was 25 frames/second. The speakers wore a white t-shirt and no make-up. The speakers were asked to produce the vowels in an infant-friendly style, gazing at the camera.

They were instructed to repeat the vowel for a minute and a half, with the same duration and intensity throughout. A flashing light on a screen instructed speakers to repeat the vowel every ~2 seconds to maintain similar inter-stimulus interval between repetitions. After each repetition, the speakers were instructed to fully close their mouth while keeping facial expressions neutral.

Match and Mismatch videos

To create an AV matching task, we first had to create AV match and AV mismatch stimuli. Both match and mismatch stimuli were created by speakers dubbing their own silent videos. To create AV match stimuli, speakers dubbed the same vowel they were watching themselves produce. For the mismatch stimuli, the speakers dubbed a different vowel from the one they were watching themselves produce (i.e., for the visually presented /i/, speakers produced /u/, and for the visually presented /u/, speakers produced /i/). The dubbed audios were recorded in a sound-attenuated room with Marantz PMD1671 recorder and Sennheiser noise-reducing microphone. Speakers were instructed to maintain similar intensity across the tokens of one vowel. The recorded audios were edited in Audacity 2.0.5 software for background noise reduction, while the intensity was normalized on 65 dB using Praat software (Boersma & Weenink, 2018).

To confirm that the auditory vowels recorded for the match and mismatch conditions were both perceived within the same intended vowel category (/i/ or /u/), the audio recordings were presented to 18 adult Spanish-Basque speakers in a categorization task. The results of the task showed that, regardless of whether the auditory vowels were recorded for the match or for mismatch videos, they were reliably categorized as vowels /i/, and /u/, respectively (minimum 98% of accuracy for each vowel category).

Finally, the visual and auditory signals were mixed using Adobe Premier Pro CS4 software to create both the match and mismatch trials. Newly recorded match and mismatch

audios were synchronized with the onset and offset of the original audios in the recorded videos such that the auditory and visual signals were synchronized. Each matching and mismatching video contained nine sequential vowel productions (different tokens) with ~ 2 seconds interval between the vowels, creating a ~30-second long video. Each video started with a natural but friendly facial expression. Duration, intensity, pitch, and inter-stimulus interval across vowel tokens within one speaker were selected to be similar, while these measures varied between speakers allowing for a natural between-speaker variation. Details on auditory measures across speakers are presented in Table 2-2.

Each mixed AV video was presented over the whole 34x27cm screen, with faces presented centrally, approximately 10cm away from left and right corners, and 3cm from upper and bottom corners. The mixed AV videos were edited to make them similar with respect to each speaker's size on the screen, brightness, and saturation. Example frames from the final videos of the two speakers are shown in Figure 2-**4.**

2.3.3   Procedure

Data collection was divided into four consecutive stages, illustrated in **Figure 2–1**; Recruitment (1), Pre-test (2), Test (3), and Post-test (4). Each of the stages will be discussed.
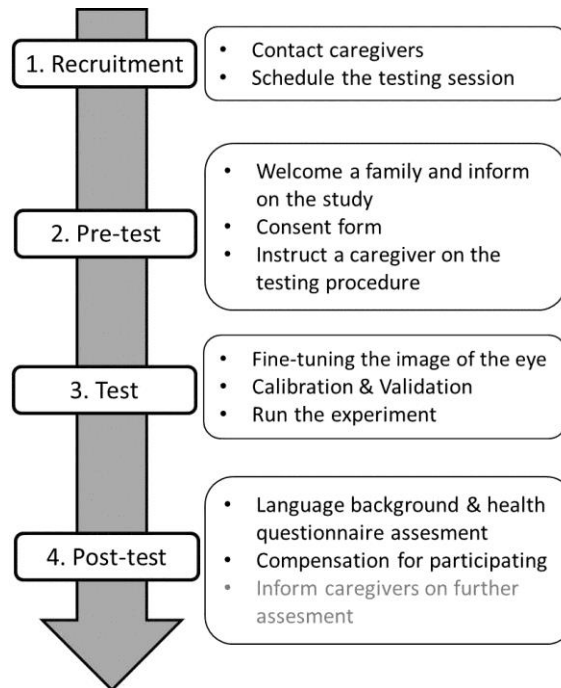

Figure 2-1. Schematic illustration of the study's procedure.

1)   Recruitment

Families were contacted in person for the first time at a local hospital, shortly after the infants were born. Families interested in the study were included in the database for further contact and were contacted over the phone one-to-two weeks prior to the experiment and were informed about the study. They were also informed that the first testing session would take place when infants turn 4.5 months, and the second one when they turn 8 months. After informing the families, the date for the testing was scheduled.

2)   Pre-test

At the age of 4.5 months, infants were assessed for the first time. The current study assessed infants' looking behavior using an eye-tracker. Details on the apparatus and eye-

tracking methodology are given in Appendix A. 1. Caregivers were informed about the study in general and how the eye-tracking technique works was explained to them. Finally, they were given a written consent form to sign (Appendix A.4)

During the experiment, infants were seated in their caregivers' laps, facing a monitor placed 55-60 cm away from them. The caregivers wore noise-canceling headphones and opaque glasses to prevent them influencing their infant's behavior (see the illustration for the testing procedure given in Figure 2-2). Caregivers were instructed not to interact with the infant during the testing session. A removable sticker with highlighted black and white contrast was placed on an infant's forehead, so that the eye-tracker could accurately detect the pupil and account for the infant's head movements.

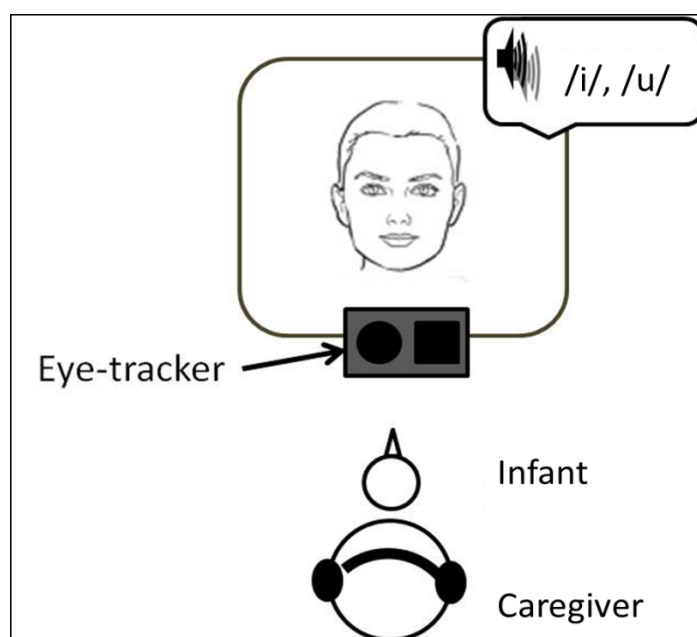*Figure 2-2.* Schematic illustration of the testing procedure. The caregiver wore headphones in order not to influence their infant's behavior. The infant is facing the display monitor, while the eye-tracker records its eye-movements.

3) <u>Test</u>

The test phase started by fine-tuning the image from the eye-tracker's camera by adjusting the lens' focus and the distance between the camera and heads of the infants. To

maintain infants' attention to the screen during this short process, an infant-friendly video was played on the screen, together with infant-friendly sounds. Once the experimenter had confirmed that the eye-tracker correctly detected the pupil and the target, the calibration started. A five-point calibration and validation system was used, with 1000ms between-calibration-points intervals. The calibration points appeared in each corner of the screen and one in the center of the screen. The location of calibration points was randomized, but the first and the last positions were the same, as recommended by the manufacturer. The calibration points were small looming targets, as recommended and provided by the manufacturer for infant data acquisition (see *Figure 2-3*). Once the experimenter confirmed the eye-gaze position on a calibration target, the next calibration target appeared on another location. Finally, if the calibration failed, it was repeated until the calibration was correct. Three different targets were used across infants, as illustrated in *Figure 2-3*. One calibration target was employed throughout the calibration process, to make sure the calibration was not modulated by possible differences between the targets. However, different targets were employed within one testing session as the drift correction targets, which will be explained further below.

Figure 2-3. Three different calibration, validation, and drift-correction targets employed in the current study.

After calibration, validation was performed. The validation is the same process as calibration, but further evaluates the accuracy of the calibration. Here, we attempted the validation with all infants that provided accurate calibration. However, if the validation was incomplete, we still ran the study, relying on the calibration process. Note that in infant eye-

tracking research, the validation is often omitted/uncomplete due to infants' limited attention and fussiness, especially in those younger than six months of age. After the calibration and validation, the experimental trials started. Each experimental trial started with an infant-friendly, small attention-getter (one of the three illustrated in *Figure 2-3*) displayed centrally on the screen, accompanied by an infant-friendly sound. The attention-getter also represented the drift correction for the eye-tracking system (correcting for small drifts in the estimation of the gaze position), by which a high eye-tracking accuracy is maintained throughout the session. When the infant's gaze at the attention-getter was registered and the drift correction was performed, the trial began.

The trial presentation was fully infant-controlled—when an infant looked away for more than two seconds, the trial ended and the attention-getter appeared on the screen. Infants' looking behavior (the duration of time they spent looking at any point on the screen) was also coded-online by an experimenter who monitored the session in a separate room. The maximum trial duration was 30 seconds. The whole experiment lasted about 20 minutes.

To record infants' overall behavior during the session, the whole testing session was also recorded with a separate GO-PRO HERO LCD camera, placed next to the monitor, recording the infants' behavior. This recording served to control whether infants were attentive during the session, whether caregivers followed the instructions, whether infants were crying, etc.

4) Post-test

After testing, caregivers had the opportunity to see the stimuli that their infant had been presented with, as well to see a video showing the infant's looking behavior. After this feedback, parents reported the infant's exposure to Spanish or Basque language via the language questionnaire that estimates the proportion of exposure to each language over time (the same questionnaire was used in Molnar, Gervain, & Carreiras, 2014). Note that we did

not assess infants exposed to any other language than Spanish and/or Basque. The details on the language questionnaire are given in Appendix A. 2. Additionally, parents reported on infant's health and general development (Appendix A. 3). All parental reports were assessed in the language the parent in question felt more comfortable with (i.e., either Spanish or Basque).

Participants were provided with an infant-friendly toy together with a "participation certificate" in return for participating in the study. Finally, caregivers were informed about the second testing session, in the infant's 8th month. The second testing session followed exactly the same procedure as the first one, except that infants were presented with a different speaker. Note that we did not conduct a second testing session with the infants whose data we excluded as explained in the section on participants.

2.3.4   Experiment design

Each infant was presented with the vowels /i/ and /u/, both in a *match* and *mismatch* condition. In the match condition, the auditory and visual signals corresponded (i.e., visual /i/ was paired with auditory /i/, and visual /u/ was paired with auditory /u/). In the mismatch condition, auditory and visual signals did not correspond (i.e., visual /i/ was paired with auditory /u/, and visual /u/ was paired with auditory /i/). The trials were grouped into two blocks: (1) vowel /i/ and (2) vowel /u/. Each block consisted of three sequentially presented matched and three mismatched trials. In total, each infant was presented with 12 trials (see **Figure 2-4**). The order of the matched and mismatched trials and the vowel blocks was counterbalanced across infants. The two speakers were presented as a between-subject factor; therefore, each infant saw only one speaker within one session.
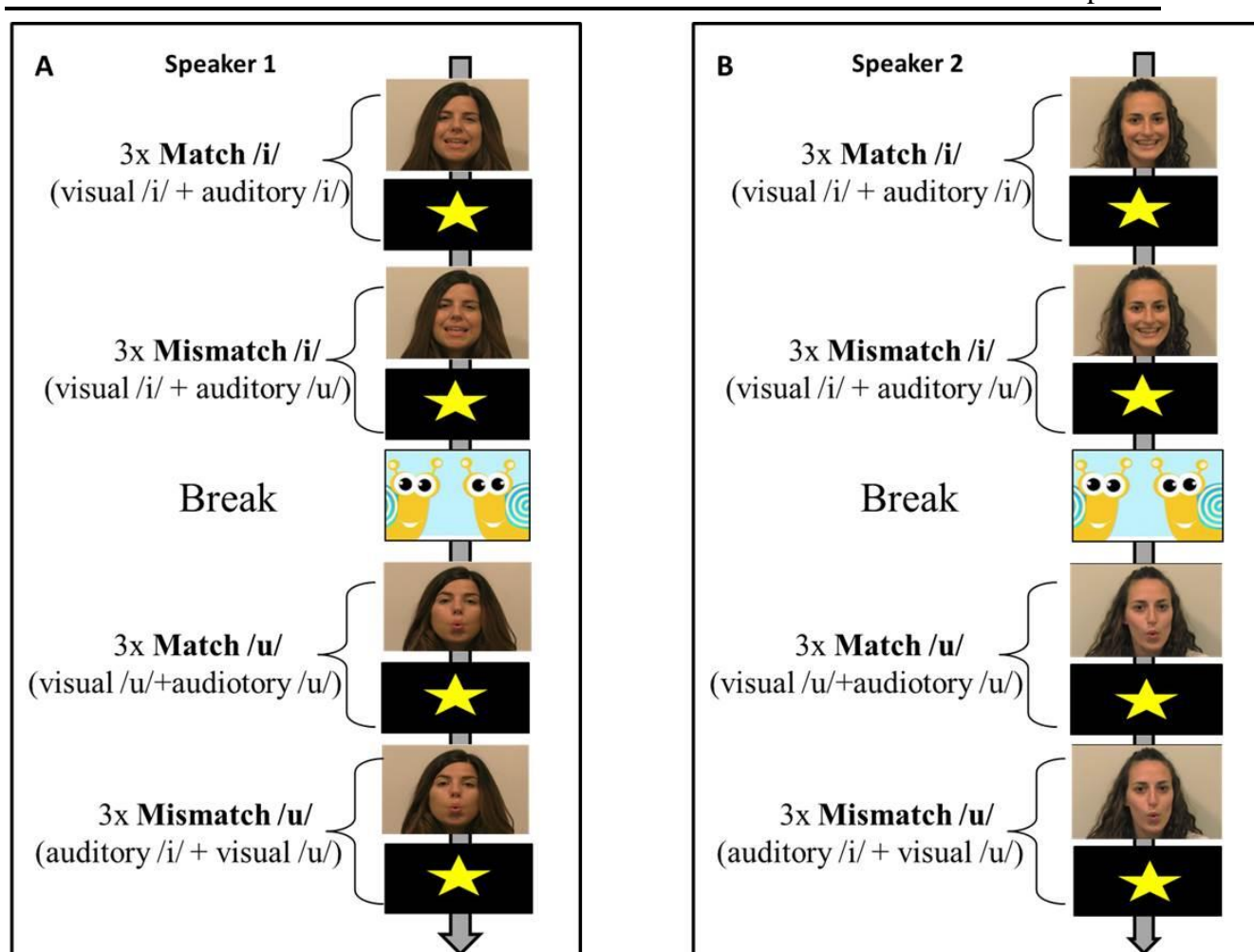
Figure 2-4. Experiment design. Infants were presented with blocks of three match and three mismatch trials for one vowel, followed by a short break, after which they were presented with another block of match and mismatch trials for the other vowel. The order of match and mismatch, as well /i/ and /u/ trials was counterbalanced. Every trial began with an attention-getter. Two speakers (example frames given) were presented as a between-subject factor, therefore each infant was presented with one speaker.

2.4   Results

2.4.1   The audiovisual matching ability in monolingual and bilingual infants

To assess infants' audiovisual matching ability (i.e., sensitivity to visual speech), we compared their looking times between AV match and mismatch trials. To this end, we calculated the time infants attended to the screen for each trial (12 in total). Following previous AV matching studies, the total looking time for each trial was calculated for each infant separately as the sum of all fixations on the entire screen recorded by the eye tracker during one trial (Altvater-Mackensen & Grossmann, 2015; Altvater-Mackensen et al., 2015).

First, to test whether infants in the current study exhibited any AV matching ability and whether bilingualism affected this ability, we compared mean looking times between match and mismatch conditions across the monolingual and bilingual infants. Looking times (in milliseconds) for each infant were averaged across match (6 trials) and mismatch (6 trials) conditions. We conducted two separate repeated-measure mixed 2x2 ANOVAs for each age group on average looking times, with *Condition* (match and mismatch) as a within-subject factor, and *Group* (monolingual and bilingual) as a between-subject factor. The results will be described separately for the 4- and 8- month-old group. Note that including *Gender* in the analyses did not produce significant main effects, or interactions with *Condition* (all *Fs* < 2.72, all *ps* > .15) in either of the age groups. Therefore, we did not include gender as a factor in further analyses.

The analysis for the 4-month old group revealed a significant effect of *Condition* $F_{(1, 82)} = 6.94$, $p = .01$, $\eta^2_G = .03$, with longer looking times to the mismatch (M = 12974, SD = 5499) than to the match condition (M = 11207, SD = 5237), as shown in the left panel of *Figure 2-5*. No effect of group or *Condition* x *Group* interaction reached significance (all *Fs*

< .36, all *ps* > .55, all η^2_G[7] < .00), suggesting that monolingual and bilingual 4-month-old infants did not differ in their AV matching ability. Conducting the same ANOVA with the 8-month old group, however, revealed no effect of condition, group, or interaction (all *Fs* < 1.4, all *ps* > .24, all η^2_G < .007), as can be seen on the right panel of ***Figure 2-5***. Thus, 8-month old infants, regardless of their language environment, spent the same amount of time looking at congruent and incongruent trials.
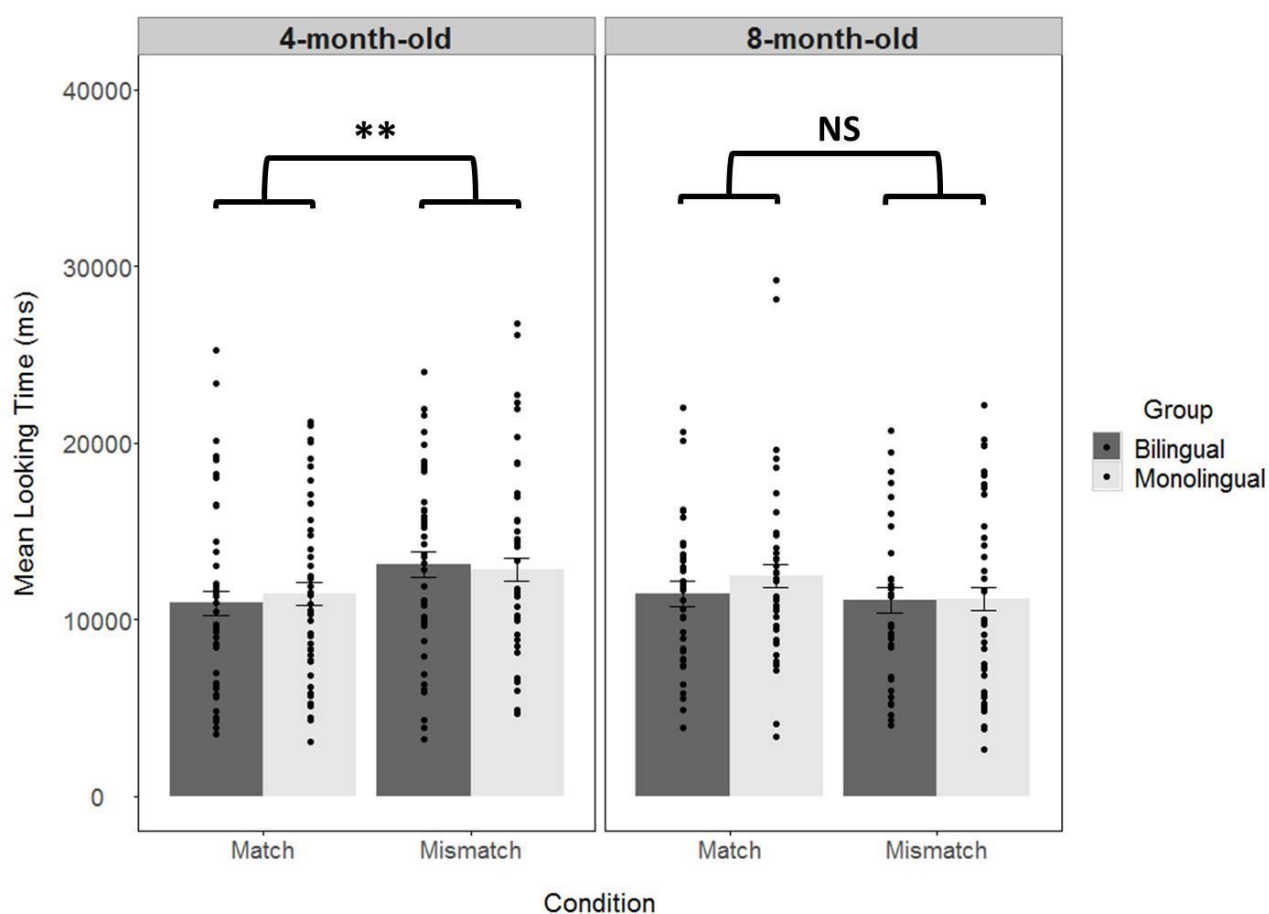
Figure 2-5. Mean looking times (in milliseconds) for match and mismatch condition across the bilingual and monolingual group, at 4 and 8 months of age. Points represent individual infants' scores. Error bars represent +/- 1 SE, asterisks indicate a significance level of ** p ≤ .01, NS indicates a non-significant effect.

---

[7]As suggested by some similar previous studies (Bakeman, 2005; Olejnik & Algina, 2003), eta squared (η^2_G) was employed to estimate the magnitude of effect size for repeated-measures and mixed designs, eta squared (η^2) was used for non-repeated ANOVAs and Cohen's d (d) for t-tests.

We also tested whether the amount of exposure to Spanish or Basque modulated AV mismatch detection within the monolingual and bilingual groups. For the monolingual group, we conducted two separate ANOVAs on mean looking times with *Language Group* (Spanish and Basque monolingual infants) as a between-subject factor, and *Condition* (match and mismatch) as a within-subject factor, separately for the two age groups. The analysis for the 4-month-old group revealed no significant interaction between the factors $(F_{(1, 40)} = .50, p = .48, \eta^2_G = .004)$. Similarly, the analysis of the 8-month-old group revealed no significant interaction $(F_{(1, 37)} = 1.29, p = .26, \eta^2_G = .010)$.

For the bilingual group, we included the amount of exposure to Basque (as measured in percentages based on the language background questionnaire) as a covariate, instead of dividing participants into three groups (Spanish-dominant, Basque-dominant, and balanced) and losing the statistical power. We conducted two separate ANCOVAs on the mean looking times with *Condition* (match and mismatch) as a within-subject factor, and percentage of exposure to Basque as a covariate. The analysis for the 4-month-old group revealed that the interaction between *Condition* and *Basque exposure* did not reach significance $(F_{(1, 40)} = .18, p = .67, \eta^2_G = .002)$. The findings for the 8-month-old group were similar $(F_{(1, 37)} = .00, p = .96, \eta^2_G = .000)$.

2.4.2   The effect of a speaker on the audiovisual matching ability

In addition to conducting analyses to test our primary question about AV matching ability in monolingual and bilingual infants, our design also allowed us to explore a secondary question. As mentioned earlier, a recent finding suggest that the visual perceptual salience of vowels interacts with the AV speech matching ability. Specifically, German-learning 5.5- to 6-month-old infants were assessed in their ability to match the auditory and visual components of the vowels /a/, /e/, and /o/. Infants exhibited sensitivity to matching AV cues (i.e., they attended longer to match over mismatch trials) for the /a/-/o/ pair, but not for the /a/-/e/ pair ((Altvater-Mackensen et al., 2015). The authors suggested that this difference was due to the fact that visually, the vowels /o/ and /a/ are more distinct than /a/ and /e/. That is, the lips are rounded for /o/ (and not for /a/), but are similarly spread for both /a/ and /e/. Thus, the lip-rounding associated with the vowel /o/ may have facilitated the detection of the mismatch between /o/ and /a/.

Findings from Altvater-Mackensen et al. (2015) lead to the interesting question of whether other naturally occurring variations in the visual features of vowels may also modulate AV matching abilities in infants. In particular, it has been observed that when it comes to the visual articulatory features of speech sound production, differences exist not only between vowels (e.g., /e/ vs. /o/, as described in Altvater-Mackensen et al., 2015), but also inter-speaker (or individually). For instance, notable individual differences have been observed for jaw height in the production of several American-English vowels (e.g., Johnson, Ladefoged, & Lindau, 1993). Inter-speaker differences in speech sound production have also been observed in caregivers, who vary greatly in their visual-articulatory characteristics when producing infant-directed speech. These inter-speaker variations in caregivers' speech have been shown to have consequences for language acquisition, such that infants of caregivers who produce more exaggerated articulatory features learn vowel categories faster (Liu, Kuhl, & Tsao, 2003).

Thus, the secondary question that we wanted to test is whether (1) visual differences between vowels and (2) visual differences between the speakers' articulation affect AV matching ability in monolingual and bilingual infants. Similarly to Altvater- Mackensen et al. (2015), we quantified the differences in the production parameters of the vowels and speakers using visual and acoustic measures. Specifically, when it comes to the visual measures, we were interested in whether speakers differ in their lip-spreading while producing the vowels /i/ and /u/. If infants' AV matching ability interacts with the between-vowel or the between-speaker differences in lip-shape, then the two factors *Visual vowel* and *Speaker* should modulate the AV matching ability in infants.

2.4.2.1 Quantifying the visual and acoustic difference across vowels and speakers

We measured the visual difference across the vowels and across the speakers similarly to Altvater-Mackensen et al. (2015). Namely, vowel articulation was measured via the horizontal (from left to right lip corner) and vertical (from upper to lower lip) lip-opening in pixels on a still video frame during a fully articulated vowel (see *Figure 2-6*, left panel). Visual inspection of the left panel in *Figure 2-6* suggests that the two vowels are distinct on horizontal lip-opening. Specifically, the vowel /i/ is produced with lips being more spread than for the vowel /u/. Furthermore, the speakers differed in how much they spread their lips while producing the vowel /i/. Specifically, Speaker 2 spreads her lips more while producing the vowel /i/ than Speaker 1, which creates a larger visual distinction between the /u/ and /i/ vowel productions in Speaker 2, as compared to Speaker 1.

The acoustic characteristics of the two vowels, as measured by first (F1) and second (F2) formant frequency (see *Figure 2-6*, right panel), demonstrated that the two vowels formed two very distinct acoustic categories, especially on the F2 dimension (vowel backness). Although the speakers seemed to be acoustically similar in their production of the vowel /u/, they appeared to be slightly different for the vowel /i/, in line with the visual

features of the vowels described above. Speaker 2 seemed to produce the vowel /i/ with lower

F1 and F2 features, suggesting that Speaker 2 produced the vowel /i/ with a slightly higher

and less frontal tongue position than Speaker 1. Furthermore, overall, Speaker 2 produced

vowels with a higher mean pitch and slightly shorter (~20 ms) tokens of /i/ than Speaker 1

(see Table 2-2).
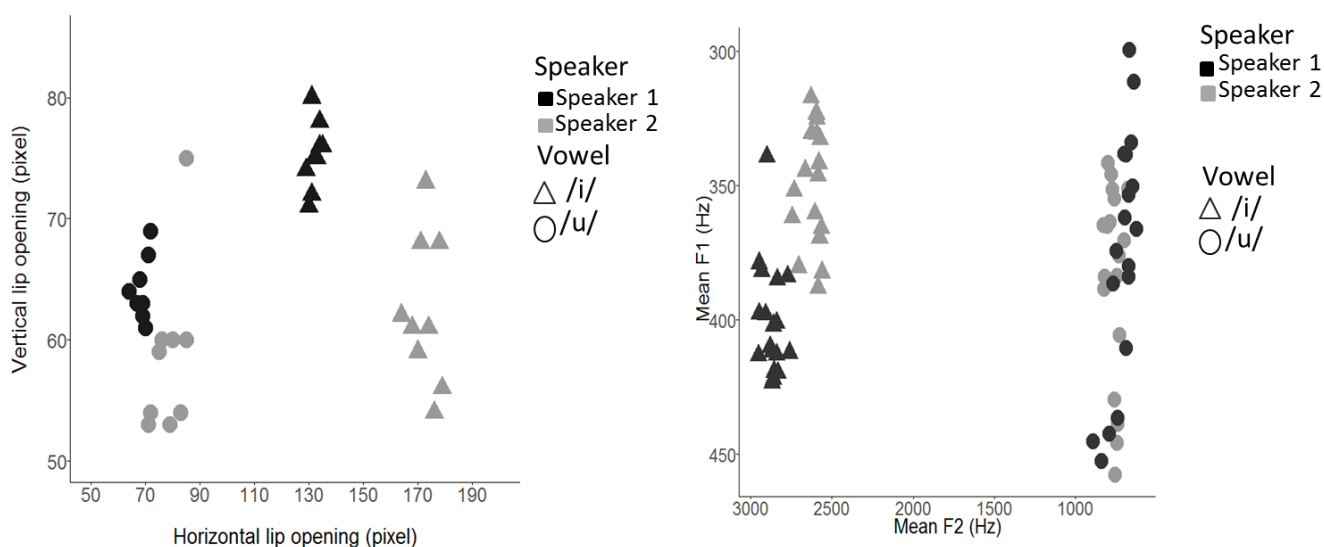
Figure 2-6. Measures of visual and acoustic vowel distinctiveness across speakers. The left panel depicts vertical and horizontal lip-opening values (in pixels) during the full articulation of each token (individual points). The right panel shows the first (F1) and second (F2) formant frequency of each token (individual points). In both panels, black points indicate Speaker 1, while the gray ones indicate Speaker 2. Circular markers indicate the vowel /u/, and triangles indicate the vowel /i/.

Table 2-2

Acoustic properties of matched and mismatched auditory stimuli across the two speakers

Condition	Speaker 1				Speaker 2			
	Match		Mismatch		Match		Mismatch	
Stimuli	Visual /i/ Auditory /i/	Visual /u/ Auditory /u/	Visual /i/ Auditory /u/	Visual /u/ Auditory /i/	Visual /i/ Auditory /i/	Visual /u/ Auditory /u/	Visual /i/ Auditory /u/	Visual /u/ Auditory /i/
Mean duration (s)	1.47	1.46	1.38	1.45	1.26	1.47	1.32	1.52)
Duration range (s)	1.32 -1.66	1.26-1.65	1.10-1.60	1.2-1.64	1.10-1.33	1.31-1.60	1.23-1.44	1.42-1.60
Mean pitch (Hz)	225.80	226.84	226.52	220.54	252.89	252.01	274.38	264.46
Pitch range (Hz)	223.62-229.26	191.80-232.60	225.45-228.00	219.50-222.50	249.30-255.60	247.30-256.80	269.50-277.31	231.91-268.40
Mean intensity (dB)	64.90	65.10	65.30	64.80	65.90	65.20	65.70	64.20
Intensity range (dB)	63.9-65.7	63.9-69.2	I64-66.2	I64.0-65.7	I65.3-66.9	I64.9-65.6	65.2-66.1	I63.1-64.7
Mean ISI (s)	2.03	1.96	2.10	1.98	1.95	1.87	1.88	1.88
ISI range (s)	1.83-2.20	1.75-2.20	1.91-2.33	1.74-2.19	1.53-2.21	1.77-1.96	1.40-2.20	1.70-2.05

To summarize, we observed that (as would be expected based on their phonetic properties) the productions of the two vowels in our stimuli set differ in lip-spreading (the vowel /i/ is produced by spreading the lips more than the vowel /i/). More interestingly, our speakers differed in their articulation. Specifically, they differed in how much they spread their lips while producing the vowels (i.e., Speaker 2 spread her lips more when articulating the vowel /i/).

2.4.2.2 Inter-speaker and inter-vowel effects on infants' AV matching ability

In order to test whether the observed visual/acoustic difference between the vowels and the speakers affected infants' AV matching ability, we included *Speaker* (Speaker 1 and Speaker 2) and *Visually presented Vowel* (/i/ and /u/) in the analysis. Two ANOVAs were run on mean looking times with *Group* (monolingual and bilingual) and *Speaker* (Speaker 1 and Speaker 2) as between-subject factors, and *Condition* (match and mismatch) and *Visually presented Vowel* (/i/ and /u/) as within-subject factors, separately for the two age groups.

The 4-month old group

The analysis of the 4-month-old group confirmed the previously reported (see Section 2.4.1) main effect of *Condition* ($F_{(1, 80)} = 6.96$, $p = .01$, $\eta^2_G = .01$). In addition, we found a significant effect of *Vowel* ($F_{(1, 80)} = 6.12$, $p = .02$, $\eta^2_G = .02$), with longer looking times to the vowel /u/ ($M = 13161$, $SD = 7590$) than to the vowel /i/ ($M = 10988$, $SD = 7645$). The interaction between *Condition* and *Speaker* ($F_{(1, 80)} = 14.13$, $p = .0003$, $\eta^2_G = .02$) was also significant, and there was a four-way interaction between Group, Speaker, Condition, and Vowel that approached significance ($F_{(1, 80)} = 3.28$, $p = .07$, $\eta^2_G = .009$; **Figure 2-7**). No other effects reached significance (all $Fs < 2.53$, all $ps > .12$, all $\eta^2_G < .009$). Note that we did not observe a main effect of Speaker, indicating that infants' visual attention was not modulated by an *overall* preference for one speaker over the other.

Following up on the *Condition* and *Speaker* interaction, pairwise comparisons (with Tukey adjustment for multiple comparisons) indicated that looking time in the match (*M* = 12802, *SD* = 7824) and mismatch (*M* = 12909, *SD* = 7624) condition did not differ for *Speaker 1* ($t_{(80)}$ = .77, *p* = .44). However, for *Speaker 2,* infants looked longer in the mismatch (M = 13770, SD = .7765) than in match (M = 9759, SD = 7070) condition ($t_{(80)}$ = 4.63, *p* < .001).

Although the 4-way interaction only approached significance, we also performed follow up analyses to see if the language groups differed depending on the vowel and speaker. The planned pairwise comparisons indicated that both monolingual and bilingual infants processed the match and mismatch conditions similarly for Speaker 1, regardless of the vowel (all *ps* > .15). However, for Speaker 2, the bilingual group looked longer during the mismatch condition for both vowels ($t_{vowel/i/}$ = 2.01, *p* = .04; $t_{vowel/u/}$ = 2.51, *p* = .01), whereas for the monolingual group the difference was significant for /i/, but not for /u/ ($t_{vowel/i/}$ = 3.12, *p* = .002; $t_{vowel/u/}$ = .52, *p* = .59). The results are illustrated in Figure 2-**7**.

Figure 2-7. Mean looking times (in milliseconds) in the 4.5-month-old group for the match and mismatch condition, for the vowels /i/ and /u/ and for the two speakers, across the bilingual and monolingual group. Points represent individual infants' scores. Error bars represent +/- 1 SE, * indicates a significance level of * $p \leq .05$; ** indicates a significance level of ** $p \leq .01$; NS indicates a non-significant effect.

The 8-month-old group

The analysis for the 8-month-old group revealed no significant main effects or interactions (all $Fs < 1.53$, all $ps > .22$, all $\eta^2_G < .008$), suggesting that looking times in the older group were not modulated by condition, group, speaker, or by the vowel the infants were presented with. The results are illustrated in ***Figure 2-8***.

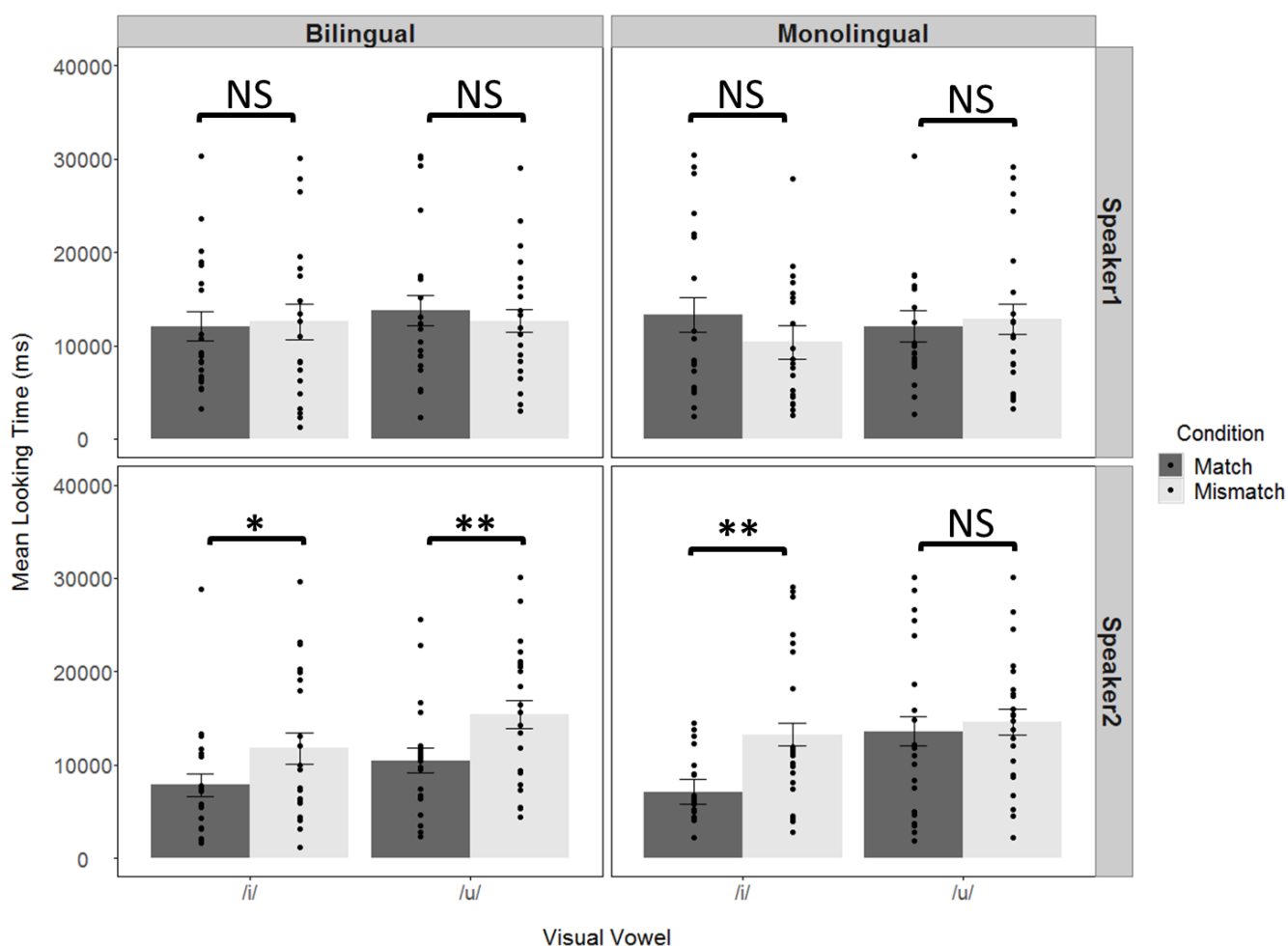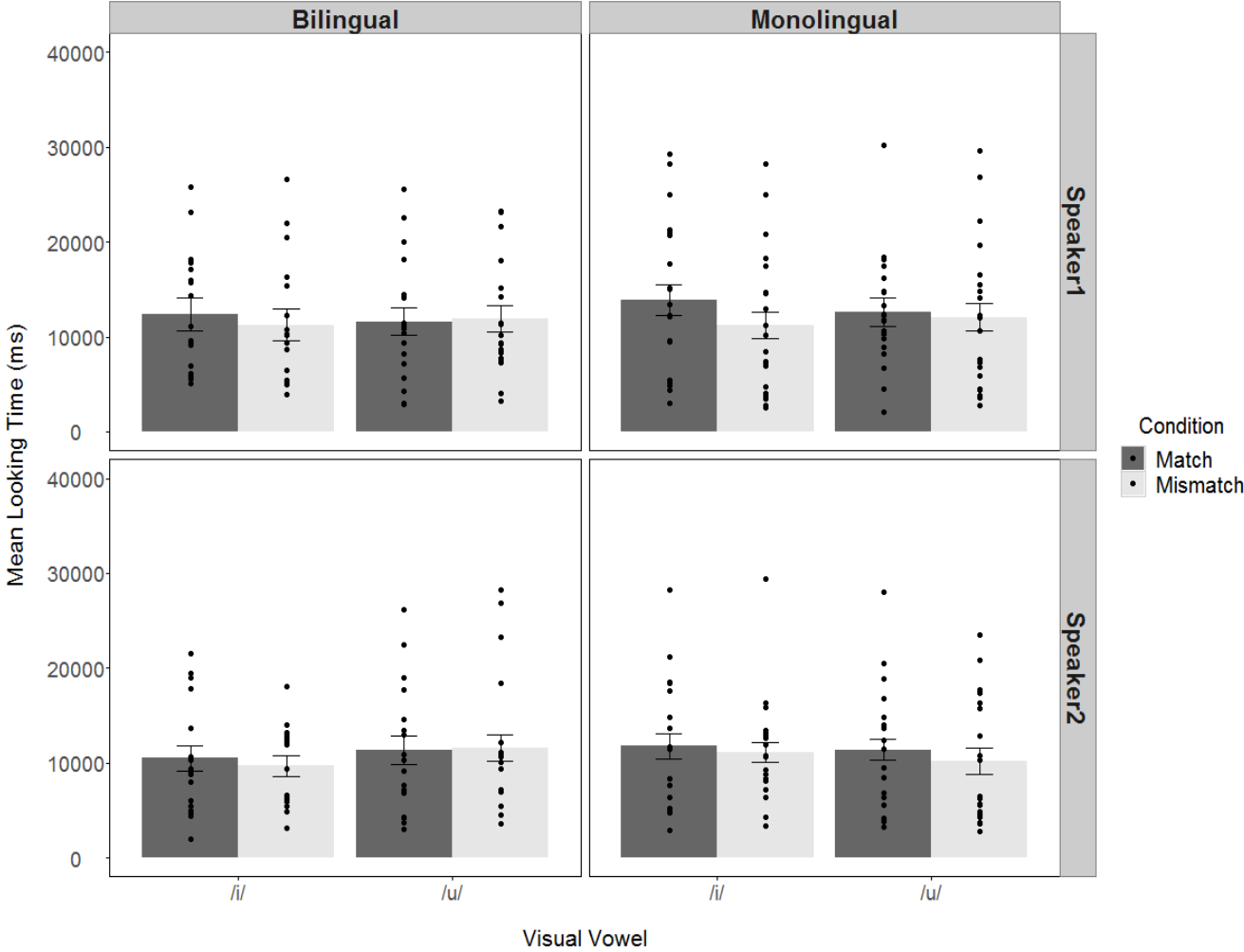***Figure 2-8***. Mean looking times (in milliseconds) in the 8-month-old group for the match and mismatch condition across the bilingual and monolingual group, for the vowels /i/ and /u/ and for the two speakers. Points represent individual infants' scores. Error bars represent +/- 1 SE.

2.4.3   Summary of results

In general, we observed as in prior studies (Table 1-1) that 4.5-month-old infants detected AV vowel mismatch. However, this ability was not modulated by bilingual experience. Further analyses demonstrated that AV matching is observed in one stimuli set (i.e., in Speaker 2), but not in the other, confirming that AV processing is shaped by the vowels' visual and acoustic properties specific to a given speaker. In addition, we observed that for Speaker 2, 4.5-month-old bilinguals detected the AV mismatch for both vowels /i/ and /u/, whereas monolingual infants demonstrated this ability only for the vowel /i/ (although this difference between the patterns displayed by the monolingual and bilingual groups only approached significance). Finally, our 8-month-old group did not demonstrate AV vowel matching ability. This finding was unaffected by language group, vowel, or speaker.

2.5   Discussion

In this chapter, we addressed the question of whether bilingualism affects infants' sensitivity to visual speech, as measured with AV matching ability in 4.5- and 8-month-old infants. The following sections discuss our findings for each age group.

The 4.5-month-old group

We predicted that both monolingual and bilingual infants would demonstrate the AV matching ability. Generally, our findings (across both speakers) are in line with this prediction. Similar to our findings, previous studies in 4.5-month-old infants (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Yeung & Werker, 2013) reported that infants at this age demonstrated the ability to match the auditory and visual speech information. Moreover, we predicted that if bilingualism affects visual speech processing at this age (as proposed by Pons et al., 2015), the bilingual infants should demonstrate greater sensitivity to AV mismatch than the monolingual infants. However, we did not observe that bilingualism (across speakers) modulated AV matching ability.

51

In the current experimental setup, the AV mismatch ability was exhibited through a mismatch preference (i.e., attending longer to the AV mismatch stimuli than to the AV match). A previous study that used a similar paradigm reported the opposite looking pattern (Altvater-Mackensen et al., 2015). It is unclear why infants in the current study preferred the mismatch stimuli. It is a possibility that the difference in the stimuli property elicited the difference in behavior across studies. In the current study, infants were presented with vowels that are more distinct from one another than those in Altvater-Mackensen et al. In line with this speculation, there is evidence that particularly large AV mismatch stimuli (i.e., stimuli that are perceived as AV impossible by adults) elicited longer looking times than AV matched events in infants (Tomalski et al., 2013). Therefore, it is possible that in the current study the AV mismatch trials with more distinct vowels elicited a behavior similar to the AV impossible trials presented in Tomalski et al. (2013). Related to infants' attentional processing, we observed that for Speaker 2 infants attended *the least* to the AV match /i/ trial (see Figure 2-**)** in comparison to the other three trial types. Such looking times pattern might indicate that when infants are presented with an AV event that is perceptually salient (here, a more spread lip-shape and higher pitch in the AV match /i/ in Speaker 2), processing of the event is facilitated. Consequently, the attention (and time) necessary to encode the stimuli is decreased. Even though it is not clear why infants demonstrated the AV mismatch preference, our findings together with another study (Streri et al., 2016) indicate that infants' direction in AV matching task varies at this early age (i.e., sometimes demonstrating a match, sometimes a mismatch preference).

Our next analysis addressed whether the AV matching ability in infants was modulated by the inter-speaker and inter-vowel variation in the visual salience of articulatory cues, as has been observed for inter-vowel variation. A recent study observed that visual perceptual differences (i.e., mouth shape) associated with different vowels modulated infants' AV vowel matching ability (Altvater-Mackensen et al., 2015). Here, we tested whether visual

perceptual *inter-speaker differences* (narrow vs. wide lip-spreading across speakers) while articulating the same vowel also affects young infants' AV vowel matching ability. The two speakers in the current experiment differed in how much they spread their lips while articulating the vowels. Specifically, Speaker 2 spread her lips more than Speaker 1 (see Section 2.4.2.1). Reflecting this visual difference across the speakers, we observed that infants exhibited AV matching ability in response to Speaker 2, but not to Speaker 1. Thus, we observed that infants' AV matching ability is modulated by inter-speaker variation in the visual salience of articulatory cues.

Altvater-Mackensen et al. (2015) suggested that the lip-rounding feature of the vowel /o/ might provide a more perceptually prominent cue for detecting AV mismatch in /a/-/o/ vowel contrast, than the lack of such rounding in the vowel /e/ when considering the /a/-/e/ contrast. Because we observed no evidence of the AV mismatch detection in Speaker 1, the visual/acoustic contrast between lip-spread vowel /i/ and lip-rounded vowel /u/ might not always be salient enough for infants to detect AV mismatch. Instead, we observed that the wider/greater lip-spreading in Speaker 2 appears to be the prominent perceptual cue that aids infants in detecting the AV mismatch. In addition, we observed that Speaker 2 produced vowels in a higher pitch than Speaker 1 (see Table 2-2), possibly boosting the acoustic perceptual salience for infants and further facilitating the AV mismatch detection. Indeed, pitch has been shown to be an important factor in driving infants' preference for infant-directed speech (A. Fernald & Simon, 1984; H. Fernald & Kuhl, 1987). Note that the current study did not address whether the acoustic or the visual perceptual salience independently modulates the AV matching ability in infants. Further research is needed to address this issue.

Finally, we observed a trend supporting the interaction between Language group, Visual vowel, Speaker, and Condition. Specifically, we observed that when presented with Speaker 2, but not Speaker 1, bilingual infants detected the AV mismatch for both vowels while monolingual infants detected it only for /i/. In other words, in response to a speaker that

produced more salient perceptual cues (here, Speaker 2), bilinguals detected the AV mismatch

for both vowels, but monolingual infants detected it only for the most prominent cues (here,

the vowel /i/). Although this finding is only a trend, it is interesting to consider it in relation to

our questions and predictions about the effect of bilingualism on AV speech processing. Some

studies suggest that bilingualism affects AV speech processing (Birules, Bosch, Brieke, Pons,

& Lewkowicz, 2018; Pons et al., 2015), especially around 4 months of age (Pons et al., 2015).

Here, we can speculate that bilingual exposure in 4.5-month-old Spanish-Basque infants

increased the AV matching ability for both vowels, while their monolingual peers AV

matched only the most salient stimuli (here the vowel /i/). Chapter 3 will address whether

bilingual or monolingual infants attended more the mouths of the speakers as a consequence

of AV mismatch sensitivity.

The 8-month-old group

In the current study, we found no evidence that 8-month-old infants detected AV

mismatch. Furthermore, we found no evidence that bilingualism, speaker or vowel affected

AV matching ability in the 8-month-olds. Therefore, our findings did not support our

prediction that 8-month-olds should attend longer to AV incongruent trials, as has been

observed previously (Danielson et al., 2017; Mercure et al., 2018; Tomalski et al., 2013).

However, previous studies relied on selective attention to the mouth as a measure of AV

mismatch detection. In fact, in Tomalski et al. (2013), looking times to the whole face (i.e.,

AV match or mismatch condition) did not reveal a difference between AV match and

mismatch conditions. A difference was only evident in looking times to the mouth. Therefore,

it is possible that sensitivity to the AV mismatch might be reflected in looking times to the

mouth of a speaker. Chapter 3 will address this possibility.

Another reason why 8-month-olds did not reveal AV matching ability could be related

to developmental changes occurring at this age. Namely, younger infants (i.e., younger than 6

months of age) usually exhibit a familiarity preference (i.e., attending more to familiar vs. novel events), whereas older infants exhibit a novelty preference (e.g., Houston-Price & Nakai, 2004). Therefore, it is possible that infants at 8 months of age undergo changes in their preference patterns (i.e., from mismatch to match) that might nullify the AV matching effect at this age. Around this age, some infants might show mismatch, but others, a matching preference, resulting in a null effect for the group as a whole. Our longitudinal data (Chapter 4) could answer more directly whether infants' AV matching ability changes its direction between 4 and 8 months of age.

Relevant to the main question of the thesis, we found no difference between the monolingual and bilingual groups in AV mismatch detection. Previous studies on visual language discrimination in 8-month-old infants demonstrated that bilingual infants are more sensitive to visual speech information than their monolingual peers (Sebastián-Gallés et al., 2012; Weikum et al., 2007). Our results did not support these findings. Possibly, the difference might have occurred because of stimuli difference across these studies. In particular, unlike previous studies, the current study did not present infants with visual speech only (without auditory input). Instead, we presented much more complex stimuli that involved matching and mismatching auditory and visual speech. Under these circumstances, no differences were observed in the looking behavior of monolingual and bilingual infants. In fact, in a study with complex AV stimuli (i.e. fluent speech), no difference between monolingual and bilingual behavior was observed at 8 months of age (Pons et al., 2015; Tsang et al., 2018), which is in line with our findings.

Interim summary

In summary, the findings from Chapter 2 are in line with previous investigations demonstrating AV mismatch detection ability in infants at 4.5 months of age (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Yeung & Werker, 2013). Further, our findings demonstrated that inter-speaker differences in articulation within the same vowel pair affected AV match detection ability in infants. Similar findings have been reported for within speaker differences produced for different vowels (Altvater-Mackensen et al., 2015).

Relevant to the main goal of the thesis, we found no straightforward evidence that bilingualism affected the AV mismatch detection at 4.5 or 8 months of age when infants' overall looking time is considered during the trials (e.g., Altvater-Mackensen & Grossmann, 2015; Altvater-Mackensen et al., 2015; Yeung & Werker, 2013). In the next Chapter, we will assess whether bilingual experience shapes infants' face scanning during match and mismatch trials by measuring infants' selective attention to the eyes vs. the mouth during the same AV matching task.

3 Chapter 3

The development of selective attention to the mouth in monolingual and bilingual infants during audiovisual vowel processing. A cross-sectional study

Chapter 2 addressed infants' sensitivity to visual speech measuring audiovisual (AV) matching ability in 4.5- and 8-month old monolingual and bilingual infants. Overall, we observed no difference between monolinguals' and bilinguals' AV matching ability, measured by their overall looking times to match and mismatch conditions. However, it remained unclear whether bilinguals attended more to the mouth of a speaker, as has been demonstrated before (Pons et al., 2015). The current chapter addressed which part of the face infants attended to (i.e., to the eyes or the mouth of a speaker) when processing AV matched and mismatched vowels. To understand our findings on monolinguals' and bilinguals' attention to the mouth in the AV matching task, we will first provide a literature review on development of infants' attention to the mouth, as well as how it is modulated by AV congruency and linguistic experience.

3.1 Theoretical background

3.1.1 Definition of attention, selective attention, and the attentional shift in infancy

'Attention' is a term widely used in cognitive research (e.g., Plude, Enns, & Brodeur, 1994). Because the term can refer to different cognitive process depending on the task and the methods used, we are going to describe below the meaning of 'attention' as related to experiments presented in the current study.

Attention is a key cognitive component of learning throughout the lifespan. It is composed of several interrelated but distinguishable processes, including: orienting, alerting, arousal, inhibition, and attention termination, as well as sustained, executive, and selective attention (e.g., Colombo, 2001; Ruff & Rothbart, 2010). Given our interest in infants' attention during AV vowel processing, two attentional processes are particularly important: sustained and selective attention. Ability to maintain attention throughout a significant period of time is known as *sustained attention*. Changing direction (focus) of such sustained attention to task-relevant stimuli is known as *selective attention*. Here, we will operationalize selective attention as the focus of infants' *visual* attention to facial features, i.e. the eyes or the mouth region, measured by their looking time at those regions. Selective attention can undergo developmental changes, i.e. *attentional shifts*. For instance, infants' selective attention can change from focusing on the eyes of a speaker to focusing on the mouth.

3.1.2   Development of infants' selective visual attention to the eyes and the mouth

Several studies demonstrated that infants from birth until their second month attended more to the outline of a face than to the eyes and mouth (Hainline, 1978; Haith, Bergman, & Moore, 1977; Hunnius & Geuze, 2004; Maurer & Salapatek, 1976). Around their second month, infants shift their attention from the outline to more internal face features (i.e., eyes or mouth). For instance, 2-month-olds demonstrated the eyes preference (Haith et al., 1977; Hunnius & Geuze, 2004; Maurer & Salapatek, 1976). Between their third and fourth month, infants attended equally long to the eyes and the mouth of a face in response to a socially engaging dynamic face (Hunnius & Geuze, 2004; Wilcox, Stubbs, Wheeler, & Alexander, 2013). The literature is not clear on how selective attention develops from fourth month on. Some studies showed an increase in attention to the mouth (e.g., Hunnius & Geuze, 2004), while others showed that 9-month olds attended more to the eyes than to the mouth when presented with short videos of a talking face (Wilcox et al., 2013).

However, all of the above studies varied greatly with respect to stimuli, making it difficult to conclude how attention to the mouth develops in response to AV speech. However, two recent studies provided more answers to this question, and systematically compared infants' eye movements in response to continuous AV speech in 4-, 6-, 8-, 10-, 12- (Lewkowicz & Hansen-Tift, 2012), 14-, and 18-month-old English learning infants (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2018). Infants were presented with a female face, reciting two short (50 seconds long) monologs either in the infants' native (English) or non-native (Spanish) language. We focus here on infants' response to native speech (English) because it is most comparable to our study, which presented native vowels only. Infants exhibited two attentional shifts during their first year of life in response to native speech. First, from the eyes to the mouth: 4-month-olds preferred looking at the eyes whereas 8-month-olds preferred looking at the mouth. The second shift occurred from the mouth to the eyes: at 10 months of age, infants preferred to look at the mouth, but by 12 months they exhibited a preference toward the eyes again (Lewkowicz & Hansen-Tift, 2012). Attentional shifts between the mouth and the eyes take place during the second year. At age of 14 and 18 months, infants reallocated their attention back to the mouth of a talking speaker (Hillairet de Boisferon et al., 2018). Therefore, throughout language development, selective attention in response to AV speech shifts from the eyes to the mouth and back.

3.1.3   The intersensory redundancy hypothesis and selective attention to the mouth

The role of attending to the mouth during AV speech has been described within the framework of the "intersensory redundancy hypothesis" (e.g., Bahrick & Lickliter, 2012, 2014; Gogate & Bahrick, 1998; Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2016; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). Intersensory redundancy refers to the idea that the same perceptual information may be simultaneously available in more than one sensory modality (Bahrick & Lickliter, 2014). For instance, speech rhythm is available in the *acoustic* speech signal, but also observable *visually* in the speaker's mouth opening while

articulating. Because the acoustic and visual modalities of speech can carry the same information (i.e., rhythm), the information carried by one of the modalities can be considered redundant (e.g., Bahrick & Lickliter, 2014). The intersensory redundancy hypothesis proposes that when the human perceptual system is exposed to a stimulus available across senses (e.g., speech that is simultaneously seen and heard), the system *selectively* attends to those features that provide redundancy (Bahrick & Lickliter, 2000, 2012). According to the intersensory redundancy hypothesis, the mouth is the most salient *visual* feature because it provides access to the redundant information (e.g., rhythm Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Hence when infants shifted their attention to the mouth at 8 months of age, they were able to access the most salient visual feature to process AV speech (e.g., Lewkowicz & Hansen-Tift, 2012). A study on non-native AV speech processing in 12-month-old infants supported this view, demonstrating that when infants were exposed to their non-native language, they attended more to the mouth as an important source for speech decoding (Lewkowicz & Hansen-Tift, 2012). Additionally, adults increased attention to the mouth when the speech comprehension was challenged (Barenholtz et al., 2016).

Past research on infants' AV speech processing has studied attention to the mouth as the source of redundant information from two perspectives: (1) Relation between linguistic experience and attention to the mouth and (2) Relation between AV congruency and attention to the mouth. We discuss both in turn below.

3.1.4   Linguistic experience and selective attention to the mouth

The first perspective on infants' AV speech processing is focused on how infants' linguistic experience (i.e. monolingual vs. bilingual) modulates sensitivity to redundant speech information (i.e., the mouth). As described in section 1.4.1, bilingual language acquisition is more demanding than monolingual language acquisition, in that the bilingual linguistic environment is more variable and less consistent. It has been proposed that in such a

demanding linguistic environment, bilingual infants rely more on redundant speech cues, i.e. the mouth, to help them decode speech (Pons et al., 2015). Specifically, Pons and colleagues found that bilingual exposure in infancy affected the attentional shift to the mouth in the first year of life. In their study, Spanish-Catalan monolingual and bilingual infants were presented with native or non-native 45-second-long AV monologues. The monolingual group was presented with a monologue in their *native* language (either in Spanish or Catalan), while bilinguals were presented with a monologue in their *dominant* language (either Spanish or Catalan). Three age groups were tested: 4-, 8-, and 12-month-old infants. We focus here on the findings related to the native AV speech processing (the conditions most comparable to the work described in the current thesis). The results for the monolingual group in response to the native monologue (Spanish or Catalan) replicated previous findings with English learning monolingual infants (Lewkowicz & Hansen-Tift, 2012): increased attention to the mouth from 4 to 8 months of age, but with a later shift back to the eyes at 12 months. The bilingual group also increased their overall attention to the mouth through the development. However, the bilinguals' shift differed from that of the monolinguals. Specifically, 4-month-old monolinguals attended more to the eyes, but bilinguals attended equally to the eyes and the mouth. At 8 months, both groups attended more to the mouth than the eyes. At 12 months, the bilinguals, but not the monolinguals, looked more to the mouth. Findings from this study are illustrated in **Figure 3–1**. The authors concluded that bilingual infants rely more on the audiovisual redundant cues i.e., the articulatory/mouth region of a speaker, and that increased mouth attention in 4- and 12-month-old bilinguals might play a role in keeping their two languages apart. In particular, attending to the mouth aids in identifying language-specific articulatory features that further support language recognition.
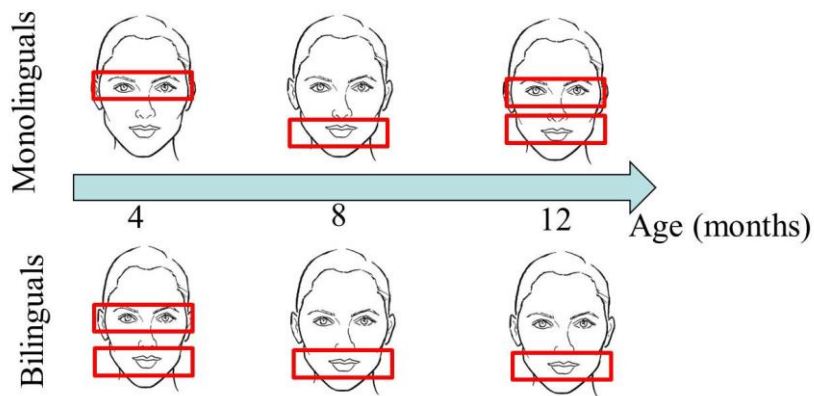
Figure 3–1. Illustration of the findings in Pons et al. (2015) in response to audiovisual native speech monologue. Areas marked in red (the eyes or mouth) reflect where infants attended longer. If both areas are marked, then infants revealed equal looking time between the eyes and mouth.

A more recent study by Ayneto and Sebastian-Galles (2017) corroborates bilinguals' greater attention to the mouth when processing dynamic, but silent faces. Spanish and Catalan monolingual and bilingual infants were presented with emotional expressions (i.e., laughing, crying, neutral), that is, stimuli that were audiovisual and dynamic, but not linguistic. The study demonstrated that 8-month-old bilingual infants attended more to the mouth than their monolingual peers, but at 12 months of age both monolingual and bilingual infants attended similarly to the mouth. The finding suggests that at least until 8 months of age, exposure to two languages leads to a greater attention to the mouth, even when AV stimuli are not linguistic.

However, a recent study by Tsang and colleagues (2018) questions the role of bilingual language experience in modulating infants' selective attention to the mouth. Monolingual (mostly exposed to English) and bilingual infants (exposed to English and various other languages, varying in their amount of exposure to the non-English language) from 6-12 months of age were presented with a female speaker talking in English. In addition, infants' expressive and receptive language skills were estimated (i.e., the Mullen Scales of Early Learning; Mullen, 1995). The study revealed two important findings. First, in line with

previous studies (Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), infants increased their attention to the mouth between 6 and 12 months of age. The second finding, however, was in contrast to prior findings on bilinguals' increased attention to the mouth (Pons et al. 2015). Surprisingly, attention to the mouth did not differ across the monolingual and bilingual groups. Attention to the mouth also did not vary as a function of the amount of exposure to L2 (note that in Pons's study the possibility that selective attention is affected by amount of L2 exposure was not addressed). Instead, Tsang et al. demonstrated that infants' attention to the mouth was positively correlated with the infants' expressive language skills. The authors interpreted that attention to the mouth in infancy might support or be the result of the development of infants' language production ability, and it is not necessarily shaped by the language environment (bilingual vs. monolingual). Similarly, some authors proposed that attention to the mouth supports production development (Hillairet de Boisferon et al., 2018; Lewkowicz & Hansen-Tift, 2012). Specifically, it has been observed that attentional shifts to the mouth co-occurred with changes in infants' language production— canonical babbling at 8 months, or vocabulary expansion around 18 months (Hillairet de Boisferon et al., 2016, 2018; Lewkowicz & Hansen-Tift, 2012; Pons et al.,2015). Thus, it has been hypothesized that attending to the mouth facilitates speech imitation/production and provides cues that aid in learning to produce speech.

3.1.5   Audiovisual congruence and selective attention to the mouth

The second line of research on attention to the mouth during infants' AV speech processing has explored attention to the mouth as an AV redundant/correspondent cue. The idea is if the mouth is the source of redundant speech information, then any disruption in AV redundancy should affect attention to the mouth. Findings on AV disruption have been mixed. In some studies, AV disruption decreased attention to the mouth. For instance, in Hillairet de Boisferon and colleagues' study (2016), looking time was measured in 4-12-month-olds who were presented with desynchronized AV continuous speech. The desynchronization disrupted

the usual looking pattern (i.e., looking longer to the mouth in a synchronized fluent speech; Lewkowicz & Hansen-Tift, 2012), but only in 10-month-old infants. At this age, infants attended for an equal amount of time to the eyes and to the mouth in response to desynchronized fluent speech.

However, studies which manipulated the disruption of the AV redundancy by presenting an incongruent AV signal, e.g. synchronously pairing an auditory syllable with a different visual syllable, demonstrated that monolingual infants between 6 and 9 months of age attended more to the mouth during AV incongruent speech (Danielson et al., 2017; Tomalski et al., 2013). In Tomalski et al., selective attention to the mouth was assessed in 6 to 7 and 8 to 9-month-old infants by presenting AV congruent and incongruent conditions. The AV congruent condition consisted of syllables that were identical in the auditory and visual domains (i.e. visual /ba/ paired with auditory /ba/), while for the AV incongruent condition the auditory and visual syllable did not match (the visual /ga/ syllable paired with the auditory /ba/). Both age groups looked longer to the mouth in the incongruent condition . Danielson et al. found that when infants were exposed to congruent and incongruent AV *non-native* phonemic contrasts, AV incongruence moderately modulated selective attention to the mouth. Specifically, three age groups, 6-, 9-, and 11-month-old English learning infants were exposed to either congruent (i.e. the visual syllable /ḍa/ paired with the auditory /ḍa/) or incongruent (i.e., the visual syllable /ḍa/ paired with the auditory /d̪a/) AV speech of a Hindu dental-retroflex contrast /ḍa/-/d̪a/. The incongruent condition moderately increased infants' attention to the mouth in the 6- and 9-month-olds, but not in the 11-month-olds. The results were interpreted as showing that the ability to detect AV incongruence declines over age with perceptual attunement to the native language.

Up to now, effects of AV incongruence and bilingualism on attention to the mouth were studied separately. A very recent study by Mercure and colleagues (2018) combined

these two factors. In their study 4-8-month-old monolingual (English-learning) and bilingual (English + various other languages) infants were assessed with an AV congruent (i.e. seeing and hearing a speaker producing a syllable /ba/ or /ga/) and an AV incongruent condition (i.e. seeing /ga/ and hearing /ba/; seeing /ba/ and hearing /ga/). The study revealed that older (6.6- to 8-month-olds) monolingual, but not bilingual infants, attend more to the mouth in the AV incongruent condition. Importantly, this monolingual increased sensitivity to the mouth is in contrast with reported increased attention to the mouth in bilingual infants (Pons et al., 2015). As described in section 1.4.4 (page 16), findings from Mercure et al. are limited by their heterogeneous bilingual sample and stimuli that might not be linguistically the same for their monolingual and bilingual sample. Indeed, the authors acknowledged that exposure to a diverse bilingual linguistic environment might hinder the AV mismatch detection in their bilingual sample. Further, the study assessed a sample that varied in age resulting in less power to detect a clear pattern of attention to the mouth at a given age. Taken together, these concerns make findings from Mercure et al. less conclusive regarding whether bilingualism affects selective attention to the mouth in response to AV congruent or AV incongruent events.

To summarize, research on infants' attention to the mouth revealed mixed findings on whether and how bilingualism modulates attention to the mouth. Increased attention to the mouth in bilingual infants has been reported (Pons et al., 2015), but has not been consistently replicated (Mercure et al., 2018; Tsang et al., 2018). The current study examined the effect of bilingualism on infants' attention to the mouth. Details on the current study are given below.

3.1.6   The current study

As previously mentioned, the current dissertation measured monolinguals' and bilinguals' sensitivity to visual speech. In the previous chapter we measured infants' AV matching ability, whereas in the current chapter we measured infants' attention to the mouth during an AV matching task. As a result of our experimental design (see 2.3.4), we measured

infants' attention to the mouth in AV match (congruent) and mismatch (incongruent) conditions (see 2.3 Method section). Recent findings suggest that AV incongruence modulates attention to the mouth—6- to 9-month-old monolinguals attended more to the mouth in response to AV incongruent syllables then to AV congruent syllables (Mercure et al., 2018; Tomalski et al., 2013), but bilinguals did not (Mercure et al., 2018). Here, by addressing limitations of previous studies on this topic (see 1.4.4), especially the limitation regarding the bilingual infants' demands, we could disentangle whether bilingualism per se modulates infants' attention to the mouth, or if this modulation is related to the task difficulty. Based on previous research, we made the predictions below.

1. *Predictions about attention to the mouth depending on language experience*

4.5-month-old group

Following previous reports that bilingual 4-month-olds attended more to the mouth than monolinguals, we predicted that that 4.5-month-old bilinguals would attend more to the mouth. Consequently, bilingual infants would be more likely to detect the AV incongruence (measured by infants' attention to the mouth) than their monolingual peers. In fact, in the previous chapter (2.4.1), we observed a trend that supports bilinguals' increased AV incongruence detection. Similarly to previous studies, we defined detection of AV incongruence as looking longer to the mouth when the AV stimulus is incongruent than when the AV stimulus is congruent (Danielson et al., 2017; Mercure et al., 2018; Tomalski et al., 2013; see more in section 3.1.5).

8-month-old group

Prior results on the effect of bilingualism in this age group are mixed (Mercure et al., 2018; Pons et al.,2015; Tsang et al., 2018). However, we predicted that due to bilinguals' advantage in visual language discrimination at this age (section 1.4.3; Sebastián-Gallés et al.,

2012; Weikum et al., 2007), bilingual infants would attend more to the mouth, and thus be more likely to detect the AV incongruence than monolingual infants.

2. Prediction about the attentional shift from the eyes to the mouth between 4 and 8 months of age

Based on previous findings on increased attention to the mouth over development (Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015; Tsang et al., 2018), specifically around 8months of age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), we predicted that both language groups in the AV congruent condition would demonstrate the attentional shift from the eyes to the mouth between 4 and 8 months of age, as described before (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015).

3.2   Method

Methods were the same as described in section 2.3.

3.2.1   Interest areas and measurements

To analyze infants' selective attention, we first had to define interest areas for analysis. Similarly to the previous studies (Hillairet de Boisferon et al., 2016, 2018; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), we selected three areas of interest (AOI): the face, the eyes, and the mouth (see **Figure 3–2**; note that the colored rectangles are for expository purposes–they were not visible during testing). The areas were rectangular and static, and were large enough to cover the relevant area (face, eyes, or mouth) throughout the entire video, despite the slight movements of the speakers. The face AOI covered the area from the chin to the top of the forehead, and from one ear to the other. The eyes AOI covered the area from the mid-nose up to above the eyebrows, and from one to the other side of the face contour. The mouth AOI covered the area from the mid-chin to the under-nose and from the left to the right mid-cheek. The AOIs (i.e., the eyes, the mouth, and face) were selected to

be similar in size across speakers and videos. See **Figure 3 2** for the size of each AOI (in

pixels) and the percentage of the face area that was covered by the eyes and the mouth AOIs.

Speaker	Vowel	Area	Pixels	% of face
1	/i/	Face	273000	NA
		Eyes	51330	18.8
		Mouth	32400	11.8
1	/u/	Face	268140	NA
		Eyes	53280	19.8
		Mouth	31212	11.6
2	/i/	Face	269180	NA
		Eyes	41316	15.3
		Mouth	31439	11.7
2	/u/	Face	226352	NA
		Eyes	46360	20.4
		Mouth	27738	12.2

Figure 3–2. Size of Area of Interests (AOIs) across the two vowels and speakers, measured in pixels. The AOIs of the eyes and the mouth are also measured as the percentage of the face area. Images represent still-example frames for the two vowels across the speakers. The face AOI is colored in orange, the eyes is in blue, and the mouth is in red.

Following previously reported analyses (Lewkowicz & Hansen-Tift, 2012; Pons et al.,

2015), we measured the relative time infants spent on the eyes and the mouth. To quantify the

relative time infants spend looking at the eyes, we calculated the proportion of looking time

(PLT) by dividing the amount of time infants spent attending to the eyes by the time they

spend attending to the whole face (See **Figure 3–3** for the calculation). Similarly, we

calculated the PLT for the mouth AOI by dividing the amount of time infants spent attending

to the mouth by the time infants spent attending to the whole face. These PLTs for the eyes

and the mouth were used to calculate the eyes-mouth (EM) preference score, calculated as the

eyes PLT minus the mouth PLT (see **Figure 3–3**). The EM preference score can have a value

between −1 (i.e., the absolute mouth preference) to +1 (i.e., the absolute eyes preference).

Zero represents equal time spent attending to the mouth and to the eyes. The EM score is

calculated for each trial and for each infant. All analyses were performed in R software (R

Core Team, 2013), using "afex" package (Singman, Bolker, Westfall, and Aust, 2018).

$$\text{PLT}_{eyes} = \frac{LT\ eyes}{LT\ face} \qquad \text{PLT}_{mouth} = \frac{LT\ mouth}{LT\ face} \qquad \text{EM} = \text{PLT}_{eyes} - \text{PLT}_{mouth}$$

Figure 3–3. Formulae for calculating the Proportion of Looking Time (PLT) for the eyes
and the mouth interest areas, and for calculating the eyes-mouth (EM) preference score. PLT
refers to the looking time (LT) for a given area.

3.3 Results

3.3.1 The effects of audiovisual congruence and language exposure on selective attention to the eyes compared to the mouth

To answer whether AV congruence (e.g., AV match vs. AV mismatch) and linguistic

experience (e.g., monolingual vs. bilingual exposure) modulated attention to the mouth, we

conducted two mixed-repeated measure ANOVAs, one for each of the age groups (4- and 8-

month-olds). Within each ANOVA, the averaged EM scores for the *Condition* (two levels:

match and mismatch) were analyzed as a within-group factor, while *Group* (two levels:

monolingual and bilingual) was a between-group factor. In contrast to previous finding (Pons

et al., 2015), the current analysis for the *4-month-old group* revealed no group differences, nor

any other effects (all $Fs < 2.51$, all $ps > .12$, all η^2_G[8]$ < .002$). The analysis for the *8-month-old group* revealed a close to significant *Group effect ($F_{(1, 73)} = 3.38$, p = .07, $\eta^2_G = .04$)*, indicating a trend for a greater EM score (i.e., greater eyes preference) in the bilingual (M = .15, SD = .43) than in the monolingual group (M = -.02, SD = .37). No other effects reached significance (all $Fs < 2.99$; all $ps > .09$, $\eta^2_G < .002$).

Thus, we did not find evidence that AV incongruence or bilingualism affects the EM preference score in the younger age group. For the older group, the weak *Group* trend suggests a slightly higher EM score in bilinguals (i.e., a greater eyes preference), but this trend did not reach significance. **Figure 3–4** presents the results on the EM score for both age groups.

---

[8]As suggested (Bakeman, 2005; Olejnik & Algina, 2003) to estimate the magnitude of effect size for repeated-measures and mixed designs eta squared (η^2_G) was employed, for non-repeated ANOVAs eta squared (η^2), and for t-tests Cohen's d (d).
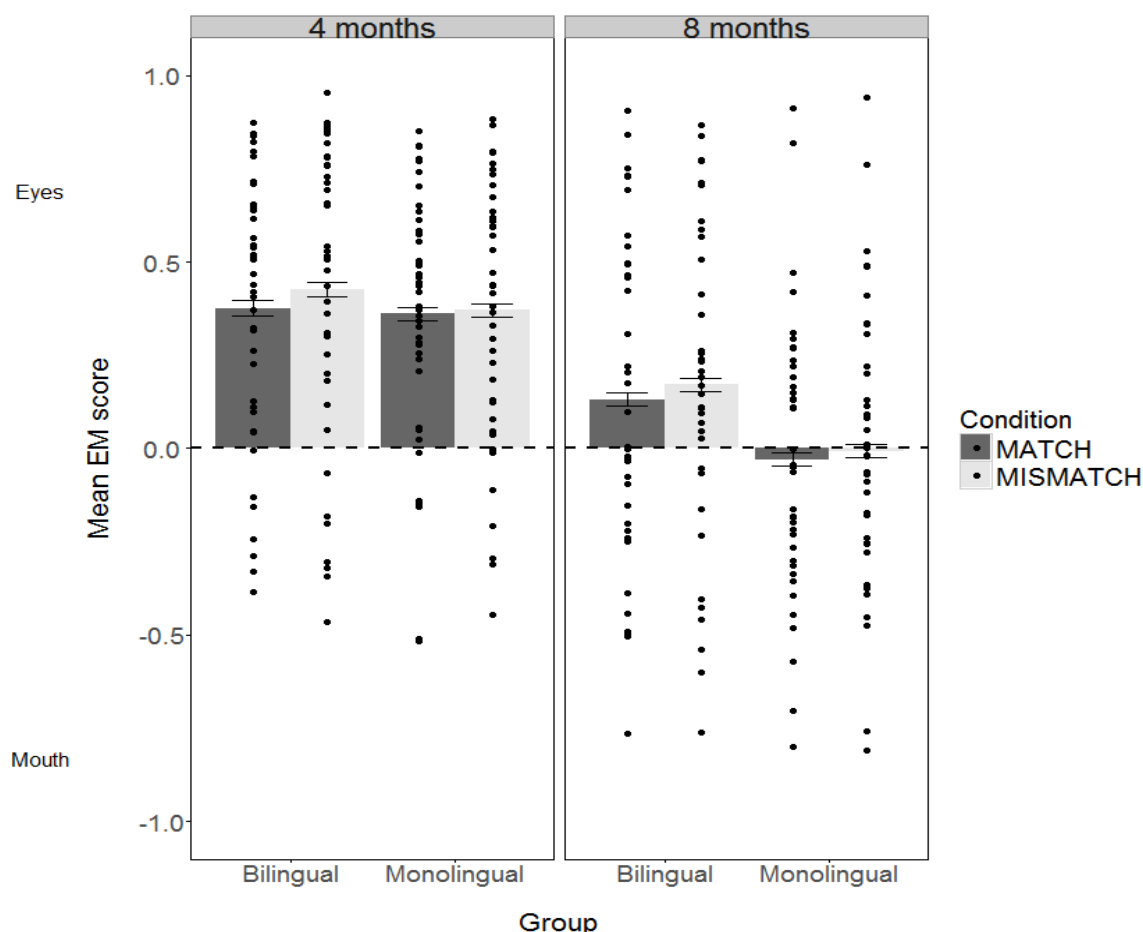
Figure 3–4. Individual eyes-mouth (EM) preference scores for the 4- and 8-month-old monolingual and bilingual infants across the conditions. Positive values reflect eyes preference (e.g., infants spent more time looking at the eyes of the speaker as opposed to the mouth of the speaker), negative reflect mouth preference. Error bars represent standard errors of the mean. Note that none of the individual scores is an outlier.

3.3.2 Attentional shift between 4 and 8months of age

The next question we addressed was whether the infants exhibited an attentional shift during development (i.e., whether infants demonstrated a decrease in attention to the eyes), as well whether any such attentional shift differed between monolinguals and bilinguals, as has been described before (Pons et al., 2015). To test whether any developmental changes in visual attention differed across language groups, we analyzed the averaged EM scores across all trials for each infant in a two-way ANOVA with *Age* (4- and 8-month-old infants) and *Group* (monolingual and bilingual). Since the previous analysis found no evidence that EM scores differ between match and mismatch trials, we did not include *Condition* as a factor.

The analysis revealed a main effect of *Age (F$_{(1, 155)}$ = 28.85, p < .0001, η^2 = .16)*[9],

demonstrating that the EM score decreased (i.e. attention to the eyes decreased) between four

(M = .38, SD = .34) and eight months of age (M = .06, SD = .40). There was also a trend for

an effect of *Group (F$_{(1, 155)}$ = 3.00, p = .09, η^2 = .02)*, reflecting a slightly greater eyes

preference for the bilingual group (as observed in 3.3.1 section too). Finally, the *Age x Group*

interaction *(F$_{(1, 155)}$ = 1.29, p =.26, η^2 = .008)* did not reach significance. These patterns are

evident in **Figure 3–4**, but to facilitate the visualization, we have collapsed results across AV

matched and mismatched condition in **Figure 3–5**.

---

[9] The relation of the effect size from the current study and the effect size in Pons et al. 2015) study is interesting from a developmental point of view. Although the two studies differ in levels of the *Age* factor (Pons et al. tested three groups of infants, while the current study tested two) and in ANOVAs structure (here we tested EM scores, while in Pons et al tested PLTs for the eyes and mouth), however, the effect size is very similar (here η^2 = .16, in Pons et al., η^2 = .11), suggesting that increase in attention to the mouth during development is a more robust effect.
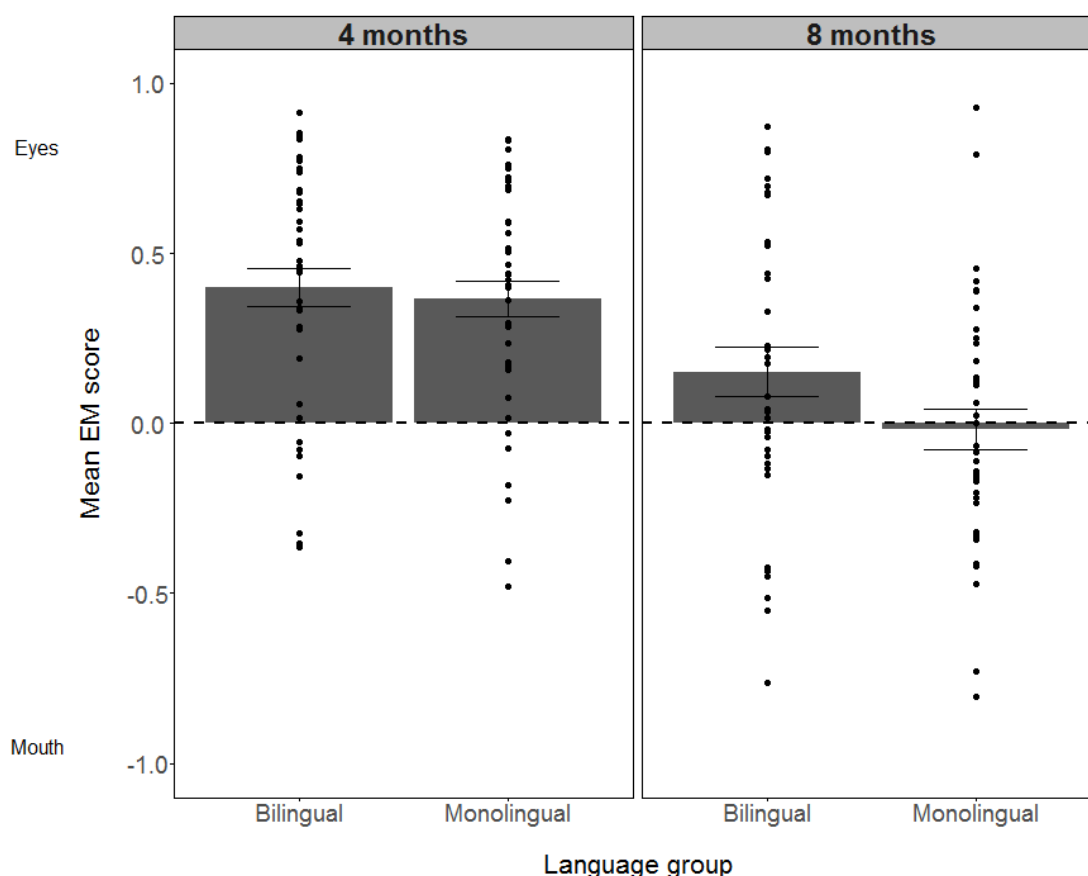
Figure 3–5. Individual eyes-mouth (EM) preference scores for monolingual and bilingual infants, across the two age groups (4 and 8 months of age). Positive values reflect eyes preference (e.g., infants spent more time looking at the eyes of the speaker as opposed to the mouth of the speaker), negative values reflect mouth preference. Error bars represent standard errors of the mean.

3.3.3 Attention to the eyes and mouth in monolingual and bilingual 4- and 8-month-old infants

The previous analysis informs us about the *general decrease* in attention to the eyes during development. Nevertheless, **Figure 3–5** suggests that at 8 months, bilinguals still showed a slight preference towards the eyes over the mouth. . However, our analyses so far have not formally tested which group(s) of infants showed a reliable eyes over mouth bias. To test this, the averaged EM scores were compared to zero, i.e., to the value reflecting equal attention to the eyes and mouth. A one sample t-test revealed that the EM score in both the 4-month-old monolingual (M = .36, SD = .33) and bilingual groups (M = .40, SD = .35)

significantly differed from zero ($t_{\text{monolingual (41)}} = 7.14$, $p < .001$, $d = 1.10$; $t_{\text{bilingual(41)}} = 7.25$, $p < .001$, $d = 1.12$), demonstrating a preference to look to the eyes (70 out of 84 infants exhibited positive EM scores, i.e. attention to the eyes). The same analysis in the 8-month-old groups showed that the EM score for the monolinguals (M = −.02, SD = .36) did not differ significantly from zero ($t_{(38)} = −31$, $p = .75$, $d = .04$), indicating that the 8-month-old monolingual group attended equally to the eyes and to the mouth (18 out of 39 exhibited positive EM score). In contrast, the EM score did significantly differ from zero in the bilingual group (M = .15, SD = .43; $t_{(35)} = 2.11$, $p = .04$, $d = .35$), indicating a preference to look to the eyes (22 out of 36 exhibited the positive EM score, see **Figure 3–4** and **3-5**).

Comparison between monolingual and bilingual infants in their mouth attention

The study by Pons et al. (2015) observed that bilingual infants at 4 and 12 months of age attended more to the mouth than their monolingual peers. Because EM score represents a preference towards the eyes over the mouth (**Figure 3–3**), we analyzed a direct measure of attention to mouth with PLTs (i.e., proportion of looking time) to the mouth. We compared averaged mouth PLTs for each infant across monolingual and bilingual 4-month-old groups (Table 3–1). The independent t-test demonstrated that the groups did not significantly differ ($t_{(82)} = −.31$, $p = .76$, $d = .06$) in their attention to the mouth. In the 8-month-old group, monolinguals demonstrated a tendency to attend more to the mouth than bilingulas, although the difference did not reach significance ($t_{(73)} = −1.39$, $p = .16$, $d = .32$). To understand the source of greater EM preference in the 8-month-old bilingual group (described in the previos section), we analyzed PLTs to the eyes between the monolingual and the bilingual group. The independent t-test demonstrated that the 8-month-old bilinguals attended to the eyes significantly more than the monolinguals ($M_{\text{bilingual}}= .44$, $SD_{\text{bilingual}} = .24$; $M_{\text{monolingual}} = .33$, $SD_{\text{monolingual}} = .20$; $t_{(73)} = 2.03$, $p = .04$, $d = .47$). Thus, 8-month-old bilingual and monolingual infants differed in their attention to the eyes, but not to the mouth (Table 3-1).

Table 3–1

The mean proportion of looking time to the mouth for the monolingual and bilingual infants across the two age groups

Age (months)	Language group	Mouth proportion
4.5	Monolingual	.15 (.14)
	Bilingual	.14 (.16)
8	Monolingual	.35 (.19)
	Bilingual	.28 (.19)

Note. Standard deviations are given in parenthesis.

3.3.4 The effect of the visual vowel and speaker on infants' selective attention

In Chapter 2, we reported that the detection of the AV mismatch was modulated by the visually presented vowel and the speaker (with 4-month-olds being more sensitive to AV mismatch in Speaker 2 than Speaker 1, and more sensitive to the mismatch for the vowel /i/ than /u/). As we suggested in Chapter 2, it is possible that 4-month-olds' AV mismatch in Speaker 2 is present and for vowel /u/, reflected in their greater attention to the mouth in AV mismatch trials (rather than in difference in overall looking times). We therefore included Speaker and Vowel as possible factors that modulate the EM score. We conducted two separate repeated-measure mixed ANOVAs for each age group on EM scores with *Condition* (match and mismatch) and *Vowel* (/i/ and /u/) as within-subject factors, and *Speaker* (Speaker 1 and Speaker 2) and *Group* (monolingual and bilingual) as between-subject factors. The results will be described separately for the 4- and 8- month-old group.

4.5-month-old group

The analysis for the 4-month-old group revealed a main effect of Speaker ($F_{(1, 80)} = 5.39$, $p = .02$, $\eta^2_G = .05$), indicating that Speaker 2 (M = .46, SD = .39) elicited higher EM score (i.e., greater eyes preference) than Speaker 1 (M = .29, SD = .36), and the *Speaker x*

Condition interaction ($F_{(1, 80)} = 6.45$, $p = .01$, $\eta^2_G = .004$). No other effects and interactions reached significance (all *Fs* < 2.5, all *ps* > .12).

To determine the source of the *Speaker x Condition* interaction (see **Figure 3–6**), post-hoc paired t-tests (using the Tukey method to control for multiple comparison showed that for *Speaker 1,* the infants exhibited a similar EM score in match (M = .30, SD = .38) and mismatch (M = .28, SD = .34) conditions ($t_{(80)} = .72$, $p = .47$). However, in response to *Speaker 2,* the EM score was greater in the mismatch (M = .50, SD = .41) than in the match (M = .43, SD = .36) condition ($t_{(80)} = 2.92$, $p = .004$).

To summarize, in the younger group of infants, the EM is modulated both by the condition and the speaker. Specifically, the mismatch condition elicited more looks to the eyes than the match condition, but only in the case when infants were presented to Speaker 2. This confirmed finding from Chapter 2, suggesting the speaker, i.e. the stimuli infants were presented with, can modulate selective attention in young infants. Specifically, infants' AV matching ability was observed with Speaker 2. Here, we observed that this ability was reflected in longer looking times to the eyes in AV mismatch vs. match condition.
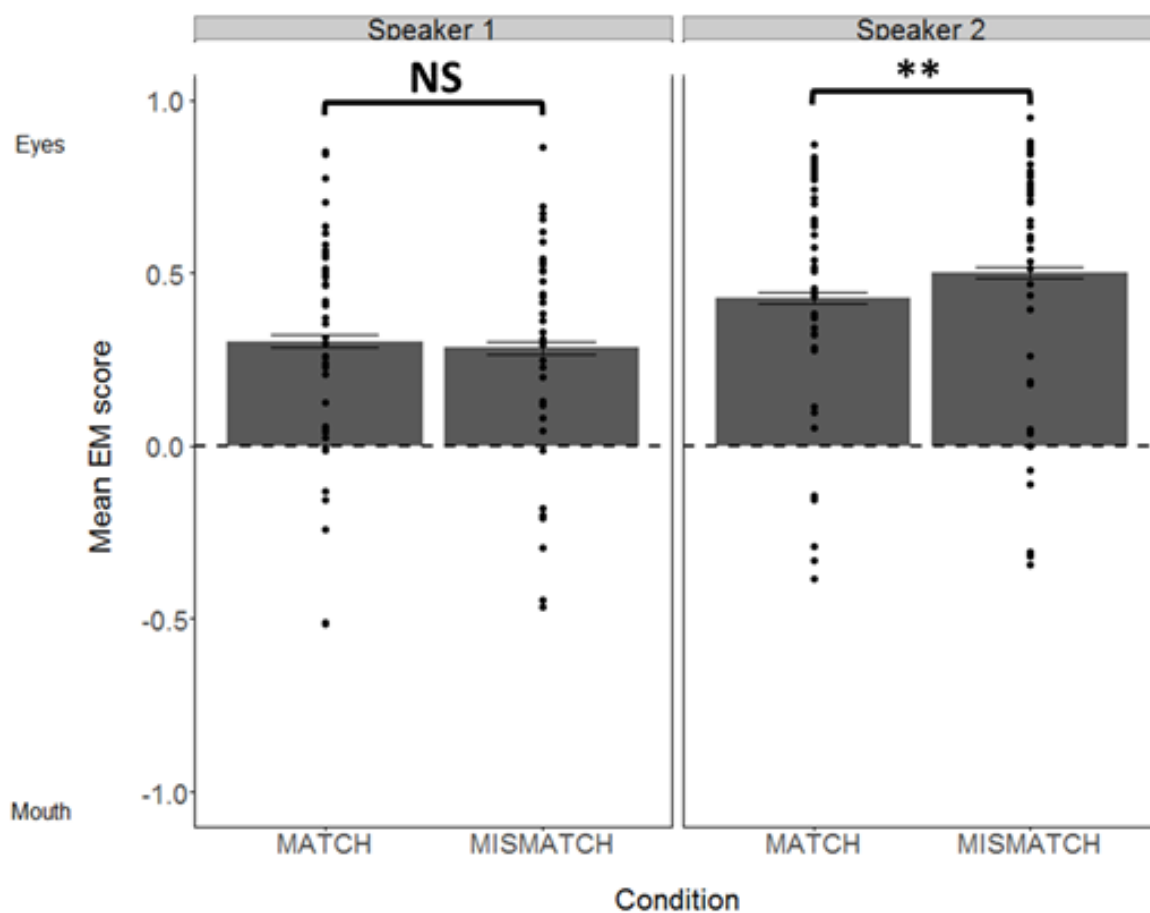
Figure 3–6. The mean eyes-mouth (EM) preference scores for the 4.5-month-old infants, across the two conditions and speakers. The positive values represent the eyes preference, the negative values represent the mouth preference, and the dashed line represents the equal looking time between the eyes and the mouth. The dots represent individual scores of each infant. Error bars represent +/- 1 SE, asterisks indicate a significance level of ** $p \leq .01$, NS indicates a non-significant effect.

8-month-old group

The analysis[10] for the 8-month-old group revealed a significant main effect of Vowel ($F_{(1, 71)} = 28.73$, $p < .001$, $\eta^2_G = .03$) indicating that the vowel /u/ (M = .14, SD = .45) elicited a lower EM score (i.e., greater mouth preference) than the vowel /i/ (M = −.01, SD = .42). The effect of Condition approached significance ($F_{(1, 71)} = 2.89$, $p = .09$, $\eta^2_G = .01$), indicating a trend for a higher EM score in the mismatch (M = .08, SD = .44) than in the match condition (M = .05, SD = .44). Finally, the analysis revealed a significant *Speaker x Vowel* interaction ($F_{(1, 71)} = 5.42$, $p = .02$, $\eta^2_G = .01$). No other effects and interactions reached significance (all Fs < 1.41, all ps > .24).

To determine the source of the *Speaker x Vowel* interaction, post-hoc paired t-tests (using the Tukey method to control for multiple comparisons) showed for both speakers that infants exhibited more negative EM scores (i.e., greater mouth preference) in response to the vowel /u/ than to the vowel /i/ ($t_{Speaker\ 1\ (71)} = 2.15$, $p = .03$; $t_{Speaker\ 2(71)} = 5.40$, $p < .0001$). **Figure 3–7** shows the EM score across the two language groups in response to the two speakers and vowels.

To summarize, in the older age group, bilingual, but not monolingual infants exhibited an eyes preference, which is modulated by the vowel and the speaker that the infants were presented with. Specifically, infants attended more to the mouth for the vowel /u/, especially with Speaker 2.

---

[10] The analysis revealed the same effect of Group as described in section 3.3.1. Specifically, the effect of Group was nearly significant ($F_{(1, 71)} = 3.30$, $p = .07$, $\eta^2_G = .04$), demonstrating a trend for a greater EM score in bilinguals (M = .15, SD = .46) than monolinguals (M = −.02, SD = .41).
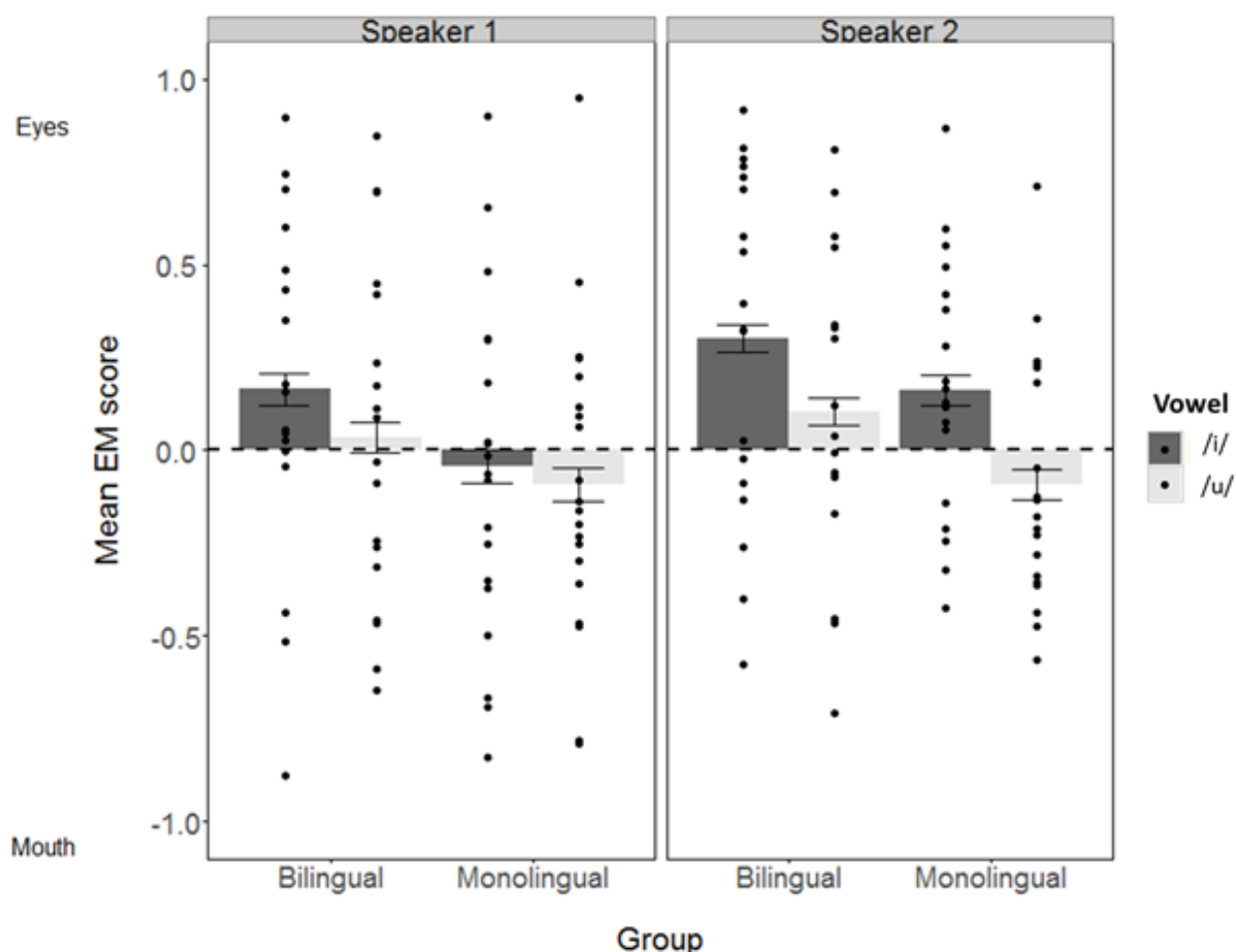
Figure 3–7. The mean eyes-mouth (EM) preference scores for the 8-month old monolingual and bilingual infants, across the two vowels, and the two speakers. The positive values represent the eyes preference, the negative values represent the mouth preference, and the dashed line represents the equal looking time between the eyes and mouth. The dots represent individual scores of each infant.

3.3.5   Interim summary of results in Chapter 3

To summarize, the analysis on selective attention to the mouth in 4- and 8-month-old monolingual and bilingual infants revealed several important findings.

First, the current study replicated the previously reported attentional shift: a decrease in attention to the eyes and an increase in attention to the mouth from 4 to 8 months of age. Second, we observed that 4-month-old infants attended more to the eyes, and this attention was not modulated by the type of language exposure (monolingual or bilingual) or by AV congruency. However, selective attention to the mouth at this age was modulated by inter-speaker differences. Specifically, Speaker 2 elicited greater attention to the eyes in response to

the mismatch condition than in response to the match condition. This finding confirms the effect of speaker described in Chapter 2, suggesting that AV speech processing in infants can be affected by the identity of the speaker who is providing the input. Third, results in the 8-month-old group indicated that bilingual, but not monolingual infants exhibited an eyes preference. Interestingly, in both monolingual and bilingual group, selective attention to the mouth was increased more when infants were presented with /u/ than /i/.

Finally, it is worth noting that although we observed an eye preference in the bilingual 8-month-old group, inspecting **Figure 3-5** reveals that overall preference (the eyes vs. the mouth) varied considerably in this age group. Some infants demonstrated a large eyes preference, some a large mouth preference, and many performed in between, suggesting that the attentional shift to the mouth as reported in Pons et al. (2015) is not universal, but might depend on the task and/or stimuli infants are presented with.

3.4   Discussion

In this chapter we assessed 4- and 8-month-old monolinguals' and bilinguals' sensitivity to visual speech by measuring their selective attention to the eyes vs. the mouth during audiovisual (AV) vowel matching task. The study revealed several important findings.

First, based on previous research (Pons et al., 2015), we predicted that 4.5-month-old bilinguals would look more to the mouth than monolinguals, and thus be more likely to detect the AV mismatch condition. However, we found no evidence that AV incongruence or bilingualism modulated 4.5-month-old infants' selective attention to the mouth. In other words, 4.5-month-olds infants' looking pattern in AV matched and mismatched condition was similar—they attended more to the eyes than to the mouth, and this result did not differ between the bilingual and monolingual group. Therefore, we did not replicate bilinguals' greater attention to the mouth (Pons et al., 2015), not even in the condition that was most similar to Pons et al study, the AV matched condition. One reason might be differences in the

tasks across the studies. Unlike Pons et al., we equalized the task demands across the groups (see discussion in section 1.5) ensuring that bilinguals were not in a more difficult task than monolinguals. Therefore, it is possible that the difference between bilinguals and monolinguals observed in Pons et al. (2015) was due to task difficulty, rather than bilingual linguistic experience per se, with more difficulty leading to more attention to the mouth in early infancy. Interestingly however, our AV incongruent condition, although arguably more difficult than our AV congruent condition, did not modulate attention to the mouth. This finding is in line with Mercure et al. (2018), in which 4- to 6.5-month-old monolingual and bilingual infants attended similarly to the mouth, regardless of AV congruency.

With regards to our finding from Chapter 2 that 4.5-month-olds detected the AV mismatch, it is puzzling how their looking pattern did not differ between AV match and mismatch conditions. In both conditions, infants attended to eyes more than to the mouth. It leads to an interesting question: *Can infants detect AV mismatch if their visual attention is directed to the eyes rather than to the mouth?* We make two speculations regarding this question. First, it is possible that the movements in the eye region that co-occurred with different articulatory movements were salient enough for young infants to detect any AV incongruence. Second, and more likely, it is possible that infants could see the mouth parafoveally, while looking at the eyes of the speaker. Further research is needed to understand this specific looking pattern at this age.

Regarding the 8-month-old group, we observed an interesting pattern. With respect to infants' attention to the mouth at this age, our findings are in line with previous studies that demonstrated no difference between monolinguals and bilinguals (Mercure et al., 2018; Pons et al., 2015; Tsang et al., 2018). However, our study also demonstrated that bilinguals attended more to the eyes, regardless of AV congruency. In relation to the previously mentioned studies (see 3.1.4), it is not clear why we observed an eyes preference in 8-month-old bilinguals. Would this eye preference reflect a delay in shifting to the mouth for

bilinguals? Or do bilingual infants attend more the social cues, such as the eyes? Or is attention to the mouth related to other abilities, such as productive vocabulary? In fact, in Tsang et al.'s study, infants' increased attention to the mouth was related to their increased expressive vocabulary (i.e., higher babbling scores). Therefore, it is possible that the infants with greater mouth preference in the current study were also more advanced in their production ability. Unfortunately, here we did not assess infants' speech production skills, because of the lack of standardized questionnaire/test assessing infants' speech production skills at 4 months of age. Certainly, further research should take into account infants' production skills in AV matching task, either by relying on parental report or by measuring infants' spontaneous speech production/vocalization. It is also worth noting that the observed bilinguals' eyes preference is not very strong (i.e., in section 3.3.1 it is a trend), therefore further research is needed to determine whether this effect is robust.

Further, our data supported our prediction regarding infants' attention shifts from the eyes to the mouth between 4 and 8 months of age. Specifically, we observed that during development, infants increased their attention to the mouth and decreased their attention to the eyes. This finding is generally in line with the previously reported increased attention to the mouth in 8-month-old infants (e.g., Lewkowicz & Hansen-Tift, 2012; Mercure et al., 2018; Pons et al.,2015; Tsang et al., 2018) with one interesting difference. In contrast to previous studies (e.g., Lewkowicz & Hansen-Tift, 2012; Pons et al.,2015), our 8-month-olds did not attend more to the mouth than to the eyes. Rather, monolinguals attended equally to the eyes and the mouth, while bilinguals demonstrated a preference for the eyes. The reason why we did not observe the attention to the mouth with our 8-month-old monolingual and bilingual sample might be due to the stimuli used in this task. The current study tested infants with linguistic stimuli, but it was far less complex than the stimuli in the previous study (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), which used continuous speech. It is possible that continuous speech draws attention to the mouth, as it has been observed that

speech, in general, attracts more attention to the mouth than a silent video (Haith et al., 1977). Therefore, it is possible that simple, repetitive stimuli such as that employed in the current study demanded less attention to the mouth than the continuous speech in previous research.

Importantly, we observed that 8-month-old infants vary in their selective attention to the eyes and the mouth. Namely, almost half of the infants (30 out of 75) attended more to the eyes than to the mouth, suggesting that attentional shift by this age does not occur in every single infant. Rather, it is a gradual shift, incorporating high individual variability. In Chapter 4, we will further study this individual variability at 8 months, examining whether it is related to infants' selective attention at 4 months.

Furthermore, following on findings from Chapter 2 about effects of speaker and vowel on infants' selective attention, we observed some interesting patterns. We found that the speaker differences reported in Chapter 2 (i.e., the inter-speaker difference affected infants' sensitivity to AV mismatch), are also reflected in visual attention to the eyes. Specifically, for Speaker 2, 4.5-month-olds attended more to the eyes in AV mismatch than match conditions, but not for Speaker 1. This finding is in contrast with previous studies where the AV mismatch was reflected in greater attention to the mouth (e.g., Mercure et al., 2018; Tomalski et al., 2013). As mentioned earlier, it is possible that the eyes provide cues to detect AV mismatch. For instance, small movements around the eyes may be different when a speaker produces vowel /i/ vs. /u/, hence providing a visual cue that can match with the auditory speech or not. A recent study supports this interpretation showing that 10-month-olds in response to AV desynchronized speech attended more to the eyes than in AV synchronized speech (Hillairet de Boisferon et al., 2016).

Finally, we observed that selective attention in monolingual and bilingual 8-month-olds was shaped by the visually presented vowel, regardless of the auditory vowel presented. Specifically, when infants saw the vowel /u/, they attended more to the mouth than when they

saw the vowel /i/. This finding is in line with previous finding that infant AV matching ability is shaped by a vowels' visual salience (Chapter 2, Altvater et al., 2015). Specifically, in Altvater et al. study, infants' AV matching was better when they were presented with a rounded vowel /o/, than with a non-rounded vowel /e/, suggesting that a rounded mouth shape might be a more visually salient cue. Similarly, in the current study the speaker's mouth shape while producing /u/ might have been a more salient cue, leading infants to allocate more attention to it. Some previous work on early perceptual biases demonstrated that infants are sensitive to shape-sound association (Pejovic & Molnar, 2017), especially when associating rounded shapes to sounds (Fort et al., 2018).

In conclusion, although we observed a general increase in attention to the mouth between 4 and 8 months of age, we found no clear evidence that bilingualism or AV incongruence affected selective attention to the mouth. Specifically, we observed that monolingual and bilingual infants at 4 months of age were attracted to the eyes of a speaker. This pattern remains even when the task is more difficult (AV incongruent condition). At 8 months, monolinguals and bilinguals did not differ in their attention to the mouth, and again that pattern did not change in the more difficult, AV incongruent condition. Together, our study expands on previous findings (Pons et al., 2015; Tsang et al., 2018) which show that that once bilingual and monolingual infants are tested in a similarly demanding task, they exhibit similar selective attention to the mouth.

Importantly, we observed large variation across individuals in the eyes-mouth preference in the 8-month-old group. Given the nature of the cross-sectional data, it is hard to determine the source of such variability. The next chapter aims at determining that source by looking at a possible relationship between early and late looking behavior, using a subset of infants who were tested at both ages (i.e., who were tested longitudinally).

4   Chapter 4

Individual variation in the development of audiovisual matching ability and selective attention to the eyes and the mouth in monolingual and bilingual infants. A longitudinal study

4.1   Introduction

In Chapters 2 and 3, we studied cross-sectionally 4.5- and 8-month-old infants' sensitivity to visual speech via audiovisual (AV) matching ability and attention to the mouth. We observed that regardless of group (i.e., monolingual or bilingual), 8-month-olds demonstrated significant individual variability with respect to the size of their eyes-mouth (EM) preference (see **Figure 3-4**); some infants demonstrated a clear eyes preference, while some demonstrated a clear mouth preference. The individual differences varied from almost absolute mouth preference (-1) to absolute eyes preference (+1), see **Figures 3-4**, **3-5**, **3-7** of the previous chapter. The goal of the present chapter is to understand the observed individual variation in the 8-month-old group. Specifically, we are interested in whether 8-month-olds' individual variation in looking behavior can be explained by their looking behavior at 4.5 months. Therefore, we ask whether the large variability in 8-month-olds might be explained by infants' individual preference. For instance, some infants might demonstrate a strong eyes preference, both at 4.5 and 8 months of age, hence their looking behavior at 8 months can be explained by their *individual* preference that is present at both ages. To test this, we analyzed a subset of our cross-sectional data that was assessed longitudinally, and compared their looking behavior on both the *AV matching ability* as well on the *EM preference*.

Longitudinal data on *AV matching ability* in infants is limited to date. A 5-year-long longitudinal study on children aged between 2-9 years demonstrated that AV speech comprehension improved with age (Bergeson, Pisoni, & Davis, 2005), but little is known about *AV matching ability* in longitudinal reports. Based on the limited longitudinal literature, it is not straightforward to predict how AV matching ability should develop between 4.5 and 8 months of age. However, based on studies that showed that integration of auditory and visual speech improves with age in children (Hockley & Polka, 2005) and in infants (Danielson et al., 2017), we predicted that infants' AV matching ability should also improve with age.

Regarding infants' EM preference, to the best of our knowledge, only one study longitudinally examined infants' EM preference during AV speech processing (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013). The authors tested infants' attention to the mouth at 6, 9, and 12 months of age, and observed a positive correlation across all age groups (i.e., between 6 and 9 months of age, between 6 and 12 months of age, and between 6 and 12 months of age). This study suggests that infants' individual preference to attend to the mouth is quite constant over development (i.e., those infants that demonstrated the mouth preference at an earlier age also demonstrated the mouth preference later on in development). However, the study also observed an overall increase in looking to the mouth with age. Therefore, we expected that on the one hand overall attention to the mouth would increase with age, and that on the other hand infants' EM preference score at 4.5 and 8 months of age would be positively correlated (i.e., that infants with a stronger individual preference to the mouth at 4.5 months would also show a stronger mouth preference at 8 months). If so, then we could explain the variability observed in the EM preference score at 8 months of age (see section 3.3) based on the EM preference at 4.5 months. Finally, if bilingualism does affect the EM preference at 4 months of age (Pons et al., 2015), we should see different developmental trajectories of the EM preference between the monolingual and bilingual group.

First, we will describe the participants in the longitudinal dataset, and then we will describe the results on the relation between 4.5 and 8 months for the *AV matching ability* and the *EM preference*.

4.2 Participants

A subset of the cross-sectional data contributed to the longitudinal data set, specifically that of infants whose data was analyzed at both 4.5 and 8 months of age. In total data from 34 infants was analyzed longitudinally. Fifteen infants were presented with Speaker 1 in the first session and with the Speaker 2 in the second session. Nineteen infants were presented with Speaker 2 in the first session and with the Speaker 1 in the second session. At the first session (at 4.5 months) 18 monolinguals (average age 4.5 months, range 131-146 days, 9 female infants; average exposure to one of the languages was 99.4 %, range 95.3-100, 10 exposed to Basque and 8 to Spanish) and 16 bilinguals (average age 4.5 months, range 133-147 days, 9 female infants; average exposure to the second language was 34.6% (range 15.3-47.2; 5 dominantly exposed to Basque, 3 to Spanish, and 8 balanced in their exposure to Spanish and Basque). At the second session (at 8 months) the layout remained the same, in other words, none of the infants changed their language group in the second session. The same 18 monolinguals (average age 8 months, range 241-252 days; average exposure to one of the languages was 98.9%, range 95-100, 10 exposed to Basque and 8 to Spanish) and 16 bilinguals (average age 8.0 months, range 237-259 days; average exposure to the second language was 33.1% (range 16.1-49; 6 dominantly exposed to Basque, 6 to Spanish, and 5 balanced in their exposure to Spanish and Basque) were tested in the second session. The full description of participants across the two speakers is given in Table 4-1.

Table 4-1

Demographic statistics of participants across the two speakers in the longitudinal data

Age group (in months)	Speaker	Nº of participants	Nº of female participants	The age range in days	Language group	Exposure to L1/L2 (%)
4.5	Speaker 1	6	5	134-144	monolingual	99.5 (98-100)
	Speaker 2	12	4	131-146	monolingual	99.3 (95.3-100)
	Speaker 1	9	4	136-142	bilingual	31 (15-44.2)
	Speaker 2	7	5	133-144	bilingual	39.1 (19.8-47.6)
8	Speaker 1	12	4	241-252	monolingual	98.3 (95-100)
	Speaker 2	6	5	244-250	monolingual	99.2 (97.3-100)
	Speaker 1	7	5	241-259	bilingual	35.6 (16.1-48)
	Speaker 2	9	4	237-250	bilingual	34.9 (15.9-49)

Note. Nº refers to a number of participants (sample size). The range of exposure to L1/L2 is given in the brackets. L1 refers to infants' dominant language, whereas L2 refers to infants' less dominant language. For monolinguals, exposure to L1 is presented.

The experimental design, stimuli, apparatus, and the procedure were the same as described in Method Section (see Section 2.3).

4.3   Results

Results will be presented separately for the AV matching ability and for the EM preference score.

4.3.1   Audiovisual matching ability

As in the cross-sectional analysis (see Section 2.3) of the AV matching ability in infants, we firstly summed all recorded fixations for each trial (12 in total) and for each infant, giving the overall looking time (in ms). Next, we averaged the looking time for each condition (match and mismatch) for each infant. Finally, we calculated the Matching Preference (MP) score to assess whether the infants' AV matching performance changed over development, as well as whether individual difference might explain the lack of the AV mismatch detection at 8 months of age. The MP score was calculated as the average looking time of the match condition minus the average looking time of the mismatch condition (see **Figure 4-1**). The

positive score reflects the matching preference; the negative score reflects the mismatch

preference. The MPs were calculated for each infant.

$$MP = LT_{match} - LT_{mismatch}$$

Figure 4-1. Formula for calculating the Matching Preference Score (MP). LT refers to the average looking times for the given condition.

To address whether there was a relation between the MPs at 4.5 and at 8 months of

age, we computed the Pearson's product-moment correlation between the two values[11].

Overall, we did not find evidence of a relation between the MP values at 4.5 and 8 months of

age, $r(32) = .01$, $p = .93$. Similar results were observed within each language group,

$r_{monolingual(16)} = .04$, $p = .86$; $r_{bilingual(14)} = -.01$, $p = .95$, demonstrating that MPs at 4.5 and 8

months of age did not significantly correlate in either the monolingual or bilingual group (see

Figure 4-2).

---

[11] Note that the findings are not affected by possible outliers. Since we aimed to preserve all valuable data in the longitudinal data set, we did not remove possible outliers; instead, we computed additional analysis to account for possible extreme values. The outlier-robust correlation from *R* package *WRS2* (Mair & Wilcox, 2017) confirmed overall non-significant correlation between MP at 4.5 and MP at 8 months, and for each language group, all $p > .74$. All the future correlations were also verified by the outlier-robust correlations.
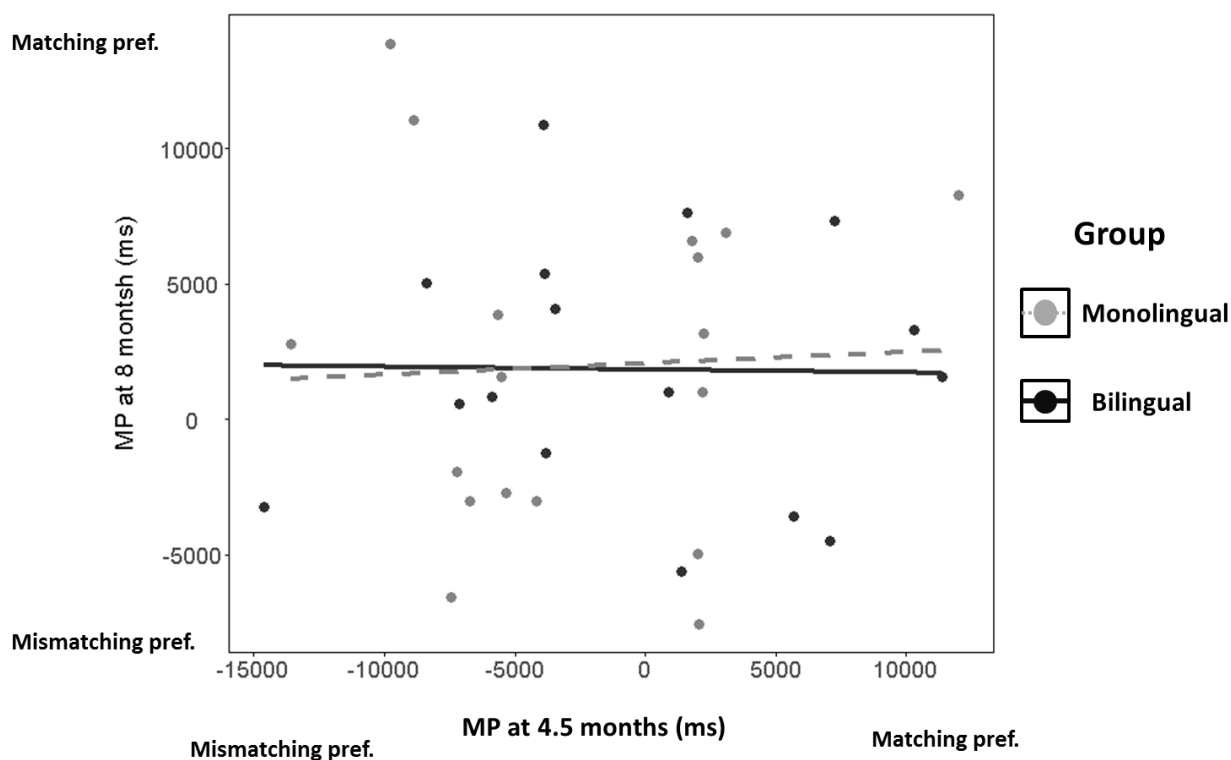
Figure 4-2. Scatter plot representing the relation between Matching Preference (MP) at 4.5 and at 8 months of age across the monolingual (in grey) and bilingual (in black) group. Positive values reflect matching, while negative reflect mismatching preference.

Previous analysis of the AV matching performance (see Chapter 2) indicated that

Speaker and *Visual Vowel* shaped the AV matching performance. Here we wanted to include

these factors and assess whether they predict the MP at 8 months of age. However, due to an

unbalanced number of infants across speakers (see Table 4-1), instead of a linear regression

we ran a mixed-model analysis using the *lmer* package in R (Bates, Mächler, Bolker, &

Walker, 2015) which minimizes the impact of unbalanced designs. Since the *speaker* factor

changed within infants at 4.5 and at 8 months, we coded a new factor–*First session speaker*

that specified what speaker infant was presented with at 4.5 months (i.e., two levels, Speaker

1 or Speaker 2). To analyze the data, MPs were averaged separately for each *visual vowel*

(i.e., /i/ and /u/), therefore each infant contributed with two data points for each of the vowels.

The final mixed-model analysis included the *MP at 8 months* as the dependent variable and

MP at 4.5 months, Group, First session speaker, and *Visual Vowel* as fixed factors (the model

94

structure allowed for interaction between all factors), while the random effect was the *by-subject intercept*. Since the model did not allow variables with very different scales, as suggested by *lmer*, the MPs at 4.5 months and MPs at 8 months were rescaled. The *p* values for the fixed effects were provided by the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2017). Inspection of the model estimates and the given *p* values indicated only a trend for the interaction between *MP at 4.5 months* and the *First session speaker* (Intercept estimate = 0.52, SE = .16, *p* = .002; *MP at 4 months* x *First session speaker* estimate = 1 .07, SE = .61, *p* = .09). This marginal interaction suggests that infants' preference shifting from mismatch at 4.5 months to match at 8 months of age for those infants who were first presented with the Speaker (**see Figure 4-3**). Further visual inspection of the intercepts for each *First session speaker* (**see Figure 4-3)** confirmed the findings from the cross-sectional data—the infants presented with Speaker 2 in their first session (at 4.5 months) tended to attend longer to the mismatch condition than the infants presented with Speaker 1. No other factors significantly contributed to the MP score at 8 months (all *ps* >.2).
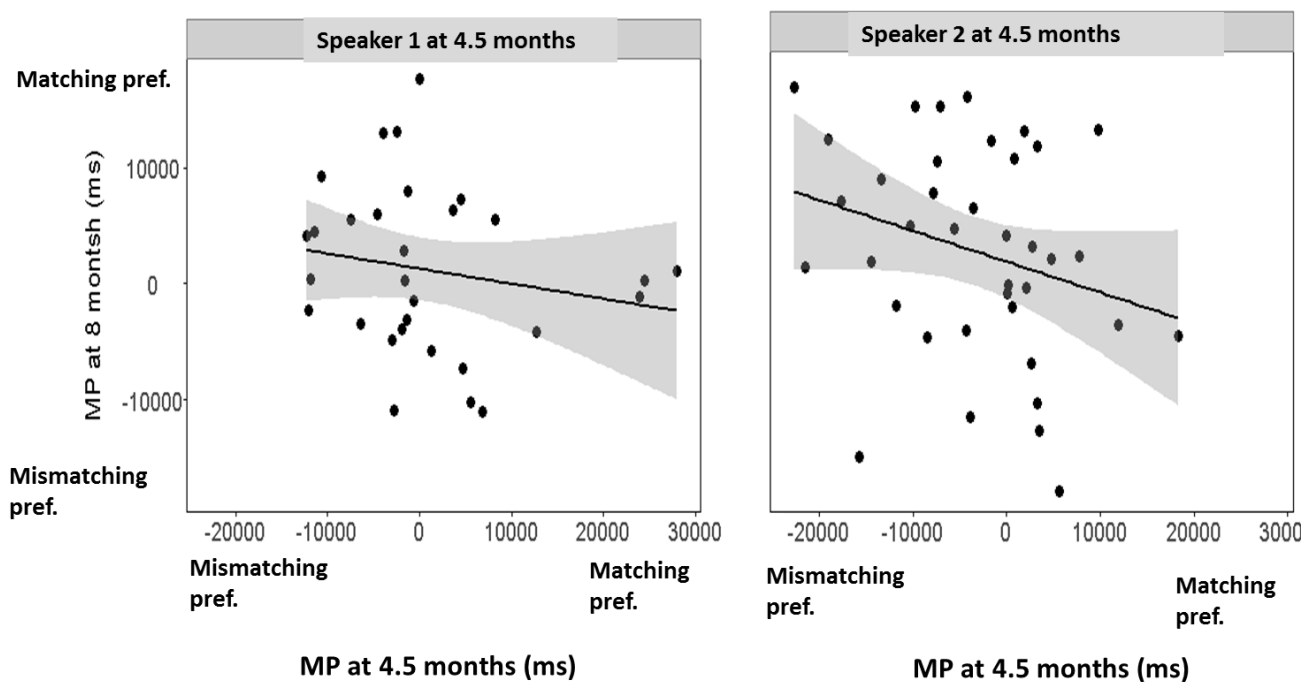
Figure 4-3. Scatter plot representing the relation between Matching Preference (MP) at 4.5 and at 8 months of age across the two speakers presented to infants at 4.5 months. Positive values reflect matching, while negative reflect mismatching preference. Shaded areas represent 95% of the confidence region.

4.3.2   Eyes-mouth preference

We calculated the average eyes-mouth (EM) preference scores (as in section 2.3) for each infant, separately for match and mismatch conditions. To assess whether individual differences might explain the wide distribution of the EM scores at 8 months of age, we computed correlations on EM scores at 4.5 months and 8 months, separately for the match and mismatch conditions, as well for each language group. Results are given in **Figure 4-4**.

For the match condition the analysis revealed a significant positive correlation in the monolingual group ($r_{(16)} = .51$, $p = .03$), and a strong trend indicating a correlation in the bilingual group ($r_{(14)} = .46$, $p = .07$). The results were similar for the mismatch condition; $r_{monolingual(16)} = .55$, $p = .01$, $r_{bilingual(14)} = .45$, $p = .08$. Note that the correlations between the monolingual and bilingual groups did not differ ($ps > .4$). Therefore, we demonstrated that,

96

regardless of the group, the EM preference, at 4.5 months was more likely to persist at 8 months. In other words, the *individual* EM was less likely to change over development (see **Figure 4-4**).
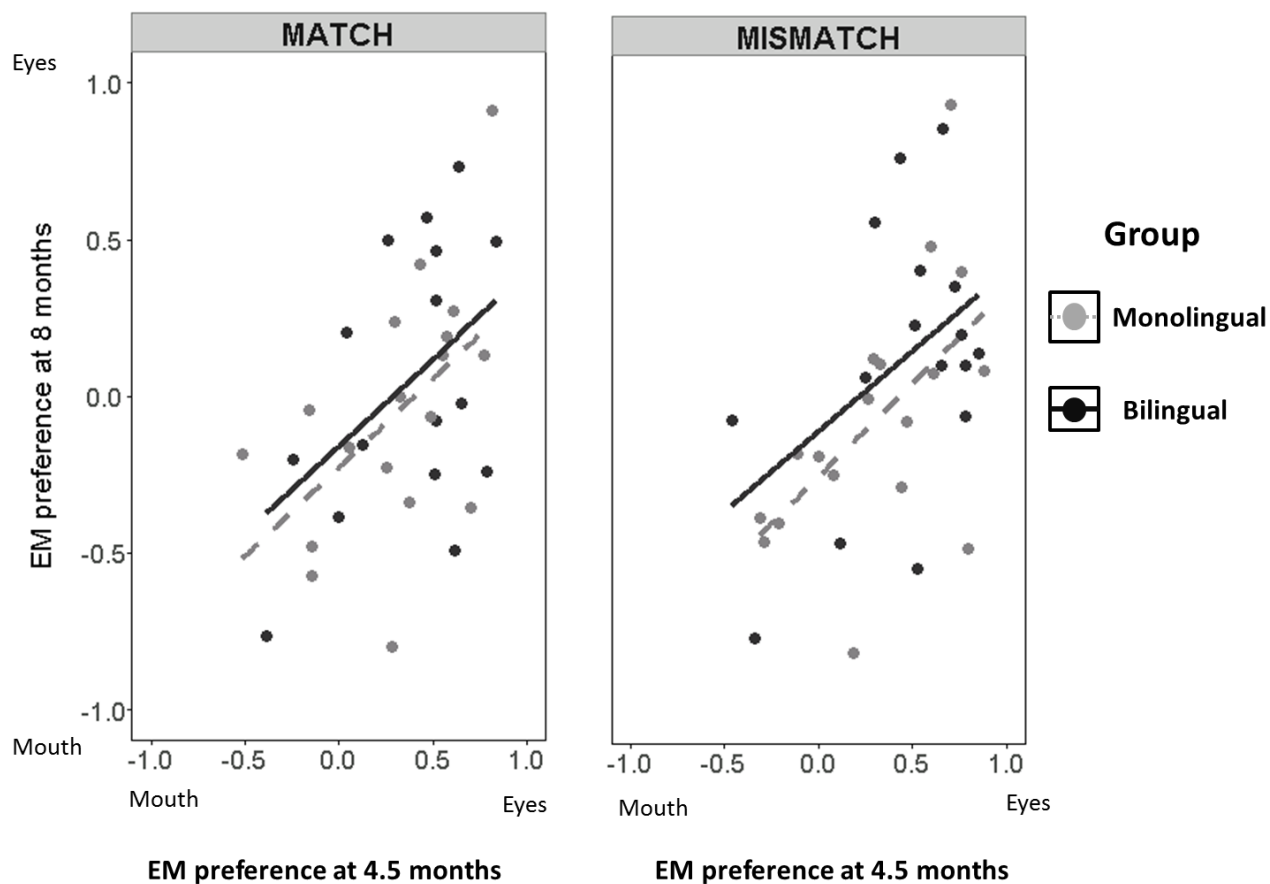

Figure 4-4. Scatter plot representing the relation between eyes-mouth (EM) preference at 4.5 and at 8 months of age across the monolingual (in grey) and bilingual (in black) group. Positive values reflect the eyes (e.g., infants spent more time looking at the eyes of the speaker as opposed to their mouth), whereas negative values reflect the mouth preference.

As in the previous analysis (see 4.3.1) we ran a linear mixed-model analysis to include the additional factors of *Visual Vowel* and *First session speaker*. The EM scores were averaged for each *Visual Vowel* (/i/ and/ u/), such that each infant contributed with two data points for each of the vowels. The model[12] included the *EM score at 8 months* as the dependent variable and the *EM score at 4 months, Group, First session speaker, Condition*

---

[12] Note that more simple models conducted separately for match and mismatch conditions, revealed the same pattern of results.

Results

and *Visual Vowel* as fixed factors (the model structure allowed for interactions between all factors), while the random effect was *by-subject intercept*. The analysis revealed that the *EM score at 4 months* significantly predicts the *EM score at 8 months* and that that relation is different across the vowels (see Table 4-2). Further visual inspection of scatterplots (see *Figure 4-5*) and correlation analyses demonstrated that although the correlations are significant for both vowels, the relationship seem to be stronger for the vowel /i/ ($r_{(66)} = .53$, $p < .001$) than for the vowel /u/ ($r_{(66)} = .36$, $p = .002$). This finding on the effect of the visual vowel is in line with the cross-sectional data (see **Figure 3–7**), suggesting that EM preference differ across vowels during development. Importantly, the *Group* and other factors did not significantly contribute to the EM score at 8 months (all *ps* >.2).

Table 4-2

Mixed-effect model's output for the fixed factors that significantly contributed to Eyes-Mouth (EM) score at 8 months of age

Fixed effects	Estimates	SE	*Lower CI*	*Upper CI*	*P*
Intercept	-.08	.08	-.22	.07	.28
EM at 4.5 months	.27	.17	.05	.46	.01
EM at 4.5 months*Visual Vowel	.17	.59	.07	.27	.004

Note. SE = standard error, CI = confidence level (95%).
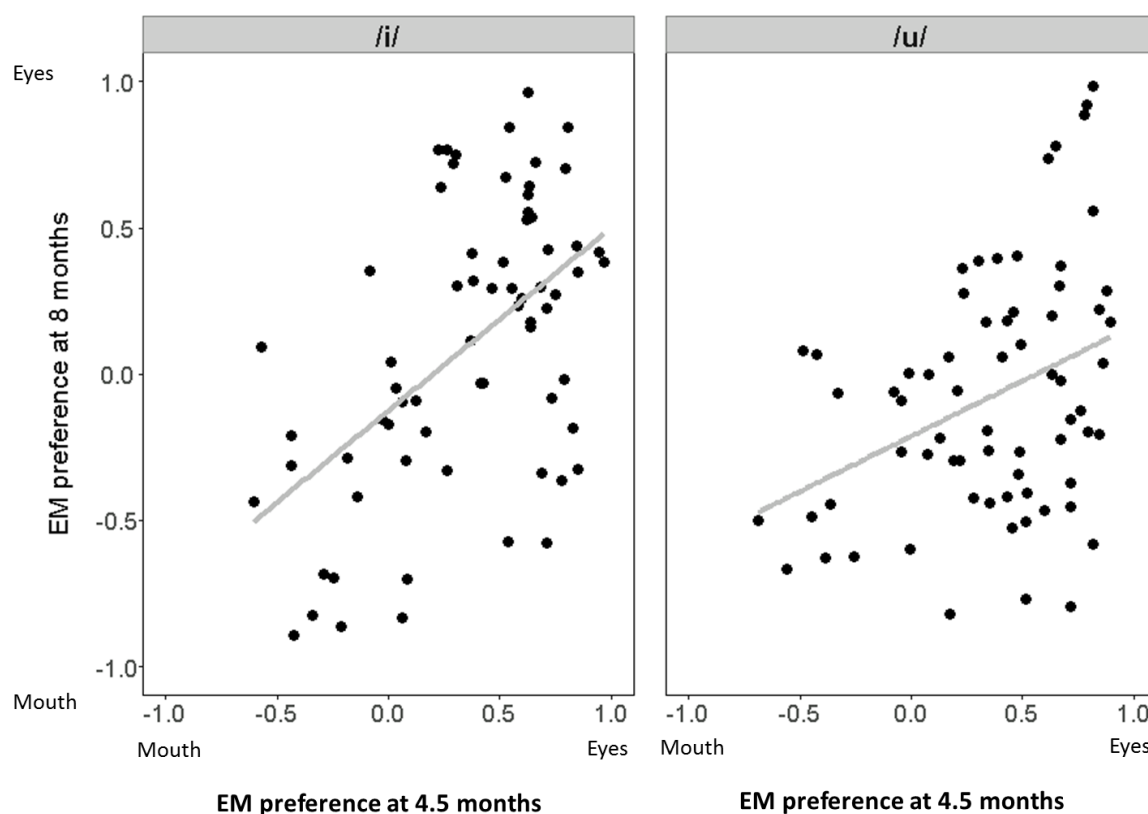
Figure 4-5. Scatter plot representing the relation between Eyes-Mouth (EM) preference at 4.5 and at 8 months of age across the two vowels. Positive values reflect the eyes preference (e.g., infants spent more time looking at the eyes of the speaker as opposed to the mouth of the speaker), negative values reflect the mouth preference.

4.4   Discussion

In this chapter, we assessed *longitudinally* the development of AV matching ability and selective attention to the eyes and the mouth in monolingual and bilingual infants between 4.5 and 8 months of age. The longitudinal data provided a unique opportunity to compare the development of AV processing within infants. Therefore, we were able to understand the individual variations in the AV matching ability and selective attention observed in the cross-sectional data. Here, we studied whether there is a relationship between early and late preferences in infants' looking behavior, specifically in the AV matching ability and the EM preference. Also, we examined whether linguistic experience (i.e. monolingual vs. bilingual language background) affected this relationship.

We found no evidence that the AV matching ability early on (here, 4.5 months) was related to the AV matching ability at 8 months. The same results were observed for both the monolingual and the bilingual groups. Further inspections revealed a trend suggesting that infants who were presented with Speaker 2 at 4.5 months of age tended to exhibit a mismatch preference (i.e. attend longer to the mismatch over the match condition), while at 8 months they tended to exhibit a match preference. Although this trend did not reach statistical significance, it is in line with the cross-sectional finding—an AV matching ability in 4.5-month-olds was observed for Speaker 2, but not for Speaker 1. The interesting trend in shifting the preference from mismatch at 4.5 months to match at 8 months of age for those infants who were first presented with the Speaker 2 is in contrast to the common pattern observed in infant visual attention studies. Specifically, previous studies on infants' visual attention demonstrated that infants younger than six months commonly exhibit a familiarity preference (i.e. attending longer to familiar/known stimuli than to novel/unknown), while after six months of age they exhibit a novel preference (e.g., Greenberg, Uzgiris, & Hunt, 1970; Houston-Price & Nakai, 2004; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982; Wetherford & Cohen, 1973). Similarly, research on AV matching ability (defined as the difference in the looking time between AV match and mismatch events) found that infants younger than six months of age attended longer to AV match events (e.g., Altvater-Mackensen & Grossmann, 2015; Altvater-Mackensen et al., 2015; Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 2003). However, some studies suggested that the familiarity preference might persist in older infants if familiarization to stimuli is brief (e.g., Rose et al., 1982). Because we did not familiarize infants with our stimuli, it is possible that this affected infants' matching preference, similarly as in studies with brief familiarization. Further research is needed to determine whether the previous familiarization affects the looking behavior in the AV matching ability tasks.

Regarding the longitudinal data on the EM preference, we demonstrated a strong relationship between early and later infant looking behavior. Specifically, we observed that even though infants' overall preference between the eyes and mouth changed with development (i.e., increased attention to the mouth in Chapter 3; see **Figure 3-4** and **3-5)** *individual* EM preference is less likely to change over development, regardless of the AV match or mismatch condition. This finding is in line with a previous study demonstrating that infants' individual attention to the mouth at the age of 9 and 12 months is predicted by their attention to the mouth at an earlier age (i.e., at 6 months of age; Tenenbaum et al., 2013). Therefore, the individual preference observed in the longitudinal data can explain the wide distribution of the EM preference at 8 months in the cross-sectional data. Note that the correlation between the EM preference at 4.5 and 8 months was moderate, suggesting that over the course of their development, infants slowly increased their attention to the mouth, rather than categorically shifting their attention from the eyes to the mouth.

Similar to the cross-sectional data, we observed that the relation of the EM score at 4.5 and 8 months of age differed depending on the visual vowels infants were presented with. In particular, the correlation seemed to be stronger for the vowel /i/ than the vowel /u/. Furthermore, visual inspection of the data suggested that more infants exhibit the eyes preference when presented with the vowel /i/, and the mouth preference when presented with the vowel /u/, as was also observed in the cross-sectional findings (3.3.4).

Finally, we observed that the EM preference develops similarly across the monolingual and bilingual groups; hence bilingualism did not modulate the development of the EM preference. Rather, our results are in line with a study demonstrating that bilingualism did not affect selective attention to the mouth (Tsang et al., 2018).

To summarize, the longitudinal data on overall *AV matching ability* did not reveal a relationship between early and late matching ability. Further analysis showed that infants' AV

matching ability might represent a dynamic process that changes over development (i.e., from mismatching to matching preference). More research is needed to address what factors affect the developmental changes in AV matching ability. More interestingly, our longitudinal study revealed that the EM preference at 4 months of age predicted the preference at 8 months of age. Therefore, the large variance in the eyes vs. mouth preference found in the 8-month-olds' cross-sectional data could be explained by infants' individual eyes-mouth preference.

5 Chapter 5

Conclusions

The current dissertation addressed to what extent monolingual and bilingual infants share strategies employed during audiovisual (AV) speech processing. Specifically, whether bilingual infants develop an increased sensitivity to visual speech cues—a proposal that is currently under debate (Pons et al.,2015; Tsang et al., 2018).

To this end we systematically assessed monolingual and bilingual infants' sensitivity to visual speech cues focusing on two parameters: (1) infants' ability to match auditory and visual vowels, and (2) infants' attention to the mouth of a speaker during the AV vowel matching task. We tested linguistically homogeneous groups of 4.5- and 8-month-old Spanish/Basque monolinguals and Spanish-Basque bilinguals with vowels (i.e., /i/ and /u/) that are shared and native across the linguistic groups. The study combined cross-linguistic and a longitudinal design.

Overall, the thesis provided evidence that monolingual and bilingual infants exhibit similar sensitivity to visual speech during AV vowel processing. This claim is supported by three main findings. First, monolinguals and bilinguals exhibited similar AV matching ability at both ages. Second, monolinguals and bilinguals did not differ in their attention to the mouth of a speaker during the AV matching task. Finally, our longitudinal data demonstrated that both monolingual and bilingual infants showed a similar trajectory in their development of AV matching ability and attention to the mouth from 4.5 to 8 months of age. Taken together, our findings demonstrated that in a testing paradigm where the linguistic stimuli are shared across the language groups under investigation, there is no clear difference between monolingual and bilingual infants in processing visual speech.

In addition to our main goal, the study contributed to research on the development of AV speech perception by demonstrating that the visual salience of the stimuli (e.g., the type of vowel) and inter-speaker differences both play a role in infants' ability to detect AV mismatch. This finding is crucial for understanding what perceptual cues affect AV matching ability. Furthermore, we extended the findings on infants' first attention shift from the eyes to the mouth between 4 to 8 months of age during AV vowel processing. We demonstrated that 4-month-olds attended longer to the eyes when scanning a talking face, whereas 8-month-olds attended less to the eyes, but more to the mouth. Finally, our longitudinal study revealed that the eye-mouth preference at 4 months of age predicted the preference at 8 months of age. This finding emphasized the role of individual variability in interpreting the mechanisms underlying AV speech processing and face scanning in infancy.

5.1   Open questions & future directions

The current study employed vowels as stimuli that are native to both Spanish/Basque monolingual and Spanish-Basque bilingual infants, and in fact our study is the first attempt to equalize the task and the stimuli across monolingual and bilingual group. However, infants' exposure to AV speech is certainly far more complex than processing isolated vowels. Our study was not designed to address infants' processing of AV continuous speech—certainly a more ecological linguistic stimuli. Therefore, it is unclear whether current findings can be generalized to the processing of continuous speech. Future research is needed to address monolinguals' and bilinguals' sensitivity to visual speech with stimuli that are linguistically more complex, but at the same time equally demanding across the groups. A potential future next step could be examining monolingual and bilingual infants processing words that are shared across the languages (e.g., cognates).

Further, more research is needed to determine *when* bilingualism plays a role in the development of AV speech processing. Our findings suggest that bilingualism may not affect

infants' AV speech processing in the first eight months of life. However, the adult studies show that bilingualism does impact adult AV speech processing (e.g., Hillairet de Boisferon et al., 2016; Marian et al., 2018; Soto-Faraco et al., 2007). Recent evidence suggested that bilingual linguistic experience starts to shape AV speech processing around 15 months of age in bilinguals learning rhythmically close languages, as Spanish and Catalan (Birules et al., 2018). Considering our current sample, Spanish and Basque are considered rhythmically similar languages, with within-class rhythmic differences at the phrasal level (Molnar et al., 2016; Molnar, Gervain, et al., 2014), therefore, it is possible that in our sample, bilingual-monolingual differences in AV speech processing also emerge at a later age (i.e., at 15 months). Future research is needed to systematically study when/whether bilingualism starts to affect AV speech processing across various different types of language groups.

Overall, the current study together with other recent studies on AV speech processing in monolingual and bilingual infants have made fundamental contributions and raised several crucial research questions relevant to developmental frameworks of language acquisition within the fields of Linguistics, Psychology, and Cognitive Science. Our findings have contributed to the literature emphasizing that visual aspects of speech play a role in early speech perception development. Importantly, the study indicated that sensitivity to visual speech is affected by age, visual salience of the stimuli, language background, task difficulty, as well as stimulus value. Undoubtedly, answering the new questions arising out of the current dissertation will get us closer to understanding the roots of language acquisition across linguistically diverse populations.

References

Altvater-Mackensen, N., & Grossmann, T. (2015). Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories. *Child Development*, *86*(2), 362–378. https://doi.org/10.1111/cdev.12320

Altvater-Mackensen, N., Mani, N., & Grossmann, T. (2015). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis)matches between auditory and visual speech cues. *Developmental Psychology*, *52*(2), 191–204. https://doi.org/10.1037/a0039964

Andrew Duchowski. (2007). *Eye Tracking Methodology* (Second). London: Springer.

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy*, *6*(2), 155–163. https://doi.org/10.1207/s15327078in0602_1

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

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

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In *Multisensory Development*. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bahrick, L. E., & Lickliter, R. (2014). Learning to Attend Selectively: The Dual Role of Intersensory Redundancy. *Current Directions in Psychological Science*, *23*(6), 414–420. https://doi.org/10.1177/0963721414549187

Baier, R., Idsardi, W. J., & Lidz, J. (2007). Two-month-olds are sensitive to lip rounding in dynamic and static speech events. *Proceedings of the International Conference on Auditory-Visual Speech Processing.*, paper L6–2.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

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

Barreña, A., Ezeizabarrena, M. J., & García, I. (2011). La influencia del grado de exposición a la lengua en la adquisición del euskera (8-30 meses) [The influence of the degree of exposure to the language in the early acquisition of Basque (8-30 months)]. *Infancy & Learning/Infancia Y Aprendizaje*, *34*(4), 393–408. https://doi.org/10.1016/j.cub.2015.12.035

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear and Hearing*, *26*(2), 149–164.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism (Cambridge, England)*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bialystok, E., & Werker, J. F. (2017, January 1). Editorial: The systematic effects of bilingualism on children's development. *Developmental Science*, *20*(1), e12535. https://doi.org/10.1111/desc.12535

Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2016). Early Prosodic Acquisition in Bilingual Infants: The Case of the Perceptual Trochaic Bias. *Frontiers in Psychology*, *7*(February), 1–8. https://doi.org/10.3389/fpsyg.2016.00210

Birules, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2018). Inside Bilingualism: Language Background Modulates Selective Attention to a Talker's Mouth. *Developmental Science*, e12755. https://doi.org/10.1111/desc.12755

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. (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

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*, 455–474. https://doi.org/10.1017.S0142716407070257

Byers-Heinlein, K. (2018). Spoken Word Recognition. In *The Listening Bilingual: Speech Perception, Comprehension, and Bilingualism, First Edition*.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech. *PLoS Computational Biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

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

Colombo, J., & Fagen, J. W. (1990). *Individual differences in infancy: Reliability, stability, prediction. Individual differences in infancy: Reliability, stability, prediction.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual Differences in Infant Visual Attention: Are Short Lookers Faster Processors or Feature

Processors? *Child Development*, *62*(6), 1247–1257. https://doi.org/10.1111/j.1467-8624.1991.tb01603.x

Coulon, M., Hemimou, C., & Streri, A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, *18*(5), 782–796. https://doi.org/10.1111/infa.12001

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. https://doi.org/10.1037/dev0000083

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*. https://doi.org/10.1016/j.cogdev.2017.02.004

Dorn, K., Weinert, S., & Falck-Ytter, T. (2018). Watch and listen – A cross-cultural study of audio-visual-matching behavior in 4.5-month-old infants in German and Swedish talking faces. *Infant Behavior and Development*, *52*(May), 121–129. https://doi.org/10.1016/j.infbeh.2018.05.003

Ezeizabarrena, M. J., & Barreña, A. (2013). Desarrollo lingüístico temprano en la adquisición monolingüe y bilingüe del euskera [The linguistic development of Basque in early monolinguals and bilinguals]. In J. Dolz & I. Idiazabal (Eds.), *Enseñar (lenguas) en contextos multilingües [Teaching (languages) in multilingual contexts]* (pp. 243–280).

Ezeizabarrena, M. J., & Fernandez, I. G. (2018). Length of utterance, in morphemes or in words?: MLU3-w, a reliable measure of language development in early basque. *Frontiers in Psychology*, *8*(JAN), 2265. https://doi.org/10.3389/fpsyg.2017.02265

Fennell, C., & Lew-Williams, C. (2017). Early bilingual word learning. In *Early Word Learning* (pp. 110–122). https://doi.org/10.4324/9781315730974

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fernald, H., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherse speech. *Infant Behaviour and Development*, *10*, 279–293.

Flege, J. E., & Eefting, W. (1987). *Production and perception of English stops by native Spanish speakers*. *Journal of Phonetics* (Vol. 15).

Gervain, J. (2015). Plasticity in early language acquisition: the effects of prenatal and early childhood experience. *Current Opinion in Neurobiology*, *35*, 13–20. https://doi.org/10.1016/J.CONB.2015.05.004

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*(1), 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory Redundancy Facilitates Learning of Arbitrary Relations between Vowel Sounds and Objects in Seven-Month-Old Infants.

Journal of Experimental Child Psychology, *69*(2), 133–149.
https://doi.org/10.1006/jecp.1998.2438

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye Tracking in Infancy Research.
Developmental Neuropsychology, *35*(1), 1–19.
https://doi.org/10.1080/87565640903325758

Greenberg, D., Uzgiris, I. C., & Hunt, J. M. (1970). Attentional preference and experience:
III. Visual familiarity and looking time. *Journal of Genetic Psychology*, *117*(1), 157–
168. https://doi.org/10.1080/00221325.1970.10533942

Hainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns
by infants. *Journal of Experimental Child Psychology*, *25*(1), 90–115.
https://doi.org/10.1016/0022-0965(78)90041-3

Haith, M. M. (2004). Thematic Collection : Commentaries Progress and Standardization in
Eye Movement Work With Human Infants. *Infancy*, *6*(2), 257–265.
https://doi.org/10.1207/s15327078in0602

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye Contact and Face Scanning in Early
Infancy Author (s): Marshall M . Haith , Terry Bergman and Michael J . Moore.
Science, *198*(4319), 853–855.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2016). Selective
attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and
linguistic experience. *Developmental Science*, *20*(3), 1–12.
https://doi.org/10.1111/desc.12381

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 in the Second Year of Life.
Journal of Experimental Child Psychology, *172*, 189–200.
https://doi.org/10.1016/j.jecp.2018.03.009

Hockley, N. S., & Polka, L. (2005). A developmental study of audiovisual speech perception
using the McGurk paradigm. *The Journal of the Acoustical Society of America*, *96*(5),
3309–3309. https://doi.org/10.1121/1.410782

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language
exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.
https://doi.org/10.1017/S0305000910000759

Houston-Price, C., & Nakai, S. (2004a). Infant and Child Development Theory and Method in
Developmental Research Distinguishing Novelty and Familiarity Effects in Infant
Preference Procedures, *13*, 341–348. https://doi.org/10.1002/icd.364

Houston-Price, C., & Nakai, S. (2004b). Infant and Child Development Theory and Method in
Developmental Research Distinguishing Novelty and Familiarity Effects in Infant
Preference Procedures. *Inf. Child Dev*, *13*, 341–348. https://doi.org/10.1002/icd.364

Hualde, J. I. (1991). *Basque Phonology*. London: Routledge.

Hualde, J. I., Olarrea, A., Escobar, A. M., Travis, C. E., Escobar, A. M., & Travis, C. E. (2010). *Introducción a la lingüística hispánica. Introducción a la Lingüística Hispánica*.

Hunnius, S., & Geuze, R. H. (2004). Developmental Changes in Visual Scanning of Dynamic Faces and Abstract Stimuli in Infants: A Longitudinal Study. *Infancy*, *6*(2), 231–255. https://doi.org/10.1207/s15327078in0602_5

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*(2 Pt 1), 701–714. https://doi.org/10.1121/1.406887

Keough, M., Derrick, D., & Gick, B. (2019). Cross-Modal Effects in Speech Perception. *Annual Review of Linguistics*, (July 2018), 1–18.

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by Eye and Ear: Infants Perceive Visual Prosody in Point-Line Displays of Talking Heads. *PLoS ONE*, *9*(10), e111467. https://doi.org/10.1371/journal.pone.0111467

Kitamura, C., & Kim, J. (2010). Infants match auditory and visual speech in schematic point-light displays. *International Conference on Audio-Visual Speech Processing*.

Kubicek, C., De Boisferon, A. H., Dupierrix, E., Pascalis, O., Loevenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual german and french fluent speech in infancy. *PLoS ONE*, *9*(2), e89275. https://doi.org/10.1371/journal.pone.0089275

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

Kuhl, P., & Meltzoff, A. (1984). The Intermodal Representation of Speech in Infants. *Infant Behavior and Development*, *7*(3), 361–381. https://doi.org/10.1016/S0163-6383(84)80050-8

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Larraza, S., Molnar, M., & Samuel, A. (2019). *Language-specific experience does not modify the perception of all speech sounds.* Manuscript to be resubmitted.

Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2016). Vocabulary size, translation equivalents, and efficiency in word recognition in very young bilinguals. *Journal of Child Language*, *43*(4), 760–783. https://doi.org/10.1017/S0305000915000252

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

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

Liu, H.-M., Kuhl, P., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1–F10.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*(4590), 1347–1349. https://doi.org/10.1126/science.6828865

Mair, P., & Wilcox, R. R. (2017). Robust statistical methods in R using the WRS2 package. *Journal of Statistical Software*, forthcoming. https://doi.org/10.18637/jss.v000.i00

Marian, V., Hayakawa, S., Lam, T. Q., & Schroeder, S. R. (2018). Language experience changes audiovisual perception. *Brain Sciences*, *8*(5), 1–14. https://doi.org/10.3390/brainsci8050085

Maurer, D., & Salapatek, P. (1976). Developmental Changes in the Scanning of Faces by Young Infants. *Child Development*, *47*(2), 523. https://doi.org/10.2307/1128813

Maurer, D., & Werker, J. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2). https://doi.org/10.1002/dev.21177

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2018). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, (October 2017), e12701. https://doi.org/10.1111/desc.12701

Molnar, M., Alemán Bañón, J., Mancini, S., & Caffarra, S. (2018). *The processing of article-noun gender agreement in Spanish by monolingual and bilingual toddlers*. Manuscript submitted for publication.

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, *152*, 150–159. https://doi.org/10.1016/J.COGNITION.2016.03.023

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

Molnar, M., Lallier, M., & Carreiras, M. (2014). The Amount of Language Exposure Determines Nonlinguistic Tone Grouping Biases in Infants From a Bilingual Environment. *Language Learning*, *64*(s2), 45–64. https://doi.org/10.1111/lang.12069

Mugitani, R., Kobayashi, T., & Hiraki, K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior and Development*, *31*(2), 307–310. https://doi.org/10.1016/j.infbeh.2007.12.002

Mullen, E. M. (1995). Mullen Scales of Early Learning, AGS Edition: Manual and Item Administrative Books. *American Guidance Services, Inc.*, 1–92.

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

Muru, A. M., & Lee, S. A. S. (2017). Development of Phonetic Categories of Stop Consonants in Spanish-English Bilingual Children. *Clinical Archives of Communication Disorders*, *2*(1), 60–68. https://doi.org/10.21849/cacd.2017.00066

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

Navarra, J., Yeung, H. H., Werker, J., & Soto-Faraco, S. (2012). Multisensory Interactions in Speech Perception. *The New Handbook of Multisensory Processing*, 435–452.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. https://doi.org/10.1037/1082-989X.8.4.434

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, *18*(1), 1–25. https://doi.org/10.1017/S0272263100014662

Patterson, M., & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M., & Werker, J. (2002). Infants' Ability to Match Dynamic Phonetic and Gender Information in the Face and Voice. *Journal of Experimental Child Psychology*, *81*(1), 93–115. https://doi.org/10.1006/jecp.2001.2644

Patterson, M., & Werker, J. (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

Pearson, B. Z. (1998). Assessing Lexical Development in Bilingual Babies and Toddlers Vocal Development in infants and toddlers, Bilingual supplement View project. https://doi.org/10.1177/136700699800200305

Pejovic, J., & Molnar, M. (2017). The development of spontaneous sound-shape matching in monolingual and bilingual infants during the first year. *Developmental Psychology*, *53*(3), 581–586. https://doi.org/10.1037/dev0000237

Piantadosi, S. T., Kidd, C., & Aslin, R. (2014). Rich analysis and rational models: inferring individual behavior from infant looking data. *Developmental Science*, *17*(3), 321–337. https://doi.org/10.1111/desc.12083

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2-3), 227–272. https://doi.org/10.1016/0001-6918(94)90004-3

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, *26*(4), 490–498. https://doi.org/10.1177/0956797614568320

Pons, F., Lewkowicz, D. J., 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

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704–713. https://doi.org/10.1037/0012-1649.18.5.704

Ruff, H. A., & Rothbart, M. K. (2010). *Attention in Early Development: Themes and Variations. Attention in Early Development: Themes and Variations*. https://doi.org/10.1093/acprof:oso/9780195136326.001.0001

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

Soto-Faraco, S., Navarra, J., Weikum, W., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, *69*(2), 218–231. https://doi.org/10.3758/BF03193744

Streri, A., Coulon, M., Marie, J., & Yeung, H. H. (2016). Developmental Change in Infants' Detection of Visual Faces that Match Auditory Vowels. *Infancy*, *21*(2), 177–198. https://doi.org/10.1111/infa.12104

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*(1), 232–242. https://doi.org/10.1016/j.cognition.2007.12.013

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. (2013). Increased Focus on the Mouth Among Infants in the First Year of Life: A Longitudinal Eye-Tracking Study. *Infancy*, *18*(4). 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). https://doi.org/10.1017/S0305000914000725

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

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

Umbel, V. M., Pearson, B. Z., Fernandez, M. C., & Oller, D. K. (1992). Measuring Bilingual Children's Receptive Vocabularies. *Child Development*, *63*(4), 1012–1020. https://doi.org/10.1111/j.1467-8624.1992.tb01678.x

Van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00388

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

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. (2007). Visual Language Discrimination in Infancy. *Science*, *316*(5828), 2007. https://doi.org/10.1126/science.1137686

Werker, J. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*(4), 703–728. https://doi.org/10.1017/S0142716418000152

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

Werker, J., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4

Wetherford, M. J., & Cohen, L. B. (1973). Developmental Changes in Infant Visual Preferences for Novelty and Familiarity. *Child Development*, *44*(3), 416. https://doi.org/10.2307/1127994

Wilcox, T., Stubbs, J. A., Wheeler, L., & Alexander, G. M. (2013). Infants' scanning of dynamic faces during the first year. *Infant Behavior and Development*, *36*(4), 513–516. https://doi.org/10.1016/J.INFBEH.2013.05.001

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289–297. https://doi.org/10.3758/BF03199477

Yeung, H. H., & Werker, J. (2013). Lip Movements Affect Infants' Audiovisual Speech Perception. *Psychological Science*, *24*(5), 603–612. https://doi.org/10.1177/0956797612458802

A Appendix

A. 1 Eye-tracker methodology

The eye-tracking methodology is a well-established method for measuring eye movements. In the last 50 years, the eye-tracking method has been widely used in research with adults (see *Figure A-1*, panel A), while it has been less common in infant research. However, eye-tracking studies with infants in the last decade have been increasing (see *Figure A-1*, panel B). This increase probably occurred because eye-tracking methodology has significantly improved, becoming more accessible to infant researchers. Current eye-trackers mostly rely on high-speed cameras recording one's eye-gaze. These eye-trackers have two important components (e.g., Andrew Duchowski, 2007). First is a *high-speed camera* that records the eye in high resolution, and the second is *an illuminator*, transmitting infra-red light to the eye (see *Figure A-2*). Infra-red light is invisible to the human eye, hence the light does not distract participants during the task.
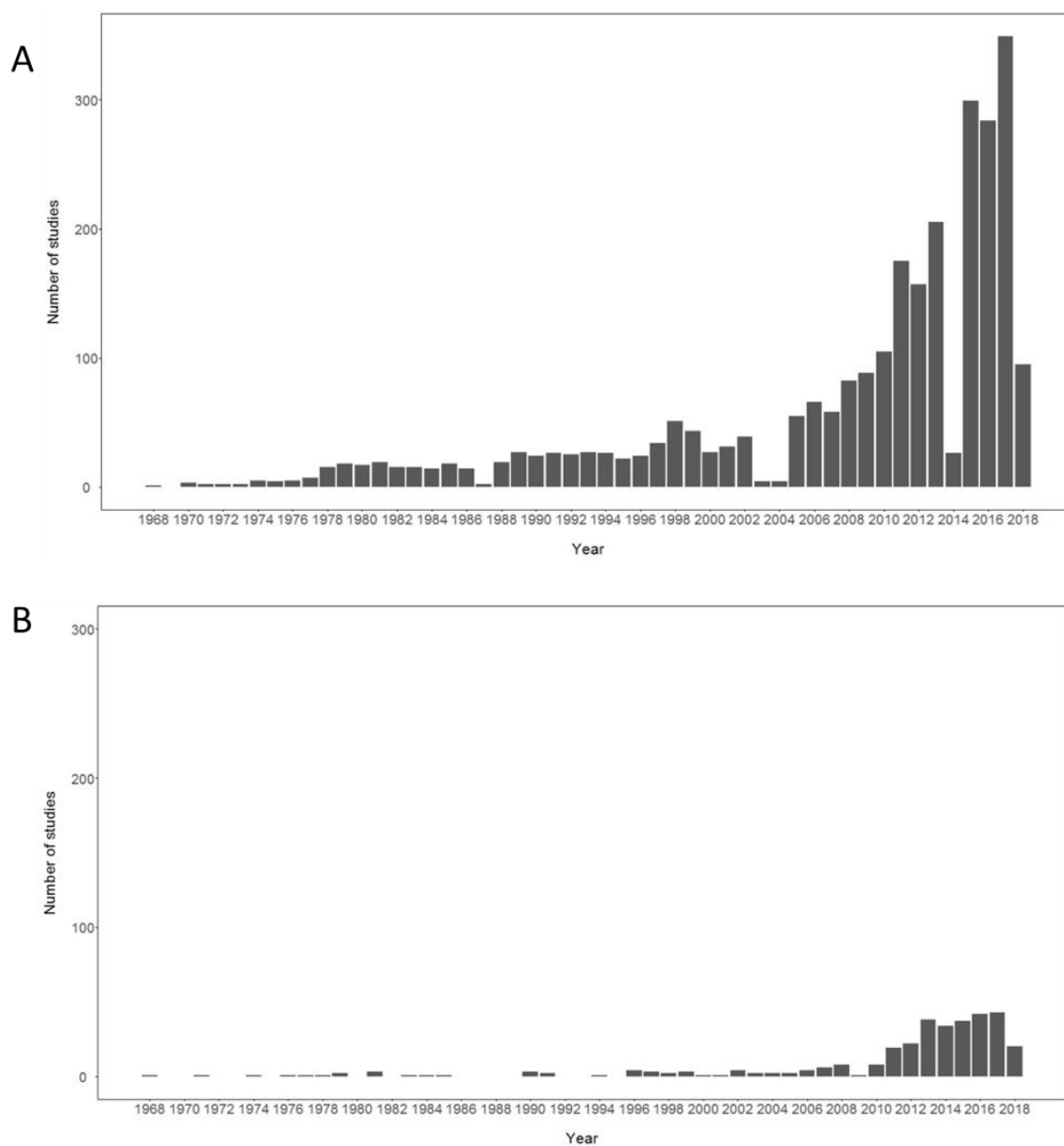
Figure A-1. Number of articles published per year and indexed in the PubMed database using eye-tracking methodology with adults (Panel A) and infants (Panel B). The data is indexed from 1968 to 2018. Retrieved July 15, 2018, from *https://www.ncbi.nlm.nih.gov/pubmed/*).

The technology of modern video-based combined pupil/corneal reflection in eye-trackers can simply be explained as follows. The illuminator with an infra-red filter transmits the infra-red light in participants' direction, reaching the eye, specifically the cornea. The cornea is a transparent membrane that covers the frontal part of the eye, covering the iris, the pupil and the lens (e.g., Andrew Duchowski, 2007; Gredebäck, Johnson, & von Hofsten, 2010). Importantly, because of its transparent properties and gel-like structure, the cornea transmits, but also reflects light. This corneal reflection is recorded by the high-speed and high-resolution camera, and the recorded eye-image is later processed by algorithms built into the eye-tracker. The algorithms also process the position of the pupil, by recognizing the darkest (in some eye trackers the brightest) rounded object on the eye-image as a pupil. Therefore, most of the modern eye trackers track two objects, the pupil, and the corneal reflection. *Figure A-2* schematically illustrates the eye-tracker system tracking the corneal reflection.
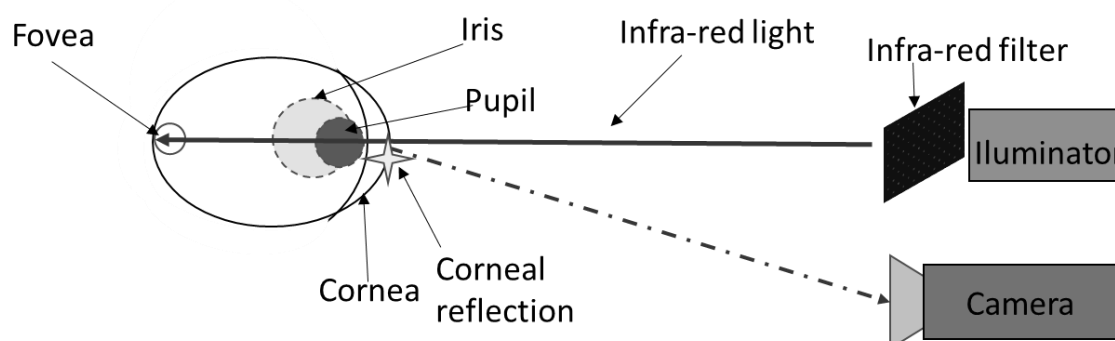

Figure A-2. The schematic illustration of a video-based eye-tracker system tracking the corneal reflection. The illuminator transmits infra-red light that reaches the eye, and through the cornea and the pupil, reaches the fovea. The cornea reflects a portion of the infra-red light, producing the corneal reflection, which is tracked by the camera.

Furthermore, built-in eye-tracking algorithms can detect the direction of one's eye-gaze (i.e., the eye-gaze position on a screen) based on the relative position of the pupil and the corneal reflection. For instance, as represented in *Figure A-3*, panel A, if the corneal reflection is positioned on a bottom-left portion of the pupil, the eye-gaze is in the top-right

direction. The direction of the eye-gaze for every tracking sample is transformed into x and y

coordinates relative to a screen, providing the exact location on a screen one is gazing at for

each time sampled. Based on this mechanism, we can calculate the dwell-time of a gaze, its

direction, and velocity (Haith, 2004). Later these parameters are used to calculate specific

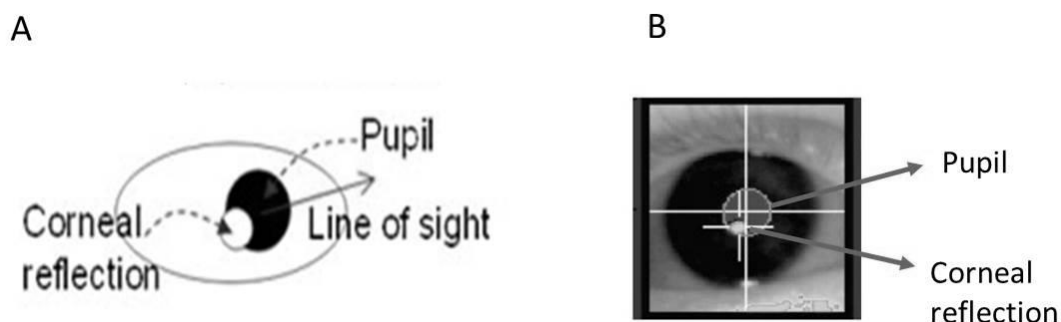eye-tracking measures as fixations, saccades, blinks, etc.

Figure A-3. Illustration of the relative position of the corneal reflection and pupil. Panel A
illustrates the eye-gaze direction based on the relative position of the corneal reflection and
pupil. The image is retrieved and adapted from
http://www.fujitsu.com/global/about/resources/news/press-releases/2012/1002-02.html. Panel
B represents the real image of the tracked eye, with the pupil and the corneal reflection
detected (adapted from EyeLink User Manual version 1.5.2).

A 1.1. <u>Eye-tracking in infancy: Similarities and differences with adult eye-tracking</u>

Although the logic and technology behind eye-tracking in infancy and adults rely on a

similar mechanism, there are few differences. First, to account for the head motion, instead of

stabilizing an infant's head with a chin-rest (as it is common practice with adults), a remote

(free-head) eye-tracking system is used[13] (Aslin & McMurray, 2004; Gredebäck et al., 2010).

Some remote eye-trackers detect infants' head position using a high-contrasted target (black

and white circles, that indeed appears like a target) placed on the infant's forehead. A built-in

algorithm calculates the relative distance between the eye and the target, therefore, when an

infant moves her/his head, the eye-tracker can easily detect the position of the eye just by

---

[13] There are also eye trackers that place a cap-like or eyeglass-like apparatus on infants' heads.
However, the majority of infant eye tracking studies use eye trackers that rely on a target.

detecting the target. Of course, remote eye-tracking systems cannot account for gross head movements, when the head moves beyond the camera's "visual field". Considering that infants often make large movements, infant eye-tracking suffers from greater data loss than adult. Moreover, calibration in infant eye-tracking is different than in adults. Calibration is important for eye-tracking because it determines the exact eye-gaze position relative to the screen. Usually, the more calibration targets that are employed, the better the estimation of eye-position. However, with infants, a long calibration is basically impossible, so fewer calibration targets are used. Consequently, calibration in infants is less accurate than with adults, yielding poorer data quality than in adults. Finally, infant and adult eye movements are very different in their neuromotor functioning, therefore for infant eye-tracking, it is necessary to adapt the equipment, specifically the illuminators, lens, and built-in algorithms, to make them appropriate for the specific properties of infants' eyes (Gredebäck et al., 2010).

Although eye-tracking in infants requires more effort, the data have advantages. First, eye-tracking has a high spatial and temporal resolution, meaning we can obtain precise data on *when* and *where* an infant is looking at, therefore we can examine what is in the focus of infants' visual attention on a very fine time scale (less than a second). Next, eye-tracking does not require hand coding of infants' looking behavior, making eye-tracking less time consuming than traditional looking techniques. Eye-tracking also allows for designs where the presentation of stimuli is conditioned by where the infant looks (i.e., a new stimulus will appear on a screen if an infant looks to a certain position).

A 1.2. Apparatus

The current study assessed the infant looking behavior using monocular eye gaze data (i.e., the eye-gaze sampling assessed from one eye) obtained from an EyeLink 1000 LCD Arm Mount eye-tracker (manufactured by SR Research) with integrated LCD screen, accuracy within 0.5 degrees, and a 500 Hz sample rate. As recommended by the Research for infant data acquisition, the eye-tracker was set to track both the pupil and the corneal

reflection, using a target sticker as a head position referent. A camera lens of 16mm was used

with an infrared illuminator of 940nm. *Figure A-4* schematically represents the eye-tracker

system used in the study. Auditory stimuli were played over two JBL-duet speakers placed

behind and on the sides of the screen, with 65-70 dB intensity. Visual stimuli were delivered

using the eye-tracking Acer AL1717 17″ monitor in 1024x768 resolution with 60 Hz

refreshing rate and with Windows XP.

The experiment was programmed in Experiment Builder Version 1.10.1385. The data

was processed in the DataViewer program (version 2.6.1), developed by SR Research.
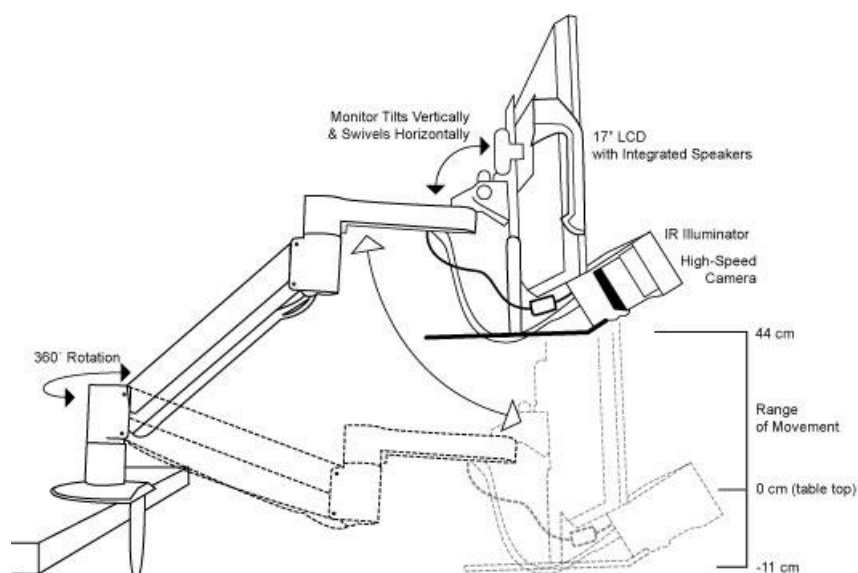

Figure A-4. Schematic representation of the EyeLink 1000 Arm Mount system used for the data acquisition. The image retrieved from EyeLink® 1000 Plus Installation Guide Version 1.0.9.

A. 2 Language questionnaire

Langauge questionnaire

Unique ID		Date of the experiment		1
Date of birth		Subject code		
Weeks		Father or mother/caregiver		
Days				
Siblings (how many)		Age of the siblings		

Father's family   **Between parents**   **Mother´s family**

Grandma   Father   gender of the baby   Mother   Grandma

Grandpa   Siblings   Grandpa

(1) How much time each person spends with the baby and in which langauge talks to the baby

Person	langauge(es)	Starting from which a In weeks	Days/Week		Hours/day		Total no of hours		
			M-F	S-S	M-F	S-S	M-F	S-S	
Mother									
Father									
Between parents									
Grandma (mother's)									
Grandpa (mother's)									
Grandma (father's)									
Grandpa (father's)									
Daycare									
Nanny									
Relative									
Relative									
Other									

Where do you live?

Do you attend some activities with your infant, like swimming lessons, masages? In which langauge?

Days per day

(4) TV/Radio

language

How many hours per day

(5) When you read to your baby

In which language you read

How much per day   minutes per day

How many days per week

(6) Excursions

Have you bee with your baby abroa more than 2 weeks

State

(7)parental estimate on language exposure

EUSK	
CAST	

A. 3 Health questionnaire

HEALTH QUESTIONNAIRE			
Subject code			
Health history			
Pregnancy date		Gestation date	
Notes on pregnancy (illness of the mother, medication, anxiety, vomit)			
Have you suffered from post-partum depression?		Medication	
C-section, forceps		Breast	Months
Birth weight			
Age of crawling		First steps	
Hospitalisation	Reason		
Has he suffered from ear infection?			
How often has the baby suffered from ear infection?			
At what age?	For how long?		
Did the baby receive medication for the ear infection?			
Has any of the familiy members of the baby had any language-related issues?			
(Problems to talk, late reading, problems understanding speech, problems writing, problems reading,etc.)			
Please, detail the issues and who suffered from it:			
Baby´s brothers or sisters			
Baby´s parents			
Baby´s grandparents			
Baby´s ncles/aunts			
Baby´s cousins			

A. 4   Consent form

BASIC INFORMATION AND INFORMED CONSENT FORM

BASIC INFORMATION AND CONSENT FORM

BABYLAB

PERSONAL INFORMATION OF THE PARTICIPANT	
Baby´s name and surname	
Name and surname of the legal guardian (Person legally in charge of the baby)	
Database ID	
Date of birth	SEX (M , F):
Date	

- This study is for research purposes only, the data gathered here will not be used for diagnostic or clinical purposes.
- No participant's identity will be revealed in any publication derived from this study.
- Parents or legal guardians are allowed to withdraw their consent at any time.
- Any question regarding this study, the technique used, or its objectives, may be asked at any time.
- BCBL researchers or responsible people will inform parents or legal guardians about the abovementioned details and about any other relevant information.
- BCBL researchers or responsible people will ask for your consent prior to the experiment.

AUTHORIZATION FOR THE PROCESSING OF PERSONAL DATA

We inform you that your personal data, or the data of the person you represent, will be processed and incorporated into files owned by the BCBL for their use in the management and administration of the centre and in studies on cognition, brain and language. The consent of the person concerned will legitimize the processing of data. These data will be kept indefinitely and may be granted, anonymously, to institutions or individuals with similar purposes, except for the legally established provisions. Your data will also be used to send you information about the BCBL's activities or publications, using the appropriate means.

☐ I do not wish to receive information on BCBL's publications or activities

The BCBL grants the confidentiality of your data, following the obligations set by the established laws. You have the right to access, amend, and delete your data and consent, according to L.O 15/ 1999, December the 13th, by writing an email to info@bcbl.eu or a post mail to Paseo Mikeletegi, 69 – 2ª - 20009 SAN SEBASTIÁN, attaching a copy of your ID..

After reading all of the above, I DECLARE that I have been properly informed of the details of the study, and I hereby authorize it.

Signature of the father/mother/guardian   Signature of the researcher/RA who informed

I´d like to be contacted again to take part in more studies: YES ☐ / NO ☐