EMPIRICAL STUDY

Impact of Bilingualism on Infants' Ability to Learn From Talking and Nontalking Faces

Mathilde Fort [iD],[a,b,c] Alba Ayneto-Gimeno,[a] Anira Escrichs,[a] and Nuria Sebastian-Galles[a]

[a]Universitat Pompeu Fabra, [b]Université Grenoble Alpes, and [c]Université Savoie Mont Blanc

To probably overcome the challenge of learning two languages at the same time, infants raised in a bilingual environment pay more attention to the mouth of talking faces than same-age monolinguals. Here we examined the consequences of such preference for monolingual and bilingual infants' ability to perceive nonspeech information coming from the eyes or the mouth region of talking faces. Using a learning procedure, we recorded 15-month-olds' and 18-month-olds' gaze while watching, at each trial, a speaker producing a sentence systematically followed by a nonspeech movement (eyebrow raise vs. lip protrusion). Differences were obtained for infants in the eyebrow-raise condition. While 15-month-old monolinguals and 18-month-old bilinguals learned to anticipate the eyebrow-raise movement before its appearance, 15-month-old bilinguals did not (i.e., they continued to look at the mouth region). Thus, bilingualism appears to impact not only how infants explore talking faces but also how they learn from them.

Keywords audiovisual; learning; attention; early language acquisition; infancy; bilingualism; talking faces

Introduction

Most of the time, humans perceive speech in an audiovisual fashion. This means that since their earliest stages of development, infants not only hear but also see the face of their social partners talking. Adults benefit from the presence of this visual speech information to decode the speech signal, notably when it is

---

Correspondence concerning this article should be addressed to Mathilde Fort, Center for Brain and Cognition, Departament de Tecnologies de la Informació i les Comunicacions, Speech Acquisition and Perception group, Universitat Pompeu Fabra, Edifici Mercè Rodoreda 24.313, Carrer Ramon Trias Fargas, 25–27, 08005 Barcelona, Spain. E-mail: mathilde.fort@upf.edu

1

uttered in noisy environments (Sumby & Pollack, 1954) or in second language (Navarra & Soto-Faraco, 2007). Infants also show some abilities to process visual speech and associate it with its corresponding auditory consequence (Bristow, Dehaene-lambertz, Mattout, Soares, & Gliga, 2009; Kubicek, De Boisferon et al., 2014; Kubicek, Gervain et al., 2014; Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Pons & Lewkowicz, 2014; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009).

Even if this ability remains rudimentary in infancy (e.g., see Lewkowicz, 2014, for a review), infants are already able to rely on the visual speech information to facilitate phonetic acquisition (Teinonen, Aslin, Alku, & Csibra, 2008; Ter Schure, Junge, & Boersma, 2016) and word learning processes (Havy, Foroud, Fais, & Werker; 2017; Weatherhead & White, 2017). This audiovisual gain is probably due to the fact that visual speech carries articulatory information—provided by the mouth area—that is highly redundant with the corresponding auditory speech signal (Summerfield, 1987). In line with this claim, adults prefer looking at the eyes region of a face talking in their native language, to directly access social and emotional cues (Buchan, Paré, & Munhall, 2007), but they switch their attention to the mouth region when the speaker talks in an unfamiliar language (Barenholtz, Mavica, & Lewkowicz, 2016) or when the acoustic speech signal is noisy (Buchan, Paré, & Munhall, 2008).

Infants, who have not mastered their native language yet, show the same shift from the eyes to the mouth of the speaker during their first year of life. For instance, Lewkowicz and Hansen-Tift (2012) demonstrated that when watching a face talking in their native or foreign language, 4-month-old monolingual infants look preferentially at the eyes region of the speaker. At 8 and 10 months of age, however, they shift their attention to the mouth region of a speaker, both for the native and the nonnative language. This first shift of attention might be notably triggered by the onset of canonical babbling, which is generally observed around 7–8 months of age (de Boysson-Bardies, Halle, Sagart, & Durand, 1989). To produce their first speech-like sounds, infants could benefit from seeing the audiovisual redundant cues provided by the talking mouth of their caregivers. At 12 months, monolinguals' preference for the mouth decreases for the native language (they look the same amount of time to the eyes and the mouth region), while it remains constant when they look at a face talking in a foreign language. This second shift, specifically observed for the native language, has been interpreted by the authors to be triggered by the fact that at this age, infants no longer need to strongly rely on the mouth region of the speaker's face as they have built their native phonetic categories (Kuhl,

Williams, Lacerda, Stevens, & Lindblom, 1992; Pons et al., 2009; Werker & Tees, 1984).

While the direct link between native phonetic learning and preference for the mouth region of the speaker has not been tested yet, other evidence showed that this preference for the mouth region of the speaker during the first year of life is correlated with the infants' ability to detect a mismatch between auditory and visual speech (Altvater-Mackensen, Mani, & Grossmann, 2015; Kushnerenko et al., 2013) and productive vocabulary development in their second year of life (Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2015; Young, Merin, Rogers, & Ozonoff, 2009). Thus it indicates infants' scanning pattern of talking faces is—at least partly—related to language expertise (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017).

In line with this claim, recent research has shown that bilingualism also modulates the preference for the mouth region of talking faces. At 4 and 12 months of age, infants growing up in a bilingual environment pay more attention to the mouth region than their monolingual peers (Pons, Bosch, & Lewkowicz, 2015). The same pattern of results is observed at 8 months when presenting bilingual and monolingual infants with nontalking communicative faces (Ayneto & Sebastian-Galles, 2017). Taking into account that in daily face-to-face situations, communicative faces usually speak, both of these results suggest that bilingual infants might have developed this increased preference for the mouth region of communicative faces as an adaptive mechanism to overcome the challenge of learning two languages at the same time (for reviews, see Costa & Sebastián-Gallés, 2014; Werker, 2012). Indeed, one remarkable fact about bilingual infants is that they seem to learn their two native languages at a similar rate than monolingual infants acquire their single mother tongue, with reduced exposure to each language. For instance, they start babbling (Oller, Eilers, Urbano, & Cobo-Lewis, 1997), build their native phonemic categories for consonants (Burns, Yoshida, Hill, & Werker, 2007; Sundara, Polka, & Molnar, 2008), or even start to produce their first words at about the same age as monolingual infants (Genesee & Nicoladis, 2008; Patterson & Pearson, 2004).

One essential prerequisite to learn two different language systems at the same time is to be able to keep their two native languages apart from each other. Interestingly, data show that while monolingual and bilingual infants show comparable abilities to aurally discriminate between different languages (Bosch & Sebastián-Gallés, 2001b; Byers-Heinlein, Burns, & Werker, 2010; Molnar, Gervain, & Carreiras, 2014), they perform better than their monolingual peers to discriminate languages on the basis of the sole visual speech stream (Sebastian-Galles, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007).

Weikum et al. compared 6- and 8-month-old French-English bilinguals and English bilingual infants using a discrimination-habituation procedure. In a first phase, infants were habituated to silent video clips of three bilingual French-English speakers reciting sentences in one language (e.g., English). Then, in the test phase, infants saw the same speakers reciting different sentences. For half of the trials, the language was switched (e.g., to French) whereas for the other half, the language remained the same (English). Results showed that at 6 months, both bilinguals and monolinguals increased their looking time for the switch trials, showing that they could discriminate languages visually. At 8 months of age, however, only French-English bilingual infants could differentiate their native languages (i.e., French vs. English) while monolingual infants (raised in a French or English environment) could not.

Interestingly, Sebastian-Galles et al. (2012) tested 8-month-old Spanish-Catalan bilingual and monolingual infants using Weikum et al.'s stimuli and procedure. Again, only bilingual infants were able to notice the difference, indicating that they do not need prior experience to be able to visually discriminate between languages. Thus, it may well be possible that the increased preference for the mouth region of dynamic faces observed for bilingual infants by Pons et al. (2015) and Ayneto and Sebastian-Galles (2017) have helped them to focus on the relevant articulatory information to discriminate beween languages in Weikum et al (2007) and Sebastian-Galles et al. (2012). In this study, we further explore the impact of such attentional strategy on monolingual and bilingual infants' ability to process information coming from the eyes or the mouth region of talking faces.

Converging the evidence reviewed above thus suggests that the more challenging the language learning situation is (e.g., bilingualism, second language acquisition, noise), the more adults and infants seem to rely on the redundant audiovisual cues provided by the talking mouth of their social partner. From a language learning and processing perspective, paying attention to the mouth of the talker appears thus to be a good strategy (Munhall & Johnson, 2012). From a larger communicative perspective, however, a visual preference for the mouth region can make it more difficult to process other cues coming from the rest of the face, in particular, important social cues present in the eyes region. Indeed, the eye region also affords social (e.g., eye-gaze direction) and emotional (e.g., eyebrow movements) information (Csibra, 2010) that are essential to the understanding of the whole situation of communication. The eyebrow movements also provide a wide range of information, notably essential for the recognition of certain emotions (Ekman, 1979) or of certain component of visual speech prosody[1] (Esteve-Gibert,

Prieto, & Pons, 2015). Accordingly, when asked to identify the emotion of a face talking in their native language, adults prefer directing their attention at the eyes region of the speaker (Buchan et al., 2007). Some studies even suggest that infants' ability to process information coming from the eyes region of a nontalking face is related to their language learning skills during their second year of life. For instance, Tenenbaum et al. (2015) showed that 12-month-olds' ability to follow the gaze of a nontalking actor engaged in a social task was positively correlated with their vocabulary size at 18 and 24 months of age (see also Brooks & Meltzoff, 2005, 2008, for related work).

In the same line of argument, other researchers found that gaze cuing boosts early word learning in infancy (Houston-Price, Plunkett, & Duffy, 2006; Moore, Angelopoulos, & Bennett, 1999). For instance, Houston and colleagues presented 15-month-olds with new object–label associations. At each trial, the infants saw two images presented side by side while hearing one target word (test phase). Then in the learning phase, the auditory target word was repeated while the picture of a head turned, cueing the side where the corresponding target object appeared (gaze cue, Experiment 1), or the target object starts moving (saliency cue, Experiment 2). Results showed that only infants in the gaze cueing condition learned to increase their looking time at the target object during the test phase, showing that this information, as opposed to increased saliency, somehow helped infants to memorize new lexical entries and/or form novel object–label associations.

Thus relevant information from both the eyes and the mouth region of talking and nontalking faces is relevant for language learning. Perceivers thus need to learn to adapt their attentional strategy very quickly when exploring communicative faces, as a function of relevance of each source of information over time. However, this specific ability might be especially challenging for young perceivers for two main reasons. First, infants are still in the process of learning their native language and thus might still need to rely on the audiovisual redundant speech cues provided by the mouth of the speaker to decode the speech stream. Second, the neural circuitry involved in attention and cognitive control is not fully mature at such young ages (Berger, Tzur, & Posner, 2006; Colombo, 2001). While stimulus-driven attention—that is, the ability to orient attention as a function of the exogenous saliency of the stimulus—is mature since 6 months of age, the anatomical structures underlying goal-directed attention[2]—that is, the ability to voluntarily orient attention as a function of the task-relevant property of the stimulus—only starts to be functional at this age and keeps on improving during childhood (see de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016, for a review). This lack of goal-directed attention maturity could

prevent infants from learning to switch from one region of a speaker's face to the other as fast as needed.

The first goal of this study is to test whether infants in their second year of life rely more on the eyes or on the mouth region of a face talking in their native language. Indeed, while infants became expert at processing their native phonemes, they still have a lot to learn to master their native language, as they just start, for instance, to produce their first words (Vihman, 1996). Focusing on the mouth of their caregivers could still afford them useful redundant audiovisual speech cues to reproduce the speech they hear from others. Second, we investigate the consequences of such preference on infants' ability to perceive additional nonspeech information coming from the eyes or the mouth region. The third goal of this study is to test how bilingualism may affect the ability to learn to switch from one region of the speaker's face to another.

We thus conducted this study with 15- and 18-month-old infants, at an age where the information coming from both the eyes and the mouth region has been shown to correlate with language performance (see below). We recorded infants' gaze fixations while they watched and listened to a female speaker reciting short sentences (Speech event). At the end of each Speech event, the speaker produced a Nonspeech movement by either systematically raising both of her eyebrows (eyebrow-raise [EB] condition) for half of the participants or protruding her lips (lip-protrusion [LP] condition) for the other half. First, if infants still need to rely on the redundant audiovisual cues provided by the speaker's mouth, we predict that they should look preferentially at this region over the eyes region, especially during the Speech event. Second, we predict that if the type of Nonspeech movement influences the way infants explore talking faces, infants should orient their visual attention more toward the eyes region in the EB condition or toward the mouth region in the LP condition during the Speech event, learning to anticipate the locus of appearance of the recurrent Nonspeech movement over time/as time progresses (goal-directed attention). Otherwise, we should only observe a change in gaze behavior during the production of the Nonspeech movement after the Speech event (stimulus-driven attention). Finally, to test whether bilingualism has an impact on how infants pay attention to talking faces and switch from one location to the other, we also manipulated participants' language environment (monolingual vs. bilingual). We conducted two experiments, testing 15- (Experiment 1) and 18-month-old (Experiment 2) infants. In Experiment 1, we predicted that bilinguals' increased preference for the mouth region of a speaker should lead them, as compared to monolinguals,

to exhibit a smaller effect of the EB condition and/or bigger effect of the LP condition.

Experiment 1

Method

Participants

Forty 15-month-old healthy full-term monolingual infants (15M) were recruited from private clinics in Barcelona (age range: 429–481 days; mean: 459 days; 20 girls). Infants were all raised in middle or mid-upper socioeconomic status families. Parents filled in a language questionnaire adapted from Bosch and Sebastián-Gallés (2001a). Monolingual infants were exposed to Catalan or Spanish at least 85% of the time. Twenty of them participated in the EB condition (10 girls, 15 Catalans monolinguals, mean language exposure to the dominant language $= 97\%$, $SD = 4$), while the other 20 infants were tested in the LP condition (10 girls, 15 Catalan monolinguals, mean language exposure to the dominant language $= 98\%$, $SD = 3$). On average, the 15M in the EB condition did not differ from the ones in the LP condition in terms of language exposure and age (all $p > .05$). The data from 17 additional infants were excluded from the final analysis due to the total looking time to the screen being less than 50% (6), number of trials fewer than 19 (5),[3] and failure to calibrate the eye tracker (6).

Forty 15-month-old healthy full-term Spanish-Catalan bilingual infants (15B) were recruited from the same private clinics as the 15M (range: 438–480 days; mean: 458 days; 20 girls). Twenty infants participated in the EB condition (9 girls, 16 Catalan dominant, exposure to the dominant language $=$ 68%, $SD = 9$), while the 20 others were assigned to the LP condition (11 girls, 14 Catalan dominant, exposure to the dominant language $= 65\%$, $SD = 11$). Importantly, the 15B in both conditions were matched in age and language exposure between each other (all $p > .05$). They were also equivalent in age to both groups of monolingual infants (all $p > .05$). The data from 22 more infants were excluded from the final analysis due to the total looking time to the screen being less than 50% (4), number of trials fewer than 19 (7), failure to calibrate (5), and experimental error (3).

Stimuli and Recordings

The Speech events (19 in Catalan and 19 in Spanish; see the Supporting Information online for more details) consisted of video recordings of six-syllable-long sentences lasting between 1,180 and 2,200 milliseconds (e.g., "Cada día canto" *Everyday I sing*). The mean duration and the mean intensity of the video
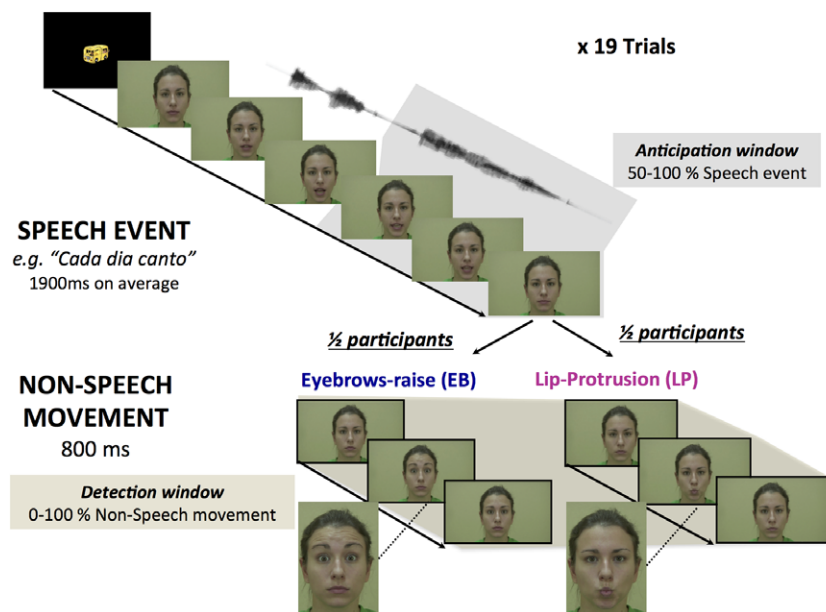
Figure 1 Schematic representation of the procedure and the windows of analysis used for the test stimuli used in Experiment 1. In Experiment 2, only the eyebrow-raise condition was presented. The peak of each of the Nonspeech movements are zoomed in the figure for clarity. [Color figure can be viewed at wileyonlinelibrary.com]

clips were similar in Catalan and in Spanish (all $t < 1$; see the Supporting Information online for more details).

We used two different types of Nonspeech movements. In the EB condition, the speaker raised her eyebrows while in the LP condition the speaker protruded her lips without opening her mouth (cf. Figure 1, bottom part). Each video clip we used for each Nonspeech movement lasted 1,880 milliseconds. We made sure that the duration before and after the maximum amplitude of the EB/LP movement (i.e., highest position of the eyebrows, maximum protrusion of the lips in the sagittal plane) was similar between conditions (EB vs. LP) and identical between languages (Catalan vs. Spanish; see the Supporting Information online for more details).

A bilingual Spanish-Catalan Caucasian female speaker was asked to produce both the speech and nonspeech stimuli. She kept her head in the same position (avoiding rotation on the axial, sagittal, or coronal plan) for the whole duration of the recordings. The speech stimuli were recited in an emotionally neutral adult-directed-speech manner and were first recorded in Spanish and

then in Catalan, separately from the nonspeech stimuli. All the video clips were then digitized with the Adobe Premiere CS3 3.2.0 software to obtain MPEG video files. The intensity of the soundtrack of each Speech event was normalized using the Audacity 2.0.5 software. We then combined the Speech events with the Nonspeech movements in the same video file using Adobe Premiere, to make sure that there was no random loading time between the presentation of the video clips of the Speech event and the Nonspeech movement. Because the speaker position remained constant during the whole recording session, the transition between the Speech event and the Nonspeech movement videos were almost imperceptible. There was no time break between the Speech event and the Nonspeech movement, except that the speaker always finished the Speech event and started the Nonspeech movement with her mouth and lips closed. Examples of our stimuli can be found on the OSF repository (https://osf.io/mwfhc/files/).

Procedure

The participants were tested in a quiet room while sitting on their parents' lap, about 60 centimeters away from a 1080 × 1920–pixel screen. The stimuli were presented with custom-made software using the MATLAB 7.11 Psychtoolbox and Tobii Analytics Software Development Kit (Tobii Analytics SDK). The visual component of the video clip was displayed at the same resolution as the screen (1080 × 1920 pixels), at a frequency of 25 frames/second, whereas the auditory component was displayed at a frequency of 44,100 Hz at 65 dB. Infants' eye movements were recorded by a Tobii TX300 stand-alone eye tracker (Tobii Technology AB, Danderyd, Sweden) at a sampling rate of 300 Hz. The videos were played in the infant's native or dominant language. We used a 5-point calibration procedure: one central point and one at each corner of the screen. To make the procedure infant friendly, each of the five calibration points was pointed out on the screen by the presentation of a small ball displayed synchronously with an environmental sound. The calibration was completed when at least three valid points were obtained. The experiment started right after the calibration.

The experiment consisted of 20 trials: one dummy trial followed by 19 test trials. Before each trial an audiovisual attention getter was presented. It consisted of a modified version of a looping Tobii animation of a bus growing and shrinking in synchrony with the presentation of an artificial "boing" sound (cf. Figure 1). Infants were required to look at this attention getter for at least 1 second to launch a new trial. The dummy trial consisted in a video of the speaker introducing herself by producing the Catalan or Spanish version of

the English sentence *Hi, how are you?* while smiling at the infant. The next 19 test trials were videos of the speaker producing a Speech event followed by a Nonspeech movement where the speaker either systematically raised her eyebrows for half of the participants (EB condition) or protruded her lips for the other half (LP condition; see Figure 1). Infants were randomly assigned to the EB or to the LP condition.

Each test trial (Speech event + Nonspeech movement) lasted between 3 to 4 seconds. The order of Speech events was pseudo-randomized, so that across infants, each video clip was seen the same amount of time between trials 1 and 6, trials 6 and 12, and trials 12 and 19. The video clips for the Speech events were always played in the infant's native or dominant language.

Preprocessing and Definition of the Areas of Interest (AOIs)

To determine which part of the talker's face infants was looking at, we divided each video clip into two rectangular AOIs, one around the eyes and one around the mouth, using the same facial landmarks as Lewkowicz and Hansen-Tift (2012). Because the speaker's position was constant across all the videos, we used only one eyes and one mouth AOI. Then, with a custom-made program (MATLAB 7.11 Tobii Analytics SDK), we determined for each trial whether each data point collected by the eye tracker was inside or outside the two defined AOIs. Because the Speech events were of different length, we then normalized these data (using the resample function in MATLAB) for each trial across the duration of the Speech event. To make all things equal, we also performed the same operation separately for the Nonspeech movements. We set up the normalization to get one data point for every 10% of the duration of each event type (i.e., Speech events and Nonspeech movements). We finally computed for each of these 10% time steps the proportion of total looking time (PTLT) each infant spent looking at each AOI. To do so, we divided the total amount of looking time at each AOI by the total looking time at the whole face as performed by, for instance, Lewkowicz and Hansen-Tift (2012) and Pons et al. (2015).

Results

We performed three separate analyses of infants' looking behavior. First, we assessed the general preference for the mouth or the eyes region of the speaker both for the last 50% of the Speech event[4] and during the presentation of the Nonspeech movements. Second, we focused on the last 50% of the Speech event to examine the anticipation of the Nonspeech movement averaged across trials and over the course of the experiment (from trial 1 to trial 19). In the third

Table 1 Descriptive statistics for Experiments 1 and 2

			Last 50% of the Speech event		Nonspeech movement	
			Eyebrow raise	Lip-protrusion	Eyebrow raise	Lip-protrusion
Exp. 1	**15M**	Eyes AOI	.31(.06)	.12(.03)	.48(.05)	.16(.05)
		Mouth AOI	.60(.07)	.77(.05)	.34(.05)	.69(.06)
		Difference	*−.29*	*−.65**	*.14*	*−.53**
	15B	Eyes AOI	.18(.05)	.17(.04)	.43(.06)	.18(.04)
		Mouth AOI	.77(.05)	.71(.05)	.44(.05)	.70(.05)
		Difference	*−.59**	*−.54**	*−.01 n*	*−.52**
Exp. 2	**18B**	Eyes AOI	.31(.04)	–	.55(.04)	–
		Mouth AOI	.56(.05)	–	.32(.04)	–
		Difference	*−.25**	–	*.23**	

Note. Mean PTLT scores and standard errors from the mean (in parentheses) for the Eyes and the Mouth AOIs averaged across the last 50% of the Speech event (Anticipation window of analysis) and the whole duration of the Nonspeech movement (Detection window of analysis), as a function of each Nonspeech movement (eyebrow raise, lip protrusion) in Experiment 1. In Experiment 2, only the eyebrow-raise condition was presented. Asterisks (*) represent significant t-test comparisons when testing the Eyes-Mouth PTLTs scores against zero (i.e., an absence of preference between the eyes and the mouth area).

analysis, we assessed the overall effect of the Nonspeech movement during its presentation, to make sure that infants could detect both Nonspeech movements (i.e., EB and LP).

Assessment of the General Preference for the Mouth Region of the Speaker
We averaged the Eyes and Mouth PTLTs for each participant and separately for each window of analysis, namely, during the last 50% of the Speech event and the whole duration of the Nonspeech movement (cf. Table 1). We then subtracted the mean Eyes PTLTs from the mean Mouth PTLTs to obtain a preference score (as in Lewkowicz & Hansen-Tift, 2012; see Figure 2 and 3). We submitted these Eyes-Mouth PTLT scores obtained for each group to a paired *t* test against zero (i.e., signaling an absence of preference). The summary of the significant comparisons is indicated by asterisks in Table 1. During the last 50% of the Speech event, a general preference for the mouth region over the eyes region of the speaker was found to be significant both for the 15B in both the EB, $t(19) = 6.27$, $p < .001$, Cohen's $d = −1.97$, and the LP conditions, $t(19) = 6.48$, $p < .001$, Cohen's $d = −2.04$. For the 15M the preference for the mouth region was almost significant for the EB condition, $t(19) = 1.9$, $p = .08$, and clearly significant in the LP condition, $t(19) = −8.22$, $p < .001$, Cohen's $d = −2.58$. During the Nonspeech movement, both 15B and 15M showed a greater preference for the mouth region in the LP condition,
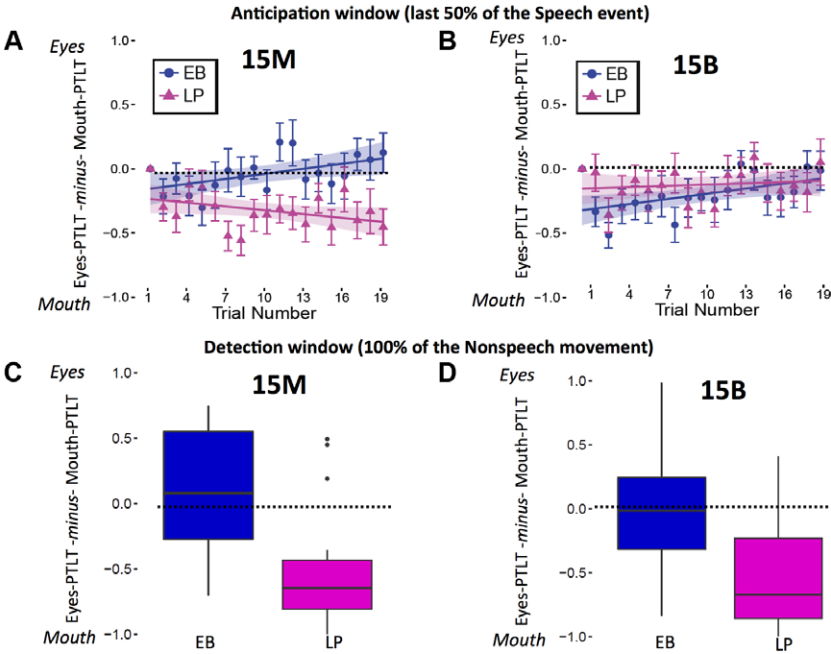
Figure 2 (Top) Mean Eyes PTLT-Mouth PTLT baseline-corrected scores averaged for the last 50% of the Speech event (anticipation window) from trials 1 to 19 for the (a) 15-month-old monolinguals (15M) and (b) 15-month-old bilinguals (15B) in Experiment 1. Individual data points represent the observations averaged across participants for each condition at each trial. Solid lines represent the regression lines of the growth curve analysis for these individual data points. Shaded ribbons indicate 95% confidence intervals based on the linear regression lines. Error bars represent standard errors from the mean. (Bottom) Boxplots of the Eyes PTLT-Mouth PTLT scores for the whole duration of the Nonspeech movements (detection window) for the (c) 15M and (d) 15B in Experiment 1. The horizontal line in the boxplot represents the median. The upper and lower portions of the box above and below the median represent the first and third quartiles, respectively. The whiskers represent 1.5 times interquartile range. (Both panels) The eyebrow-raise (EB) condition is in blue, the lip-protrusion (LP) condition is in pink. Positive scores indicate a preference for the eyes region over the mouth region. [Color figure can be viewed at wileyonlinelibrary.com]

$t(19) = 5.7, p < .001$, Cohen's $d = -1.79$, and $t(19) = -5.38, p < .001$, Cohen's $d = -1.71$, respectively. In the EB condition they looked a similar amount of time at the eyes and mouth regions of the speaker, $t < 1$ and $t(19) = 1.17, p = .25$, respectively.
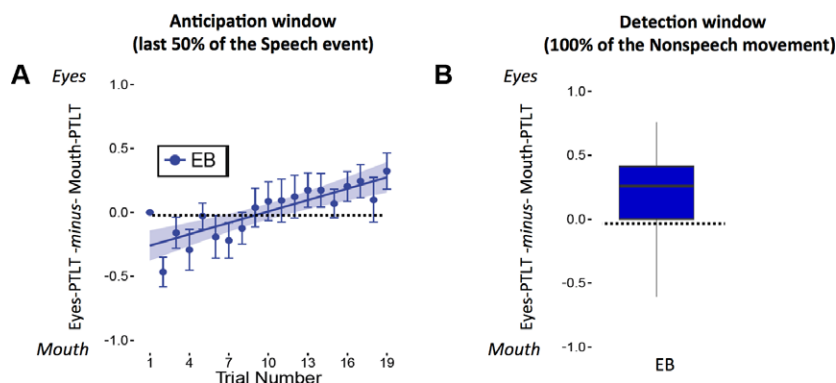
Figure 3 (a) Mean Eyes PTLT-Mouth PTLT baseline-corrected scores averaged for the last 50% of the Speech event (anticipation window) from trials 1 to 19 for the 18-month-old bilinguals (18B) in Experiment 2 in the eyebrow-raise (EB) condition. (b) Boxplots of the Eyes PTLT-Mouth PTLT scores for the whole duration of the EB movement (detection window) for 18B in Experiment 2. [Color figure can be viewed at wileyonlinelibrary.com]

Anticipation of the Nonspeech Movement During the Last 50% of the Sentence
To explore whether 15M and 15B differ in their ability to anticipate the Nonspeech movement right before its appearance, we first considered these Eyes-Mouth PTLT preference scores for the last 50% of the Speech event averaged across trials. We then submitted these preference scores to an analysis of variance (ANOVA) with Language Group (15M, 15B) and Nonspeech movement (EB, LP) as between-participant factors. The analysis yielded an interaction between Language Group and Nonspeech movement, $F(1,76) = 4.37$, $p < .05$, $\eta^2 p = .05$. Planned comparisons revealed that 15M and 15B did not differ in the LP condition, $F < 1$. In the EB condition, however, 15M significantly looked less at the mouth region than 15B, $F(1,38) = 3.7$, $p = .05$, $\eta^2 p = .05$.

To make sure that the differences between 15M and 15B were due to a learned anticipation during the last 50% of the Speech event, we further explored infants' preference scores over the course of the experiment (from trial 1 to trial 19). For each participant, we considered the Eyes-Mouth PTLTs at each trial and baseline corrected the initial preference for the eyes region or the mouth region of the speaker at trial 1 (when they still had no exposure to the Nonspeech movement) on the Eyes-Mouth PTLTs of the next trial (2 to 19). To do so, we used the following formula.

If $EMTn > EMT_1$

$$EMbcTn = \frac{EMTn - EMT1}{1 - EMT1}$$

If $EMTn < EMT_1$

$$EMbcTn = \frac{EMTn - EMT1}{1 + EMT1}$$

where *EMbcTn* represents the baseline-corrected Eyes-Mouth PTLT output and *EMT1* and *EMTn* are the Eyes-Mouth PTLT at trial 1 and trial *n*, respectively. The baseline-corrected Eyes-Mouth PTLTs averaged across participants are presented in Figure 2a and 2b: A positive score indicates an increased preference for the eyes area while a negative one indicates an increased preference for the mouth area.

Growth curve analysis (Mirman, 2014) was used on these baseline-corrected scores to analyze the evolution over trials of the 15M's eyes and mouth preference. The overall learning curves were modeled with first-order linear polynomials and fixed effects of Nonspeech movement (EB, LP) on all time terms (i.e., Trial Number). The LP condition was treated as baseline and parameters were estimated for the EB condition. The model also included random effects of participants on all time terms. The fixed effects of Nonspeech movement were added individually and their effects on model fit were evaluated using model comparisons. Improvements in model fit were evaluated using –2 times the change in log likelihood, which is distributed as χ^2 with degrees of freedom equal to the number of parameters added. All analyses were carried out in R version 3.0.2 (R Core Team, 2012) using the lme4 package (version 1.0–5; Bates, Maechler, Bolker, & Walker, 2015). Appendix S3 in the Supporting Information online shows the fixed-effect parameter estimates and their standard errors along with *p* values estimated using Satterthwaite's approximation for degrees of freedom for the last 50% of the Speech event. For the 15M, the effect of the Nonspeech movement on the intercept did not improve model fit on its own, $\chi^2(1) = 3.51, p = .06$, but significantly improved it on the Trial Number term, $\chi^2(1) = 4.99, p < .05$. For the 15B, the effect of Nonspeech movement on the intercept did not improve model fit, $\chi^2 < 1$, nor did the effect of Nonspeech movement on the Trial Number term, $\chi^2(1) = 2.11, p = .14$. In summary, no significant effects were observed for the last 50% of the Speech event in the 15Bs.

Detection of the Nonspeech Movement
To make sure that 15M and 15B could detect the EB or the LP movement (cf. Figure 2c and 2d), we submitted these Eyes-Mouth PTLT scores averaged

across the whole duration of the Nonspeech movements (detection window) to a Language group × Nonspeech movements ANOVA. The analyses only yielded a significant effect of Nonspeech movement type, $F(1,76) = 33.9$, $p < .001$, $\eta^2 p = .31$. Neither a main effect of Language Group nor an interaction was significant, both $F < 1$. These results indicate that as predicted, both 15M and 15B similarly detected both Nonspeech movements by orienting their gaze to the eyes region or the mouth region of the speaker in the EB and LP conditions, respectively.

Discussion

In Experiment 1, our results showed that as compared to the LP condition, 15M in the EB condition could disengage from the mouth of the speaker to anticipate the occurrence of the Nonspeech movement in her eyes region. However, same-age bilinguals (15B) did not show the same pattern of results. Further statistical comparisons showed that 15M and 15B only differed for the last 50% of the Speech event in the EB condition (critical condition) but not in the LP condition (control condition). In the latter, the absence of difference between 15M and 15B might have been masked by a floor effect on their performances, given that both groups exhibited a strong preference for the mouth region (see the general discussion for more on this finding). In the EB condition, however, only 15M changed their pattern of exploration of talking faces during the Speech event over the course of the experiment. First, this finding is in line with Pons et al. (2015), indicating that bilinguals show more attention to the mouth region of talking faces than their monolingual peers. The results are also in line with more general finding in the literature showing different language learning developmental trajectories between monolingual and bilingual infants (for reviews, see Costa & Sebastián-Gallés, 2014; Werker, 2012). Indeed, probably to deal with the more complex nature of their dual-language input, bilingual infants differ from monolinguals in the time course of language learning, notably in building some native phonetic categories (Bosch & Sebastian-Galles, 2003; Garcia-Sierra et al., 2011; Sebastián-Gallés & Bosch, 2009), transitorily showing distinct behavior than their monolingual peers at a certain age and for certain tasks (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011). In other cases, bilingual infants clearly show a different pattern of results over time, maintaining their sensitivity to lexical stress (Abboub, Bijeljac-Babic, Serres, & Nazzi, 2015; Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012) or, as already discussed, their ability to visually discriminate between languages (Sebastian-Galles et al., 2012; Weikum et al., 2007). In Experiment 2, we further explore whether older

bilingual infants show a difference in the time course of their exploration of talking faces or simply maintain their preference for the mouth of the speaker regardless of the EB movement. We thus focused on the critical EB condition and examined whether 18-month-old bilinguals (18B) could anticipate its appearance during the last 50% of the Speech event.

Experiment 2
Method
Participants

Twenty 18-month-old healthy, full-term bilingual infants (18B) participated in the EB condition (range: 531–555 days; mean: 545 days; 10 girls, 14 Catalan dominant, exposure to the dominant language = 65%, $SD = 9$). Infants were recruited from the same database and following the same criteria as in Experiment 1. No infant had participated in Experiment 1. The data from four more infants were excluded from the final analysis due to the total looking time to the screen being less than 50% (3) and failure to calibrate (1).

Stimuli and Procedure

The stimuli and procedure were the same as in Experiment 1, except that participants were only tested in the EB condition.

Results and Discussion

We performed the same computations of the raw data collected by the eye tracker as in Experiment 1. The results are presented in Table 1 and Figure 3. During the last 50% of the Speech event, 18B showed a preference, as 15M and 15B did, for the mouth region over the eyes region of the speaker, $t(19) = 3.65, p < .005$, Cohen's $d = -.97$. During the EB movement, 18B looked longer at the eyes region than at the mouth region, $t(19) = 2.99, p < .01$, Cohen's $d = .98$, indicating they could successfully detect it. To compare the performance of 18B to that of 15M and 15B for the last 50% of the Speech event, we ran two *t* tests on the Eyes-Mouth PTLTs. Results showed that 18B did not differ from 15M in the EB condition, $t < 1$. However, they looked significantly less at the mouth region of the speaker than did 15B in the EB condition, $t(38) = 2.65, p < .05$, Cohen's $d = .85$.

Finally, we investigated the learned anticipation during the last 50% of the Speech event, over the course of the experiment. As shown in Appendix S3 in the Supporting Information online, the growth curve analysis showed that the effect of yrial number on the intercept was significant. This means that during the last 50% of the Speech event, 18B's preference for the mouth of the speaker linearly decreased from trials 1 to 19. Thus, these results show that 18B could,

like 15M, anticipate the EB movement of the speaker before its appearance, namely, during the Speech event. We further discuss the implications of these findings in the general discussion.

General Discussion

In this study, we developed an original design to measure the impact of infants' attentional strategies on their ability to perceive information coming from talking (Speech event) and nontalking (Nonspeech movement) faces. At each trial, infants were presented with a video clip of a speaker producing a Speech event followed by a Nonspeech movement (EB vs. LP). First, we showed that when presented with faces talking in their native language, both monolingual (at 15 months) and bilingual infants (at 15 and 18 months) pay more attention to the mouth region than the eyes region of the face (i.e., during the Speech event). Second, we observed that at 15 month of age, monolingual infants in the EB condition anticipated the locus of the Nonspeech movement before its appearance, namely during the Speech event. However, same-age bilingual infants only showed a change in their gaze pattern during the production of the Nonspeech movement, namely, after the Speech event (Experiment 1). At 18 months, bilinguals could anticipate the EB movement in a similar way as 15-month-old monolinguals (Experiment 2).

LP Condition and General Mouth Preference

Recall that when presented with the same type of stimuli, adults, who have mastered their native language, prefer focusing on the upper part of a face talking in their native language, allowing them to process socioemotional information coming from the eyes region (Barenholtz et al., 2016; Buchan et al., 2008). Our results in this study are the first to show that in their second year of life, infants have not mastered enough of their native language and thus seem to rely on the audiovisual redundancy between auditory and visual speech streams when perceiving a face talking in their native language. Importantly, this increased preference for the mouth region might have masked the measure of anticipation of the LP movement during the Speech event or a greater anticipation in bilingual infants, as predicted in the introduction.

While it may well be the case that native phonetic learning could be one important language computation responsible for orienting infants' attention on the mouth region of talking faces during their first year of life (Lewkowicz & Hansen-Tift, 2012), it cannot explain the mouth preference observed in 15- and 18-month-olds in this study. Indeed, at this age, they became expert at processing the phonetic contrast of their native language (Kuhl et al., 1992;

Pons et al., 2009; Werker & Tees, 1984). However, the acquisition of vocabulary knowledge could be responsible for maintaining infants' visual attention on the mouth region of talking faces. At 15 months, infants are at a period of great lexical acquisition (that lasts till late childhood) and notably at the very onset of word production (Vihman, 1996). Productive vocabulary size is usually small (estimated around 50 entries on average) although great interindividual variability is observed until at least 24 months of age (range: 50–550 words). Thus, at 15 months, watching the redundant audiovisual cues provided by the mouth of the speaker could enhance the saliency of the whole speech signal, facilitating lexical encoding in memory (Bahrick & Lickliter, 2014), in turn improving infants' ability to imitate/reproduce their first words. In line with this claim, two longitudinal studies suggested that infants' preference for the mouth of a speaker during their first year of life positively correlated with their productive vocabulary development later on (Tenenbaum et al., 2015; Young et al., 2009). Further research is thus needed to establish the role of attention to the eyes region and to the mouth region of talking faces and its relation to typical and atypical lexical acquisition at later stages of childhood.

EB Condition

Previous findings by Pons et al. (2015) and Ayneto-Gimeno & Sebastian-Galles (2017) have shown that during their first year of life, bilingual infants look more at the mouth region of talking and nontalking faces than their monolingual peers. Here we found the same pattern of results at a later stage of development but we also went one step further: We showed that during their second year of life (15 months of age), bilingual infants kept looking at the mouth region of the speaker rather than disengaging from it, preventing them from anticipating the appearance of additional information coming from the eyes region (EB movement).

So which mechanisms could be responsible for the difference observed between monolingual and bilingual infants at this age? First, our data support the hypothesis that the trial-to-trial anticipation observed during the Speech event at 15 (for monolinguals) and 18 (for bilinguals) months of age is, by its nature, supported by goal-driven attention. However, it is highly improbable that the absence of anticipation during the Speech event is caused by a cognitive control delay in 15-month-old bilinguals. Previous literature showed that bilingual infants would outperform their monolingual peers in terms of executive functions or cognitive control (Kovács & Mehler, 2009a, 2009b). Here we argue that when it comes to processing talking faces, the load of learning two languages at the same time has led bilinguals to develop different attentional

strategies than their monolingual peers. While monolingual infants could afford decreasing their mouth preference while the speaker is still talking, same-age bilinguals chose to prioritize the information provided by the mouth until the end of the Speech event, gazing toward the eyes region solely when detecting the beginning of the EB movement. This strategy is probably a well-adapted mechanism that allows bilinguals to become as efficient as their monolingual peers at acquiring both of their native languages. Thus, one the one hand, having a greater and direct daily experience with the mouth region of their speaking caregivers could lead them, as compared to their monolingual peers, to certain cognitive advantages in performing tasks requiring great expertise in processing speech and mouth-related cues. This could explain the fact that, for instance, bilingual infants, during their first year of life, are better at detecting and memorizing differences between two languages on the basis of the visual speech information alone (Sebastian-Galles et al., 2012; Weikum et al., 2007). During their second year of life, this increased attention to the mouth region could also support their lexical development at pace with monolinguals. Indeed, bilingual children produce their first words at the same time as monolinguals, 12 to 13 months (Patterson & Pearson, 2004), and their rate of acquisition falls within the same range as that reported for their monolingual peers, as long as both languages are taken into consideration for bilinguals (Pearson, Fernández, & Oller, 1993).

On the other hand, our results indicate that, at least at some point in their development, bilinguals' bias toward talking faces prevents them from antici-pating the EB movement while the face is still talking. This finding opens the possibility that when it comes to learning information from the eyes region of a talking face, bilinguals show a different time course of the development of this ability than monolingual infants. Of course, it is possible to attend to and process information in the peripheral vision (Richards, 2005). It is thus possible that while bilinguals maintain their attention on the mouth region of talking faces, they are also able to perform some low-level processing from cues displayed in some other part of the face (e.g., detection of eye-gaze direc-tions that is supported by subcortical structures; see Senju & Johnson, 2009, for more details). However, whether they can process and learn from physically less salient but nonetheless informative signals such as the EB movement used in this study remains unknown. Indeed, it is important to note that even if the EB movement in this study was not informative per se, EB movements usually provide useful information for the recognition of certain components of visual speech prosody that infants learn to process audiovisually during infancy (e.g., Esteve-Gibert et al., 2015). Further studies should examine this issue.

New Perspectives for Language Acquisition

Last but not least, this study opens new possibilities and areas of investigation for the field of early language acquisition. We managed to capture how language learning processes unfold over the time course of the same experiment, an approach that is almost never used in infant research, probably due to the great statistical dispersion around the mean that is observed in infant data. In the present study, great variability in the attentional strategies of exploration of talking faces was notably observed across participants of the same age and language background. To our knowledge, only one study already reported a similar pattern of results in infants in their first year of life (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013). In this work, we show that during the second year of life, extracting a developmental norm of how infants scan the talking faces of their caregivers remains a difficult task, in line with emerging evidence suggesting that this interindividual variability could reliably predict language and social development later on. By applying growth learning curve analysis (Mirman, 2014), we thus could track infants' learning on a trial-by-trial basis with a method that is both sensitive to small effect sizes and highly robust to samples with significant variability within and across participants. We believe that developing and applying such statistical approaches to infant research would greatly benefit the field, notably by diminishing the probability of Type 1 and Type 2 errors, adding to the emerging incentive observed in infant research (e.g., Tsuji, Bergmann, & Cristia, 2014).

Conclusion

This study first shows that in their second year of life, infants still need to rely on the redundancy between auditory and visual speech information when perceiving a face talking in their native language. This suggests the possibility that these redundant audiovisual cues support not only phonetic learning during the first year of life but also lexical development (especially the production of new words) at later stages of infancy and childhood. Second, we show that bilingualism strongly constrains infants' development of visual exploration of talking faces and that it impacts how they learn from them.

<div align="right">Final revised version accepted 17 October 2017</div>

Notes

1 Auditory speech prosody refers to the use of pitch, loudness, tempo, and rhythm in speech to convey information about the structure and meaning of an utterance.

Visual speech prosody refers to the visual correlates of these auditory variations, such as eyebrow, head, and hand movements.

2   For simplification purposes we consider stimulus-driven versus goal-directed attention (Corbetta & Shulman, 2002) as a model framework for this study and consider other distinctions (endogenous vs. exogenous attention; Petersen & Posner, 2012) as synonymous. For more discussion about this topic, see de Diego-Balaguer et al. (2016).

3   As we measured anticipation over time, infants had to go through the whole experiment—that is, complete the 19 experimental trials.

4   Even if our hypothesis was focused on the last 50% of the Speech event, we ran the same analyses on the whole duration of the Speech event. None of the analyses led to quantitatively different results or different conclusions.

References

Abboub, N., Bijeljac-Babic, R., Serres, J., & Nazzi, T. (2015). On the importance of being bilingual: Word stress processing in a context of segmental variability. *Journal of Experimental Child Psychology*, *132*, 111–120. https://doi.org/10.1016/j.jecp.2014.12.004

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *14*, 395–401. https://doi.org/10.1111/j.1467-7687.2010.00989.x

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*, 191–204. https://doi.org/10.1037/a0039964

Ayneto, A., & Sebastian-Galles, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science*, *20*: e12446. https://doi.org/10.1111/desc.12446

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

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: linear mixed-effects models using Eigen and S4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, A., Tzur, G., & Posner, M. I. (2006). Infant brains detect arithmetic errors. *Proceedings of the National Academy of Sciences*, *103*, 12649–12653. https://doi.org/10.1073/pnas.0605350103

Bijeljac-Babic, R., Serres, J., Höhle, B., Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE7*, *2*, e30843. https://doi.org/10.1371/journal.pone.0030843

Bosch, L., & Sebastián-Gallés, N. (2001a). Early language differentiation in bilingual infants. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 71–93). Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.1.05bos

Bosch, L., & Sebastián-Gallés, N. (2001b). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49. https://doi.org/10.1207/S15327078IN0201_3

Bosch, L., & Sebastian-Galles, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*, 217–243. https://doi.org/10.1177/00238309030460020801

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*(5), 905–921. https://doi.org/10.1162/jocn.2009.21076

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

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220. https://doi.org/10.1017/S030500090700829X

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

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

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

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

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755

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

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language*, *25*, 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, *16*, 1–17. https://doi.org/10.1017/S0305000900013404

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, *7*, 1–15. https://doi.org/10.3389/fpsyg.2016.00044

Ekman, P. (1979). About brows: Emotional and conversational signals. In D. von Cranach, M. Foppa, K. Lepenies, & W. Ploog (Ed.), *Human ethology* (pp. 169–249). Cambridge, UK: Cambridge University Press. https://doi.org/10.2307/40538147?ref=no-x-route:d64912790719373b34a8db686a88913d

Esteve-Gibert, N., Prieto, P., & Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behavior and Development*, *38*, 126–129. https://doi.org/10.1016/j.infbeh.2014.12.016

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., et al. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*, 546–557. https://doi.org/10.1016/j.wocn.2011.07.002

Genesee, F., & Nicoladis, E. (2008). Bilingual first language acquisition. In *Blackwell Handbook of Language Development* (pp. 324–342). Blackwell Publishing. https://doi.org/10.1002/9780470757833.ch16

Havy, M., Foroud, A., Fais, L., & Werker, J. F. (2017). The role of auditory and visual speech in word learning at 18 months and in adulthood. *Child Development*. https://doi.org/10.1111/cdev.12715

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

Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, *95*, 27–55. https://doi.org/10.1016/j.jecp.2006.03.006

Jesse, A., & Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *Journal of Experimental Child Psychology*, *145*, 1–10. https://doi.org/10.1016/j.jecp.2015.12.002

Kovács, A. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 6556–6560. https://doi.org/10.1073/pnas.0811323106

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

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). https://doi.org/10.1371/journal.pone.0089275

Kubicek, C., Gervain, J., Hillairet de Boisferon, A., Pascalis, O., Lœvenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior and Development*, *37*(4), 644–651. https://doi.org/10.1016/j.infbeh.2014.08.010

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Source: Science, New Series*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Kushnerenko, E., Tomalski, P., Ballieux, H., Ribeiro, H., Potton, A., Axelsson, E. L., et al. (2013). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *European Journal of Neuroscience*, *38*, 3363–3369. https://doi.org/10.1111/ejn.12317

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

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

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Chapman and Hall / CRC.

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*, 326–337. https://doi.org/10.1111/infa.12041

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, *35*, 60–68. https://doi.org/10.1037/0012-1649.35.1.60

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

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*, 4–12. https://doi.org/10.1007/s00426-005-0031-5

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, *24*, 407–425. https://doi.org/10.1017/S0305000997003097

Patterson, J. L., & Pearson, B. Z. (2004). Bilingual lexical development: Influences, contexts, and processes. In B. A. Goldstein (Ed.), *Bilingual language development*

and disorders in Spanish-English speakers. (pp. 77–104). Baltimore: Paul H
Brookes Publishing.

Patterson, M. L., & Werker, J. F. (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

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in
bilingual infants and toddlers: Comparison to monolingual norms. *Language
Learning*, *43*, 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain:
20 years after. *Annual Review of Neuroscience*, *35*, 73–89.
https://doi.org/10.1146/annurev-neuro-062111-150525

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

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech
synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, *149*,
142–147. https://doi.org/10.1016/j.actpsy.2013.12.01

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*, 10598–10602.
https://doi.org/10.1073/pnas.0904134106

R Core Team. (2012). R: A language and environment for statistical computing.
Retrieved from http://www.r-project.org

Richards, J. E. (2005). Development of covert orienting in young infants. In L. Itti, G.
Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 82–88). Brighton, MA:
Elsevier. https://doi.org/10.1016/B978-012375731-9/50018-5

Ross, L. a, Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you
see what I am saying? Exploring visual enhancement of speech comprehension in
noisy environments. *Cerebral Cortex*, *17*, 1147–1153.
https://doi.org/10.1093/cercor/bhl024

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012).
A bilingual advantage in visual language discrimination in infancy. *Psychological
Science*, *23*, 9, 994–999. https://doi.org/10.1177/0956797612436817

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of
vowel contrasts in bilingual infants: Is the distributional account all there is to it?
Developmental Science, *12*, 874–887.
https://doi.org/10.1111/j.1467-7687.2009.00829.x

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and
development. *Trends in Cognitive Sciences*, *13*, 3, 127–134.
https://doi.org/10.1016/j.tics.2008.11.009

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*, 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*, 850–855.

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, *18*, 534–553. https://doi.org/10.1111/j.1532-7078. 2012.00135.x

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. 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*, 1173–1190. https://doi.org/10.1017/S0305000914 000725

Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*. 525. https://doi.org/10.3389/fpsyg.2016.00525

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science*, *9*, 661–665. https://doi.org/10.1177/1745691614552498

Vihman, M. (1996). *Phonological development: The origins of language in the child. Applied language studies*. Cambridge, MA: Blackwell Publishing.

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

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

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*, 1866–1878. https://doi.org/10.1121/1.390988

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

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Sentences Used as Speech Events in Experiments 1 and 2
Appendix S2. Durations for Each Nonspeech Movement Used for Each Language in Experiment 1.
Appendix S3. Results of the Growth Curve Analysis for the Anticipation Window in Experiments 1 and 2.