Brief Report

Effect of infant bilingualism on audiovisual integration in a McGurk task

Evelyne Mercure [a,b,c,*], Peter Bright [d], Isabel Quiroz [b], Roberto Filippi [c]

[a] Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK
[b] School of Psychology, Birkbeck, University of London, London WC1E 7HX, UK
[c] UCL Institute of Education, University College London, London WC1E 6BT, UK
[d] School of Psychology and Sport Science, Anglia Ruskin University, Cambridge CB1 1PT, UK

ARTICLE INFO

ABSTRACT

Infants growing up in an environment where more than one language is spoken tend to follow the early milestones of early language development. This is an impressive achievement given that they are learning two languages while receiving reduced exposure to each of these languages compared with monolingual infants. This increased variability in their linguistic environment may lead to adjustments in the way bilingual infants process visual and auditory speech. This study aimed to clarify the influence of infant bilingualism on the development of audiovisual speech integration. Using eye tracking and a McGurk paradigm, we studied face scanning patterns when 7- to 10-month-old infants were viewing articulation of audiovisually congruent and incongruent syllables. We found that monolingual infants decreased their attention to the mouth and increased their attention to the eyes of speaking faces when presented with incongruent articulation, typically leading to the McGurk illusion during adulthood. In bilingual infants, no differences in face scanning patterns were observed between audiovisually congruent and incongruent articulation, suggesting that the increased variability in their speech experience may lead to more tolerance to articulatory inconsistencies. These results suggest that the development of audiovisual speech perception is influenced by infants' language environment.

* Corresponding author at: Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK.
E-mail address: e.mercure@gold.ac.uk (E. Mercure).

Introduction

Infants as young as 2 months can successfully associate the auditory aspect of speech with its visual articulation movements (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003). However, despite the convincing demonstration of this early integration, audiovisual speech perception is a complex skill that develops throughout the first year of life and beyond (Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012).

One paradigm that has been used to study audiovisual integration is the McGurk illusion. In this illusion discovered during the 1970s (McGurk & MacDonald, 1976), adults are presented with incongruent audiovisual speech and report perceiving a syllable that is neither the one presented auditorily nor the one presented visually but rather a fusion of both into the closest syllable. Interestingly, participants do not usually notice the audiovisual incongruence. Certain combinations of syllables have been known to lead to the McGurk illusion (referred to in this article as *fusible incongruent;* e.g., visual /ga/ with auditory /ba/, leading to the illusory fusion as /da/ or /ẟa/), whereas other combinations do not typically lead to this illusion (referred to as *non-fusible incongruent;* e.g., visual /ba/ with auditory /ga/, leading to the perception of the audiovisual incongruence and a non-natural percept such as /bga/). Research with a habituation paradigm (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997) or electrophysiology (Kushnerenko, Teinonen, Volein, & Csibra, 2008) has revealed that infants from 4 months of age process non-fusible incongruent articulation differently from congruent articulation and fusible incongruent articulation. These findings are compatible with the idea that infants could also perceive the McGurk illusion, but it was also found that audiovisual integration is not as strong or consistent in infants as in adults (Desjardins & Werker, 2004).

During their first year of life, infants change their scanning patterns for speaking faces. Indeed, it has been observed that infants shift the focus of their attention from the eyes to the mouth of speaking faces from 4 to 8 months of age (Lewkowicz & Hansen-Tift, 2012; Mercure et al., 2019; Morin-Lessard, Poulin-Dubois, Segalowitz, & Byers-Heinlein, 2019; Pejovic, Yee, & Molnar, 2021; Pons, Bosch, & Lewkowicz, 2015; Tsang, Atagi, & Johnson, 2018). This eye-to-mouth shift corresponds to a developmental period of intense phonological learning during infancy reflected in advances in both speech perception and production (Best, 1993; Curtin & Werker, 2007; Lang et al., 2019; Lee, Jhang, Relyea, Chen, & Oller, 2018). The "language expertise hypothesis" suggests that attention to the mouth at this stage is key to language development (Tsang et al., 2018), but the evidence to support this hypothesis is not always reliably strong (Morin-Lessard et al., 2019). After this period, some studies report that looking time to the mouth decreases (Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2019), whereas other studies find that it remains stable until 5 years of age (Morin-Lessard et al., 2019). Moreover, at 12 months of age, a focus on the eyes, and not the mouth, of speaking faces was associated with increased communicative and social skills (Pons et al., 2019).

Most of the studies on face scanning patterns have investigated the perception of natural speech, but a similar shift from eyes to mouth has also been observed when infants view audiovisually congruent and incongruent syllables (Danielson, Bruderer, Kandhadai, Vatikiotis-Bateson, & Werker, 2017; Mercure et al., 2019; Tomalski, Ribeiro, et al., 2013). These studies revealed that face scanning patterns not only differ with age but also differ depending on whether the syllable being viewed is audiovisually congruent or incongruent. Tomalski, Ribeiro, et al. (2013) demonstrated that looking times to the mouth differed when infants aged 6 to 9 months viewed articulation of audiovisually congruent and incongruent syllables. Longer looking times to the mouth were observed for fusible incongruent articulation compared with both non-fusible incongruent and congruent articulation. This pattern was modulated by an effect of age in which 6- and 7-month-olds looked at the mouth more for the fusible versus non-fusible incongruent articulations, whereas no difference was observed between the two types of incongruent articulations in 8- and 9-month-olds. Both patterns of results are incompatible with the McGurk illusion given that younger infants were processing the fusible incongruent articulation differently than congruent articulation and older infants did not scan faces differently for fusible and non–fusible articulation. It is important to note that this study included both

monolingual and bilingual infants, some of whom had no regular exposure to English in their home environment. It is unclear how the infants' language experience influenced the findings of this study. Furthermore, Danielson et al. (2017) demonstrated that face scanning patterns differ when infants view a non-native phonological contrast. Indeed, they observed that 6- and 9-month-olds tended to increase their looking time to the mouth when viewing incongruent non-native articulation (Hindi dental vs. retroflex syllables), whereas 11-month-olds did not show any difference in face scanning patterns for congruent versus incongruent syllables.

Increased attention to the eyes or decreased attention to the mouth when presented with audiovisually incongruent native syllables at 6 to 9 months of age is associated with better language skills at 14 to 16 months (Kushnerenko et al., 2013). This suggests that infants who present more mature language development can shift their attention away from unhelpful visual articulation movements and focus on the social information present in the eyes instead. Studying infants with different language experiences is another way of assessing potential links between language learning and scanning patterns for speaking faces. Indeed, it has been observed that infants raised in a bilingual environment are more sensitive to visual articulation than monolinguals (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012). They may also focus on the mouth more than monolinguals when viewing speaking faces (Pons et al., 2015) or faces displaying non-speech dynamic movements, including crying and laughing (Ayneto & Sebastian-Galles, 2017). However, increased attention to the mouth in bilinguals compared with monolinguals has not always been reliably found (Mercure et al., 2019; Morin-Lessard et al., 2019; Tsang et al., 2018). One explanation for this discrepancy in results is that increased attention to the mouth may be restricted to bilinguals exposed to two very similar languages such as Spanish and Catalan (Birulés, Bosch, Brieke, Pons, & Lewkowicz, 2019). This effect might not generalize to bilingual infants exposed to two spoken languages with less similarity in phonology and rhythm, such as French and English, or to infants exposed to two languages in different sensory modality, such as infants with deaf parents exposed to a spoken language and a signed language.

Few studies have assessed the impact of bilingual experience on the processing of audiovisual incongruence. In a prior study, we found that monolingual and bilingual infants exposed to English and another spoken language (unimodal bilinguals) increased their looking time to the mouth of speaking faces from 4 to 8 months of age, whereas infants with deaf parents exposed to English and British Sign Language (bimodal bilinguals) did not (Mercure et al., 2019). This suggests that the eye-to-mouth shift from 4 to 8 months is not a simple process of maturation but rather is dependent on the type and amount of audiovisual speech experience infants have accumulated. Moreover, monolingual infants aged 6 to 8 months were observed to increase their looking time to the mouth of speaking faces when presented with audiovisually incongruent articulation compared with audiovisually congruent articulation. No difference was observed between fusible and non-fusible incongruent articulation, a pattern of result nonsupportive of the perception of the McGurk illusion. Younger monolingual infants and both groups of bilingual infants (unimodal and bimodal) displayed no influence of audiovisual congruence on their face scanning patterns. This suggests that experience of audiovisual speech shapes the ability of infants to detect and react to audiovisual incongruences. It may be that bilingual infants, both unimodal and bimodal, have more experiences of variability in articulation and therefore are more tolerant of audiovisual incongruences. It remains unclear whether the difference in sensitivity to audiovisual incongruences is a transient phase or whether it is a more developmentally stable pattern. To address this question, the current study compared monolingual and unimodal bilingual infants in a slightly older age group.

The first aim of this study was to compare the sensitivities of 7- to 10-month-old monolingual and bilingual infants to audiovisual incongruences. In Mercure et al.'s (2019) study, audiovisual speech incongruences influenced face scanning patterns in monolingual infants aged 6 to 8 months, but not in bilingual infants of the same age or in younger monolingual and bilingual infants. Unimodal bilinguals showed a trend toward an increase in looking times to the mouth for incongruent articulation. Given increased variability in their language environment, they may require a few more months of audiovisual speech experience to reach a similar pattern to monolingual infants. The current study tested the hypothesis that this sensitivity to audiovisual incongruences would remain observable in monolingual infants aged 7–10 months (as also observed by Tomalski et al., 2013, in infants aged

6–9 months processing native contrasts and by Danielson et al., 2017, in infants aged 6–9 months processing non-native contrasts). It also tested the hypothesis that sensitivity to audiovisual speech incongruence would emerge for native contrasts in bilingual infants aged 7–10 months. The current study used the experimental design presented by Tomalski et al. (2013) and by Mercure et al. (2019), which allows comparing not only congruent versus incongruent articulation but also fusible incongruent versus non-fusible incongruent articulation. If scanning patterns for fusible incongruent articulation differ from those for non-fusible incongruent articulation, but not for those from congruent articulation, the pattern of results can be considered to be compatible with the perception of the McGurk illusion. This was not the case in younger infants in Mercure et al. (2019) or in Tomalski et al. (2013) in an older sample of infants with varied language experience. The current study assessed the compatibility of the results with the McGurk illusion in a group of 7- to-10-month-old contrasting monolinguals and bilinguals.

A secondary aim of this study was to test the developmental trajectories of scanning patterns for faces articulating syllables from 7 to 10 months of age. Based on prior literature, it was expected that the sharing of attention to the mouth and eyes of speaking faces would be relatively stable across this age range. This study also tested the prior finding that bilingual infants look at the mouth of talking faces (Pons et al., 2015) longer than monolinguals. The current study tested this prediction in a mixed group of bilinguals, most of whom experienced languages that were not similar in rhythm and phonology. If increased looking time to the mouth generalized to this mixed group of bilinguals compared with monolinguals, it would support the idea that this effect is not strictly restricted to bilinguals experiencing very similar languages such as Spanish and Catalan.

Method

Participants

A total of 35 infants aged 7 to 10 months contributed data. A further 21 infants participated in the study but could not be included due to equipment malfunction or failure to calibrate ($n = 6$), experimenter error ($n = 1$), or failure to reach looking criteria ($n = 14$). Infants were from two different groups with different language experience: 18 monolinguals exposed to English (12 female; mean age = 8.8 months) and 17 bilinguals with hearing parents regularly and frequently exposed to both English and one or more additional spoken languages (6 female; mean age = 8.4 months). There was no significant effect of age between groups, $t(33) = 1.43$, $p = .162$. Bilinguals were exposed to English on average 51.6% of the time (minimum 10%–maximum 90%, $SD = 26.3$). Additional languages of exposure included Mandarin, Farsi, French, Sinhala, Vietnamese, Polish, Dutch, German, Flemish, Samali, Greek, Spanish, and Italian.

Infants were recruited from the Infants were recruited from the Birkbeck Babylab database of volunteers database of volunteers. Infants were born at term (37–42 gestational weeks) except for 1 monolingual infant born at 36 weeks for whom a corrected age was used. Parents reported no hearing or vision problems as well as no serious developmental or physical conditions. Most families came from the greater London area. Families were reimbursed for their travel expenses and were offered a baby T-shirt and certificate of participation. This study was approved by the University College London and Birkbeck ethics committees. Parents offered written consent after the procedure was described to them and they had an opportunity to ask questions.

Procedure

Participants were invited to participate in a larger research protocol investigating the effects of early bilingualism on preverbal communication and attention. The protocol began with the presentation of three eye-tracking tasks presented in Tobii Studio (Tobii, Stockholm, Sweden): the McGurk task reported here, an "attention to faces" task (Mercure et al., 2018), and a gaze-following task. This was followed by seven short eye-tracking tasks in a different experimental setup that are not presented in this article. The whole protocol usually required 1 to 1.5 h for each family, including resting, napping,

and feeding breaks. During the McGurk task, infants sat on their parent's lap in a dimly lit room about 60 cm from the Tobii T120 eye tracker (17-inch diameter, screen refresh rate 60 Hz, eye-tracking sampling rate of 60 Hz, spatial accuracy < 1°). The protocol began with calibration of the infant gaze position using colorful animations and a five-point routine. Infant behavior was monitored during the study using a camera and Tobii Studio Live Viewer. When infants were distracted, their attention was occasionally brought back to the screen by shaking a rattle behind the screen. After completion of both eye-tracking protocols or during breaks, parents were asked to complete questionnaires about their infant's family context, language experience, and medical history.

Stimuli

Short videos of a female native English speaker articulating /ba/ or /ga/ were presented. These stimuli have been used and described in detail in prior studies (Kushnerenko et al., 2008, 2013; Mercure et al., 2019; Tomalski, Moore, et al., 2013). Five experimental conditions were presented: (1) congruent audiovisual /ba/; (2) congruent audiovisual /ga/; (3) fusible incongruent: audio /ba/ and visual /ga/, which is associated with the illusory McGurk effect in adults and perceived as /da/ or /δa/; (4) non-fusible incongruent: audio /ga/ and visual /ba/, which is associated with a non-natural percept in adults such as "bga"; and (5) silent articulation: visual /ba/ or /ga/ without any auditory information. The auditory track of one stimulus was dubbed onto the visual track of another to create the incongruent conditions. Sound onset was at 360 ms from the start of the stimulus, auditory syllables lasted 300 ms, and visual syllables lasted 760 ms. Ten repetitions of the same stimulus were presented to form a trial, which lasted 7.6 s. The face in the video covered approximately 14° × 22° of visual angle. Infants viewed 10 trials (2 of each condition) in a fixed order, which began and ended with presentations of the congruent conditions. A colorful animation was presented at the center of the screen before each trial and was ended by the experimenter when infants focused on it.

Data analysis

The analysis strategy follows those elaborated in a previous study to allow direct comparison of results (Mercure et al., 2019). In each trial, the mean looking time was extracted in regions of interest using Tobii Studio. These regions were identical to those used in Mercure et al. (2019) and were defined as (1) the eyes region (oval shape of maximum dimensions 285 × 128 pixels), (2) the mouth region (oval shape of maximum dimensions 171 × 142 pixels), and (3) the entire face region (oval shape of maximum dimensions 332 × 459 pixels). Trials were excluded when infants looked at the entire face for less than a cumulative 3 s, and only infants with at least 7 good trials out of 10 were included for analyses. These criteria were identical to those used in Mercure et al.'s (2019) study

Table 1
Mean looking times to areas of interest (mouth, eyes, and face) followed by mouth-to-face and eyes-to-face ratios for monolinguals and bilinguals.

	Mouth		Eyes		Face		Mouth-to-face		Eyes-to-face	
	M	B	M	B	M	B	M	B	M	B
Congruent /ba/	4.22 (1.69)	3.51 (140)	1.75 (1.60)	1.69 (1.52)	6.58 (0.54)	5.98 (1.07)	0.64 (0.24)	0.60 (0.25)	0.27 (0.25)	0.28 (0.25)
Congruent /ga/	4.39 (1.47)	3.66 (1.11)	1.45 (1.24)	1.50 (1.06)	6.25 (0.86)	5.84 (0.73)	0.70 (0.22)	0.63 (0.20)	0.24 (0.20)	0.25 (0.18)
Fusible	4.18 (1.52)	3.83 (0.84)	1.36 (1.18)	1.54 (1.02)	6.13 (0.67)	6.06 (0.91)	0.68 (0.22)	0.64 (0.15)	0.22 (0.20)	0.25 (0.16)
Non-fusible	3.43 (1.48)	4.07 (0.89)	1.95 (1.09)	1.24 (0.91)	6.14 (0.94)	6.16 (0.80)	0.57 (0.24)	0.67 (0.14)	0.33 (0.20)	0.20 (0.13)
Silent	3.82 (1.48)	3.79 (0.96)	1.25 (1.11)	1.16 (0.85)	5.42 (1.19)	5.47 (0.97)	0.70 (0.21)	0.70 (0.14)	0.24 (0.21)	0.20 (0.15)

Note. M, monolinguals; B, bilinguals. Looking times are measured in seconds. Numbers in parentheses represent standard deviations.

but led to a higher rejection rate based on looking criteria (29% vs. 16% of infants). This is probably because older infants were more likely to get bored with the stimuli and look away. An average of 9.4 trials were included (SD = 0.9), and there was no difference between groups, $t(33)$ = 0.51, p = .616. Mean mouth-to-face and eyes-to-face ratios were calculated for each participant in each condition (see Table 1 for summary data).

Results

Effect of audiovisual congruency on mouth-to-face and eyes-to-face ratios

A mixed analysis of variance (ANOVA) was performed with two regions (mouth-to-face and eyes-to-face ratios), five stimuli (congruent /ba/, congruent /ga/, fusible incongruent, non-fusible incongruent, and silent), and two groups (monolinguals and bilinguals) (see Fig. 1). Infants spent a larger proportion of their time looking at the mouth than looking at the eyes of speaking faces [region main effect: $F(1, 33)$ = 43.715, p < .001, η^2 = .570]. There was also a significant Region × Stimulus interaction, $F(4, 30)$ = 3.184, p = .027, η^2 = .298, and a significant Region × Stimulus × Group interaction, $F(4, 30)$ = 5.363, p = .002, η^2 = .417. This three-way interaction was followed by separate ANOVAs on each group.

In monolinguals, a significant effect of region was found, $F(1, 17)$ = 16.350, p = .001, η^2 = .490, as well as a significant Region × Stimulus interaction, $F(4, 14)$ = 6.781, p = .003, η^2 = .660. On the mouth-to-face ratio, the impact of stimulus was significant, $F(4, 14)$ = 8.641, p = .001, η^2 = .712. In response to the non-fusible condition, monolinguals looked at the mouth significantly less than for all the other conditions (all ps < .001). The impact of stimulus was also significant on the eyes-to-face ratio, $F(4, 14)$ = 4.750, p = .012, η^2 = .576. The non-fusible incongruent condition was the only one driving significant effects on post hoc comparisons. An increased eyes-to-face ratio was observed for the non-fusible condition compared with congruent /ga/, $t(17)$ = 3.185, p = .005, fusible incongruent, $t(17)$ = 4.046, p = .001, and silent articulation, $t(17)$ = 3.626, p = .002.

In bilinguals, a 2 (Region) × 5 (Stimulus) ANOVA revealed a main effect of region, $F(1, 16)$ = 33.555, p < .001, η^2 = .677, but no effect of stimulus, $F(4, 13)$ = 1.054, p = .417, η^2 = .245, and no Region × Stimulus interaction, $F(4, 13)$ = 1.340, p = .307, η^2 = .292.

The same pattern of results was obtained when these analyses were performed on absolute looking times to the eyes and mouth regions instead of ratio measures (see online supplementary material).
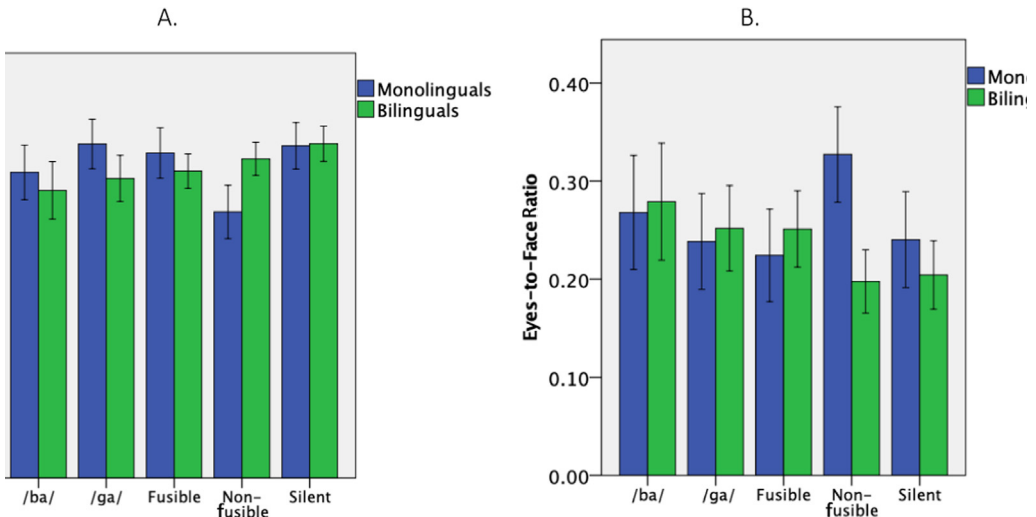

Fig. 1. Mouth-to-face (A) and eyes-to-face (B) ratios in each experimental condition for monolingual infants (blue; left bars) and bilingual infants (green; right bars). Error bars represent standard errors.

Looking time to the entire face and to the mouth of talking faces

A univariate ANOVA was used to compare looking time to the entire face across groups. The groups did not differ in their general attention to the faces presented in this task, $F(1) = 0.210$, $p = .427$, $\eta^2 = .019$. The same analysis on the mouth-to-face ratio also revealed no group differences, $F(1) = 0.014$, $p = .906$, $\eta^2 < .001$, suggesting that monolinguals and bilinguals did not differ in their visual attention to the mouth of speaking faces in the current task.

Correlation of mouth-to-face and eyes-to-face ratios with age

To assess the developmental differences in face scanning patterns with age, correlations were performed between age and the mouth-to-face and eyes-to-face ratios in each group (see Fig. 2). No Kendall's tau nonparametric correlation was found to be significant (monolinguals eyes-to-face ratio: $r = .085$, $p = .622$; mouth-to-face ratio: $r = -.124$, $p = .472$; bilinguals eyes-to-face ratio: $r = .037$, $p = .837$; mouth-to-face ratio: $r = -.066$, $p = .711$), suggesting stable face scanning patterns from 7 to 10 months of age in both monolingual and bilingual infants. The same results were obtained when performing these analyses on absolute looking times (see supplementary material for these analyses as well as analyses of potential outliers and their lack of impact on these findings).

Discussion

The primary aim of this study was to compare the influence of audiovisual incongruences on face scanning patterns of monolingual and bilingual infants from 7 to 10 months of age. It was predicted that sensitivity to audiovisual incongruences would emerge for bilinguals in this age group and would also remain observable for monolinguals. The results did not fully support this hypothesis. The audiovisual condition significantly influenced face scanning patterns in monolinguals but not in bilinguals. Interestingly, the effect observed in monolinguals was different from the one observed in slightly younger infants. In Mercure et al. (2019), it was observed that 6- to 8-month-old monolinguals increased their looking times to the mouth in response to both fusible and non-fusible incongruent articulations. In the current study, 7- to 10-month-old monolinguals demonstrated reduced attention to the mouth, and increased attention to the eyes, in the case of non-fusible articulation compared
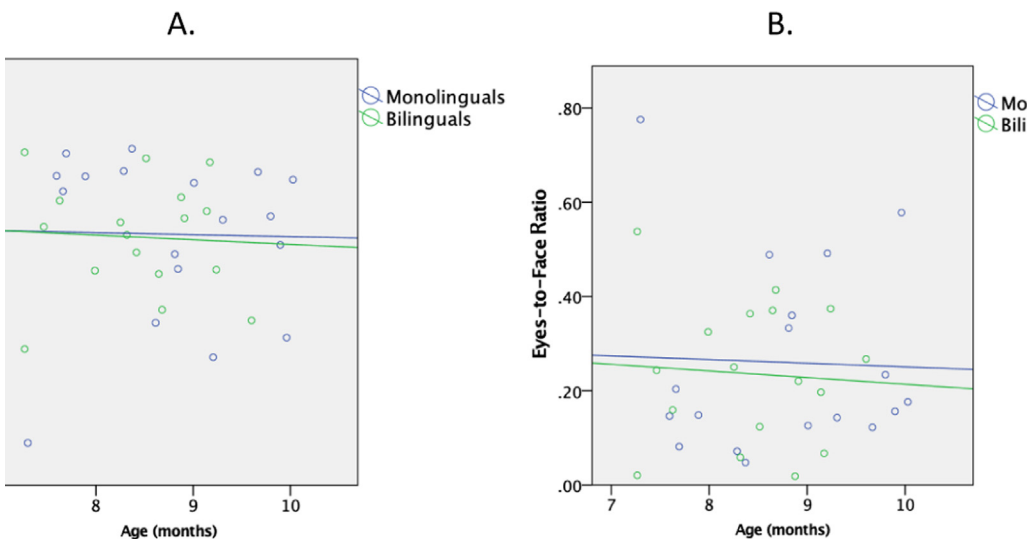

Fig. 2. Correlations of mouth-to-face ratio (A) and eyes-to-face ratio (B) with age in each group.

with all (or most) other conditions. This may represent a more mature face scanning pattern in which attention is shifted away from unhelpful mouth movements not matching the sound heard and directed to the eyes instead. Kushnerenko and colleagues (2013) observed that infants who displayed this face scanning pattern in response to incongruent articulation from 6 to 9 months of age demonstrated better language skills at 14 to 16 months. Interestingly, in the current study, the fusible articulation did not elicit any differences in scanning patterns from congruent or silent articulation but differed from non-fusible incongruent articulation in terms of mouth-to-face and eyes-to-face ratios. One explanation for these findings is that monolingual infants may have experienced the McGurk illusion and not perceived the audiovisual incongruence of the fusible condition, as adults usually report. These findings are congruent with behavioral evidence (Burnham & Dodd, 2004; Rosenblum et al., 1997) and electrophysiological evidence (Kushnerenko et al., 2008) compatible with the McGurk illusion during infancy. However, these findings differ from those of Tomalski et al. (2013), who observed no difference in looking times to the mouth for the fusible versus non-fusible articulation in 8- and 9-month-olds. It may be that the inclusion of bilingual infants, some of whom were not exposed to English in their home environment, eliminated an effect that may have been present in the monolinguals of this sample. Tomalski et al. (2013) do not present separate data for monolinguals and bilinguals and do not mention the number of monolinguals and bilinguals, so it is impossible to fully assess this interpretation. The findings of the current study also differ from those of Danielson et al. (2017). Indeed, those authors observed increased looking times to the mouth for incongruent articulation, whereas the current study observed decreased looking times to the mouth for incongruent articulation. It is important to note that Danielson and colleagues presented a non-native contrast with a slight change in place of articulation (dental vs. retroflex), whereas the current study used a visually obvious native contrast with a bilabial versus velar place of articulation. It may be that infants noticing the subtle audiovisual incongruence in the non-native contrast were intrigued by its novelty, whereas infants noticing the audiovisual incongruence of the visually salient native contrast were more likely to disregard the mouth as unreliable. Only a comparison of native and non-native contrasts of closer and farther places of articulation could assess this possibility.

Contrary to monolingual infants, bilingual infants did not modify their face scanning patterns in response to the different audiovisual stimuli presented. In this respect, it is impossible to tell from these data whether the bilingual infants noticed any audiovisual incongruence and whether they experienced the McGurk illusion. These results are similar to those observed in Mercure et al. (2019), where bilingual infants with deaf and hearing parents demonstrated no significant difference in their face scanning patterns for audiovisually congruent and incongruent articulation. One explanation for these findings is that bilingual infants are likely to experience more variable audiovisual speech, including more than one language and potentially foreign accented speech, which in turn could lead to more tolerance to inconsistent articulation. This could explain why bilingual infants did not reliably modify their face scanning patterns when presented with different audiovisual articulation conditions. The current study extends prior findings in establishing that this difference between monolingual and bilingual infants is replicable in a slightly older age group.

A secondary aim of the current study was to assess the developmental trajectory of scanning patterns in response to faces articulating syllables. No influence of age was found on these patterns, suggesting that after a steep increase in attention to the mouth from 4 to 8 months of age these patterns reach a plateau and remain relatively stable from 7 to 10 months. These findings are compatible with prior studies suggesting a stable developmental period in face scanning patterns (Morin-Lessard et al., 2019), but not with findings of increased mouth looking in this time window (Lewkowicz & Hansen-Tift, 2012; Tsang et al., 2018). These results are more difficult to compare with studies contrasting 4-, 8-, and 12-month-olds (Pons et al., 2015). Moreover, this study observed no difference in general attention to the mouth between monolingual and bilingual infants in response to faces articulating syllables. These findings join a now growing body of studies failing to find this impact of language background on face scanning patterns (Mercure et al., 2019; Morin-Lessard et al., 2019; Tsang et al., 2018). Increased attention to the mouth has been observed in bilingual infants compared with monolingual infants at 8 and 12 months of age for non-speech movements (Ayneto & Sebastian-Galles, 2017) and at 4 and at 12 months, but not at 8 months, for speech movements (Pons et al., 2015). It may be that this effect is developmentally transient and would not have been observed for speech

in infants from 7 to 10 months. Moreover, this effect might only be present in infants exposed to very similar languages such as Catalan and Spanish (Ayneto & Sebastian-Galles, 2017; Birulés et al., 2019; Pons et al., 2015), where increased attention to visual articulation may help to differentiate languages of similar rhythm and phonology. In the current study, infants were exposed to English and any other spoken language, with very few pairs being as similar in rhythm and phonology as Catalan and Spanish. Alternatively, this nonsignificant group difference could be attributed to the broad definition of bilingualism used in this study. Indeed, infants with 10% to 90% exposure to English were included as bilinguals. It may be that differences in face scanning patterns only exist in bilinguals receiving a more balanced exposure to their languages. However, this possibility is less likely given the lack of relationship of face scanning patterns with language exposure in the current data (see supplementary material).

Taken together, the results of the current study do not demonstrate any impact of bilingualism on general face scanning patterns but suggest that bilingual language experience can influence the development of audiovisual speech perception. The more complex task of learning two phonological systems during infancy may lead to increased tolerance to audiovisual incongruences in the second half of the first year of life.

Acknowledgments

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2021.105351.

References

Ayneto, A., & Sebastian-Galles, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science, 20* e12446.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. De Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289–304). New York: Springer Science + Business Media.

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2019). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science, 22* e12755.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*, 204–220.

Curtin, S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 579–599). Oxford, UK: Oxford University Press.

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development, 42*, 37–48.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*, 187–203.

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

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

Kushnerenko, E., Tomalski, P., Ballieux, H., Ribeiro, H., Potton, A., Axelsson, E. L., ... Moore, D. G. (2013). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *European Journal of Neuroscience, 38*, 3363–3369.

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., ... Marschik, P. B. (2019). Canonical babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports, 6*(3), 111–118.

Lee, C.-C., Jhang, Y., Relyea, G., Chen, L., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development, 50*, 140–153.

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

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science, 22* e12701.

Mercure, E., Quiroz, I., Goldberg, L., Bowden-Howl, H., Coulson, K., Gliga, T., ... MacSweeney, M. (2018). Impact of language experience on attention to faces in infancy: Evidence from unimodal and bimodal bilingual infants. *Frontiers in Psychology, 9.* https://doi.org/10.3389/fpsyg.2018.01943.

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

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,* 237–247.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6,* 191–196.

Pejovic, J., Yee, E., & Molnar, M. (2021). June 1). Eyes can tell: Attention to the eyes and the mouth during audiovisual vowel processing in monolingual and bilingual infants. *PsyArXiv.* https://doi.org/10.31234/osf.io/xjcr6.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science, 26,* 490–498.

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

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59,* 347–357.

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,* 994–999.

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 207–228). New York: Oxford University Press.

Tomalski, P., Moore, D. G., Ribeiro, H., Axelsson, E. L., Murphy, E., Karmiloff-Smith, A., ... Kushnerenko, E. (2013). Socioeconomic status and functional brain development—Associations in early infancy. *Developmental Science, 16,* 676–687.

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,* 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.