



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



The redeployment of attention to the mouth of a talking face during the second year of life

Anne Hillairet de Boisferon^a, Amy H. Tift^a, Nicholas J. Minar^b,
David J. Lewkowicz^{c,*}

^a Department of Psychology, Florida Atlantic University and Florida Atlantic University High School Research Program, Boca Raton, FL 33314, USA

^b Institute for the Study of Child Development, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA

^c Department of Communication Sciences and Disorders, Northeastern University, Boston, MA 02115, USA



ARTICLE INFO

Article history:

Received 4 May 2017

Revised 18 March 2018

Available online 5 April 2018

Keywords:

Selective attention

Infants

Speech perception

Eye-tracking

Audiovisual speech

Vocabulary

ABSTRACT

Previous studies have found that when monolingual infants are exposed to a talking face speaking in a native language, 8- and 10-month-olds attend more to the talker's mouth, whereas 12-month-olds no longer do so. It has been hypothesized that the attentional focus on the talker's mouth at 8 and 10 months of age reflects reliance on the highly salient audiovisual (AV) speech cues for the acquisition of basic speech forms and that the subsequent decline of attention to the mouth by 12 months of age reflects the emergence of basic native speech expertise. Here, we investigated whether infants may redeploy their attention to the mouth once they fully enter the word-learning phase. To test this possibility, we recorded eye gaze in monolingual English-learning 14- and 18-month-olds while they saw and heard a talker producing an English or Spanish utterance in either an infant-directed (ID) or adult-directed (AD) manner. Results indicated that the 14-month-olds attended more to the talker's mouth than to the eyes when exposed to the ID utterance and that the 18-month-olds attended more to the talker's mouth when exposed to the ID and the AD utterance. These results show that infants redeploy their attention to a talker's mouth when they enter the word acquisition phase and suggest that infants rely on the greater perceptual salience of redundant AV speech cues to acquire their lexicon.

© 2018 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail address: d.lewkowicz@northeastern.edu (D.J. Lewkowicz).

Introduction

The acquisition of speech and language during infancy is a protracted developmental process. On the perception/processing side, it consists of months of tuning to the sounds and sights of the native language(s). The tuning is driven by learning and differentiation as well as by perceptual narrowing. The former process results in the acquisition of novel speech and language skills, whereas the latter process results in a decline in responsiveness to non-native categories of information (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014). On the production side, acquisition of speech and language consists of the emergence of increasingly more complex and functionally flexible vocalizations. This includes squeals, vowel-like sounds, and growls at around 3 or 4 months of age (Oller et al., 2013), canonical babbling sounds at around 6 months (Oller, 2000), single words at the start of the second year of life, and multiple words by 17–20 months (Fenson et al., 1994; Nelson, 1974).

What factors might contribute to the growth of speech perception and speech production capacity during infancy? Studies of infant–adult interaction have found that infant perception and production of speech is affected by the cues that infants' interlocutors provide to them during social interactions (Goldstein, King, & West, 2003; Goldstein & Schwade, 2008; Kuhl, 2007). Many such interactions consist of face-to-face contact between infants and their interlocutors. This raises the possibility that selective attention to different parts of a social partner's face—especially the interlocutor's mouth—contributes to the acquisition of speech and language. Indeed, this possibility is supported by findings indicating that infants begin to deploy their selective attention to a talker's mouth sometime between 6 and 8 months of age (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2016; Lewkowicz & Hansen-Tift, 2012; Merin, Young, Ozonoff, & Rogers, 2007; Pons, Bosch, & Lewkowicz, 2015; Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013; Tenenbaum et al., 2015).

In one of the studies examining infant selective attention to talking faces, Lewkowicz and Hansen-Tift (2012) investigated the relative amount of time that 4-, 6-, 8-, 10-, and 12-month-old monolingual, English-learning infants and adults attend to a talker's eyes and mouth and whether such selective attention is affected by language familiarity and/or manner of speech. To do so, they measured eye gaze while participants watched videos during which an actor could be seen and heard uttering a monologue either in the participants' native language or in a non-native language spoken in either an infant-directed (ID) or adult-directed (AD) manner. Findings indicated that, regardless of manner of speech and language spoken, 4-month-olds attended more to the talker's eyes, 6-month-olds attended equally to the eyes and mouth, and 8- and 10-month-olds attended more to the mouth. Findings also indicated that 12-month-olds did not attend more to the mouth in response to native speech but that they attended more to the mouth in response to non-native speech. Finally, the findings showed that adults attended more to the talker's eyes regardless of language spoken. Lewkowicz and Hansen-Tift (2012) drew three conclusions from these results. First, they concluded that the initial attentional shift to the talker's mouth by 8 and 10 months of age reflects the emergence of an endogenous selective attention mechanism and noted that the shift occurs at around the same time that canonical babbling begins emerging. Second, they concluded that the attentional shift enables infants to gain direct access to synchronous, redundantly specified, and thus highly salient audiovisual (AV) speech cues. Finally, they concluded that access to redundant AV speech cues is likely to facilitate acquisition of initial native speech forms (i.e., native-language phonological representations as well as motor speech forms) because multisensory speech cues are known to be perceptually more salient than unisensory speech cues.

A study by Pons et al. (2015) provided additional evidence that multisensory redundancy plays a role in infant selective attention to AV speech. These investigators compared monolingual and bilingual infants' response to talking faces and found that, overall, bilingual infants attended more to the mouth than did monolingual infants. Specifically, they compared Spanish- and Catalan-learning monolingual infants' response to native and non-native AV speech with bilingual (Spanish–Catalan) infants' response to native (i.e., dominant) and non-native speech. Results paralleled the results

reported by Lewkowicz and Hansen-Tift (2012). Specifically, Pons et al. found that monolingual and bilingual infants shifted their attention from the eyes at 4 months of age to the mouth at 8 months of age. They also found that at 12 months of age monolingual infants looked equally at the talker's eyes and mouth but that bilingual infants attended more to the talker's mouth than to the eyes, that they attended more to the mouth than did monolingual infants, and that they attended more to the mouth regardless of whether the talker spoke in the native or non-native language.

Lewkowicz and Hansen-Tift's (2012) conclusion that their findings reflect infant selective attention to multisensory redundancy per se is consistent with a large body of findings documenting the power of multisensory redundancy across different species and across the lifespan. Overall, this body of findings demonstrates that multisensory, as opposed to unisensory, inputs increase perceptual salience (Ernst & Bühlhoff, 2004) and neural responsiveness (Musacchia & Schroeder, 2009; Stein & Stanford, 2008; Wallace & Stein, 2001), facilitate perceptual detection and discrimination (Barenholtz, Mavica, & Lewkowicz, 2016; Partan & Marler, 1999; Rowe, 1999; Sumbly & Pollack, 1954; Summerfield, 1979; van Wassenhove, Grant, & Poeppel, 2005; Von Kriegstein & Giraud, 2006), and enhance learning and memory (Bahrick & Lickliter, 2012; Barenholtz, Lewkowicz, Davidson, & Mavica, 2014; Shams & Seitz, 2008; Thelen, Talsma, & Murray, 2015).

If multisensory redundancy plays a role in infant selective attention to talking faces during the first year of life, might it continue to play such a role into the second year of life? The studies to date have suggested that attention to a talker's mouth facilitates phonological tuning but, of course, access to the redundant AV speech cues inherent in a talker's mouth is useful for much more than phonological tuning. This is illustrated by findings that adults attend more to a talker's mouth when speech is at a low volume (Lansing & McConkie, 2003), when they need to perform phoneme discrimination while learning a second language (Hirata & Kelly, 2010), when they have access to facial motion conveying linguistic information (Vö, Smith, Mital, & Henderson, 2012), when auditory speech is degraded (Sumbly & Pollack, 1954), when speech is difficult to understand (Reisberg, McLean, & Goldfield, 1987), and when they need to perform a difficult language detection task (Barenholtz et al., 2016).

Given that adults rely on the highly salient redundant AV speech cues available in a talker's mouth to maximize processing of speech and language, it is theoretically possible that greater attention to a talker's mouth may help infants older than 12 months to acquire their first words. For this to be the case, however, it would also need to be the case that infants' attention to a talker's mouth varies as a function of the types of AV speech cues available and their specific developmental experience. Indeed, two sets of findings indicate that this is the case. As indicated earlier, Lewkowicz and Hansen-Tift (2012) found that 12-month-olds attend more to a talker's mouth when the talker speaks in a non-native language and Pons et al. (2015) found that bilingual 12-month-olds attend more to a talker's mouth than do monolingual 12-month-olds regardless of whether the AV speech is in their dominant language or in a non-native language. These findings indicate that infants deploy their selective attention to a talker's mouth as a function of language familiarity and developmental experience. If so, and given that new words are unfamiliar to preverbal infants, it is theoretically reasonable to expect that infants older than 12 months of age might redeploy their attention to a talker's mouth when they begin acquiring their first words.

Our general prediction leads to several specific predictions. First, we predicted that infants would redeploy their selective attention to a talker's mouth by 14 months of age—a time when the pace of word learning begins to pick up—and that they would continue to focus on the talker's mouth at 18 months of age when they find themselves in the midst of what has been called the lexical explosion (Fenson et al., 1994; Nelson, 1974). Although this prediction is partly supported by findings from a study by Chawarska, Macari, and Shic (2012) of infant selective attention during the second year of life, unfortunately these investigators reported no statistical comparison of the difference between looking at the eyes and mouth. Therefore, it is still not known whether infants attend more to a talker's mouth during the second year of life. Second, we predicted that infants would deploy greater attention to a talker's mouth when speaking in an ID manner than when speaking in an AD manner. This prediction is based on the fact that ID speech elicits more infant attention (Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990). In addition, ID speech is easier to process because it consists of shorter utterances, longer pauses, and better articulated sounds and because it has a higher overall pitch, a slower cadence, and contains exaggerated intonation contours (Fernald, 1985; Fernald & Kuhl,

1987; Liu, Kuhl, & Tsao, 2003; Trainor & Desjardins, 2002). Of particular importance with respect to word acquisition during infancy, ID speech also is known to facilitate infant word segmentation (Fernald & Mazzie, 1991; Thiessen, Hill, & Saffran, 2005), word recognition (Singh, Nestor, Parikh, & Yull, 2009), and word learning (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). Finally, we predicted that there may be a positive relation between attention to a talker's mouth at 18 months of age and vocabulary size. This prediction is based on previously reported findings of a link between attention to a talker's mouth at an earlier age and productive vocabulary at a later age (Chawarska et al., 2012; Tenenbaum et al., 2015; Young, Merin, Rogers, & Ozonoff, 2009). Here, we wondered whether there is also a contemporaneous relationship between attention to a talker's mouth and productive vocabulary.

To test our predictions, we tracked eye gaze in 14- and 18-month-old infants while they watched and heard a person talking. To determine whether early experience with a particular language and/or the manner in which a person talks affect where infants deploy their selective attention, the talker spoke in either a native language (English) or a non-native language (Spanish) and in either an ID or AD manner. Finally, to determine whether selective attention to the talker's mouth and productive vocabulary are correlated, we used the MacArthur–Bates Communicative Development Inventory (Fenson et al., 1994) to measure productive vocabulary.

Method

Participants

A total of 91 full-term healthy infants (62 boys), who had a birth weight ≥ 2500 g, an Apgar score ≥ 7 , a gestational age ≥ 37 weeks, no history of ear infections, and no motor, language, or other behavioral delays according to parental report, contributed data in the current study. All infants were raised in a mostly monolingual, English-speaking environment, defined as greater than 80% exposure to English according to parental report. The sample of 91 infants who contributed data in this study consisted of two distinct age groups: a group of 14-month-olds ($n = 47$; $M_{\text{age}} = 60.1$ weeks, $SD = 1.2$) and a group of 18-month-olds ($n = 44$; $M_{\text{age}} = 78.3$ weeks, $SD = 1.5$). We tested an additional 14 infants, but they did not contribute data because of equipment failure ($n = 4$), experimental error ($n = 1$), failure to look at the face for a minimum of 4 s ($n = 2$), or fussing/inattentiveness ($n = 7$).

Apparatus and stimuli

Participants were tested in a sound-attenuated and dimly illuminated room and were seated approximately 70 cm from a 19-in. computer monitor. Most of the infants sat in an infant seat, and those who refused sat in their parent's lap. We recorded point of gaze with an Applied Science Laboratories Eye-Trac Model 6000 eye-tracker (Bedford, MA, USA) operating at a sampling rate of 60 Hz. We used the corneal reflection technique and the participant's left eye to monitor pupil movements.

To permit comparisons with earlier studies, the stimulus materials used here were the same four multimedia movies presented in the Lewkowicz and Hansen-Tift (2012) study. During each of the movies, infants could see and hear the face of one of two monolingual female actors reciting a prepared monologue in her native language. The duration of each movie was 50 s. The actor spoke in English in two of the movies and in Spanish in the other two movies. In addition, the actor spoke in either an ID or AD manner in each language. When the actor spoke in an ID manner, she spoke in a prosodically exaggerated fashion that was characterized by a slow tempo, high pitch excursions, and continuous smiling. When the actor spoke in an AD fashion, she spoke in the way that adults usually speak to one another.

Procedure

Calibration was attempted first, and data were kept if an infant was successfully calibrated to at least five calibration points (the four corners and center of the monitor). During the calibration phase,

infants saw a looming/sounding round object sequentially pop up at nine locations determined by a 3×3 grid across the screen. If insufficient data were collected to complete the calibration, the missing calibration points were repeated up to three times. Once calibration was completed, the experiment continued with the presentation of one of the four movies (i.e., each infant was given a single 50-s test trial). The English-language movie was presented to 23 infants at 14 months of age and 21 infants at 18 months of age. The Spanish-language movie was presented to 24 infants at 14 months of age and 23 infants at 18 months of age. For each language, participants were assigned randomly to the ID or AD version of the monologue. Thus, language and manner of speech were between-participants variables.

The eye-tracking data were collected using GazeTracker software. Fixations were defined as looking at a circular area of 40 pixels in diameter for at least 100 ms. We created two principal areas of interest (AOIs) corresponding to the actor's eyes and mouth, as well as a face AOI (see Fig. 1). The eye AOI was defined by an area demarcated by two horizontal lines, one above the eyebrows and the other through the bridge of the nose, and two vertical lines, one at the edge of the actor's hairline on the left side of her face and the other at the edge of the actor's hairline on the right side of her face. The mouth AOI was defined by an area demarcated by two horizontal lines, one located between the bottom of the nose and the top lip and the other running through the center of the chin, and two vertical lines, each of which was located halfway between the right and left corners of the mouth and the edge of the face on each side. Each AOI was intentionally bigger than the eyes and mouth so as to allow for the slight head motions and mouth movements made by the actors when they talked, and each was created in accordance with the specifications reported in [Lewkowicz and Hansen-Tift \(2012\)](#). Finally, the face AOI encompassed the entire face.

To compare looking at the eyes with looking at the mouth, we computed proportion of total looking time (PTLT) scores for each of these two AOIs for each participant by dividing the amount of time they looked at each AOI by the total amount of time they looked at any portion of the face.

Although the principal aim of our study was to investigate the deployment of selective attention to the eyes and mouth of a talker producing AV speech, we also took the opportunity to determine whether 18-month-old infants' relative distribution of attention to the eyes and mouth might be associated with the development of expressive vocabulary. To this end, we asked the parents to complete the MacArthur–Bates Communicative Development Inventory (CDI) Toddler form Part IA (“Words Children Use”; 680 words) designed for use with children between 16 and 30 months of age.



Fig. 1. Still image of one of the faces presented in the study and an outline of the eye and mouth areas of interest.

Results

We analyzed the PTLT scores with a mixed, repeated-measures analysis of variance (ANOVA), with AOI (eyes or mouth) as a within-participant factor and age (14 or 18 months), language (English or Spanish), and manner of speech (ID or AD) as between-participants factors. Partial eta-squared values were calculated to determine effect size. According to Richardson (2011), partial eta-squared values of approximately .01, .06, and .14 indicate small, medium, and large effects, respectively. The ANOVA revealed that the main effect of language spoken was not significant, that the AOI \times Language interaction also was not significant, $F(1, 83) = 1.80, p = .18, \eta_p^2 = .02$, and that language did not interact with any of the other factors. The ANOVA yielded a significant main effect of AOI, $F(1, 83) = 51.71, p < .001, \eta_p^2 = .38$, which was attributable to greater overall looking at the mouth ($M = 37.7\%, SD = 22.5$) than at the eyes ($M = 16.0\%, SD = 14.9$). This main effect of AOI was, however, qualified by a significant AOI \times Manner-of-Speech interaction, $F(1, 83) = 5.68, p = .019, \eta_p^2 = .06$, which was due to greater looking to the mouth than to the eyes in the ID condition ($M_{\text{Mouth}} = 42.2\%, SD = 24.1, M_{\text{Eyes}} = 11.8\%, SD = 10.8$) than in the AD condition ($M_{\text{Mouth}} = 33.6\%, SD = 20.3, M_{\text{Eyes}} = 20.0\%, SD = 17.1$). Moreover, regardless of the differential looking across the two manner-of-speech conditions, planned comparisons showed that infants looked more at the mouth than at the eyes in both the ID speech condition, $F(1, 83) = 44.34, p < .001, \eta_p^2 = .35$, and the AD speech condition, $F(1, 83) = 11.96, p < .001, \eta_p^2 = .13$.

In addition to the significant effects described above, the overall ANOVA indicated that there was a significant AOI \times Manner-of-Speech \times Age interaction, $F(1, 83) = 6.06, p = .016, \eta_p^2 = .08$. This interaction can be seen in Fig. 2. To further investigate the source of this three-way interaction, we conducted planned comparison analyses of the PTLT scores. These analyses indicated that the three-way interaction was due to the fact that the two age groups distributed their looking to the mouth and eyes differently in the ID and AD conditions. The 14-month-old infants looked more at the mouth than at the eyes in the ID speech condition ($M_{\text{Mouth}} = 45.1\%, SD = 26.1, M_{\text{Eyes}} = 10.0\%, SD = 11.0$), $F(1, 83) = 27.66, p < .001, \eta_p^2 = .25$, but not in AD speech condition ($M_{\text{Mouth}} = 26.8\%, SD = 16.4\%, M_{\text{Eyes}} = 23.8\%, SD = 19.0$), $F(1, 83) < 1$. In contrast, the 18-month-olds looked more at the mouth than at the eyes in both the ID speech condition ($M_{\text{Mouth}} = 39.2\%, SD = 22.1, M_{\text{Eyes}} = 13.5\%, SD = 10.6$), $F(1, 83) = 17.21, p < .001, \eta_p^2 = .13$.

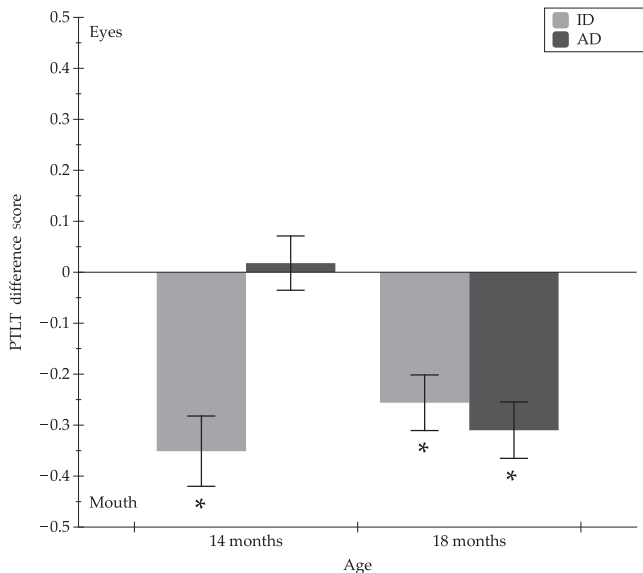


Fig. 2. Mean proportion of total looking time (PTLT) difference scores (PTLT to eyes minus PTLT to the mouth) as a function of age and manner of speech (collapsed over the two languages). Error bars represent standard errors of the mean, and asterisks indicate statistical significance. AD, adult-directed manner; ID, infant-directed manner.

.17, and the AD speech condition ($M_{\text{Mouth}} = 41.4\%$, $SD = 21.9\%$, $M_{\text{Eyes}} = 15.7\%$, $SD = 13.8\%$), $F(1, 83) = 17.71$, $p < .001$, $\eta_p^2 = .18$.

We also conducted an analysis to determine whether attention generally was affected by any of the factors of interest in the current study. To do so, we conducted a three-way ANOVA on the total amount of attention to the face, with age, language, and manner of speech as the three between-participants factors. This analysis indicated that the amount of attention devoted to the face differed across the two age groups ($M = 24.4$ s, $SD = 10$ for the 14-month-olds and $M = 19$ s, $SD = 10.6$ for the 18-month olds), $F(1, 83) = 5.55$, $p = .021$, $\eta_p^2 = .06$. Neither manner of speech nor language affected overall attention to the face, indicating that infants attended equally to the Spanish and English monologues at both ages and that they did so regardless of whether they were recited in an ID or AD manner.

Finally, we performed a chi-square test of independence to determine whether expressive vocabulary scores depended on infant distribution of attention to a talker's face. The analysis was conducted on the expressive vocabulary scores obtained by the 18-month-olds ($n = 41$). Here, 2 participants for whom parents did not complete the CDI questionnaire and 1 infant whose score was 3.6 standard deviations away from the sample mean were removed. Based on their CDI scores, we classified the remaining 18-month-olds into two expressive vocabulary groups: those who reached the 50-word level (high level) and those who did not (low level). We chose the 50-word cutoff point based on studies showing that a marked acceleration in vocabulary can occur when infants have acquired at least 50 words (Fenson et al., 1994; Nelson, 1981). Infants were also classified based on their relative attention to the eyes and mouth. For each infant, we computed an eyes–mouth index (EMI) by dividing the amount of gaze to the eyes by the total amount of time the infant looked to either the eyes or the mouth (i.e., $\text{Eyes}/[\text{Eyes} + \text{Mouth}]$). This measure has been previously used in this type of research (Merin et al., 2007; Tenenbaum et al., 2015; Young et al., 2009). An EMI $< .50$ means that the infant looked relatively more at the mouth, whereas an EMI $> .50$ means that the infant looked relatively more at the eyes. Fig. 3 shows the relationship between the number of words produced at 18 months of age and the EMI index. The chi-square test indicated that there was no statistically significant relationship between these two variables, $\chi^2(1, N = 41) = 1.42$, $p = .23$. The chi-square-based measure of association indicated a relatively small association between the two variables ($\phi = .19$).

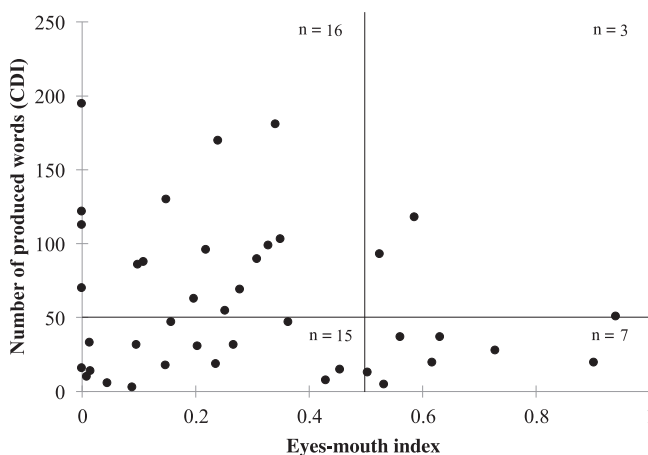


Fig. 3. Scatter plot showing the relationship between the number of words produced as measured by the Communicative Development Inventory (CDI) and the eyes–mouth index. The superimposed frames represent the four cells of the chi-square contingency table.

Discussion

The current study examined selective attention to talking faces in monolingual 14- and 18-month-old English-learning infants. Based on the fact that infants acquire their initial lexicon during the second year of life, and based on the fact that highly salient AV speech cues are located in a talker's mouth, we expected that infants at these two ages would deploy more of their selective attention to a talker's mouth than to the eyes. Our findings were consistent with our predictions. We found that infants attended more to the talker's mouth than to eyes, that they did so regardless of whether the talker spoke in a native or non-native language, and that they did so in response to ID speech at 14 months of age and in response to both ID and AD speech at 18 months.

The current findings are interesting in the context of several previous studies that have investigated developmental changes in monolingual and bilingual infants' selective attention to a talker's eyes and mouth during the first year of life. In two of these studies, monolingual English-learning infants' selective attention to a talker's face was measured while the talker could be seen and heard speaking in either a native or non-native language and in either an ID or AD manner (Hillairet de Boisferon et al., 2016; Lewkowicz & Hansen-Tift, 2012). In another one of these studies, monolingual and bilingual Catalan- and/or Spanish-learning infants' selective attention to a talker's face was investigated (Pons et al., 2015). The findings from the studies of monolingual English-learning infants indicated that, regardless of manner of speech, 4-month-olds attended more to the talker's eyes, 8- and 10-month-olds attended more to the talker's mouth regardless of language spoken, and that 12-month-olds attended equally to the talker's eyes and mouth when the talker spoke in the infants' native language but that they attended more to the talker's mouth when the talker spoke in a non-native language. The findings from the study examining monolingual and bilingual infants' selective attention to Catalan and/or Spanish showed that the monolingual infants exhibited the same attentional shifts as the English-learning infants in the Lewkowicz and Hansen-Tift (2012) study. Finally, these findings also showed that bilingual 4-month-olds attended equally to the talker's eyes and mouth, bilingual 8-month-olds attended more to the talker's mouth, and that bilingual 12-month-olds attended more to the talker's mouth regardless of whether the talker spoke in their dominant language or in a non-native language.

As noted by Lewkowicz and Hansen-Tift (2012), the initial attentional shift to a talker's mouth at 8 months of age is correlated with the emergence of two related behavioral skills. The first is endogenous selective attention, which for the first time enables infants to voluntarily deploy their attention to points of interest in their environment. The second is canonical babbling, which signals the onset of an explicit interest in speech. Given the emergence of these two behavioral skills, Lewkowicz and Hansen-Tift (2012) suggested that infants now begin taking advantage of the redundant, and thus highly salient, AV speech cues located in a talker's mouth to acquire both their native phonology and the motor templates for producing native speech sounds. Extending this interpretation to bilingual infants, Pons et al. (2015) suggested that the reason why bilingual infants deploy more attention to a talker's mouth at 12 months of age, and that they do so regardless of whether the talker speaks in their dominant language or a non-native language, is because they rely on the greater salience of redundant AV speech cues to identify distinct language-specific features and thus keep apart their two languages.

The current findings from 14- and 18-month-old infants extend the previous findings from younger infants' selective attention to talking faces. They indicate that once infants enter the word acquisition phase during the second year of life, they begin attending more to a talker's mouth again. Interestingly, like 8- and 10-month-olds who begin attending more to a talker's mouth regardless of language spoken, 14- and 18-month-olds also attend more to a talker's mouth regardless of language spoken. Of course, in the case of the older infants, the most likely reason for their greater deployment of attention to the talker's mouth is that now they are attempting to access redundant AV segmental cues that define both the words that constitute fluent speech utterances and the motor templates that define specific speech articulations. This conclusion is supported by findings that 18-month-olds' lexical representations are not just auditory but visual as well (Havy, Foroud, Fais, & Werker, 2017) and that visual speech cues can facilitate speech segmentation in adults (Lusk & Mitchel, 2016).

In addition to finding that 14- and 18-month-old infants attended more to a talker's mouth than to the eyes, we found that ID speech was generally more effective in eliciting attention to the talker's mouth. That is, we found that ID speech elicited more relative attention to the mouth than to the eyes (42% vs. 12%) than did AD speech (34% vs. 20%). We also found that 14-month-olds attended more to the mouth only when the talker spoke in ID speech but that 18-month-olds did so regardless of whether the talker spoke in ID or AD speech. The results from the younger infants are not surprising because studies suggest that ID speech is especially important when infants are acquiring new speech production skills. For example, the more parents use ID speech when they communicate with their 11- and 14-month-olds, the more their infants babble (Ramírez-Esparza, García-Sierra, & Kuhl, 2014). Presumably, at 14 months of age, infants rely on the perceptual attributes associated with ID speech to obtain the clearest information about the words that they are beginning to imitate. By this logic, the fact that the 18-month-olds attended more to the talker's mouth regardless of whether she spoke in an ID or AD manner suggests that, by this age, infants have become sufficiently expert at extracting relevant speech information and that they no longer need to rely exclusively on the greater salience of ID speech. This interpretation is consistent with findings that 17- to 20-month-olds depend less on the greater perceptual salience of ID speech because they are sensitive to the fine phonetic details that distinguish words (Werker & Fennell, 2004).

Finally, we failed to find a statistically significant link between infant selective attention to a talker's mouth and vocabulary size. This is interesting given that studies have found a link between speech segmentation (Singh, Reznick, & Xuehua, 2012), vowel discrimination (Tsao, Liu, & Kuhl, 2004), and matching of auditory and visual vowels (Altwater-Mackensen & Grossmann, 2016), respectively, and subsequent language development. Of course, we did not assess these sorts of perceptual abilities and subsequent language development. Therefore, the difference between the previous findings and ours is not surprising because the previous studies measured different underlying processes than we did. In addition, our failure to find a link between attention to the mouth and vocabulary size is interesting in the context of findings from studies that have found a link between infant attention to a talker's mouth at 6 or 12 months of age and children's vocabulary size between 18 and 24 months (Tenenbaum et al., 2015; Young et al., 2009). One possible reason for our failure to find such a link is that whereas the previous studies examined it over a relatively large age span, we assessed it contemporaneously and at an older age. This suggests that the relationship between attention to the mouth during early infancy and later vocabulary might be stronger than the relationship between these two variables at the same age during the second year of life.

Our failure to find a link between greater attention to a talker's mouth and vocabulary size suggests that infants do not coordinate the extraction of the redundant AV cues related to whole words and the acquisition of those words at the same age. This is probably because these two skills are just emerging and because infants have not yet mastered each skill sufficiently well to take advantage of the former in the service of the latter. This interpretation is based on the assumption that, by the second year of life, infants rely on attention to the mouth to detect cues that index whole words rather than phonemes and that they use it to extract segmentation cues whenever they attempt to extract the words and the motor articulation templates that are embedded in their interlocutors' fluent AV speech. It may be that it takes time before infants are able to take advantage of the information gained through access to the redundant AV word cues when learning new words. Therefore, it may be that the benefits of gaining access to redundant AV speech cues at 18 months of age might not be observed until later (e.g., at 24 months) simply because it takes this long for infants to master the first skill and use it in the service of the second skill.

One way of testing this interpretation is to test infants in a word-learning task while measuring selective attention to the mouth. If our interpretation is correct, then selective attention to the mouth should not be correlated with word learning at 18 months of age but it should be correlated with word learning at 24 months. Crucially, it should be noted that, regardless of whether our interpretation is correct with respect to the link between selective attention to the mouth and productive vocabulary at 18 months, this does not invalidate the previously reported findings of a link between attention to a talker's mouth in infants younger than 18 months and their productive vocabulary at later ages. In fact, when the previously reported positive link between attention to the mouth and subsequent vocabulary size is considered together with our failure to find such a

link, the most reasonable conclusion is that attention to the mouth may serve multiple functions during early development. As suggested earlier, during the first year of life, attention to the mouth is likely to facilitate phonological/phonetic tuning. In contrast, during the second year of life, attention to the mouth is likely to facilitate word learning. Crucially, it should be noted that no direct empirical support for either of these conclusions has been published to date and, therefore, that future studies should put them to test.

Despite the fact that the current study focused on infant selective attention to a talker's mouth, there is little doubt that a social partner's eyes are equally important as a source of information. For example, joint attention cues produced by adults during their social interactions with infants are especially important. These cues can facilitate language learning (Corkum & Moore, 1995) and can provide infants with contextual cues that can help them determine the relevant relationships between linguistic tokens and their referents and the meaning of others' language (Tomasello & Farrar, 1986). Indeed, several studies have shown that joint attention skills (i.e., the ability to follow the direction of a social partner's gaze) emerge gradually between 6 and 18 months of age and that they are a significant predictor of expressive and receptive vocabulary during the second year of life (Morales et al., 2000; Mundy & Gomes, 1998). These findings, together with our findings of greater attention to a talker's mouth at 14 and 18 months of age, raise interesting questions for future studies about the complex relationship between selective attention to joint attention cues in a talker's eyes and AV speech cues in a talker's mouth. Indeed, when our findings are compared with those from other studies, it becomes apparent that the types of cues available to infants in a particular study play an important role in selective attention. For example, in our study, the actor looked directly into the camera while reciting a monologue. In contrast, in other studies, either live actors have been involved or an actor was presented in a video in which the actor specifically provided joint-attention cues by looking to one side or the other (Tenenbaum et al., 2015). Given that the information conveyed by the eyes in our stimulus materials was minimized, infants' perception of the talker's communicative intentions was most likely significantly reduced. As a result, unlike in the Tenenbaum et al. (2015) study, where an actor explicitly provided joint attention cues, the actor did not provide such cues in our study.

Together, our findings and those from the Tenenbaum et al. (2015) study suggest that our failure to find a link between looking to the mouth and vocabulary size may be due, in part, to the fact that we did not provide joint attention cues. At the same time, however, there is no question that attention to a talker's mouth alone is involved in speech and language processing and that, in some cases, it actually facilitates processing. This is evident in findings from adult studies (Barenholtz et al., 2016; Rosenblum, 2008; Sumby & Pollack, 1954; Summerfield, 1979, 1992) and, as noted earlier, in findings from infant studies showing that infants' attention to a talker's mouth depends on their prior linguistic experience (i.e., either learning a single specific language or learning more than one language). Thus, it is likely that attention to a talker's mouth also contributes to word acquisition. This is especially likely because greater attention to a talker's mouth provides infants with the opportunity to associate articulatory motions with speech sounds and to imitate sounds and words. As noted earlier, however, it may be that the benefits of selective attention to redundant AV speech cues during the word-learning phase of language acquisition are not evident contemporaneously because infants need to master one skill to use it together with the other one.

In conclusion, the current findings indicate that monolingual English-learning infants redeploy their attention to a talker's mouth when they enter the word acquisition phase of language development. Our findings also suggest that selective attention to a talker's mouth during the second year of life enables infants to acquire their first words by providing them with access to maximally salient AV speech cues. Of course, our failure to find a contemporaneous link between attention to the mouth and vocabulary size requires caution with respect to this conclusion. To confirm it, future studies would need to demonstrate that this link holds between attention to the mouth during the first half of the second year of life and vocabulary size at a later age. Given that we did not investigate this link, and given that our sample size was relatively small, a test of this putative relationship awaits the results of future studies.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jecp.2018.03.009>.

References

- Altwater-Mackensen, N., & Grossmann, T. (2016). The role of left inferior frontal cortex during audiovisual speech perception in infants. *NeuroImage*, *133*, 14–20.
- Bahrack, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–206). Oxford, UK: Oxford University Press.
- Barenholtz, E., Lewkowicz, D. J., Davidson, M., & Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, *21*, 1345–1352.
- 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.
- Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry*, *53*, 903–913.
- Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Lawrence Erlbaum.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.
- Fenson, L., Dale, P., Reznick, S. J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5, Serial No. 242).
- Fernald, A. (1985). Five-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.
- Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese. *Infant Behavior and Development*, *10*, 279–293.
- Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*, 209–221.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 8030–8035.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523.
- Havy, M., Foroud, A., Fais, L., & Werker, J. F. (2017). The role of auditory and visual speech in word learning at 18 months and in adulthood. *Child Development*, *88*, 2043–2059.
- Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2016). Selective attention to the source of audiovisual speech in infancy: The role of temporal synchrony and linguistic experience. *Developmental Science*. Advance online publication. <http://doi.org/10.1111/desc.12381>.
- Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310.
- Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, *10*, 110–120.
- Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, *65*, 536–552.
- Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*, 292–315.
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*, 470–478.
- 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.
- Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*, F1–F10.
- Lusk, L. G., & Mitchel, A. D. (2016). Differential gaze patterns on eyes and mouth during audiovisual speech segmentation. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.00052>.
- Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*, 185–201.
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*, 154–178.
- Merin, N., Young, G. S., Ozonoff, S., & Rogers, S. J. (2007). Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants. *Journal of Autism and Developmental Disorders*, *37*, 108–121.
- Morales, M., Mundy, P., Delgado, C. E., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*, 283–298.
- Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*, 469–482.
- Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics, and perceptual functions of multisensory interactions in auditory cortex. *Hearing Research*, *258*, 72–79.
- Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, *81*, 267–285.
- Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology*, *17*, 170–187.

- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6318–6323.
- Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*, 539–545.
- Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*, 1272–1273.
- 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.
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*, 880–891.
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum.
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*, 135–147.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.
- Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*, 921–931.
- Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*, 411–417.
- Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*, 654–666.
- Singh, L., Reznick, S. J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*, 482–495.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.
- Summerfield, Q. (1979). Use of visual information in phonetic perception. *Phonetica*, *36*, 314–331.
- Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *335*, 71–78.
- 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.
- Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*, 1173–1190.
- Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148–160.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*, 53–71.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463.
- Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*, 335–340.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067–1084.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 1181–1186.
- Vö, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*. <https://doi.org/10.1167/12.13.3>.
- Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*(10), e326.
- Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, *21*, 8886–8894.
- Werker, J. F., & Fennell, C. T. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 79–109). Cambridge, MA: MIT Press.
- Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*, 798–814.