Infants deploy selective attention to the mouth of a talking face when learning speech

David J. Lewkowicz[1] and Amy M. Hansen-Tift

Department of Psychology, Florida Atlantic University, Boca Raton, FL 33431

Edited by Charles Gross, Princeton University, Princeton, NJ, and approved December 15, 2011 (received for review September 7, 2011)

The mechanisms underlying the acquisition of speech-production ability in human infancy are not well understood. We tracked 4–12-mo-old English-learning infants' and adults' eye gaze while they watched and listened to a female reciting a monologue either in their native (English) or nonnative (Spanish) language. We found that infants shifted their attention from the eyes to the mouth between 4 and 8 mo of age regardless of language and then began a shift back to the eyes at 12 mo in response to native but not nonnative speech. We posit that the first shift enables infants to gain access to redundant audiovisual speech cues that enable them to learn their native speech forms and that the second shift reflects growing native-language expertise that frees them to shift attention to the eyes to gain access to social cues. On this account, 12-mo-old infants do not shift attention to the eyes when exposed to nonnative speech because increasing native-language expertise and perceptual narrowing make it more difficult to process nonnative speech and require them to continue to access redundant audiovisual cues. Overall, the current findings demonstrate that the development of speech production capacity relies on changes in selective audiovisual attention and that this depends critically on early experience.

human infants | multisensory perception | speech acquisition | cognitive development

Although the development of speech perception during human infancy and the mechanisms underlying it are now relatively well understood (1, 2), the development of speech production is not as well understood (3). Despite this imbalance, it is clear that the emergence of speech production depends on infants' ambient linguistic environment, its structure, and the contingent nature of social interactions (1). This is evident in findings showing that 12–20-wk-old infants imitate simple vowels (4), 8–10-mo-old infants' babbling sounds reflect their specific linguistic environment (5), and that 9.5-mo-old infants learn new vocal forms (i.e., canonical syllables) from their mothers' contingent responses to their babbling sounds (3). What mechanisms might facilitate the acquisition of speech production capacity during infancy? One possible mechanism might be the deployment of selective attention to an interlocutor's vocal tract during social interactions. Such a mechanism can provide infants with direct access to the tightly coupled and highly redundant patterns of auditory and visual speech information (6–8) and enable them to profit from the fact that audiovisual speech is perceptually more salient than auditory-only speech (9–11).

When might infants first begin to focus their attention on a talker's mouth? Studies of selective attention have shown that between birth and 6 mo of age infants first attend more to the eyes and later less so but never more to the mouth. For example, even though newborns do not respond to faces as communicatively and socially meaningful objects (12), they look more at the eye region (13). Older, 3–11-wk-old, infants also look more at the eyes and, even when they hear the face talking, they still look 10 times longer at the eyes than at the mouth (14). Interestingly, by 6 mo of age, infants begin to increase their looking at the mouth regardless of whether they see (15) and/or hear the face talking (16) but, even then, they still do not look more at the mouth than at the eyes.

The findings from the selective attention studies suggest that if infants do begin to focus their attention on the talker's mouth, it must be later than 6 mo of age because it is then that infants first begin to engage in canonical babbling (17). Greater attention to a talker's mouth at this point would be advantageous from the standpoint of imitation. Second, attention to the mouth and its deformations during the canonical babbling stage can provide infants with the most direct access to the redundant audiovisual speech cues available there (6–8) and can permit them to profit from the increased salience that multisensory redundancy provides in general (18, 19) and the redundancy that audiovisual speech provides in particular (8–11). Third, after 6 mo of age infants become increasingly more motivated to produce vocalizations (20) and this, in turn, is likely to make them increasingly more interested in the source of audiovisual speech. Finally, endogenous attention begins to emerge after 6 mo of age (21, 22) and this enables infants to voluntarily direct their attention to the source of audiovisual speech for the first time.

If infants do make the predicted shift to the talker's mouth after 6 mo of age, then to profit maximally from the redundant audiovisual information available there, they should be able to perceive the multisensory coherence of such information. Studies have shown that infants can, in fact, perceive various types of intersensory relations (23, 24). For example, starting at birth infants can perceive low-level (i.e., intensity-based and synchrony-based) audiovisual relations (25, 26) and, as they grow, infants begin to exhibit increasingly more sophisticated multisensory perceptual abilities. These include the ability to detect the temporal coherence of faces and voices (27, 28), the ability to associate faces and voices (24, 28–31), and the ability to match and integrate the auditory and visual attributes of speech (32–35). Critically, the developmental emergence of multisensory perceptual abilities depends on early experience with coordinated audiovisual inputs. For example, developing animals and humans who are deprived of early visual input exhibit neural and behavioral deficits in multisensory integration (36, 37). Similarly, children with autism spectrum disorder who tend to pay less attention to people's faces during early development (38) exhibit audiovisual speech integration deficits later in childhood (39–41).

The effects of deprivation suggest that attention to the source of speech not only may be advantageous but, in fact, essential for the continuing acquisition of multisensory perceptual skills and speech production abilities in infancy. This may be especially important after 6 mo of age when speech perception and production begin to interact and when infants first begin to produce canonical babbling sounds that, for the first time, start to resemble speech-like sounds (17, 20). Two sets of findings support this conclusion. First, infants who look more at their talking

Author contributions: D.J.L. designed research; A.M.H.-T. performed research; D.J.L. and A.M.H.-T. analyzed data; and D.J.L. and A.M.H.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: lewkowic@fau.edu.

mother's mouth at 6 mo score higher on expressive language, size of vocabulary, and socialization measures at 24 mo of age (42). Second, adults look more at the mouth when they are exposed to audiovisual speech in noise (43) and when they see and hear an ambiguous soundtrack (44).

Although the initial predicted shift to the mouth after 6 mo of age is theoretically reasonable, the focus on the mouth is likely to begin diminishing a few months later. This is because infants acquire sufficient expertise in their native language by 12 mo of age (20, 34, 45) and, thus, by this age, they are less likely to require direct access to redundant audiovisual speech information. In addition, the eyes of social partners provide crucial social and deictic perceptual cues that are essential for further cognitive development (46) and, infants need to and do, in fact, begin to discover the value of such cues at around 12 mo of age (46, 47). Thus, we predicted that infants would begin to decrease their looking at the mouth around 12 mo of age. Critically, we also predicted that they would do so in response to native but not nonnative speech. This is because the ability to perceive the attributes of nonnative speech declines as infants become experts in their native language and as their sensitivity to nonnative speech narrows (2, 23). Once their sensitivity has narrowed, infants begin to find it harder to process what is now foreign speech to them and, as a result, are likely to continue to focus on the mouth of a person reciting nonnative speech so as to take full advantage of audiovisual redundancy and the greater salience that it offers.

To test our three predictions, we tracked eye gaze in monolingual, English-learning 4-, 6-, 8-, 10-, and 12-mo-old infants and adults while they watched videos of female talkers. In experiment 1, participants watched a video of a native (i.e., English) monologue, whereas in experiment 2, participants watched a video of a nonnative (i.e., Spanish) monologue.

Results

Experiment 1. All participants watched a video while a native English-speaking female was seen and heard reciting a prepared English monologue. She recited the monologue either in an infant-directed (ID) manner that consisted of the type of prosodically exaggerated speech that infants find particularly attractive (48), or in an adult-directed (AD) manner that is typical of normal adult speech. We manipulated the manner-of-speech variable to determine whether the greater prosody of ID speech might play a role in infant selective attention to audiovisual speech. To determine how much time participants spent gazing at the eyes and the mouth, respectively, we created two principal areas of interest (AOIs) on the face of the talker, one around the eyes and the other around the mouth, and monitored participants' point-of-gaze (POG) with an eye tracker.

We calculated the proportion-of-total-looking-time (PTLT) that participants spent looking at each AOI, respectively, by dividing the total amount of looking directed at each AOI by the total amount of looking at any portion of the face. A repeated-measures ANOVA, with AOI (eyes, mouth) as a repeated-measures factor and age (4, 6, 8, 10, and 12 mo and adults) and prosody (ID, AD) as between-subjects factors, yielded two significant findings. The first was a significant prosody × AOI interaction [$F(1, 98) = 5.86$, $P < 0.025$], which was attributable to greater overall looking at the mouth during ID speech and greater looking at the eyes during AD speech. Because the manner-of-speech variable did not interact with participants' age, it had no bearing on the interpretation of the second and principal finding. This finding indicated that there was a significant AOI × age interaction [$F(5, 98) = 9.09$, $P < 0.001$]. Fig. 1 depicts this interaction by showing PTLT difference scores (calculated by taking the difference between eye-PTLT scores and mouth-PTLT scores for each participant) as a function of age. As can be seen, an initial attentional shift occurred between

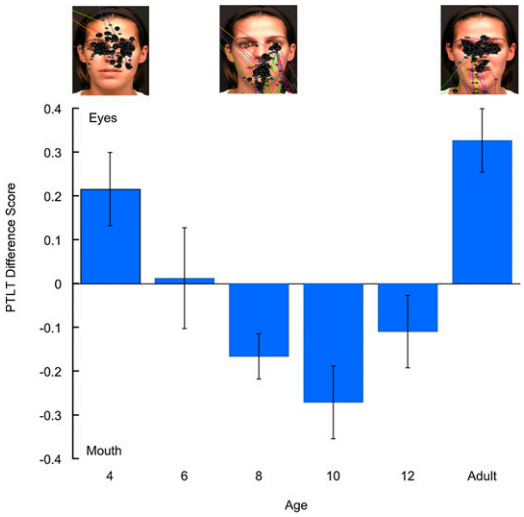

Fig. 1. PTLT difference scores as a function of age in response to the English monologue. Error bars represent SEMs. The screen shots above the graph are representative scan patterns of the talker's face in a 4- and an 8-mo-old infant and in an adult (each black dot represents a single visual fixation).

4 and 8 mo of age and a second attentional shift began between 10 and 12 mo of age and was completed by adulthood. We carried out planned comparison tests to determine at what age participants exhibited a significant preference for one of the two regions. Results indicated that 4-mo-old infants looked longer at the eyes [$F(1, 98) = 6.90$; $P < 0.025$], 6-mo-old infants looked equally at the eyes and the mouth [$F(1, 98) = 0.22$; not significant], 8- and 10-mo-old infants looked longer at the mouth [$F(1, 98) = 4.09$ and $P < 0.05$; $F(1, 98) = 11.22$ and $P < 0.01$, respectively], 12-mo-old infants looked equally at the eyes and mouth [$F(1, 98) = 2.09$; not significant], and that adults looked longer at the eyes [$F(1, 98) = 21.19$; $P < 0.001$].

Because our predictions regarding selective attention shifts were specifically concerned with developmental changes during infancy, we reanalyzed the data with the same ANOVA but without the adult data. Once again, we found a significant AOI × age interaction [$F(4, 79) = 5.18$; $P < 0.001$]. Planned comparison tests indicated that the 4-mo-olds looked longer at the eyes [$F(1, 79) = 6.58$; $P < 0.025$], the 6-mo-olds looked equally at the two regions [$F(1, 79) = 0.21$; not significant], the 8- and 10-mo-olds looked longer at the mouth [$F(1, 79) = 3.90$ and $P = 0.05$; $F(1, 79) = 10.69$ and $P < 0.01$, respectively], and that the 12-mo-olds looked equally at the eyes and mouth [$F(1, 79) = 1.99$; not significant].

Finally, to determine whether overall attention varied across age, we analyzed the total amount of looking at the face with a one-way ANOVA, with age and prosody as the two between-subjects factors. The age effect was not significant [$F(4, 79) = 0.34$; not significant] indicating that the pattern of shifting attention was not attributable to differences in overall attention. The prosody effect was significant [$F(1, 79) = 7.60$; $P < 0.01$], with infants looking more during ID speech (23.03 s) than during AD speech (17 s), indicating that ID speech was more salient overall.

In sum, when monolingual, English-learning infants were exposed to native audiovisual speech, they exhibited evidence of the two predicted attentional shifts between 4 and 12 mo of age. That is, between 4 and 8 mo of age, infants decreased their looking at the eyes from an average of 36% to an average of 17% of total looking time, whereas they increased their looking at the mouth from an average of 15% to an average of 36% of total looking time. This is consistent with our prediction that ac-

quisition of speech-production ability can profit from a shift in attention toward the source of audiovisual speech. In addition, between 10 and 12 mo of age infants decreased their looking at the mouth from an average of 38% to an average of 28% of total looking time (they looked at the eyes 19% of total looking time at 10 mo and 18% at 12 mo). This is consistent with our second prediction that the emergence of native-language expertise should reduce the need for direct access to redundant audiovisual speech information by the end of the first year of life.

Experiment 2. Prior studies have shown that young infants can perceive native as well as nonnative audible and visible speech attributes and that they can integrate them but that older infants can only perceive and integrate native ones (34, 45, 49). If the results from experiment 1 reflect infant processing of audiovisual speech per se then perceptual narrowing should affect the timing of the second attentional shift. Specifically, infants who have acquired expertise in their native language and, therefore, whose perceptual sensitivity has narrowed, should fail to exhibit the second attentional shift when exposed to nonnative speech. This is presumably because they now need access to the redundant audiovisual speech information to disambiguate what has by this time become unfamiliar speech. Thus, we expected that 12-mo-old monolingual, English-learning infants would continue to focus their attention on the mouth when presented with Spanish. To test this prediction, we used the identical procedures as in experiment 1 and tested separate and new groups of monolingual, English-learning 4-, 6-, 8-, 10-, and 12-mo-old infants and a new group of monolingual, English-speaking adults. This time, however, we presented a movie of a native Spanish speaker reciting a Spanish version of the monologue presented in experiment 1. Once again, the speaker recited the monologue either in the ID or AD style.

As Fig. 2 shows, we replicated the finding of the first shift in experiment 1 and, again, found that infants shifted their looking from the eyes to the mouth between 4 and 8 mo of age. In addition, and in keeping with our prediction that specific early linguistic experience is likely to affect the second attentional shift, we found that the 12-mo-old infants looked longer at the mouth. The mixed, repeated-measures ANOVA of the infant and adult PTLT scores indicated that the prosody × AOI interaction was marginally significant [$F(1, 97) = 3.41$; $P = 0.067$] and that the AOI × age interaction was significant [$F(5, 97) = 9.13$; $P < 0.001$]. Planned comparison tests showed that the 4-mo-old infants looked longer at the eyes [$F(1, 97) = 9.30$; $P < 0.01$], the 6-mo-old infants looked equally at the eyes and the mouth [$F(1, 97) = 1.17$; not significant], the 8-, 10-, and 12-mo-old infants looked longer at the mouth [$F(1, 97) = 5.43$ and $P < 0.025$; $F(1, 97) = 12.25$ and $P < 0.001$; $F(1, 97) = 16.05$ and $P < 0.001$, respectively] and that adults looked longer at the eyes [$F(1, 97) = 3.94$; $P < 0.05$].

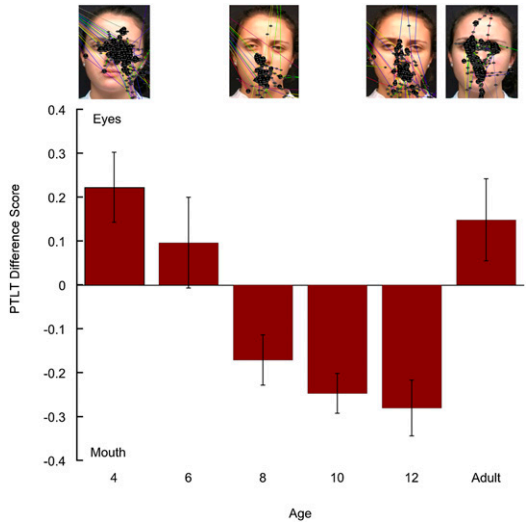

Fig. 2. PTLT difference scores as a function of age in response to the Spanish monologue. Error bars represent SEMs. The screen shots above the graph are representative scan patterns of the talker's face in a 4-, an 8-, and a 12-mo-old infant and in an adult.

The same ANOVA, but without the adult data, yielded a significant AOI × age interaction [$F(4, 80) = 10.89$; $P < 0.001$]. Planned comparisons indicated that the 4-mo-olds looked longer at the eyes [$F(1, 80) = 10.90$; $P < 0.01$], the 6-mo-olds looked equally at the eyes and mouth [$F(1, 80) = 1.37$; not significant], and that the 8-, 10-, and 12-mo-olds all looked longer at the mouth [$F(1, 80) = 6.36$ and $P = 0.025$; $F(1, 80) = 14.35$ and $P < 0.001$; and $F(1, 80) = 18.80$ and $P < 0.001$, respectively]. Finally, there was a marginal decrease in total duration of looking as a function of age, [$F(4, 80) = 2.14$; $P = 0.08$], with looking decreasing from an average of 23.4 s at 4 mo to 16.45 s at 12 mo (the average amount of looking collapsed over age was similar to that in experiment 1 and equaled 19.46 s). This decrease probably reflects the increasing effects of perceptual specialization for native speech and a consequent decline of interest in nonnative speech.

As in experiment 1, we found an attentional shift between 4 and 8 mo of age that consisted of decreased looking at the eyes (from an average of 39–18% of total looking time) and increased looking at the mouth (from an average of 17–36% of total looking time). In contrast to experiment 1, however, and in line with our predictions, we also found that 12-mo-old infants continued to look longer at the mouth (42% of total looking) than at the eyes (12% of total looking time). Thus, growing expertise for native speech as well as perceptual narrowing both lead infants to continue to focus their attention on the source of audiovisual speech, presumably to access the redundant speech signal. Given that the adults in this experiment still preferred the eyes, the 12-mo findings suggest that the second shift for nonnative speech begins later in development than the shift for native speech.

Arousing Effects of Non-Native Speech. Research with adults has shown that pupil dilation is affected by a person's state of arousal in response to visual (50) and auditory stimulation (51) and by the interest value of a stimulus (52). Infants also exhibit pupil dilation in response to novel stimulation (53). Thus, we hypothesized that infants might begin to exhibit greater pupil dilation in response to nonnative audiovisual speech but not to native speech just as they begin to acquire native-language expertise. To test this prediction, we extracted the average pupil size over the entire test session for each infant and then compared these data across age separately for the infants exposed to the English and the Spanish monologues, respectively, and separately for the eye and the mouth AOIs, respectively. For each of these analyses, we used a one-way ANOVA with prosody and age as the between-subjects factors.

We found no differences in average pupil size in responses to the English monologue but found that average pupil size increased as a function of age in response to the Spanish monologue. As Fig. 3 shows, this was the case, both when infants were looking at the mouth [$F(4, 80) = 3.87$; $P < 0.01$] and at the eyes [$F(4, 80) = 3.10$; $P < 0.025$] of the Spanish speaker (prosody did not have an effect). Moreover, the increase in average pupil size first occurred at 8 mo and remained at that level up through 12 mo. This increase corresponds with the initial attentional shift to the mouth found in experiment 2 and suggests that infants were reacting to the novel attributes of what by this age begins to be unfamiliar speech. Multisensory perceptual narrowing is known

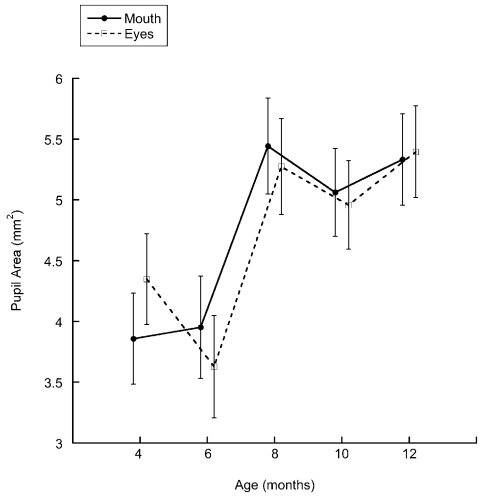

Fig. 3. Average pupil size during the test session in infants exposed to the Spanish monologue as a function of age.

to begin by around 8 mo of age (29). As a result, the pupil size data for the Spanish monologue probably reflect this narrowing and the increasingly greater novelty of the nonnative audiovisual input.

Importantly, the increase in average pupil size cannot be attributed to changes in luminance because all infants were tested under identical conditions and with the identical stimuli. Similarly, the monologue difference cannot be attributed to luminance differences because the movies of the English and Spanish monologue were produced under identical lighting conditions. Finally, the fact that the increase in pupil size occurred only when infants were exposed to nonnative audiovisual speech indicates that the attentional shift observed at 8 mo in both experiments cannot be ascribed to a generalized increase in arousal.

Discussion

Starting at 6 mo of age infants begin to produce simple speech-like syllables which eventually become incorporated and transformed into children's first meaningful communicative signals (17). Studies have shown that the development of speech-production capacity is facilitated by imitation, language-specific experience, and social contingency (3–5). Obviously, selective attention must mediate all of these effects because infants would not be able to imitate others' speech productions nor take advantage of their linguistic environment and the contingent nature of their caregivers' responses unless they paid close attention. The best place for infants to focus their attention is the orofacial cavity of their interlocutors because this is where they can gain direct access to redundant and highly salient audiovisual speech information that specifies the native-speech forms that they are trying to master. Based on this reasoning, we expected that infants would begin to attend to the mouth of a talker precisely when they begin to produce their first speech sounds at around 6 mo of age. Furthermore, we expected that once infants acquire native-language expertise by 12 mo of age they would no longer focus on the mouth when exposed to native speech but that they would continue to focus on the mouth when exposed to nonnative speech. To test these predictions, we presented videos of a female talker who could be seen and heard speaking either in the participants' native or nonnative language and recorded their eye gaze. Like in earlier studies (13–16), we found that 4-mo-old infants looked more at the eyes and that 6-mo-old infants looked equally at the eyes and mouth. As predicted, we also found that by 8 mo of age infants shifted their attention to the mouth of the talker regardless of whether the person was speaking in their native or nonnative language. Also, as predicted, we found that infants began to shift their attention away from the mouth by 12 mo of age when the talker spoke in the infants' native language. Finally, as predicted, we found that 12-mo-old infants, like 8- and 10-mo-old infants, continued to look more at the talker's mouth when they were exposed to nonnative speech.

There are four related reasons for the specific developmental timing of the initial shift in looking from the eyes to the mouth of a talker. First, starting at 6 mo, infants begin to produce canonical, speech-like syllables (17) and a focus on the mouth of a talker can help them learn how to produce such syllables through imitation. Second, the communicative signals that infants usually encounter in their daily lives are not isolated auditory speech sounds. Rather, they are concurrent auditory and visual speech signals that are correlated over time, redundant, and therefore perceptually more salient than auditory-only speech signals (6–11). This is especially true in en face interactions. Because of its increased salience, audiovisual speech and its source becomes especially attractive to infants when they first begin learning how to speak. Third, as infants begin to babble, their motivation to produce speech sounds increases (17, 20). Finally, endogenous attentional mechanisms begin emerging around 6 mo of age (21, 22), and this makes it possible for infants, for the first time, to voluntarily and flexibly control their focus of attention. As a result, they can now begin to direct their attention to the redundant audiovisual signals located in the mouth region. Overall, there are two consequences of the initial attentional shift: (*i*) it enables infants to look and listen while the orofacial cavity of their interlocutor produces the speech forms that they are now interested in reproducing, and (*ii*) it helps infants maximize the opportunity for learning these speech forms.

The finding that the initial eye-to-mouth attentional shift occurred during the same age range regardless of whether the speech presented was native or nonnative suggests that monolingual, English-learning infants do not have a preference for English over Spanish at 8 mo of age. This is consistent with reports that native-language expertise emerges gradually over the second half of the first year of life, with sensitivity to some phonetic categories, such as vowels, declining early (i.e., 6 mo) and sensitivity to other categories such as consonants declining a few months later (2, 23). Thus, the initial shift most likely reflects infants' attempt to extract the physical audiovisual attributes of the speech signal without regard to language identity. Nonetheless, even though infants at 8 mo may not yet prefer one language over the other, the pupil-size data suggest that once infants begin to focus on the mouth of the speaker, they begin to detect the native vs. nonnative speech difference.

The start of the second attentional shift at 12 mo in response to native audiovisual speech and its absence in response to nonnative speech are consistent with two related experience-driven developmental processes that represent two sides of the same coin: growth of perceptual expertise, on the one hand, and perceptual narrowing, on the other (20, 23, 29, 34, 45). The first process is the natural result of increasing experience with native auditory, visual, and audiovisual inputs, whereas the second is the result of the relative absence of nonnative perceptual inputs. Thus, on the one hand, the start of a shift away from the mouth and, as suggested by the adult data, in the direction of the eyes in 12-mo-old infants exposed to native audiovisual speech is consistent with the emergence of native-language expertise. This newly acquired expertise makes it possible for infants to begin producing their first meaningful words (54) and reduces their need to have direct access to the redundant audiovisual speech cues inherent in their interlocutors' orofacial cavity. As a result, they are now free to start exploring the various social and deictic cues available in their social partners' eyes that signal shared meanings, beliefs, and desires (46, 47). On the other hand, the

absence of the shift in response to nonnative speech is consistent with the process of perceptual narrowing. This process reduces infants' sensitivity to nonnative audiovisual speech and their ability to perceive it as a unified multisensory event (34). Because of this, when 12-mo-old infants are exposed to nonnative speech, they are now faced with the task of processing what has now become foreign speech. To overcome the greater difficulty of this perceptual task they, like younger infants as well as adults who find themselves in a difficult auditory speech perception task (43, 44), revert to a reliance on redundant audiovisual information to disambiguate an unclear speech signal.

It should be noted that the findings from the 12-mo-old infants' response to native speech only suggest the beginnings of a shift back to the eyes, not its completion. Nonetheless, there is little doubt that these findings are evidence of an eventual shift to the eyes. This is based on three types of evidence. First, the adults in experiment 1 looked more at the eyes indicating that a shift to the eyes occurs sometime between 12 mo and adulthood. Second, studies on the development of joint attention indicate that infants begin to notice their social partners' direction of eye gaze between 14 and 18 mo of age (46). Finally, and most relevant to the question of a possible attentional shift to the eyes, it has been found that 2-y-old typically developing children prefer to look at the eyes of talking faces (55). Thus, the most likely developmental scenario is that infants complete the shift back to the eyes by the second year of life.

Finally, the current findings have important implications for understanding the atypical development of infants at risk for autism spectrum disorder. Studies have found that 1-y-old children diagnosed with autism look less at faces (38) and that 2-y children diagnosed with autism look more at the mouth of a talker unlike typically developing children who look more at the eyes (55). These findings indicate that infants at risk for autism spectrum disorder follow a different developmental path. In light of the current study, the most likely developmental path for these infants might be that whenever they happen to encounter a person who is talking to them, they may look at the person's eyes when they are 4 mo of age and then, like typically developing infants, might discover the greater perceptual salience of the mouth of a talking face and shift their attention to the mouth by 8 mo of age. Alternatively, they may discover the mouth as a key source of audiovisual information sometime later (i.e., during the second year of life). In either case and in contrast to typically developing infants, however, infants at risk for autism may subsequently fail to shift their attention to the eyes for two principal reasons. First, they may be too captivated by the greater physical salience of the audiovisual stimulation inherent in talking faces and, second, because like older children with autism (39–41), they may fail to perceive the multisensory coherence of audiovisual speech. As a result, they are unable to extract the communicative meanings inherent in others' audiovisual speech and do not develop the perceptual expertise for native speech that can ultimately free them to shift their attention to the eyes to access the social cues that are so critical for subsequent cognitive development. Whether this developmental scenario is an accurate depiction of what occurs in infants at risk for autism, it is clear that an attentional shift to the eyes is critical. Because of this, it has been proposed that a failure to look at the eyes (i.e., greater looking at a talker's mouth) may be diagnostic of autism in early development (55). Although this may be the case for 24-mo-old children, the findings from the current study indicate that this cannot be the case for infants. Because 8–10-mo-old typically developing infants look more at a talker's mouth when she speaks in a native language and because 8–12-mo-old infants look more at her mouth when she speaks in a nonnative language, less looking at the eyes and greater looking at the mouth during this part of infancy is, if anything, diagnostic of typical development.

In conclusion, the findings from the current study show that selective attention to the mouth and eyes of a talking person undergoes dramatic changes during the second half of the first year of life and that the developmental timing of these changes is affected by specific early linguistic experience. The shift in attentional focus to the mouth at 8 mo of age corresponds with the emergence of speech production in infancy and indicates that the development of speech perception and production consists of several months of sustained attention to the source of audiovisual speech. Moreover, our findings indicate that the period between 8 and 12 mo of age is critical because typically developing infants rely on redundant audiovisual speech cues to acquire their speech production abilities during this time. Findings from studies of patients with congenital dense cataracts who have been deprived of early visual input have shown that these patients exhibit multisensory processing deficits (36), as well as audiovisual speech perception deficits (56). This suggests that for these patients, the audiovisual speech perception deficits are probably attributable to inadequate access to the redundant audiovisual signals during early life and is consistent with our findings. In contrast, however, the problem may be different for infants at risk for autism spectrum disorder despite the fact that they also exhibit audiovisual integration deficits later in life. Their problem may be that they have an intersensory integration problem to begin with, the etiology of which is currently indeterminate, and that this problem is not ameliorated by adequate access to audiovisual information during early life. Whether this conclusion is correct or not, it is clear that further studies of infants at risk for autism are needed. Nonetheless, it is clear that attention to a talker's mouth during the second half of the first year of life corresponds to the emergence of speech production ability during typical development and, thus, suggests that access to the redundant audiovisual cues available in the mouth is critical for normal development.

Materials and Methods

Participants. The sample consisted of a total of 179 full-term infants (birth weights, ≥2500 g; APGAR scores, ≥7; gestational age, ≥37 wk). Eighty-nine of these infants (45 boys) participated in experiment 1, and 90 participated in experiment 2 (51 boys). All infants were raised in a mostly monolingual environment. With the exception of three infants who heard English >80% of the time according to parental report, all remaining infants heard English >90% of the time. We tested an additional 76 infants but did not include them in the final sample because of their failure to complete the experiment because of fussiness or inattentiveness (33), failure to calibrate, which was either attributable to the infant being uncooperative or the eye tracker not being able to find the pupil (39), or equipment failure (4). The participants in experiment 1 consisted of separate groups of 4-mo-old (n = 19; mean age, 16.9 wk; SD = 0.58 wk), 6-mo-old (n = 16; mean age, 26.1 wk; SD = 0.7 wk), 8-mo-old (n = 17; mean age, 34.4 wk; SD = 0.6 wk), 10-mo-old (n = 17; mean age, 43 wk; SD = 0.6 wk), and 12-mo-old (n = 20; mean age, 52.2 wk; SD = 0.9 wk) infants, as well as 21 adults (8 males; mean age, 21 y; SD = 6.5 y). The participants in experiment 2 consisted of separate groups of 4-mo-old (n = 19; mean age, 16.8 wk; SD = 0.58 wk), 6-mo-old (n = 15; mean age, 26.4 wk; SD = 0.7 wk), 8-mo-old (n = 17; mean age, 34.2 wk; SD = 0.5 wk), 10-mo-old (n = 20; mean age, 43.1 wk; SD = 0.6 wk), and 12-mo-old (n = 19; mean age, 52.2 wk; SD = 0.8 wk) infants, as well as 19 adults (8 males; mean age, 20.5 y; SD = 5 y).

Apparatus and Stimuli. Participants were tested in a sound-attenuated and dimly illuminated room and were seated ~70 cm from a 19 inch computer monitor. Most of the infants were seated in an infant seat and those who refused sat in their parent's lap. We calibrated eye gaze using a 3 × 3 grid of small looming/sounding dots equidistant from each other and referenced to each of the four corners of the monitor. Participants watched a single 50 s multimedia movie of one of two female actors, each of whom was a native speaker in her respective language, reciting a prepared monologue. One of the actors recited an English version of the monologue, whereas the other recited the Spanish version of the monologue, and each produced two prosodically different versions. One version was recited in a prosodically exaggerated manner (ID speech) and was characterized by slow tempo, high

pitch excursions, and continuous smiling. The other version was recited in a prosodically neutral manner (AD speech). Point of gaze was recorded by monitoring pupil movements with an ASL (Applied Science Laboratories) Eye-trac Model 6000 eye-tracker (sampling rate: 60 Hz) using the corneal reflection technique and using the participant's left eye.

Procedure. Participants in each group were assigned randomly to an ID or an AD movie. Calibration was attempted first and data were kept if an infant was successfully calibrated to five or more calibration points. Once calibration was completed, the test movie started. To determine the amount of time participants spent fixating the eyes and mouth, respectively, we defined two AOIs. 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 boundaries, 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 boundaries each of which was located halfway between the right and left corners of the mouth and the edge of the face on each side, respectively.

ACKNOWLEDGMENTS. We thank Kelly McMullan and Nicholas Minar for their assistance and Asif A. Ghazanfar for helpful suggestions. This work was supported by National Science Foundation Grant BCS-0751888 and Grant R01HD057116 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (to D.J.L.).

1. Kuhl PK (2007) Is speech learning 'gated' by the social brain? *Dev Sci* 10:110–120.
2. Werker JF, Tees RC (2005) Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev Psychobiol* 46:233–251.
3. Goldstein MH, Schwade JA (2008) Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol Sci* 19:515–523.
4. Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. *J Acoust Soc Am* 100:2425–2438.
5. De Boysson-Bardies B, Hallé P, Sagart L, Durand C (1989) A crosslinguistic investigation of vowel formants in babbling. *J Child Lang* 16:1–17.
6. Yehia H, Rubin P, Vatikiotis-Bateson E (1998) Quantitative association of vocal-tract and facial behavior. *Speech Commun* 26:23–43.
7. Munhall KG, Vatikiotis-Bateson E (2004) Spatial and temporal constraints on audiovisual speech perception. *The Handbook of Multisensory Processes*, eds Calvert GA, Spence C, Stein BE (MIT Press, Cambridge, MA), pp 177–188.
8. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLOS Comput Biol* 5(7):e1000436.
9. Rosenblum LD, Johnson JA, Saldaña HM (1996) Point-light facial displays enhance comprehension of speech in noise. *J Speech Hear Res* 39:1159–1170.
10. Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
11. Summerfield Q (1979) Use of visual information for phonetic perception. *Phonetica* 36:314–331.
12. Simion F, Leo I, Turati C, Valenza E, Dalla Barba B (2007) How face specialization emerges in the first months of life. *Prog Brain Res* 164:169–185.
13. Cassia VM, Turati C, Simion F (2004) Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychol Sci* 15:379–383.
14. Haith MM, Bergman T, Moore MJ (1977) Eye contact and face scanning in early infancy. *Science* 198:853–855.
15. Hunnius S, Geuze RH (2004) Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy* 6:231–255.
16. Merin N, Young GS, Ozonoff S, Rogers SJ (2007) Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison Infants. *J Autism Dev Disord* 37:108–121.
17. Oller DK (2000) *The Emergence of the Speech Capacity* (Lawrence Erlbaum, Mahwah, NJ).
18. Lewkowicz DJ, Kraebel K (2004) The value of multimodal redundancy in the development of intersensory perception. *Handbook of Multisensory Processing*, eds Calvert G, Spence C, Stein B (MIT Press, Cambridge), pp 655–678.
19. Bahrick LE, Lickliter R, Flom R (2004) Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr Dir Psychol Sci* 13:99–102.
20. Jusczyk PW (1997) *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).
21. Colombo J (2001) The development of visual attention in infancy. *Annu Rev Psychol* 52:337–367.
22. Richards JE, Reynolds GD, Courage ML (2010) The neural bases of infant attention. *Curr Dir Psychol Sci* 19:41–46.
23. Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cogn Sci* 13:470–478.
24. Lewkowicz DJ (2000) The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychol Bull* 126:281–308.
25. Lewkowicz DJ, Leo I, Simion F (2010) Intersensory perception at birth: Newborns match non-human primate faces & voices. *Infancy* 15:46–60.
26. Lewkowicz DJ, Turkewitz G (1980) Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Dev Psychol* 16:597–607.
27. Lewkowicz DJ (2000) Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Dev* 71:1241–1257.
28. Lewkowicz DJ (2010) Infant perception of audio-visual speech synchrony. *Dev Psychol* 46:66–77.
29. Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci USA* 103:6771–6774.
30. Brookes H, et al. (2001) Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant Child Dev* 10:75–82.
31. Vouloumanos A, Druhen M, Hauser M, Huizink A (2009) Five-month-old infants' identification of the sources of vocalizations. *Proc Natl Acad Sci USA* 106: 18867–18872.
32. Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. *Science* 218:1138–1141.
33. Patterson ML, Werker JF (2003) Two-month-old infants match phonetic information in lips and voice. *Dev Sci* 6:191–196.
34. Pons F, Lewkowicz DJ, Soto-Faraco S, Sebastián-Gallés N (2009) Narrowing of intersensory speech perception in infancy. *Proc Natl Acad Sci USA* 106:10598–10602.
35. Rosenblum LD, Schmuckler MA, Johnson JA (1997) The McGurk effect in infants. *Percept Psychophys* 59:347–357.
36. Putzar L, Goerendt I, Lange K, Rösler F, Röder B (2007) Early visual deprivation impairs multisensory interactions in humans. *Nat Neurosci* 10:1243–1245.
37. Wallace MT, Perrault TJ, Jr., Hairston WD, Stein BE (2004) Visual experience is necessary for the development of multisensory integration. *J Neurosci* 24:9580–9584.
38. Osterling JA, Dawson G, Munson JA (2002) Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Dev Psychopathol* 14: 239–251.
39. Irwin JR, Tornatore LA, Brancazio L, Whalen DH (2011) Can children with autism spectrum disorders "hear" a speaking face? *Child Dev* 82:1397–1403.
40. Smith EG, Bennetto L (2007) Audiovisual speech integration and lipreading in autism. *J Child Psychol Psychiatry* 48:813–821.
41. Bebko JM, Weiss JA, Demark JL, Gomez P (2006) Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *J Child Psychol Psychiatry* 47:88–98.
42. Young GS, Merin N, Rogers SJ, 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. *Dev Sci* 12:798–814.
43. Vatikiotis-Bateson E, Eigsti I-M, Yano S, Munhall KG (1998) Eye movement of perceivers during audiovisual speech perception. *Percept Psychophys* 60:926–940.
44. Lansing CR, McConkie GW (2003) Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Percept Psychophys* 65:536–552.
45. Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7:49–63.
46. Langton SRH, Watt RJ, Bruce I, I (2000) Do the eyes have it? Cues to the direction of social attention. *Trends Cogn Sci* 4:50–59.
47. Moore C, Corkum V (1998) Infant gaze following based on eye direction. *Br J Dev Psychol* 16:495–503.
48. Fernald A, Simon T (1984) Expanded intonation contours in mothers' speech to newborns. *Dev Psychol* 20:104–113.
49. Weikum WM, et al. (2007) Visual language discrimination in infancy. *Science* 316: 1159.
50. Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45:602–607.
51. Partala T, Jokiniemi M, Surakka V (2000) *Pupillary Responses to Emotionally Provocative Stimuli* (ACM Press, New York), pp 123–129.
52. Libby WL, Jr., Lacey BC, Lacey JI (1973) Pupillary and cardiac activity during visual attention. *Psychophysiology* 10:270–294.
53. Gredebäck G, Melinder A (2010) Infants' understanding of everyday social interactions: A dual process account. *Cognition* 114:197–206.
54. Fenson L, et al. (1994) Variability in early communicative development. *Monogr Soc Res Child Dev* 59:1–173.
55. Jones W, Carr K, Klin A (2008) Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch Gen Psychiatry* 65:946–954.
56. Putzar L, Hötting K, Röder B (2010) Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restor Neurol Neurosci* 28: 251–257.