Developmental Science

Developmental Science (2016), pp 1–12 DOI: 10.1111/desc.12381

Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience

Anne Hillairet de Boisferon,[1] Amy H. Tift,[2] Nicholas J. Minar[2] and David J. Lewkowicz[1]

1. *Department of Communication Sciences and Disorders, Northeastern University, USA*
2. *Department of Psychology, Florida Atlantic University, USA*

Abstract

Previous studies have found that infants shift their attention from the eyes to the mouth of a talker when they enter the canonical babbling phase after 6 months of age. Here, we investigated whether this increased attentional focus on the mouth is mediated by audio-visual synchrony and linguistic experience. To do so, we tracked eye gaze in 4-, 6-, 8-, 10-, and 12-month-old infants while they were exposed either to desynchronized native or desynchronized non-native audiovisual fluent speech. Results indicated that, regardless of language, desynchronization disrupted the usual pattern of relative attention to the eyes and mouth found in response to synchronized speech at 10 months but not at any other age. These findings show that audio-visual synchrony mediates selective attention to a talker's mouth just prior to the emergence of initial language expertise and that it declines in importance once infants become native-language experts.

Research highlights

- After 6 months of age, when infants enter the canonical babbling phase, they shift their attention from a talker's eyes to a talker's mouth, presumably to benefit from the redundant and highly salient audiovisual speech cues located there.
- We investigated whether attention to the talker's mouth during the babbling phase is due to the redundant nature of synchronous audiovisual speech and to linguistic experience.
- We tracked gaze behavior in groups of 4-, 6-, 8-, 10-, and 12-month-old infants while they watched videos of desynchronized native and non-native audiovisual speech.
- Regardless of language, desynchronization eliminated the preference for the mouth found in response to synchronized speech at 10 months of age.

These results demonstrate that the temporal coherence of fluent audiovisual speech mediates infants' preference for a talker's mouth at a point when native-language expertise is emerging and that its importance declines shortly after native-language expertise emerges.

Introduction

During most social interactions, infants hear and see their interlocutors talking to them and, as a result, they experience audiovisual rather than auditory speech. In general, audiovisual speech is more salient because it consists of overlapping and highly redundant auditory and visual information (Chandrasekaran, Trubanova, Stillittano, Caplier & Ghazanfar, 2009; Munhall & Vatikiotis-Bateson, 2004; Sumby & Pollack, 1954; Summerfield, 1979; Yehia, Rubin & Vatikiotis-Bateson, 1998). Adults are known to benefit from the redundancy inherent in audiovisual speech by automatically integrating it (McGurk & MacDonald, 1976; Rosenblum, 2008). Of course, to benefit from audiovisual speech redundancy, one needs to direct one's attention to a social partner's face and, especially, to the mouth where concurrent auditory and visual cues can be accessed most directly. Indeed, adults typically do direct their attention to a talker's mouth when they are exposed to talking faces (Lansing & McConkie, 2003; Võ, Smith, Mital & Henderson, 2012; Barenholtz, Mavica, & Lewkowicz, 2016). When they do, they not only automatically benefit from the greater perceptual salience of

Address for correspondence: David J. Lewkowicz, Department of Communication Sciences & Disorders, Northeastern University, 360 Huntington Avenue 503BK, Boston, MA 02115, USA; e-mail: d.lewkowicz@neu.edu

audiovisual speech but also from the specialized neural mechanisms that facilitate the processing of multisensory as opposed to unisensory signals (Schroeder, Lakatos, Kajikawa, Partan & Puce, 2008; van Wassenhove, Grant & Poeppel, 2005).

Given the processing advantage that audiovisual speech offers over auditory speech for experienced perceivers, it is likely that it plays an important role in the acquisition of speech and language during infancy. Of course, this prediction requires evidence that infants actually attend to the source of audiovisual speech when exposed to an interlocutor. Indeed, two recent studies have provided such evidence. The first of these studies (Lewkowicz & Hansen-Tift, 2012) presented either native or non-native audiovisual speech (i.e. a video of a talker) to 4-, 6-, 8-, 10-, and 12-month-old monolingual infants and investigated whether they attended selectively to the talker's eyes or mouth. Results indicated that 4-month-olds attended more to the eyes, that 6-month-olds attended equally to the eyes and mouth, that 8- and 10-month-olds attended more to the mouth, and that 12-month-olds no longer attended more to the mouth when exposed to native audiovisual speech but that they continued to do so when exposed to non-native speech. The second of these studies (Pons, Bosch & Lewkowicz, 2015) replicated the initial findings and, in addition, showed that bilingual infants deployed more of their attention to a talker's mouth than did monolingual infants. Together, these findings demonstrate that once infants reach the canonical babbling stage when they become more interested in speech production, they begin directing their attention to the redundant audiovisual speech cues located in a talker's mouth. Lewkowicz and Hansen-Tift (2012) suggested that the greater focus on a talker's mouth facilitates speech and language acquisition because it enables infants to gain direct access to the most salient attributes of the speech signal. The findings from Pons *et al.* (2015) support this conclusion by showing that bilingual infants rely even more on audiovisual speech redundancy than do monolingual infants and suggest that bilinguals rely on it to overcome the challenge of acquiring two languages.

Here, we asked the following question: What specific redundancy cues might help focus infants' attention on a talker's mouth? The answer to this question requires recognition of the complexity of audiovisual speech as well as infants' limited capacity to process such speech because of their neural immaturity and relative lack of perceptual experience. Everyday fluent audiovisual speech is specified by a hierarchy of increasingly more complex perceptual cues (Lewkowicz & Ghazanfar, 2009). At the lowest level of the hierarchy, audiovisual speech is specified by the concurrent onsets and offsets of its audible and visible attributes (i.e. their temporal synchrony) and by their equivalent dynamic variations in intensity. At the next level of the hierarchy, audiovisual speech is specified by several intersensory equivalence cues that specify the temporal dynamics of vocal tract action, including equivalent audible and visible duration, tempo, and prosody. Finally, at the highest level of the hierarchy, audiovisual speech is specified by various categorical amodal attributes such as a talker's gender, affect, and identity.

Of course, if infants are to benefit from the various forms of audiovisual redundancy, they must be able to perceive multisensory coherence. Indeed, evidence indicates that infants become capable of perceiving multisensory coherence during the first year of life. In general, they begin life by detecting low-level multisensory coherence cues and gradually begin to detect increasingly more complex ones as they grow and as they acquire perceptual experience. For example, at birth, infants exhibit the ability to perceive amodal intensity cues (Lewkowicz & Turkewitz, 1980) as well as temporal synchrony cues (Lewkowicz, Leo & Simion, 2010). The most likely reason why they detect audio-visual (A-V) temporal synchrony cues early in life is because these cues are relatively simple and easy to detect. Nonetheless, A-V synchrony cues are very powerful because they can scaffold the perception of multisensory coherence, regardless of the specific nature of the information. Indeed, studies have found that young infants are sensitive to the temporal synchrony of many different types of auditory and visual stimuli including flashing lights and beeping sounds, moving and sounding objects, and auditory and visual speech attributes (Bahrick, 1983, 1988; Dodd, 1979; Lewkowicz, 1986, 1992a, 1992b, 1996, 2000b, 2003, 2010; Morrongiello, Fenwick & Nutley, 1998; Scheier, Lewkowicz & Shimojo, 2003). Together, these findings provide empirical support for the theoretical view that once infants can detect the co-occurrence of auditory and visual information, they can proceed to learning about the unity of their multisensory world at a more complex level of specificity.

Despite the perceptual power that A-V synchrony cues provide to a developing, immature, and relatively inexperienced organism, they are clearly not sufficient to learn about the complex nature of the multisensory world. Obviously, infants must begin to discover the higher-level multisensory coherence cues fairly quickly if they are to learn about their world. Indeed, by 2 months of age infants already exhibit the ability to perceive the amodal character of the audible and visible attributes of isolated phonemes even in the absence of synchrony cues (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003) and by 5 months they begin to automatically integrate audible and visible speech syllables (Rosenblum, Schmuckler & Johnson, 1997). By 7 to 8 months of age they

begin to perceive amodal affect (Walker-Andrews, 1986) and gender (Patterson & Werker, 2002; Walker-Andrews, Bahrick, Raglioni & Diaz, 1991), and by around 12 months of age they begin to perceive the amodal character of fluent audiovisual speech (Lewkowicz, Minar, Tift & Brandon, 2015) and amodal language identity (Lewkowicz & Pons, 2013).

Given that A-V temporal synchrony is important during infancy and, given that early linguistic experience affects responsiveness to synchronized auditory and visual speech (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014; Werker & Tees, 2005), here we investigated the role that these two factors play in infants' selective attention to the eyes and mouth of a talking face. We began by making two related theoretical assumptions. First, we assumed that the shift to the talker's mouth observed by 8 months of age may be mediated specifically by the temporally synchronous nature of typical audiovisual speech because this makes such speech highly salient. Second, we assumed that as infants grow and acquire perceptual experience, they begin to discover the other, higher-level, properties of audiovisual speech, and that because of this, they may cease relying as much on A-V temporal synchrony cues for processing native audiovisual speech but that they may continue relying on them for the processing of non-native audiovisual speech. The latter is likely because perceptual narrowing renders non-native audiovisual speech relatively unfamiliar and presumably leads infants to deploy more attentional resources to the visible and audible streams of information in an attempt to disambiguate it.

To investigate the role of A-V synchrony and linguistic experience in selective attention to audiovisual speech, we presented videos of talkers producing desynchronized audiovisual speech utterances to infants ranging in age from 4 to 12 months of age. The utterances were in the infants' native language (Experiment 1) or in a non-native language (Experiment 2). During the presentation of the utterances, we tracked infants' eye gaze to determine how much attention they deployed to the talkers' eyes and mouth. We then compared the findings from the current study with the findings from the study by Lewkowicz and Hansen-Tift (2012) in which infants were exposed to synchronized audiovisual speech. This enabled us to determine whether A-V synchrony and early experience affect infant selective attention to different parts of talking faces.

Experiment 1

In this experiment, we investigated whether, and at what age, synchrony-based redundancy might be involved in attracting attention to the mouth of a talker speaking in the infants' native language. To do so, we tracked eye gaze in separate groups of 4-, 6-, 8-, 10-, and 12-month-old monolingual, English-learning infants while they saw and heard desynchronized streams of audible and visible English speech.

Method

Participants

In all, 93 infants (54 boys) contributed data in this experiment. All infants were full-term, healthy, and had no history of ear infections according to parents' report (birth weight, ≥ 2500 g; APGAR score, ≥ 7; gestational age, ≥ 37 weeks). All infants were raised in a mostly monolingual, English-speaking, environment, defined as greater than 80% exposure to English according to parental report. An additional 33 infants were tested but excluded for failure to complete the experiment because of fussiness or inattentiveness (10), failure to calibrate either because the infant was uncooperative or the eye tracker could not find the pupil (15), equipment failure (6), experimental error (1), or parent interference (1).

The participants consisted of separate groups of 4-month-olds (n =20; mean age, 17.1 weeks; SD = 0.7 week), 6-month-olds (n = 20; mean age, 26 weeks; SD = 0.6 week), 8-month-olds (n = 16; mean age, 34.1 weeks; SD = 0.8 week), 10-month-olds (n = 19; mean age, 43.2 weeks; SD = 0.5 week), and 12-month-olds (n = 18; mean age, 52.2 weeks; SD = 0.6 week) infants.

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 recorded point of gaze with an Applied Science Laboratories Eye-trac Model 6000 eye-tracker operating at a sampling rate of 60 Hz. We used the corneal reflection technique and the participant's left eye to monitor the infants' pupil movements.

The stimulus materials consisted of the same two multimedia movies presented in Experiment 1 in the Lewkowicz and Hansen-Tift (2012) study except that here the auditory and visual speech streams were desynchronized and that the movies consisted of the first 30 s of the original movies. During each of the two movies, infants could see the face of a monolingual actor and hear her reciting a prepared monologue in her native English. In one version of the movie, the actor spoke in an infant-directed (ID) fashion (i.e. in a prosodically

exaggerated manner with a slow tempo, high pitch excursions, and continuous smiling) while in the other movie she spoke in an adult-directed (AD) fashion (the way adults usually speak to one another). To desynchronize the audiovisual speech presented here, we moved the auditory speech stream ahead of the visual stream by 666 ms. Prior studies have shown that this degree of temporal A-V desynchronization is discriminable by infants as young as 4 months of age (Lewkowicz, 2010; Pons & Lewkowicz, 2014). In addition, we moved the audible speech stream ahead of the visual one – as opposed to the reverse – to minimize the predictive visual cues that are normally available in everyday audiovisual speech where mouth movements usually precede phonation.

Procedure

Calibration was attempted first and data were kept if an infant was successfully calibrated to at least five calibration points (this included the four corners and the 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, participants were presented with a single 30 s movie of the female actor and data were kept if infants accumulated a minimum of 4 seconds of looking. No infants were excluded in this experiment based on this criterion. Participants were assigned randomly to the ID or AD version of the monologue.

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 50 ms. We created two areas of interest (AOIs) corresponding to the actor's eyes and mouth, respectively. 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 each corner of the mouth and the edge of the face on that side. Each AOI was intentionally bigger than the eyes and mouth, respectively, so as to allow for the slight head and mouth movements made by the actors when they talked.

The dependent measure was the amount of time participants looked at each AOI. To compare looking at the eyes and mouth, we computed proportion-of-total-looking-time (PTLT) scores for each of the two AOIs for each participant by dividing the amount of time they looked at each AOI, respectively, by the total amount of time they looked at the face.

Results

To determine whether responsiveness differed as a function of prosody and/or across age, we analyzed the PTLT scores with a mixed, repeated-measures analysis of variance (ANOVA), with AOI (eyes, mouth) as a within-subjects factor and age (4, 6, 8, 10, and 12 months) and prosody (ID, AD) as between-subjects factors. The ANOVA revealed a significant prosody × AOI interaction [$F(1, 83) = 4.35$, $p = .040$, $\eta_p^2 = .05$], which was attributable to greater overall looking at the mouth (M = 30.1%, SD = 22.8%) than the eyes (M = 19.3%, SD = 19.4%) during ID speech. Critically, the prosody factor did not interact with the participants' age [$F(4, 83) < 1$, *ns*, $\eta_p^2 = .035$] and, thus, had no bearing on the principal age-based hypothesis under test. Finally, and as expected, we found that there was a significant AOI × Age interaction [$F(4, 83) = 3.75$, $p = .007$, $\eta_p^2 = .15$].

To further investigate the source of the AOI × Age interaction, we conducted planned comparison analyses of the PTLT scores at each age, respectively. These analyses indicated that neither the 4- nor the 6-month-old infants exhibited significant differences in the amount of looking directed at the eyes and mouth [$F(1, 83) = 1.29$, $p = .26$, $\omega^2_{(\psi)}$[1] = .007 and $F(1, 83) = 2.75$, $p = .10$, $\omega^2_{(\psi)} = .042$, respectively], that the 8-month-old infants looked longer at the mouth [$F(1, 83) = 8.54$, $p = .004$, $\omega^2_{(\psi)} = .19$], and that neither the 10- nor the 12-month-old infants exhibited significant differences in the amount of looking directed at the eyes and mouth [$F(1, 83) = 0.51$, $p = .48$, $\omega^2_{(\psi)} = .013$ and $F(1, 83) = 2.84$, $p = .096$, $\omega^2_{(\psi)} = .049$, respectively]. The PTLT scores for each AOI at each age can be seen in Table 1, and Figure 1 shows the same data but in terms of mean PTLT difference scores which were computed by subtracting the mouth-PTLT score from the eye-PTLT score for each participant, respectively, and then by computing the average PTLT difference score for each age group, respectively. Positive scores indicate greater looking at

[1] See Keppel and Wickens (2004) for this effect-size measure for planned comparisons.

the eyes while negative scores indicate greater looking at the mouth.

To ensure that overall attention did not vary across age and, thus, that our principal results were not affected by this factor, we analyzed the total amount of looking at the face with a two-way ANOVA with age and prosody as the two between-subjects factors. This analysis showed that the age effect was not significant [$F(4, 83) = 1.35$, $p = .26$, $\eta_p^2 = .061$], indicating that the different patterns of attention found at the different ages were not attributable to differences in overall attention. In addition, this same analysis showed that the prosody effect was not significant [$F(1, 83) < 1$, *ns*, $\eta_p^2 < .001$], indicating that the infants attended equally to ID and AD audiovisual speech across the different ages.

Table 1 *Mean proportion-of-total-looking-time (PTLT) scores and (SD) for the eye and mouth areas-of-interest (AOI) as a function of age and in response to native, desynchronized, audiovisual speech in Experiment 1 and in response to non-native, desynchronized, audiovisual speech in Experiment 2*

	Experiment 1		Experiment 2	
	Eye AOI	Mouth AOI	Eye AOI	Mouth AOI
4-month-olds	.28 (.23)	.18 (.17)	.29 (.18)	.18 (.16)
6-month-olds	.33 (.20)	.21 (.22)	.35 (.26)	.21 (.21)
8-month-olds	.12 (.12)	.36 (.20)	.17 (.18)	.40 (.19)
10-month-olds	.21 (.18)	.26 (.21)	.28 (.19)	.36 (.17)
12-month-olds	.20 (.18)	.30 (.19)	.11 (.11)	.46 (.17)

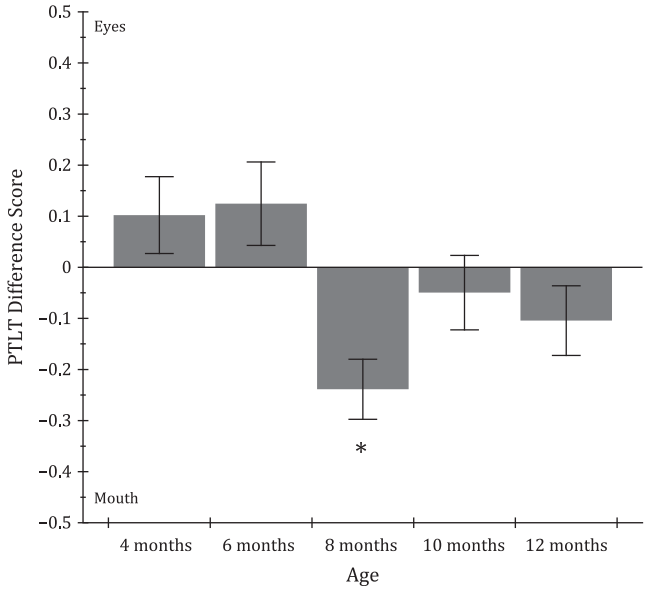

Figure 1 *Mean PTLT difference scores as a function of age in response to the native, English monologue. Error bars represent SEMs and the asterisk indicates a statistically significant result.*

Discussion

The findings from this experiment indicated that monolingual, English-learning infants distributed their selective attention to a talker's eyes and mouth differently across the different ages. Of greatest interest was the fact that the overall developmental pattern of relative attention devoted to the eyes and mouth differed from the pattern obtained by Lewkowicz and Hansen-Tift (2012). Specifically, at 4 months, the infants in the current study exhibited no difference in looking to the eyes and mouth, whereas the infants in the prior study looked more at the eyes than at the mouth. At 6 months, the infants in the current study looked equally long at the mouth and eyes and so did the infants in the prior study. At 8 months, the infants in the current study looked longer at the mouth than the eyes as did the infants in the prior study. At 10 months, the infants in the current study looked equally long at the mouth and eyes, whereas the infants in the prior study looked more at the mouth than at the eyes. Finally, at 12 months, the infants in the current study exhibited no differential looking at the mouth and eyes as did the infants in the prior study.

To determine whether the findings from the current study differed from those in the Lewkowicz and Hansen-Tift study (2012), we compared the data from the two age groups where we found differences with separate mixed, repeated-measures ANOVAs, with Study (2) and Prosody (2) as the between-subjects factors and AOI (2) as the within-subjects factor. We expected the Study × AOI interaction to be significant if the outcomes in the two studies were different. The ANOVA indicated that this interaction was not significant at 4 months of age [$F(1, 35) = 1.30$, $p = .26$, $\eta_p^2 = .036$] but that it was significant at 10 months of age [$F(1, 32) = 4.31$, $p = .046$, $\eta_p^2 = .12$]. As can be seen in Figure 2, the significant interaction at 10 months was due to less looking at the mouth when the audiovisual monologue was desynchronized than when it was synchronized. A planned comparison of looking at the mouth across the two studies confirmed that this difference was significant [$F(1, 32) = 6.51$, $p = .015$, $\omega^2_{(\psi)} = .07$].

The different developmental patterns obtained across the two studies indicate that temporal A-V synchrony cues play a role in infant selective attention to talking faces during the canonical babbling stage but not before. Specifically, it does not appear that synchrony influences attention at 4 months of age because even though the 4-month-olds in our experiment did not look more at the

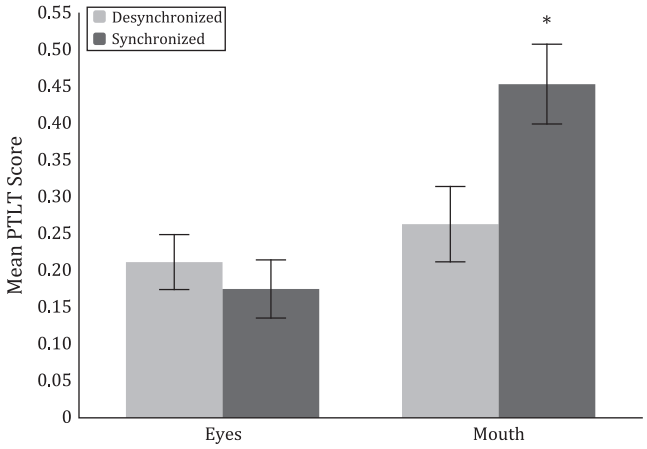

Figure 2 *Mean PTLT scores for attention directed at the eyes and mouth, respectively, in the 10-month-old infants in Experiment 1 as a function of the temporal relation between the audible and visible native speech streams. Asterisk indicates a statistically significant result.*

eyes than the mouth, whereas the 4-month-olds in the Lewkowicz and Hansen-Tift (2012) study did, the difference across the studies was not statistically significant. Also, our findings from 8-month-old infants indicated that A-V synchrony is not the principal determinant of infant preference for the talker's mouth at this age because they, like the infants in the Lewkowicz and Hansen-Tift's (2012) study, looked longer at the talker's mouth. Only our findings from the 10-month-old infants provided evidence that synchrony-based redundancy is an important determinant of infant preference for the talker's mouth because, unlike the 10-month-olds in the Lewkowicz and Hansen-Tift (2012) study, our 10-month-olds no longer attended more to the mouth than to the eyes. Finally, the findings from the 12-month-olds, like the findings from the Lewkowicz and Hansen-Tift (2012) study, indicated that at this age infants no longer prefer the mouth. This suggests that once infants acquire sufficient experience with their native speech, they no longer need to focus as much of their attention on the mouth to access the redundant audiovisual speech information located there. Moreover, the fact that the 12-month-olds responded similarly whether audiovisual speech was synchronized or not suggests that synchrony-based redundancy does not mediate the deployment of selective attention at this age.

Experiment 2

Lewkowicz and Hansen-Tift (2012) found that the developmental pattern obtained in response to non-native audiovisual speech was identical to that obtained in response to native speech except that the 12-month-old infants who were exposed to non-native speech continued to deploy greater attention to the mouth than to the eyes. This was interpreted as reflecting the onset of audiovisual speech processing as a communicative event per se and of the concurrent negative effects of perceptual narrowing on infants' ability to process what has by this age become unfamiliar speech (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014; Werker & Tees, 2005). Given this interpretation, might the continued focus on the mouth at 12 months of age in response to non-native audiovisual speech mean that infants of this age are benefiting from synchrony-based redundancy to disambiguate the speech signal? Similarly, does the 8- and 10-month-olds' greater focus on the mouth reflect these infants' reliance on synchrony-based redundancy?

To answer these questions, we desynchronized the non-native audiovisual monologue presented by Lewkowicz and Hansen-Tift (2012) and presented it to 4-, 6-, 8-, 10-, and 12-month-old infants. As in Experiment 1, one reasonable expectation was that the desynchronization may change the way infants – especially at 8, 10, and 12 months of age – deploy their selective attention to the talker's eyes and mouth if they rely on A-V synchrony as the principal redundancy cue. Alternatively, given that synchrony-based redundancy did not have an effect on responsiveness to native speech at 8 and 12 months of age but did at 10 months in Experiment 1, it may be that responsiveness to non-native speech also might not depend on synchrony at some ages. For example, like the infants in the Lewkowicz and Hansen-Tift (2012) study, the 12-month-olds in the current experiment might still attend more to the mouth so as to gain simultaneous access to the audible and visible speech streams to, presumably, disambiguate what have become unfamiliar streams of unisensory information. Importantly, however, it should be noted that at this age infants do not perceive the multisensory coherence of non-native audiovisual speech when the audible and visible speech streams are desynchronized (Lewkowicz *et al.*, 2015). Therefore, if the 12-month-olds in the current experiment continue to devote more attention to the mouth then this will indicate that synchrony-based redundancy cues do not mediate attention to the mouth.

Method

Participants

In all, 81 infants (42 boys) initially contributed data in this experiment. All infants were full-term, healthy, and

had no history of ear infections according to parental report (birth weight, ≥ 2500 g; APGAR score, ≥ 7; gestational age, ≥ 37 weeks). All infants were raised in a mostly monolingual environment, meaning that their language exposure to English exceeded 80% according to parental report. An additional 49 infants were excluded because of failure to meet the 4 s looking criterion (4), fussiness or inattentiveness (12), failure to calibrate because the infant was uncooperative or the eye tracker could not find the pupil (18), equipment failure (13), experimental error (1), or parent interference (1).

The participants consisted of separate groups of 4-month-olds (n =13; mean age, 16.9 weeks; SD = 0.7 weeks), 6-month-olds (n = 18; mean age, 26.1 weeks; SD = 0.6 weeks), 8-month-olds (n = 18; mean age, 34.3 weeks; SD = 0.6 weeks), 10-month-olds (n = 13; mean age, 43.4 weeks; SD = 0.6 weeks), and 12-month-olds (n = 19; mean age, 52.4 weeks; SD = 0.8 weeks).

Apparatus, stimuli and procedure

We used the identical procedures that we used in Experiment 1 except that this time we presented movies of a native Spanish speaker reciting the Spanish version of the monologue presented in Experiment 1 either in the ID or AD style. These Spanish movies were the same as those that were presented by Lewkowicz and Hansen-Tift (2012) in their second experiment except that here we only presented the first 30 s of the original movie. In addition, we desynchronized the auditory and visual speech streams by 666 ms, with the auditory speech stream leading the visual speech stream.

Results

We used a mixed, repeated-measures ANOVA to analyze the PTLT scores, with AOI (eyes, mouth) as a within-subjects factor and Age (4, 6, 8, 10, and 12 months) and Prosody as between-subjects factors. The ANOVA indicated that the AOI × Age interaction was significant [F(4, 71) = 7.44, p < .001, η_p^2 = .30] and that no other effects were significant. The AOI × Age interaction can be seen in Figure 3.

To probe the data further, we conducted planned comparison tests. These tests showed that the 4- and 6-month-old infants did not exhibit differential looking at the eyes and mouth [F(1, 71) = 1.62, p = .21, $\omega^2_{(\psi)}$ = .023, and F(1, 71) = 3.93, p = .051, $\omega^2_{(\psi)}$ = .075, respectively], that the 8-month-olds looked more at the mouth than the eyes [F(1, 71) = 7.49, p = .008, $\omega^2_{(\psi)}$ = .15], that the 10-month-olds did not exhibit differential looking at the eyes and mouth [F(1, 71) < 1, ns, $\omega^2_{(\psi)}$ = .029], and that

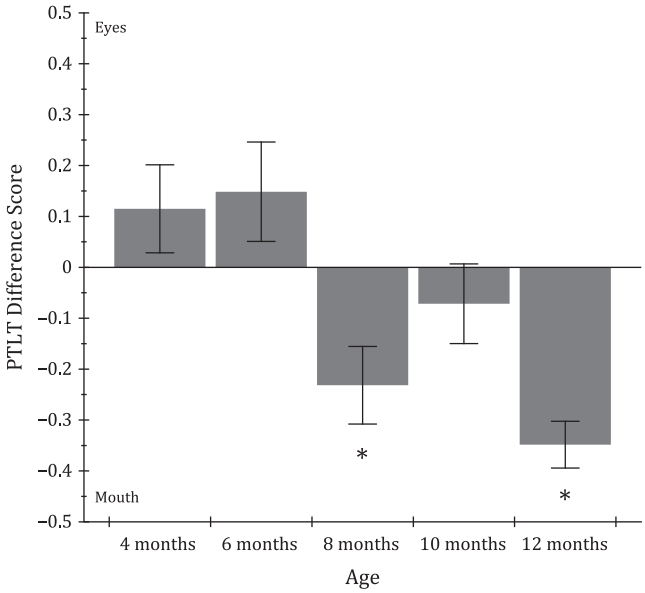

Figure 3 *Mean PTLT difference scores as a function of age in response to the non-native, Spanish monologue. Error bars represent SEMs and the asterisks indicate a statistically significant result.*

the 12-month-olds looked longer at the mouth than the eyes [F(1, 71) = 21.99; p < .001, $\omega^2_{(\psi)}$ = .36]. The PTLT scores for this Experiment are depicted in Table 1.

To rule out the possibility that overall attention may have varied as a function of age and/or prosody and, thus, that it might have affected the results, we analyzed the total amount of looking at the face with a two-way ANOVA, with Age and Prosody as the between-subjects factors. Results of this analysis revealed that the different patterns of attention found at the different ages were not attributable either to differences in overall attention across age [F(4, 71) < 1, ns, η_p^2 = .035] or to differences in responsiveness across the two prosody conditions [F(1, 71) < 1, ns, η_p^2 < .001].

Discussion

As in Experiment 1, we found that the relative amount of selective attention that monolingual, English-learning infants deployed to the eyes and mouth of a talker producing non-native desynchronized audiovisual speech differed across the first year of life. In addition, and as in Experiment 1, we found that the overall pattern of responsiveness across development was not the same as that found by Lewkowicz and Hansen-Tift (2012) in their study of infant response to synchronized non-native audiovisual speech. First, our 4-month-old infants did

not exhibit a preference, whereas the infants in the prior study looked more at the eyes than the mouth. Second, neither the 6-month-old infants in our study nor the infants in the prior study exhibited differential looking at the eyes and mouth. Third, the 8-month-old infants in our study as well as those in the prior study looked longer at the mouth than the eyes. Fourth, the 10-month-old infants in our study did not exhibit differential looking at the eyes and mouth, whereas the infants in the prior study looked longer at the mouth than the eyes. Finally, our 12-month-old infants as well as the 12-month-old infants in the prior study looked longer at the mouth than the eyes.

To further compare the findings in the current study with those from the Lewkowicz and Hansen-Tift (2012) study, as in Experiment 1, we used separate mixed, repeated-measures ANOVAs with Study (2) and Prosody (2) as the between-subjects factors and AOI (2) as the within-subjects factor. Again, we only found significant differences in the 10-month-old infants. That is, whereas the Study × AOI interaction was not significant at 4 months of age [$F(1, 28) = 0.75$, *ns*, $\eta_p^2 = .025$], it was significant at 10 months of age [$F(1, 29) = 4.82$, $p = .036$, $\eta_p^2 = .14$]. As can be seen in Figure 4, this interaction was due to more looking at the eyes when the audiovisual monologue was desynchronized than when it was synchronized. A planned comparison of looking at the eyes across the two studies confirmed that this difference was significant [$F(1, 29) = 11.50$, $p = .002$, $\omega^2_{(\psi)} = .06$].

Again, as in Experiment 1, even though our 4-month-olds did not look more at the eyes than the mouth whereas the infants in the Lewkowicz and Hansen-Tift (2012) study did, this was not a significant difference. The findings from our 8-month-old infants indicated that, at this age, A-V synchrony does not mediate selective attention to the mouth of a person uttering non-native audiovisual speech because they, like the 8-month-olds in the Lewkowicz and Hansen-Tift (2012) study, also looked longer at the mouth than the eyes. Had synchrony played an important role in attention at this age, the 8-month-olds would not have been expected to attend more to the mouth when A-V synchrony was disrupted. The findings from our 10-month-olds indicate that A-V synchrony does play an important role in the preference for the talker's mouth found by Lewkowicz and Hansen-Tift (2012) in response to non-native speech. That is, like the 10-month-olds in Experiment 1, and unlike the 10-month-olds who are exposed to synchronized non-native speech, our 10-month-olds no longer looked more at the talker's mouth than the eyes when the synchrony of non-native speech was disrupted. In fact, our 10-month-olds increased the amount of time they looked at the eyes when non-native speech was desynchronized relative to when it was synchronized. Finally, similar to the 12-month-olds in Experiment 1, our 12-month-olds were not affected by desynchronization in that they did not differ from the 12-month-olds' response to synchronized non-native speech in the Lewkowicz and Hansen-Tift (2012) study. That is, our 12-month-olds also devoted more attention to the mouth than the eyes and they did so despite the fact that the non-native audiovisual speech was desynchronized. Given that 12–14-month-olds do not perceive the coherence of desynchronized non-native audiovisual speech (Lewkowicz *et al.*, 2015), the findings from our 12-month-olds suggest that they may have been trying to decode the unisensory information in each modality without regard to its multisensory coherence.

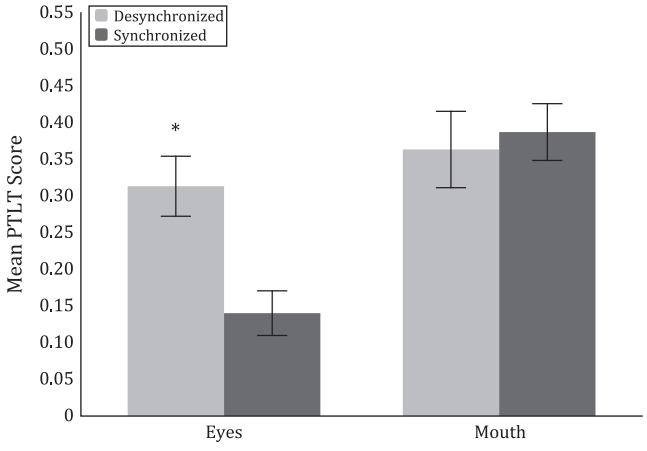

Figure 4 *Mean PTLT scores for attention directed at the eyes and mouth, respectively, in the 10-month-old infants in Experiment 2 as a function of the temporal relation between the audible and visible non-native speech streams. Asterisk indicates a statistically significant result.*

General discussion

Previous studies of eye gaze behavior in response to talking faces have found that infants begin to shift their attention from a talker's eyes to the talker's mouth as they enter the second half of the first year of life (Haith, Bergman & Moore, 1977; Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Pons *et al.*, 2015; Tenenbaum, Shah, Sobel, Malle & Morgan, 2013; Young, Merin, Rogers & Ozonoff, 2009). In the current study, we investigated two questions related to this behavior. First, given that A-V synchrony plays a major role in multisensory perception in infancy (Bahrick & Lickliter, 2012; Lewkowicz, 2000a, 2014; Lewkowicz & Ghazanfar, 2009), we asked whether A-V synchrony

plays a role in infant selective attention to a talker's mouth by disrupting it. Second, given that responsiveness to auditory, visual, and audiovisual speech changes during the first year of life and given that this includes perceptual narrowing of responsiveness to native versus non-native speech (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2014; Werker & Tees, 2005), we asked whether early experience affects the role that A-V synchrony plays in infant selective attention to different parts of a talker's mouth.

Like Lewkowicz and Hansen-Tift (2012), who presented synchronized audiovisual speech, we found that 4- and 6-month-old infants did not attend more to a talker's mouth when audiovisual speech was desynchronized. When the data from the previous study and ours are considered together they show that 4- and 6-month-old infants do not attend more to the mouth regardless of whether the speech is native or not and regardless of whether the audible and visible speech streams are synchronized or not. What is particularly interesting about these findings is that the 4- and 6-month-old infants did not attend to the talker's mouth even though the mouth was moving and despite the fact that they are sensitive to motion (Kaufmann, Stucki & Kaufmann-Hayoz, 1985; Stucki, Kaufmann-Hayoz & Kaufmann, 1987). Thus, it appears that 4–6-month-old infants are not as interested in audiovisual speech per se as are older infants.

In contrast to the 4–6-month-olds, we found that 8-month-olds focused their attention on the talker's mouth. This is in line with findings from other studies with infants older than 6 months of age (Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Pons *et al*., 2015; Tenenbaum *et al*., 2013; Young *et al*., 2009) and suggests that the attentional shift to the mouth by 8 months of age reflects the emergence of an explicit interest in audiovisual speech. If so, then it is reasonable to infer that this attentional shift is likely to facilitate the acquisition of new speech forms because it makes it possible for infants to gain direct access to the highly salient redundant audiovisual speech cues that are normally available in a talker's mouth. Crucially, however, the fact that 8-month-olds focused their attention on the talker's mouth regardless of whether the audiovisual speech was native or not indicates that, at this age, infants do not yet possess sufficient linguistic expertise to distinguish between these two types of speech. This is in line with findings from studies showing that infants only become native-language experts and begin to perceive amodal language identity at the end of the first year of life (Lewkowicz & Pons, 2013; Werker & Tees, 2005).

The findings from the 8-month-olds are also interesting because, in the aggregate, the findings from the Lewkowicz and Hansen-Tift (2012) study and the current one indicate that 8-month-olds attend more to the mouth regardless of whether the audiovisual speech emanating from it is synchronized or not. At first blush, this might seem inconsistent with findings that 3-month-old infants prefer synchronized over desynchronized audiovisual speech (Dodd, 1979) and that 8-month-old infants can detect the desynchronization of fluent audiovisual speech (Pons & Lewkowicz, 2014). It should be noted, however, that the task demands differed across the different studies. In the current study, infants passively viewed and listened to a talking face. In contrast, in the other studies infants had to actively choose a particular audiovisual event over another or had to discriminate between different audiovisual events.

In contrast to the findings from the 4-, 6-, and 8-month-olds, the findings from the 10-month-olds indicated that A-V temporal synchrony does mediate attentional responsiveness at this age. Specifically, whereas Lewkowicz and Hansen-Tift (2012) found that 10-month-olds look more at the mouth when exposed to synchronized native and non-native audiovisual speech, we found that 10-month-olds do not attend more to the talker's mouth when they are exposed to either native or non-native desynchronized audiovisual speech. Furthermore, direct comparisons of our data with those from the Lewkowicz and Hansen-Tift (2012) study revealed that this was due to less looking at the mouth in response to native desynchronized speech and to more looking at the eyes in response to non-native desynchronized speech. One way to interpret the greater looking at the eyes in response to non-native speech is that this may reflect the emergence of an understanding of the social meaning of eye contact by this age (Brooks & Meltzoff, 2005). On this account, the 10-month-olds looked more at the eyes because asynchronous non-native speech was somewhat perplexing to them. Presumably, by looking more at the talker's eyes they were attempting to disambiguate a 'confusing' linguistic event. Needless to say, this interpretation is purely speculative and requires further scrutiny.

The data from the 12-month-olds showed that these infants did not attend more to the mouth when the talker produced native speech and that they did attend more to it when she produced non-native speech. This replicates the findings reported by Lewkowicz and Hansen-Tift (2012) for this age group. The data from this age group also indicated that desynchronization did not change responsiveness. This suggests that A-V synchrony no longer mediates attentional responsiveness to audiovisual speech at this age. Crucially, however, it should be noted that this conclusion only applies to a free viewing/listening situation. This is because when infants have to

make an explicit choice based on temporal A-V synchrony cues and when audiovisual speech is not native, they do rely on synchrony to make a match. This is illustrated by findings that 12–14-month-olds can match streams of native auditory and visual fluent speech even if they are desynchronized but that they do not match streams of non-native auditory and visual fluent speech if they are desynchronized (Lewkowicz *et al.*, 2015). Thus, infants no longer rely on A-V synchrony cues at the end of the first year of life in a multisensory matching task when the multisensory inputs are familiar but they continue to rely on them when the inputs are unfamiliar.

Overall, when the current findings and those from the Lewkowicz and Hansen-Tift (2012) study are considered together, the following developmental picture emerges. Prior to the emergence of canonical babbling (at 4 and 6 months of age), infants are less interested in speech production and, because of this, they do not focus their attention on a talker's mouth. In contrast, once infants find themselves in the midst of the canonical babbling stage (between 8 and 10 months of age), they become interested in speech and begin to focus their attention on the most reliable and salient source of speech, namely a talker's mouth. Our findings suggest, however, that 8-month-old infants' speech processing abilities are not sufficiently developmentally advanced to enable them to process fluent speech as a meaningful linguistic signal. This conclusion is based on the fact that they continue to focus on a talker's mouth regardless of whether the audiovisual speech is synchronized or not. It appears that their attention is captured by the greater amount of overall stimulation in the mouth region and less by the temporal congruency of the audible and visible speech streams. Importantly, however, this is only true in the case of the deployment of selective attention in a free viewing/listening situation because studies have found that 8-month-olds can detect the difference between synchronous and asynchronous audiovisual speech (Pons & Lewkowicz, 2014). Unlike at 8 months of age, by 10 months of age, infants appear to be tracking the temporal alignment of the audible and visible streams of fluent audiovisual speech because they no longer prefer the talker's mouth when audiovisual speech is desynchronized. Finally, by 12 months of age infants no longer seem to track the temporal alignment of native audible and visible speech in a free viewing/listening situation but they may still track the temporal alignment of non-native speech even if they exhibit no disruption of the preference for the talker's mouth when speech is desynchronized.

The current findings raise an interesting question. What defines multisensory redundancy? Previously, we indicated that the greater perceptual salience of audiovisual as opposed to auditory or visual speech is due to two types of multisensory relations. One relation derives from the synchronous onsets and offsets of audible vocalizations and visible mouth, face, and head movements and the other derives from the correlation of the continuous temporal dynamics of audible and visible speech (Chandrasekaran *et al.*, 2009; Munhall, Jones, Callan, Kuratate & Vatikiotis-Bateson, 2004; Munhall & Vatikiotis-Bateson, 1998; Rosenblum, 2008; Rosenblum *et al.*, 1997; Yehia *et al.*, 1998). Previously, we also indicated that the influence of relatively low-level onset/offset A-V synchrony cues begins to decline in relation to higher-level multisensory redundancy cues by the end of the first year of life (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009). Therefore, it is theoretically reasonable to ask whether and when A-V synchrony cues drive infants' selective attention to audiovisual speech. Here, we found that temporal A-V synchrony plays a role in infant selective attention to fluent audiovisual speech but that this is only the case at 10 months of age. This is surely due to developmental changes in perceptual processing as well as to the fact that multisensory redundancy is not a unitary phenomenon. For example, the multisensory redundancy that specifies a single audiovisual syllable is not the same as the redundancy that specifies fluent audiovisual speech. With particular regard to synchrony-based A-V redundancy cues and infants' selective response to talking faces, the temporal correlation of the audible and visible speech streams of fluent audiovisual speech is specified by a set of hierarchically organized and nested perceptual cues. That is, the auditory and visual streams of fluent audiovisual speech correspond not only in terms of their global onsets and offsets but also in terms of their overall rhythmic/prosodic structure, tempo, and intensity. Moreover, the auditory and visual streams of fluent audiovisual speech consist of episodic components (phonemes, words, sentences) and the auditory and visual attributes of each of those components have equal durations. Potentially, infants may attend to any of these multisensory redundancy cues depending on their developmental status. Whether, when, and why they do so is currently an open question.

Acknowledgements

We thank Kelly Henning for her assistance. This study was supported by Grant R01HD057116 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to DJL. The work reported here was performed when DJL was at Florida Atlantic University.

References

Bahrick, L.E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, **6**, 429–451.

Bahrick, L.E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, **59**, 197–209.

Bahrick, 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: Oxford University Press.

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.

Brooks, R., & Meltzoff, A.N. (2005). The development of gaze following and its relation to language. *Developmental Science*, **8**, 535–543.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, **5**, e100436.

Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, **11**, 478–484.

Haith, M.M., Bergman, T., & Moore, M.J. (1977). Eye contact and face scanning in early infancy. *Science*, **198**, 853–855.

Hunnius, S., & Geuze, R.H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: a longitudinal study. *Infancy*, **6**, 231–255.

Kaufmann, F., Stucki, M., & Kaufmann-Hayoz, R. (1985). Development of infants' sensitivity for slow and rapid motions. *Infant Behavior and Development*, **8**, 89–98.

Keppel, G., & Wickens, T.D. (2004). *Design and analysis: A researchers handbook* (4th edn.). Englewood Cliffs, NJ: Prentice Hall.

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

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. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development*, **9**, 335–353.

Lewkowicz, D.J. (1992a). Infants' response to temporally based intersensory equivalence: the effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior and Development*, **15**, 297–324.

Lewkowicz, D.J. (1992b). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics*, **52**, 519–528.

Lewkowicz, D.J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 1094–1106.

Lewkowicz, D.J. (2000a). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological Bulletin*, **126**, 281–308.

Lewkowicz, D.J. (2000b). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development*, **71**, 1241–1257.

Lewkowicz, D.J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, **39**, 795–804.

Lewkowicz, D.J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, **46**, 66–77.

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.

Lewkowicz, D.J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: newborns match non-human primate faces and voices. *Infancy*, **15**, 46–60.

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

Lewkowicz, D.J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, **37** (2), 90–94.

Lewkowicz, D.J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: auditory–visual intensity matching. *Developmental Psychology*, **16**, 597–607.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 229–239.

Maurer, D., & Werker, J.F. (2014). Perceptual narrowing during infancy: a comparison of language and faces. *Developmental Psychobiology*, **56**, 154–178.

Morrongiello, B.A., Fenwick, K.D., & Nutley, T. (1998). Developmental changes in associations between auditory-visual events. *Infant Behavior and Development*, **21**, 613–626.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science*, **15**, 133–136.

Munhall, K.G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 123–139). Hove, East Sussex: Psychology Press/Erlbaum (UK) Taylor & Francis.

Munhall, K.G., & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In G.A. Calvert, C. Spence & B.E. Stein (Eds.), *The handbook of*

multisensory processes (pp. 177–188). Cambridge, MA: MIT Press.

Patterson, M.L., & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, **22**, 237–247.

Patterson, M.L., & Werker, J.F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, **81**, 93–115.

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

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. doi:10.1177/0956797614568320

Pons, F., & Lewkowicz, D.J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, **149**, 142–147.

Rosenblum, L.D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, **17**, 405.

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

Scheier, C., Lewkowicz, D.J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, **6**, 233–244.

Schroeder, C., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, **12**, 106–113.

Stucki, M., Kaufmann-Hayoz, R., & Kaufmann, F. (1987). Infants' recognition of a face revealed through motion: contribution of internal facial movement and head movement. *Journal of Experimental Child Psychology*, **44**, 80–91.

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

Summerfield, A.Q. (1979). Use of visual information in phonetic perception. *Phonetica*, **36**, 314–331.

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.

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.

Võ, 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**, 3.

Walker-Andrews, A.S. (1986). Intermodal perception of expressive behaviors: relation of eye and voice? *Developmental Psychology*, **22**, 373–377.

Walker-Andrews, A.S., Bahrick, L.E., Raglioni, S.S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, **3**, 55–75.

Werker, J.F., & Tees, R.C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology. Special Issue: Critical Periods Re-examined: Evidence from Human Sensory Development*, **46**, 233–234.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, **26**, 23–43.

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.

Received: 8 April 2015
Accepted: 9 October 2015