1 Introduction

1.1 Background

When we talk with one another, we not only hear each other's voice, but typically also see one another speaking. It has long been known that both heard and seen speech are important in adult speech processing, where information in either modality has the potential to modify perception in the other modality (e.g., Sumby & Pollack, 1954; McGurk & Macdonald, 1976). While an increasing number of studies have explored the extent to which speech perception is multisensory in young infants, there are still many unanswered questions. First and foremost is the question of whether information in one modality can affect processing in the other, and how this might influence and/or be influenced by the timing of the sensitive period for perceptual attunement, the process in the first year of life by which infants' discrimination of non-native speech contrasts declines and their discrimination of native speech contrasts improves (see Werker & Gervain, 2013 for a review). The current set of studies addresses these issues.

We asked two specific questions. First, using sounds with which infants were unfamiliar, we tested whether or not young infants are sensitive to the congruence between the auditory and visual information in the speech signal and, if so, whether their sensitivity is independent of experience with specific sound-sight pairings from the native language. We explored the possibility that such sensitivity, if revealed, might decline in tandem with perceptual attunement. Second, we asked whether experimental exposure to congruent versus incongruent audiovisual speech can alter subsequent auditory-only speech perception, and possibly reveal sensitivity to non-native auditory distinctions beyond the age at which infants typically discriminate non-native sounds.

1.2 Perceptual attunement

From a young age, infants auditorily discriminate many of the similar consonant sounds used across the world’s languages, regardless of whether such sounds are used to contrast meaning between two words (phonemically) in the language(s) that the child hears. For example, at six to eight months of age, both English- and Hindi-learning infants discriminate between the voiced dental and retroflex consonants of Hindi ([d̪] and [ɖ], respectively), though no such phonemic distinction exists in English, and English-speaking adults exhibit no such discrimination (Werker et al., 1981; Werker & Tees, 1984; Werker & Lalonde, 1988). However, by the time they are nine months old, English-learning infants exhibit reduced discrimination of non-native consonantal phonemic distinctions. By 11 months, auditory discrimination of many non-native consonantal phonemes has declined even further, while discrimination of native phonemes has improved (Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010).

This pattern of decline in sensitivity to non-native consonant contrasts and improvement in sensitivity to native contrasts across the first year of life is called *perceptual attunement*. Similar findings have emerged for discrimination of tone distinctions (Mattock & Burnham, 2006; Yeung, Chen, & Werker, 2013), and even for the discrimination of handshape distinctions in visual-only sign language (Palmer, Fais, Golinkoff, & Werker, 2012) and for discrimination of articulatory configurations in silent visual-only speech (Weikum et al., 2007; Sebastián-Gallés et al., 2012). The same pattern is seen for perception of vowel distinctions, but may develop earlier than for consonants (Polka & Werker, 1994). The consistency in the timing of this pattern of change, particularly for perception of consonant contrasts, suggests a critical or sensitive period in development between six and 12 months of age, during which the speech input plays an especially important role in changing perceptual sensitivities (Doupe & Kuhl, 1999; Kuhl, 2010; Friederici & Wartenburger, 2010; Werker & Tees, 2005; Maurer & Werker, 2014; Werker & Hensch, 2015).

1.3 Audiovisual speech perception

Although the bulk of research in speech perception—and in perceptual attunement—has been conducted by investigating the role of individual modalities, the audiovisual nature of speech perception has nevertheless been well attested in adults. A commonly observed piece of evidence in support of a multisensory view of speech perception is adults' robust ability to speechread: to use visual information from an interlocutor's eyes and mouth to aid in perceiving speech in noise (Sumby & Pollack, 1954; MacLeod & Summerfield, 1987; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998; Grant & Seitz, 2000). Even more evidence comes from the imposition of *incongruent* visual information onto the auditory speech signal. Under certain conditions, when adult listeners are presented with simultaneous auditory and visual signals that conflict with each other (e.g., a visual /ba/ and an auditory /ga/), an entirely different illusory percept arises (adults report perceiving /da/), a phenomenon known as the McGurk effect (McGurk & Macdonald, 1976; Massaro, Cohen, & Smeele, 1996; Rosenblum & Saldaña, 1992, *inter alia*).

A growing body of work suggests that speech perception is audiovisual for the infant as well. Infants exhibit the same McGurk effect that adults do (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997), although perhaps less strongly

(Desjardins & Werker, 2004). Like adults, infants' auditory perception of speech in noise is improved when visual information is added (Hollich, Newman, & Jusczyk, 2005).

Much research on audiovisual processing of speech in infancy has involved cross-modal matching. When shown first a video display of two side-by-side identical faces, one articulating one syllable and the other articulating a different syllable, and then are shown the same video display accompanied by the sound for one of the syllables, infants as young as two months of age look longer to the side articulating the syllable that matches the sound that they hear (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2002, 2003; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). This evidence indicates that infants' perception of heard and seen speech is audiovisual from early in life. Moreover, in the first six months of life, infants match audiovisual speech combinations from languages with sounds that are unfamiliar to them (Walton & Bower, 1993; Pons et al., 2009; Kubicek et al., 2014), and even with pairs of non-human animal faces and their vocalizations (Vouloumanos, Druhen, Hauser & Huizink, 2009; Lewkowicz & Ghazanfar, 2006; Lewkowicz, Leo, & Simion, 2010).

Just as the perception of auditory speech attunes in the infant's first year to just those distinctions used in the native language, so too does the matching of the auditory and visual signal. By 11 months of age, infants no longer match heard and seen speech if the stimuli are from a non-native language. For example, six-month-old Spanish-learning infants look longer at a face articulating /ba/ (than a face articulating /va/), when hearing the sound /ba/, and longer at the face articulating /va/ when hearing the sound /va/, even though Spanish does not use these two sounds contrastively. However, by 11 months of age, Spanish-learning infants no longer match heard and seen /ba/ and /va/, whereas infants learning English—in which the distinction is used contrastively—continue to do so (Pons, et al, 2009; but see Kubicek et al., 2014, for possibly contrasting results with 12-month-olds). While this work could be explained solely on the basis of sensitive periods for the attunement of auditory speech perception, Pons and colleagues (2009) argue that their results may also indicate that perceptual attunement is a "pan-sensory" process. Presently, we explore this possibility further by probing whether infants detect (in)congruence in the content of dynamic, unfamiliar speech events that simultaneously provide signals in two sensory modalities (audition and vision). If infants do detect cross-modal incongruence in such a task, which would require both auditory and visual sensitivity to the contrast utilized, that would provide additional evidence that infants' speech percept is audiovisual ("pan-sensory") and that it is so independently of infants' experience with a specific language system. Furthermore, the discovery that the decline of such sensitivity to audiovisual congruence follows a different temporal trajectory than does auditory-only speech discrimination could indicate that the sensitive period for speech contrast discrimination is altered when information from more than one modality is taken into account.

In the current study, we operationalize infants' detection of audiovisual (in)congruence by focusing on infants' attention to various areas of a speaker's face while observing speech. While most infants and adults fixate on the eyes of a speaking face (Haith, Bergman, & Moore, 1977; Vatikiotis-Bateson et al, 1998; Cassia, Turati, & Simion, 2004; Hunnius & Geuze, 2004; Merin, Young, Ozonoff, & Rogers, 2006*)*, 8- to 12-month-old infants fixate preferentially on the speaker's mouth, a pattern that is even more pronounced when infants are viewing non-native speech (Lewkowicz & Hansen-

Tift, 2012; Kubicek et al., 2013). Lewkowicz and Hansen-Tift (2012) explain this effect by proposing that during the period of perceptual attunement, infants may attend to the visual information provided by the mouth of a speaking face to boost auditory perception and phonetic production. Indeed, children who attend more to their mothers' mouths in early infancy exhibit higher expressive vocabularies in toddlerhood (Young, Merin, Rogers, & Ozonoff, 2009). Moreover, at least one study has demonstrated that infants at six to 12 months of age attend to the mouth region of the face when observing incongruent audiovisual speech in their native language (Tomalski et al., 2013). Taken together, these results suggest that where infants look on the face while they perceive audiovisual speech may provide information about their perception of non-native speech, their sensitivity to audiovisual (in)congruence, and their progress in the developmental trajectory of perceptual attunement.

1.4 Visual modification of auditory discrimination

Although prior studies have suggested that infants' speech perception is audiovisual and that infants match auditory and visual content when perceiving speech, it is not known if and how visual information presented in audiovisual speech might change infants' auditory phonetic discrimination before, during, and after perceptual attunement. To the extent that the speech percept is audiovisual for the young infant, the addition of visual articulatory information to the auditory speech signal could alter this discrimination.

Indeed, Teinonen and colleagues (2008) demonstrated that pairing degraded speech sounds with photographs of visual articulations matching those sounds boosted six-month-old infants' ability to discriminate the auditory speech sounds in a later test.

Crucially, that study tested infants using native speech sounds with which they were familiar. Also, since infants in their study were presented with only congruent stimuli (albeit minimally so), it remains unclear whether it was necessary that the visually-presented mouth shapes corresponded to the sounds being tested, or whether infants' performance would have been boosted simply by the presence of *any* consistent visual correlate. Another recent study explored a similar question by attempting to boost infants' discrimination of a non-native vowel contrast by pairing visual articulatory information with auditory sounds during a distributional learning paradigm (Ter Schure, Junge, & Boersma, 2016). In that study, infants familiarized to a bimodal audiovisual distribution exhibited a moderate boost to subsequent auditory discrimination. However, as in the study conducted by Teinonen and colleagues (2008), no incongruent visual correlates were tested to determine whether content congruence of the auditory and visual signals was important. Given these results, it is possible that infants' discrimination of similar stimuli in one modality (e.g., audition) can be aided by pairing those items consistently with distinctive stimuli from an additional modality (e.g., vision), even when the link between the auditory and visual items in each pair is arbitrary. Such *acquired distinctiveness* has been shown to boost discrimination of otherwise similar stimuli (Lawrence, 1949; Hall, 1991; Norcross & Spiker, 1957; Reese, 1972), including non-native speech sounds at nine months of age (Yeung & Werker, 2009; see General Discussion).

It is important, however, to note that speech is typically perceived as a dynamic event in which the auditory and visual signals are presented both synchronously and congruently. Thus, it would be informative to determine the extent to which an alteration of infants’ auditory discrimination depends on the content congruence between auditory and visual signals. If speech perception is audiovisual from the earliest stages of life, congruent, synchronous visual information could affect subsequent auditory discrimination of these sounds differently than would incongruent information, even when the latter is presented synchronously.

1.5 Current study

The current set of studies was thus designed to test two questions. First, we asked whether and how infants detect content congruence in non-native audiovisual speech and whether this sensitivity to congruence declines in tandem with the trajectory of perceptual attunement previously established in auditory perception studies. Our second question probed whether the addition of congruent visual information would alter subsequent auditory discrimination of these same speech contrasts, possibly constituting a shift in the timing of the sensitive period for auditory speech perception. In each of the conditions described here, the Hindi dental-retroflex ([d̪]/[ɖ]) contrast was utilized. English-learning monolingual infants were sampled from three age populations: at six months, when infants auditorily discriminate the sounds used; at nine months, when perceptual attunement is underway and infants’ auditory discrimination abilities of non-native contrasts have begun to decline; and at 11 months, when perceptual attunement for speech sounds has stabilized and infants are expected to fail at discriminating these sounds.

Each of the present manipulations began by familiarizing participants to audiovisual videos of Hindi dental and retroflex syllables. Half of the infants were familiarized to incongruent, temporally aligned audiovisual speech, and the other half was familiarized to congruent, temporally aligned audiovisual speech. To address the first question, infants' familiarization data were analyzed to determine whether, as hypothesized, those familiarized to incongruent speech would exhibit a different pattern of looking to regions of the model's face as compared to those familiarized to congruent speech. A finding of greater looking to the mouth rather than the eyes while watching incongruent audiovisual speech has been demonstrated in infants viewing incongruent speech in their own language (Tomalski et al., 2013). Thus, infants' looking patterns to two anatomical regions of interest (the eyes and the mouth) were measured to determine whether infants in the incongruent familiarization group deployed a greater proportion of their visual fixations to the mouth region of the speaker's face, as compared to the infants familiarized to congruent speech. We predicted that, if detection of audiovisual congruence in unfamiliar speech declines at the same time as does auditory discrimination of unfamiliar sounds, such an effect would be observable in infants before perceptual attunement (at six months of age), attenuated for infants undergoing perceptual attunement (at nine months), and absent once attunement is complete (at 11 months).

To test the second question, following familiarization, infants were tested on discrimination of these same non-native speech sounds auditorily, with no visual information provided. It was predicted that the additional cross-modal information provided to infants by congruent audiovisual familiarization might boost subsequent auditory-only discrimination of this non-native contrast for infants undergoing perceptual attunement (at nine months). Moreover, it was hypothesized that *incongruent* audiovisual information would not produce this effect, and might in fact *alter* discrimination ability for the youngest group of infants who might be more sensitive to incongruence in unfamiliar speech.

2 Materials and methods

2.1 Sample

Infants were sampled from three different age groups from a database of families recruited from a maternity hospital in Western Canada. Parents of all infants tested reported that their children heard approximately 90-100% English; none heard a language that uses the dental-retroflex contrast phonemically, and none had been diagnosed with an audiological or speech production disorder. Infants in the first age group (before perceptual attunement) were six months old (*n* = 32; mean age = 198 days; age range = 182-225 days; 16 females). Infants in the second group (during perceptual attunement) were nine months old (*n* = 32; mean age = 269 days; age range = 256-281 days; 16 females), and infants in the third age group (after perceptual attunement) were 11 months old (*n* = 32; mean age = 329 days; age range = 308-345 days; 16 females). Additional infants were tested and excluded from final data analysis as follows: from the six-month-old sample: two infants due to experimenter error; three infants due to poor eyetracker calibration; 13 infants who did not finish the experiment due to crying or fussiness; from the nine-month-old sample: four infants due to poor eyetracker calibration; eight infants who did not finish the experiment due to crying or fussiness; from the 11-month-old sample: five infants due to poor eyetracker calibration; 11 infants who did not finish the experiment due to crying or fussiness; and three infants due to parental interference during the experiment (e.g., talking, feeding).

2.2 Stimuli

One female native speaker of Hindi was recorded to create the stimuli for these experiments. The speaker was video-recorded using a Panasonic AJ-PX270 HD camcorder and a Sennheiser MKH-416 interference tube microphone. During recording, the speaker produced triads of monosyllabic utterances consisting of a target consonant ([d̪] or [ɖ]) and a vocalic segment ([a:]) in infant-directed speech (see Figure 1). The speaker was oriented at a 45° angle from the camera, to optimize the viewer's ability to see the orofacial and head motions associated with the two stimulus syllables. For example, the retraction and raising of the tongue tip for the retroflex, [ɖ], should be produced with the jaw in a lower position and possibly slightly protruded. This may result in less jaw lowering for the following vowel, [a:], compared to that associated with the transition from the dental consonant, [d̪], to the following vowel, [a:]. Another visible difference concerns the tongue tip, which is likely to be visible for the dental consonant, but not for the retroflex.

[insert Figure 1 here]

From this raw material, experimental stimulus items were chosen from among the second items in each triad sequence, in order to control for list intonation effects. Final stimulus tokens were those that had a natural duration between 750 and 1000 ms, and which contained no abnormalities in pitch contour or phonation. Stimuli were then combined to create familiarization sequences and test sequences. Familiarization sequences each consisted of eight audiovisual tokens from the same category (audiovisually congruent [d̪a:] or [ɖa:], and audiovisually incongruent stimuli with visual [d̪a:]-audio [ɖa:], or visual [ɖa:]-audio [d̪a:]). To create incongruent audiovisual stimulus items, visual tracks of stimulus items were spliced with duration-matched auditory tracks from tokens of the opposite phonetic category (auditory [d̪a:] paired with visual [ɖa:] and auditory [ɖa:] paired with visual [d̪a:]). To ensure that the process of mismatching did not result in asynchronous audiovisual stimuli, consonant burst releases from the original video tokens were aligned with the burst releases of the incongruent, auditory token. The interstimulus interval within the familiarization sequences was 2.2 seconds, and sequences were 20 seconds in total length. Test stimuli were eight-item auditory-only sequences of two types: alternating sequences consisted of tokens from both phonetic categories, while non-alternating sequences consisted of tokens from only one category (Best & Jones, 1998). The interstimulus interval for test sequences was 2.2 seconds and the total length of each test sequence was 20 seconds.

2.3 Procedure

All participants were tested in a developmental psychology laboratory at a university in Western Canada. Infants were tested in a dimly lit, sound-attenuated room while sitting on a caregiver’s lap. The experimenter and the equipment in the experimental room were hidden from the infant’s view by dark curtains. Caregivers, who were asked not to speak to their infants, wore darkened sunglasses to avoid potential interference from their gaze on the eyetracking data, and to prevent their own responses to the stimuli from affecting the responses of the infant.

Infants were seated facing a television screen (101 cm x 57 cm) equipped with a small video camera and a Tobii Technology X60 eyetracker sampling at 60 Hz at a distance of 90 cm from the screen. Stimuli were presented using Psyscope (Cohen et al., 1993). Eytracker data were recorded using Tobii Studio (Tobii Technology, 2008), and a reference video was recorded with iMovie (Apple, Inc., 2013). Before the study, the eyetracker was calibrated using a five-point visual display with non-linguistic tones to establish each infant's eye gaze characteristics. Prior to familiarization, infants watched an animated waterwheel attention-attractor until they had fixated on the screen. Half of the infants in each age group (*n* = 16) were then familiarized to congruent audiovisual sequences of the Hindi dental and retroflex CV syllables (four dental sequences and four retroflex sequences). The other half of the infants was familiarized to incongruent audiovisual sequences. All stimuli were presented at a mean intensity level of 65 dB. Between familiarization trials, infants regained attention to a silent animated ball attention-attractor, and only proceeded to the next familiarization trial after attention was refixated on the screen. The eyetracker provided data indicating where on the screen infants were looking during familiarization, and the durations of fixation to each area of the screen were calculated and summed for each infant using Tobii Studio.

After familiarization, all infants were tested using an auditory discrimination task in which they were exposed to eight sequences of auditory test stimuli while watching a still checkerboard. Four of these sequences (*non-alternating* sequences) consisted of test tokens from one phonemic category ([ɖa:] or [d̪a:]), and four sequences (*alternating* sequences) consisted of tokens from both phonemic categories. Trials were separated by the attention-attracting ball, and infants proceeded to the next test trial when they had refixated on the ball. Alternating and non-alternating sequences alternated with one another during the test phase, and counterbalancing ensured that half of the infants began their trials with non-alternating sequences and the other half with alternating sequences. In this procedure, longer looking times to one type of trial (alternating versus non-alternating) indicates discrimination of the sound contrast (Best & Jones, 1998; Yeung &

Werker, 2009). Previous studies have revealed both alternating (Best & Jones, 1998; Mattock, Molnar, Polka, & Burnham, 2008) and non-alternating (e.g., Maye, Werker, & Gerken, 2002; Teinonen et al., 2008; Yeung & Werker, 2009) preferences, either of which is interpreted as evidence of discrimination. We thus did not predict a specific direction of preference in the current study, but rather were most interested in determining whether familiarization to congruent versus incongruent stimuli would change discrimination patterns at test as exhibited by infants' preference for alternating or non-alternating stimuli.

3 Results

Familiarization data were analyzed separately for each age group to determine on which anatomical regions of the face the infants fixated during presentation of the audiovisual videos. To code familiarization looking time data, the screen to which infants were fixated was divided into regions of interest (ROI). Although ROIs were defined using static images of the moving faces, they were large enough to cover the entire facial region in question throughout the dynamic audiovisual presentation. One region of interest corresponded to the area surrounding the model's mouth (34.24 x 18.29 cm), and the other region of interest to the area surrounding her eyes (34.24 x 9.68 cm). Mean differences of looking to the eyes minus the mouth for all ages and conditions are visualized in Figure 2. Prior to analyzing familiarization data by age group, a three-way 3 (Age Group) x 2 (Condition) x 2 (Region) mixed-effects ANOVA was fitted to an aggregate dataset containing all of the familiarization data from the three age groups. A medium-sized three-way interaction between age group, condition, and region of interest emerged ($F(2,90) = 3.64$, $p = .030$, $\eta^2_P = .07$).

Test data were analyzed to probe auditory discrimination, and specifically to determine whether congruent or incongruent audiovisual familiarization had any effect on discrimination at test, as exhibited by a difference in looking time between alternating and non-alternating stimuli sequences. Of the 768 test trials across the three age groups (32 infants in three groups completed eight test trials each), the eye tracker did not capture looking time data for 15 trials, but no two trials of the same sequence type (alternating/non-alternating) were skipped in an individual infant's dataset. In order to analyze data from all subjects, these 15 points were replaced with each infant's sequence-type-specific mean looking time. Test data can be visualized in Figure 3 as differences between looking to alternating over non-alternating trials. Prior to analyzing test data separately by each age group, data were first analyzed in pairs of trials. The first pair consisted of the first and second test trials (one alternating and one non-alternating trial); the second pair consisted of the third and fourth test trials, and so on. A four-way 3 (Age Group) x 2 (Condition) x 2 (Sequence Type) x 4 (Pair) mixed-effects ANOVA revealed a significant effect of age group ($F(2,90) = 3.32$, $p = .041$, $\eta^2_P= .07$) and of pair ($F(3,270) = 33.60$, $p < .001$), $\eta^2_P= .27$), though no other main effects or interactions emerged as significant. Subsequent test analyses were then conducted separately for each age group.

[Insert Figure 2 here]

[Insert Figure 3 here]

3.1 Six-month-olds

A two-way 2 (Condition) x 2 (Region of interest) mixed-effects ANOVA was performed on the 6-month-olds' familiarization looking time data. There were no main effects of condition ($F(1,30) = .89$, $p = .353$, $\eta^2_P= .03$) or region of interest ($F(1,30) = .36$,

p = .554, η^2_P= .01), but a medium-sized interaction between condition and region of interest nearly reached significance (F(1,30) = 3.43, p = .074, η^2_P= .10). Infants familiarized to congruent stimuli deployed a greater proportion of their looking time to the eye region of the model's face (M_{eyes} - M_{mouth} = 1.48 seconds, SD = .65) than did infants familiarized to incongruent stimuli, who deployed a greater proportion of their looking time to the mouth (M_{eyes} - M_{mouth} = -2.89 seconds, SD = .72).

A three-way 2 (Condition) x 2 (Sequence type) x 4 (Pair) mixed-effects ANOVA was performed on the six-month-olds' test data. A main effect of pair emerged (F(3,90) = 7.39, p = .001, η^2_P= .20) indicating that infants looked progressively less to the screen as the test phase continued, a typical pattern in infant looking time studies. No main effect of condition (F(1,30) = .03, p = .858, η^2_P < .01) or sequence type (F(1,30) = .02, p = .884, η^2_P < .01) emerged, but a significant interaction between condition and sequence type (F(1,30) = 5.30, p = .028, η^2_P = .15) revealed that the six-month-olds familiarized to congruent stimuli exhibited a significantly different pattern of looking during test than did the infants familiarized to incongruent stimuli. As is visualized in Figure 3, infants familiarized to congruent stimuli looked longer during alternating test trials (M_{alt} - M_{non} = .66 seconds, SD = .59), while those familiarized to incongruent stimuli looked longer during non-alternating test trials (M_{alt} - M_{non} = -.77 seconds, SD = .67), accounting for the significant interaction between condition and sequence type. Although the differences between looking times to alternating and non-alternating sequences did not differ significantly in either group of six-month-olds when considered separately ($t_{\text{congruent}}$(15) = 1.44, p = .170; $t_{\text{incongruent}}$(15) = 1.12, p = .282), the significant interaction between condition and sequence type indicates that audiovisual familiarization (congruent and incongruent, respectively) affected later auditory perception.

3.2 Nine-month-olds

A two-way 2 (Condition) x 2 (Region of interest) mixed-effects ANOVA on the nine-month-olds' familiarization data revealed no main effect of condition ($F(1,30) = 1.49$, $p = .232$, $\eta^2_P = .05$). A main effect of region of interest emerged ($F(1,30) = 7.80$, $p = .009$, $\eta^2_P = .21$), indicating that infants in both familiarization conditions looked longer to the mouth region of the model's face than to the eye region ($M_{\text{eyes}} – M_{\text{mouth}} = -3.09$ seconds, $SD = .68$). As with the 6-month-olds, there was a medium-sized interaction between condition and region that nearly reached significance ($F(1,30) = 3.19$, $p = .084$, $\eta^2_P = .10$). Although both groups of infants looked more to the mouth, infants familiarized to incongruent stimuli ($M_{\text{eyes}} – M_{\text{mouth}} = -5.07$ seconds, $SD = .69$) did so more than did infants familiarized to congruent stimuli ($M_{\text{eyes}} – M_{\text{mouth}} = -1.12$ seconds, $SD = .68$).

A three-way 2 (Condition) x 2 (Sequence type) x 4 (Pair) mixed-effects ANOVA was performed on the nine-month-olds' test data. Again, the main effect of pair emerged ($F(3,90) = 12.21$, $p < .001$, $\eta^2_P = .29$) indicating that infants looked progressively less to the screen as the test phase continued. No main effect of condition ($F(1,30) = 1.37$, $p = .251$, $\eta^2_P = .04$) or sequence type ($F(1,30) = 1.17$, $p = .288$, $\eta^2_P = .04$) emerged, nor did the crucial interaction between condition and sequence type ($F(1,30) = .02$, $p = .887$, $\eta^2_P < .01$), revealing that both groups of nine-month-olds, regardless of familiarization condition, exhibited similar patterns of looking to the alternating and non-alternating test trials.

3.3 11-month-olds

A two-way 2 (Condition) x 2 (Region of interest) mixed-effects ANOVA on the 11-month-olds' familiarization data revealed a small effect of condition that nearly reached significance ($F(1,30) = 3.82$, $p = .060$, $\eta^2_P = .11$). Infants familiarized to congruent stimuli looked at the model's face more ($M = 12.73$ seconds, $SD = 3.37$) than infants familiarized to incongruent stimuli ($M = 10.32$ seconds, $SD = 3.61$). A large main effect of region of interest also emerged ($F(1,30) = 24.56$, $p < .001$, $\eta^2_P = .45$), indicating that infants in both familiarization conditions looked longer to the mouth region of the model's face than to the eye region ($M_{\text{eyes}} – M_{\text{mouth}} = -6.11$ seconds, $SD = .69$). Crucially, the interaction between condition and region of interest did not reach significance ($F(1,30) = 2.12$, $p = .156$, $\eta^2_P = .07$), indicating that the difference in amount of looking to the eyes versus to the mouth did not differ as a function of condition in the 11-month-olds.

A three-way 2 (Condition) x 2 (Sequence type) x 4 (Pair) mixed-effects ANOVA was performed on the 11-month-olds' test data. Again, a main effect of pair emerged ($F(3,90) = 16.38$, $p < .001$, $\eta^2_P= .35$) indicating that infants looked progressively less to the screen as the test phase continued. No main effect of condition ($F(1,30) = 1.91$, $p = .178$, $\eta^2_P = .06$) or sequence type ($F(1,30) = .01$, $p = .933$, $\eta^2_P < .01$) emerged, nor did the crucial interaction between condition and sequence type ($F(1,30) = .31$, $p = .582$, $\eta^2_P = .01$), revealing that both groups of 11-month-olds, regardless of familiarization condition, exhibited similar patterns of looking to the alternating and non-alternating test trials.

4 General Discussion

In this research, we addressed two questions about infants' processing of multisensory speech sounds before, during, and/or after perceptual attunement. First, we examined whether infants are sensitive to the audiovisual congruence of seen and heard speech in an unfamiliar language. Secondly, we explored whether prior exposure to audiovisual speech sounds can influence infants' ability to discriminate purely auditory speech sounds.

Regarding our first question, we hypothesized that infants in the early stages of perceptual attunement would be sensitive to the congruence of the auditory and visual signals while viewing non-native speech, as measured by differences in their looking to distinct regions of the face during familiarization, while older infants would not exhibit such sensitivity. Indeed, analysis of the familiarization data from the six-month-olds revealed a nearly significant, medium-sized interaction between familiarization condition and region of interest (eyes versus mouth). As demonstrated by their increased visual fixation to the mouth region of a speaker's face when observing incongruent audiovisual speech, this result suggests that six-month-old infants may have detected the content congruence of heard and seen speech. These findings extend Tomalski and colleagues' (2013) finding that six- to nine-month-old infants focus attention from the eye region to the mouth region of the face when perceiving incongruent speech in their own native language. In the current study, six-month-old infants also deployed a greater proportion of their visual fixations to the mouth region of a speaker's face when watching her produce incongruent, non-native audiovisual speech than while watching her produce congruent versions of the same unfamiliar speech events. Notably, nine-month-old infants, in the midst of perceptual attunement and at an age when auditory discrimination of non-native sounds has declined, also exhibited such a moderate pattern of detection. Although nine-month-olds, as a group, looked longer to the mouth region of the model's face (a result that is consistent with the findings of Lewkowicz and Hansen-Tift (2012), and which may be a result of the emergence of productive language at around this age), those familiarized to incongruent audiovisual speech did so more than those familiarized to congruent stimuli. Finally, these findings go beyond prior work in showing that 11-month-old infants, having concluded perceptual attunement, exhibited no pattern of incongruence detection[1]. Taken together, these incongruence detection results indicate that the sensitive period for congruence detection in audiovisual speech is similar to, but may last somewhat longer into ontogeny than, the sensitive period estimated by previous research from purely auditory evaluation of speech sound discrimination.

The present finding that six- and nine-month-olds were sensitive to the (in)congruence between the auditory and visual signals in an unfamiliar language is especially striking because the onsets and offsets of the visual and auditory signals were aligned in both the congruent and the incongruent speech stimuli. This rules out the possibility that infants were sensitive to incongruence simply via detection of a temporal mismatch in the audiovisual signal. Instead, it suggests that infants are sensitive to the congruence of finer details in the acoustic and visual signals, despite having had no prior experience with these specific speech sound contrasts.

[1] It is further noteworthy that the current findings from three separate age groups are generally consistent with the age-related changes in face-scanning patterns recorded by Lewkowicz and Hansen-Tift (2012), despite the different angles at which the models' faces in the two studies were presented: 0° (directly facing) versus a 45° rotation in our study.

While it is probable that the neural architecture involved in speech perception in the infant, like in the adult (Campbell, 2008), supports links between heard and seen speech, it is difficult to explain how the mapping could be so precise without specific experience as to enable detection of the differences between congruent versus incongruent auditory-visual dental ([d̪a]) vs retroflex ([ɖa]) speech syllables. One possibility is that infants' sensitivity to audiovisual congruence is mediated by information from infants' proprioception of their own pre-verbal oral-motor movements. Even prenatally, infants engage in frequent sucking and swallowing behaviour (Arabin, 2004; Kurjak et al., 2005), which provides corresponding acoustic information (see also Choi, Kanhadhai, Danielson, Bruderer, & Werker, in press). Moreover, prior to the age at which infants were tested in the current experiments, they begin to produce primitive vocalizations (Oller, Eilers, Neal, & Schwartz, 1999), and their own oral-motor movements affect their discrimination of unfamiliar speech sounds at six months of age (Bruderer et al., 2015). Indeed, one recent study has demonstrated that 4.5-month-old infants' articulatory configurations affect their matching of heard and seen speech, an effect that varies as a function of the specific oral-motor gesture that the infant makes (Yeung & Werker, 2013). Such results advance the proposal that infants' robust audiovisual speech perception may be grounded in early sensorimotor perception (Guellaï, Streri, & Yeung, 2014). Although infants in the current studies had not experienced the specific sound-sight pairings of Hindi in a language-learning environment, their endogenous experience with their own oral-motor movements (and corresponding acoustic productions), in addition to their experience perceiving audiovisual speech in their native language, may have provided them with sufficient information to establish a mapping of the relation between heard and seen speech. This in turn may have enabled them to detect the congruence in unfamiliar audiovisual speech.

In addressing our second question, we hypothesized that exposing six-month-old infants prior to perceptual attunement to incongruent—but not congruent—visual information would change their subsequent auditory discrimination of the non-native contrast. For infants undergoing perceptual attunement (at nine months), we predicted that congruent —but not incongruent— visual information would extend the observable sensitive period for non-native phoneme discrimination and thus boost subsequent auditory discrimination of the speech sounds. Finally, we predicted that familiarization with audiovisual stimuli would not affect the subsequent auditory discrimination of the 11-month-old infants, regardless of whether the stimuli were presented congruently or incongruently.

Analysis of the six-month-olds' discrimination data provided evidence that the addition of incongruent visual information changed auditory perception at test such that infants familiarized in that condition exhibited a different pattern of looking time to alternating versus non-alternating trials than did infants familiarized to congruent stimuli. Importantly, the content of the visual information appears to be crucial: while familiarization to congruent visual information resulted in the maintenance of auditory discrimination at six months, incongruent visual information changed the pattern of discrimination at this age.-Therefore, prior to perceptual attunement, infants' auditory speech perception is altered by visual information, advancing the proposal that infants' perception of speech is audiovisual.

Contrary to our predictions, the analysis of the nine-month-old and 11-month-old test data revealed no interaction between condition and sequence type, indicating that auditory perception of non-native speech sounds may only be affected by the addition of visual information prior to perceptual narrowing. Regardless of how they were familiarized, nine- and 11-month-old infants' auditory discrimination patterns at test were not altered by familiarization to audiovisual speech.

The present findings augment a growing body of recent work aimed at better understanding sensitive periods in language learning from a multisensory perspective. Like ours, a few of these studies have similarly probed whether the addition of visual information to the speech signal would change auditory discrimination as the sensitive period for speech sound discrimination closes. For example, it was recently found that adding a visual display of a speaker producing either an [æ] or an [ɛ] to an auditory training procedure improved Dutch 8-month-old infants' sensitivity to this distinction, which they otherwise no longer discriminate at this age (Schure, Junge, & Boersma, 2016). Using a similar set of speech sounds as the ones used in these experiments, another study succeeded in changing infants' auditory discrimination after pairing the sounds with visual objects (Yeung & Werker, 2009; see also Yeung, Chen, & Werker, 2014). In that study, nine-month-old infants were familiarized to sight-sound pairings consisting of one visual novel object paired with one of the Hindi speech sounds (a voiced dental or a voiced retroflex consonant), and a second visual object consistently paired with the other Hindi speech sound. Although the sight-sound pairings were arbitrary, infants exhibited increased discrimination of the auditory speech sounds after familiarization to the object-sound pairings. However, that study used objects, not visual articulations; to our knowledge, no study to date has probed whether the congruence between seen and heard speech influences the manner in which perception is affected by the addition of visual information to the auditory signal.

Our results uniquely contribute to the understanding of how visual and auditory information interact in infant speech perception by demonstrating the differential impact of congruent versus incongruent visual articulatory gestures on auditory discrimination of speech. Unlike with arbitrary object-sound pairings, it is not simply the consistent temporal co-occurrence of a sight and a sound in audiovisual speech that can change auditory discrimination. Rather, speech perception is multisensory from early in life, and experiencing the non-arbitrary content congruence between the auditory and visual signals of speech shapes the developmental trajectory of speech perception across modalities.

5 Conclusion

The timing of perceptual attunement to the speech sounds of the native language typically follows a very consistent time course, leading to the proposal that it is best described as a sensitive or critical period in development. Here we demonstrate that infants' detection of audiovisual congruence in non-native speech follows a similar pattern of perceptual attunement, but may last somewhat longer into ontogeny than the sensitive period for unisensory speech contrasts. Moreover, these results indicate that the characteristics of auditory speech perception can be changed by pre-exposure to congruent or incongruent audiovisual speech, but only up to a point in development (about six months of age) when sensitivity to the auditory contrast remains evident. Taken together, these results suggest that our current understanding of perceptual attunement of speech can be deepened by considering sensitive periods in a richer, multisensory environment.

Acknowledgments

The authors are thankful to [names removed for review] for their contributions to data collection, manuscript proofreading, and artwork.

Funding

This work was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health [grant number removed for review] and the Acoustical Society of America [grant number removed for review].

References

Apple Inc. (2013). iMovie (Version 9.0.9) [Computer software].

Arabin, B. (2004). Two-dimensional real-time ultrasound in the assessment of fetal activity in single and multiple pregnancy. *The Ultrasound Review of Obstetrics & Gynecology, 4*(1), 37-46. doi:10.1080/14722240410001700258

Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development, 21*, 295. doi:10.1016/s0163-6383(98)91508-9

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences, 112*(44), 13531-13536. doi:10.1073/pnas.1508631112

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220. doi:10.1002/dev.20032

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1001-1010. doi:10.1098/rstb.2007.2155

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379-383. doi:10.1111/j.0956-7976.2004.00688.x

Choi, D., Kandhadai, P., Danielson, D.K., Bruderer, A.G., & Werker, J.F. (in press). Does early motor development contribute to speech perception? [Peer commentary on “Neonatal Imitation in Context: Sensory-Motor Development in the Perinatal Period” by N. Keven & K.A. Akins]. *Behavioral and Brain Sciences*.

Cohen, J., Macwhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25*(2), 257-271. doi:10.3758/bf03204507

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*(4), 187-203. doi:10.1002/dev.20033

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*(1), 567-631. doi:10.1146/annurev.neuro.22.1.567

Friederici, A. D., & Wartenburger, I. (2010). Language and brain. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 150-109. doi:10.1002/wcs.9

Grant, K. W., & Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*(3), 1197-1208. doi:10.1121/1.1288668

Guellaï, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology, 5*(812), 1-7. doi:10.3389/fpsyg.2014.00812

Haith, M., Bergman, T., & Moore, M. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853-855. doi:10.1126/science.918670

Hall, D. G. (1991). *Perceptual and Associative Learning*. Oxford: Clarendon Press.

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development, 76*(3), 598-613. doi:10.1111/j.1467-8624.2005.00866.x

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy, 6*(2), 231-255. doi:10.1207/s15327078in0602_5

Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089. http://doi.org/10.1111/j.1467-9280.2007.02029.x

Kubicek, C., Boisferon, A. H., Dupierrix, E., L Venbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development, 37*(2), 106-110. doi:10.1177/0165025412473016

Kubicek, C., Boisferon, A. H., Dupierrix, E., Pascalis, O., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS ONE, 9*(2). doi:10.1371/journal.pone.0089275

Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138-1141. doi:10.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development, 7*(3), 361-381. doi:10.1016/s0163-6383(84)80050-8

Kuhl, P. K., Tsao, F., Liu, H., Zhang, Y., & Boer, B. (2006). Language/Culture/Mind/Brain. *Annals of the New York Academy of Sciences, 935*(1), 136-174. doi:10.1111/j.1749-6632.2001.tb03478.x

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713-727. doi:10.1016/j.neuron.2010.08.038

Kurjak, A., Stanojevic, M., Azumendi, G., & Carrera, J. M. (2005). The potential of four-dimensional (4D) ultrasonography in the assessment of fetal awareness. *Journal of Perinatal Medicine, 33*(1), 46-53. doi:10.1515/jpm.2005.008

Lawrence, D. H. (1949). Acquired distinctiveness of cues: Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology, 39*(6), 770-784. doi:10.1037/h0058097

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences, 103*(17), 6771-6774. doi:10.1073/pnas.0602027103

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109*(5), 1431-1436. doi:10.1073/pnas.1114783109

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: newborns match nonhuman primate faces and voices. *Infancy, 15*(1), 46-60. doi:10.1111/j.1532-7078.2009.00005.x

Mackain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science, 219*(4590), 1347-1349. doi:10.1126/science.6828865

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-141. doi:10.3109/03005368709077786

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America, 100*(3), 1777. doi:10.1121/1.417342

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265. doi:10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367-1381. doi:10.1016/j.cognition.2007.07.002

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56*(2), 154-178. doi:10.1002/dev.21177

Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111. doi:10.1016/S0010-0277(01)00157-3

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748. doi:10.1038/264746a0

Merin, N., Young, G. S., Ozonoff, S., & Rogers, S. J. (2006). Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants. *Journal of Autism and Developmental Disorders, 37*(1), 108-121. doi:10.1007/s10803-006-0342-4

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science, 13*(3), 407-420. doi:10.1111/j.1467-7687.2009.00898.x

Norcross, K. J., & Spiker, C. C. (1957). The effects of type of stimulus pretraining on discrimination performance in preschool children. *Child Development, 28*(1), 79-84. doi:10.1111/j.1467-8624.1957.tb04833.x

Oller, D., Eilers, R. E., Neal, A., & Schwartz, H. K. (1999). Precursors to speech in infancy. *Journal of Communication Disorders, 32*(4), 223-245. doi:10.1016/s0021-9924(99)00013-1

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the first year of life. *Child Development*, *83*(2), 543-553. doi:10.1111/j.1467-8624.2011.01715.x

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development, 22*(2), 237-247. doi:10.1016/s0163-6383(99)00003-x

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*(1), 93-115. doi:10.1006/jecp.2001.2644

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*(2), 191-196. doi:10.1111/1467-7687.00271

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435. doi:10.1037/0096-1523.20.2.421

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences, 106*(26), 10598-10602. doi:10.1073/pnas.0904134106

Reese, H. W. (1972). Acquired distinctiveness and equivalence of cues in young children. *Journal of Experimental Child Psychology, 13*(1), 171-182. doi:10.1016/0022-0965(72)90017-3

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception & Psychophysics, 52*(4), 461-473. doi:10.3758/bf03206706

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59*(3), 347-357. doi:10.3758/bf03211902

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

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212. doi:10.1121/1.1907309

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*(3), 850-855. doi:10.1016/j.cognition.2008.05.009

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

Tobii Technology (2008). Tobii Studio (Version 1.7.3). [Computer software].

Tomalski, P., Ribiero, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology, 10*(5), 611-624. doi:10.1080/17405629.2012.728076

Vatikiotis-Bateson, E., Eigsti, I., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisualspeech perception. *Perception & Psychophysics, 60*(6), 926-940. doi:10.3758/bf03211929

Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences, 106*(44), 18867-18872. doi:10.1073/pnas.0906049106

Walton, G. E., & Bower, T. (1993). Amodal representation of speech in infants. *Infant Behavior and Development, 16*(2), 233-243. doi:10.1016/0163-6383(93)80019-5

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159-1159. doi:10.1126/science.1137686

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development, 52*(1), 349-355. doi:10.2307/1129249

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63. doi:10.1016/s0163-6383(84)80022-3

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*(5), 672-683. doi:10.1037/0012-1649.24.5.672

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, 46*(3), 233-251. doi:10.1002/dev.20060

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology, 66*(1), 173-196. doi:10.1146/annurev-psych-010814-015104

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition, 113*(2), 234-243. doi:10.1016/j.cognition.2009.08.010

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123-139. doi:10.1016/j.jml.2012.09.004

Yeung, H. H., Chen, L. M., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development, 85*(3), 1036-1049. doi:10.1111/cdev.12185

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*(5), 798-814. doi:10.1111/j.1467-7687.2009.00833.x

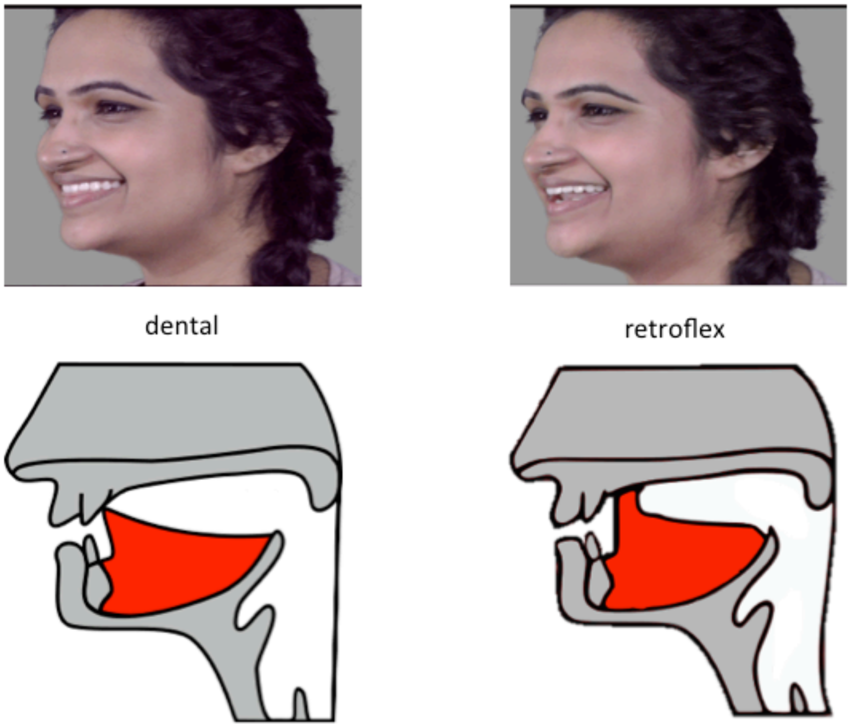

Figure 1. Still frames and corresponding schematics of model producing dental and retroflex consonants.

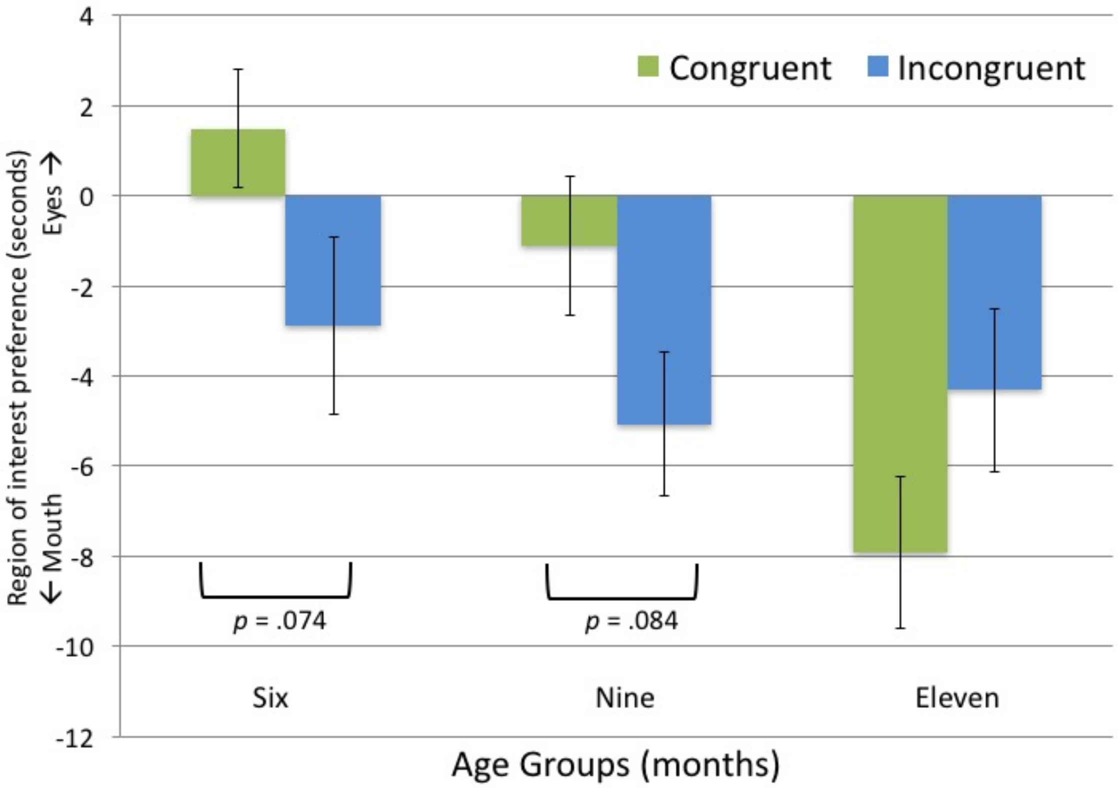

Figure 2. Differences in looking time during familiarization to the eye region of the model’s face minus the mouth region. Positive scores indicate preference for the eye region. Error bars are +/- one standard error of the mean difference in looking times.

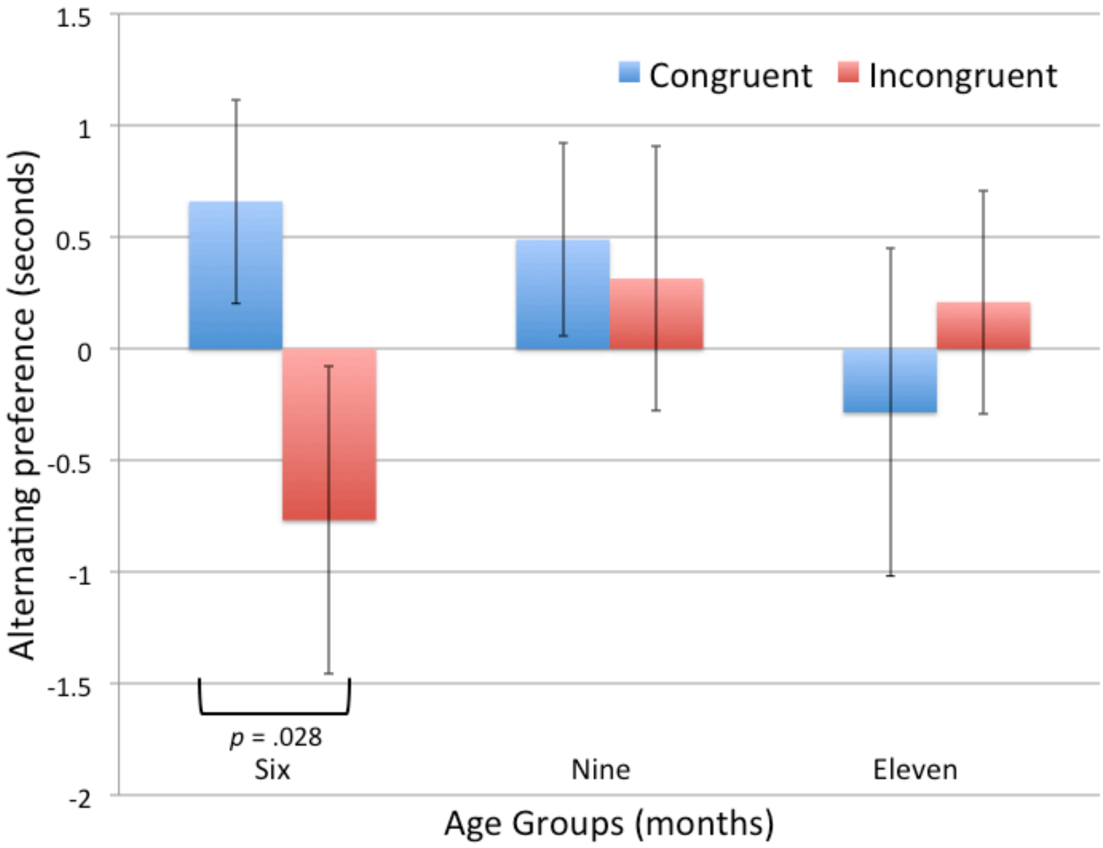

Figure 3. Differences in looking time during test to alternating minus non-alternating sequences. Positive scores indicate preference for alternating sequences. Error bars are +/- one standard error of the mean difference in looking times.