

Spring 5-8-2011

# Language-Specific Tuning of Audiovisual Integration In Early Development

Juliana Flynn

*University of Connecticut - Storrs*, jumarielynn@gmail.com

Follow this and additional works at: [http://digitalcommons.uconn.edu/srhonors\\_theses](http://digitalcommons.uconn.edu/srhonors_theses)



Part of the [Communication Sciences and Disorders Commons](#)

---

## Recommended Citation

Flynn, Juliana, "Language-Specific Tuning of Audiovisual Integration In Early Development" (2011). *Honors Scholar Theses*. 190.  
[http://digitalcommons.uconn.edu/srhonors\\_theses/190](http://digitalcommons.uconn.edu/srhonors_theses/190)

Running Head: LANGUAGE-SPECIFIC TUNING

Language-Specific Tuning of Audiovisual

Integration In Early Development

Juliana Flynn

University of Connecticut

Abstract

According to the perceptual narrowing hypothesis, older infants look longer towards speech in a native language than towards a non-native language. We presented speech in English, Spanish, and mis-matched English and Spanish speech, and recorded looking-time towards the speech. Results suggest that the synchrony of speech plays a strong role in infants' attention to speech, whereas nativeness of language does not.

### Language-Specific Tuning of Audiovisual Integration In Early Development

From birth, an infant's experience in the world is multisensory. That is, infants integrate audio, visual, and other sensory information into unified percepts. However, this was not always believed to be the case.

Initially, developmental psychologists proposed that only basic unisensory perceptual abilities were present at birth, and multisensory perception emerged gradually with experience; infants' unisensory experiences contributed to infants' integration of unisensory experiences into multisensory perception (Piaget, 1952; Birch and Lefford, 1967; both as described by Lewkowicz and Ghazanfar, 2009). This view seems relatively unsupported, however, as recent research suggests newborns integrate the visual and auditory speech information of their mothers (Sai, 2005), and likewise infants as young as four months are sensitive to multisensory synchrony in monkey vocalizations (Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastian-Galles, and Werker, 2007). Consistent with the evidence that infants are sensitive to multisensory information from birth, most studies seem supported by a second view, that basic multisensory abilities are present at birth and become increasingly differentiated or refined with experience (Gibson, 1984; as cited by Lewkowicz and Ghazanfar, 2009). The only difference in these views is the question of whether infants are born with basic multisensory perceptual abilities, and generally both views are considered progressive by proposing that perceptual ability increases with experience during development (Lewkowicz and Ghazanfar, 2009).

In contrast to these two views, a third view posits that broad basic multisensory perceptual abilities are present at birth, and these abilities are selectively narrowed according to infants' typical environment, to the effect that infants' sensitivity to environmentally irrelevant (atypical) information decreases while sensitivity to environmentally relevant (typical) information increases (Lewkowicz and Ghazanfar, 2009; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk, 1997; Werker, & Tees 1984, 2005, as cited by Lewkowicz and Ghazanfar, 2009; Pons, Lewkowicz, Soto-Faraco, and Sebastián-Gallés, 2009; Scott, Pascalis, & Nelson, 2007). This perceptual narrowing process is argued by Lewkowicz and Ghazanfar

(2009) to be a regressive process because the result is a decrease in infants' responsiveness to atypical or non-native sensory information (e.g., non-native languages, other-race faces)

Whichever view is more accurate, there is no question that infants' multisensory experiences propel the visual, auditory, and other sensory systems to develop, allowing infants to make sense of the faces, voices, and other information often present in their environments. Due to the multisensory nature of speech perception, early language learning is likely influenced by infants' perceptual development. Likewise, infants' language development may provide insight into the mechanisms involved in perceptual development by highlighting consistencies and inconsistencies between language development and the contrasting frameworks of perceptual development. By focusing on the multisensory aspects of speech perception, the goal of this study is to shed light on infant perceptual development.

Most infant studies have supported the notion that perceptual narrowing affects infants' development across a broad range of abilities such as intersensory perception (i.e., perceiving audiovisual stimuli such as talking faces), phonemic perception, face perception, and musical rhythm perception, which suggests that the narrowing process may be domain general (Scott et al., 2007; Lewkowicz and Ghazanfar, 2009; Pons et al., 2009).

In an intersensory perception study by Lewkowicz and Ghazanfar (2006), infants aged four, six, eight, and 10 months were presented with paired images of vocalizing monkey faces while listening to one of the two vocalizations. Consistent with the perceptual narrowing hypothesis, only the four and six-month-olds correctly paired (indicated by longer looking) the auditory vocalization with the correct monkey face. On the other hand, newborn infants correctly matched monkey facial gestures with vocalizations significantly better than either four-month-olds or 12-month-olds (Lewkowicz, Leo, & Simon, 2010, as described by Lewkowicz and Ghazanfar, 2009). Furthermore, Weikum and colleagues (2007, also described by Scott et al., 2007) found that English and French monolingual four- and six-month-olds discriminated both French and English visual speech (i.e., mouth and lip-movements), whereas

at eight months of age, only the bilingual French-English infants discriminated the visual speech in both languages.

In phonemic perception studies, infants younger than six months of age discriminated phonetic contrasts in both native and non-native languages, whereas the ability to discriminate non-native languages declined between six and twelve months (Kuhl et al., 1992, as described by Pons et al., 2009; Werker, and Tees, 2005, as described by Scott et al., 2007). For example, six to eight month old infants whose native language is English distinguished the Hindi phonemic contrasts /Da/ and /da/, but native English-exposed infants aged 10 to 12 months could not distinguish the non-native contrasts (Werker, and Tees, 1984, as described by Lewkowicz and Ghazanfar, 2009). Werker and Tees (1984, as described by Lewkowicz and Ghazanfar, 2009) concluded from this study that the decline in discrimination of non-native phonemic contrasts was perhaps driven by infants' frequent language experience with native contrasts, compared to rare experience with non-native contrasts. Thus, perceptual narrowing to native speech sounds is an experience-dependent process, that is, learning a specific language requires experience with that language (Lewkowicz and Ghazanfar, 2009).

Evidence from face-perception studies has suggested a similar narrowing process. In a visual paired-comparison task that involved familiarizing infants to a single picture (in this case, a monkey or human face) and next pairing the familiar face with a novel face (same species as the familiar face), infants at six months of age showed a novelty preference for both monkey and human faces, whereas nine-month-olds and adults showed a novelty preference for human faces only (Pascalis, de Haan, and Nelson, 2002, as described by Scott et al., 2007). Pascalis et al. (2007, as described by Scott et al., 2007) suggested that by nine months of age, infants' face perception has narrowed to process faces of their own species, and to ignore faces outside their species. The same result was found in infants who viewed unfamiliar race faces; three- and six-month-olds showed discrimination of other-race faces, but nine-month-olds only showed discrimination of faces within their own race (Kelly, Quinn, Slater, Lee, Ge, and Pascalis, 2007, as described

by Scott et al., 2007). Furthermore, three-month-old Caucasian infants showed greater preference for female faces than for male faces when the faces were Caucasian (i.e., same race), whereas the infants did not show a preference for female faces when the faces were Asian (Quinn, Uttley, Lee, Gibson, Smith, Slater, Pascalis, 2008, as described by Lewkowicz and Ghazanfar, 2009).

Infant studies investigating musical rhythm perception have suggested similar findings. Consistent with the perceptual narrowing hypothesis, Hannon and Trehub (2005a, 2005b, as described by Scott et al., 2007) found that North American six-month-olds noticed violations in both North American music and Bulgarian music, but infants aged 12-months and adults detected violations in the North American music, but not in the Bulgarian music. These findings suggest that after 12 months of age, infants became perceptually tuned to musical rhythm structures relevant to their own culture, to the extent that rhythm structures of other cultures were less optimally processed.

Although the mechanisms by which infants' perception narrows to their native language are unclear, Scott et al. (2007) hypothesized that the increased pruning of exuberant synaptic connections after six months of age may play a role in the narrowing of infants' perceptual abilities between six and 12 months. Likewise, Scott et al. suggested that Hebbian Learning, which works to strengthen neural circuits and synaptic connections that are used most, and weaken ones that are used least, may push infants to selectively attend to environmentally relevant information (compared to irrelevant information). On the other hand, Lewkowicz et al. (2009) suggested that perceptual narrowing may be the result of 'selective elaboration' of synapses activated most frequently and depending on the infants' experiences, rather than selective pruning of less used synapses. In support of this hypothesis, Lewkowicz et al. pointed to size increase in primates' brains (across species) long after birth, and the increasing density of axons and dendrites with development. Thus, increased sensitivity to relevant information and decreased sensitivity to irrelevant information may be the result of new neural connections, rather than a loss in connections. However, it seems more likely that perceptual narrowing to relevant environmental information is a result

of both the pruning away of synaptic connections that are activated least, as well as increasing complexity of axonal and dendritic structures that are most frequently activated.

Most studies investigating the neural correlates of perceptual narrowing have involved speech and face perception measured by event-related potentials (ERPs). ERPs measure electrical activity of simultaneously active populations of neurons recorded on the scalp, typically in response to images or sounds. Importantly, they are time-locked to significant events such as onset of stimulus presentation, and they are a reliable measure of infant neural processing (Csibra, Kushnerenko, & Grossmann, 2008; Scott et al., 2007). In a longitudinal phonemic perception study conducted by Rivera-Gaxiola and colleagues (2005, as described by Scott et al., 2007), ERPs were recorded in infants presented with native versus non-native speech contrasts at seven months and again at 11 months. The infants responded electrophysiologically to both contrasts by seven months of age, but by 11 months some infants within this group were significantly slower at processing the non-native contrasts compared to the native ones, whereas other infants in the group could process both native and non-native contrasts to a similar degree (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005, as described by Scott et al., 2007). In sum, although some of the infants' neural processing of non-native contrasts was diminished by 11 months, neural discrimination was still maintained for the other infants. This effect of non-native processing in 11-month-olds may not be strong enough to be detected through behavioral measures, whereas neural measures can point to more specific changes in processing. A similar effect was found in face perception, such that although ERPs recorded in nine-month-olds showed differential response to human and monkey faces, there was also neural differentiation of familiar and unfamiliar monkey faces (Scott, Shannon, and Nelson, 2006, as described by Scott et al., 2007). Likewise, this suggests that neural measures may inform behavioral studies by showing finer processing differences, which may suggest that perceptual narrowing is more gradual than it seems.

In tune with the idea that perceptual narrowing may be a more gradual process, many recent infant studies have provided support that experience can shape the narrowing process and allow specialization for

infrequently encountered stimuli such as other-species faces (see Scott and Monesson, 2010) and non-native speech contrasts (see Kuhl, Tsao, and Liu, 2003, as described by Lewkowicz and Ghazanfar, 2009). Werker and Tees (2005, as described by Lewkowicz, 2009) suggested that perceptual tuning to environmentally relevant information may be achieved through processes that result in gradual reorganization of neural structures, rather than permanent loss of function or structures. For example, three-month-old infants presented with faces activated brain areas typically devoted to language in adults (Tzourio-Mazoyer, De Schonen, Crivello, Reutter, Aujard, and Mazoyer 2002, as described by Scott et al., 2007). This suggests that infants may initially use different brain regions than adults, which become reorganized to different areas through experience. This further emphasizes the role of experience in the development of perception.

Given the evidence that infants tune to multiple sources of perceptual information about their language from a very early age, it is informative to consider whether infants may become sensitive to the perceptual cues (i.e., auditory and visual signals) of spoken language (native or non-native) after eight months of age (when perceptual narrowing is evidenced to begin). The majority of research on perceptual tuning has focused exclusively on auditory speech that is presented without corresponding visual information, in spite of the fact that a speaking face provides surprising amounts of information that observers incorporate into their processing of human speech. Likewise, most of this research has involved contrastive phonemic pairings (i.e., isolated speech sounds), and not fluent speech (i.e., spoken sentences), thus the results of previous studies may be failing to capture infants' perceptual capabilities, such that infants are accustomed to perceiving fluent speech, but not speech isolated into single syllables. Thus, it would be informative to present to infants fluent speech in concurrent auditory and visual modalities.

The objective of this study was to determine whether older infants, between eight and 16 months of age, remain sensitive to the auditory and visual signals of fluent speech in a native as well as a non-native language. We presented to infants fluent speech in the form of congruent and incongruent paired audio

and visual signals; that is, the auditory signals either matched the visual signals (congruent condition), or did not match the visual signals (incongruent condition). Within the congruent speech conditions, infants were presented a native language (English) and a non-native language (Spanish), while the incongruent speech conditions involved auditory and visual speech from both languages (i.e., English audio paired with Spanish visual speech, and Spanish audio paired with English visual speech). Infants' sensitivity towards the congruency of speech was measured by the length of looking-time (in seconds) towards the presentation screen.

If infants looked towards the congruent speech for longer time than towards the incongruent speech, then this may suggest that they perceived the relationship between the auditory speech sounds and the corresponding mouth and lip-movements. According to the perceptual narrowing hypothesis and research on infant speech perception, monolingually English-exposed infants aged eight months and older should be able to discriminate speech presented in a native language, but not a non-native language, that is, infants should look longer during the congruent native language segments, whereas during the congruent non-native segments infants should look for a lesser length of time. In contrast to the perceptual narrowing hypothesis, and given that previous research on infant speech perception has typically involved isolated phonemic contrasts in only one modality (i.e., either auditory or visual but not paired audiovisual)= speech) we hypothesized that infants should look longer towards the congruent speech segments compared to the incongruent speech segments, regardless of language. That is, we did not expect to find a significant difference in mean proportion looking-time towards the congruent native speech versus congruent non-native speech, whereas we did expect a significant difference between the mean proportion looking-time towards congruent versus incongruent segments overall. If this is the case, then this should present a challenge to the perceptual narrowing hypothesis by suggesting that older infants seem able to discriminate speech in a non-native language.

### **Method**

*Participants:*

Infants were recruited from towns within a 50-mile radius of Storrs, CT via birth announcements found in local papers. Most participants were Caucasian, although several races were represented. Infants were monolingually English-exposed since birth and were screened for ear infections.

A total of 12 infants aged eight to 16 months ( $M=11.92$ ,  $SD=2.67$ ) participated. The sample included 6 girls and 6 boys.

*Materials:*

A computer monitor was used to present the visual stimuli, while computer speakers were used to present the auditory stimuli. The audio for all video segments was uniformly scaled to 70 dB and played at 65-70 dB. The height and width of the monitor used was 14" by 17", and screen size was 10.5" by 13.5". A JVC 45x dynamic zoom Everio digital camcorder as well as a surveillance camera were used to record infants' looking at the computer monitor. The surveillance camera allowed experimenters to watch, in real time and in a separate room, the caregiver and infant during trial presentation. The caregiver listened to music through Bose QuietComfort noise-canceling headphones during trial presentation. Additionally, a microphone was wired to the stereo, which allowed experimenters to communicate to the caregiver through the headphones in case of infant fussiness.

The stimulus involved a video presentation (see Appendix for specific trial type information) that consisted of audiovisual spoken sentences from children's stories in English and translated into Spanish. Each story segment was read by one woman fluent in English and Spanish. The same eight story segments (each between 12-25 seconds long) were contained in each video, and each followed an 'attention-getter' silent video 10 seconds in length designed to center infants' looking behavior.

The speech segments each contained visual speech (i.e., mouth and lip-movements, see Figure 3 for still-frame of visual speech stimulus) paired to auditory speech (i.e., speech sounds). The visual speech was paired to either congruent auditory speech (i.e., same language as visual; congruent condition) or

incongruent auditory speech (i.e., different language than visual; incongruent condition). Each speech pair contained the same exact content for both visual and auditory speech. In congruent speech, the visual mouth and lip-movements was synchronous with the auditory speech, thus congruent conditions contained the same language and same content (i.e., meaning), whereas in incongruent speech, only the content was the same in the visual and auditory speech. That is, in non-matched speech, when the visual speech was in English, the auditory speech contained the same sentences as the visual speech that were translated into Spanish; likewise, when the visual speech was in Spanish, the auditory speech contained the same sentences as the visual speech that were originally in English.

Speech segment order was pseudo-randomized and counterbalanced across eight trial types so that all infants were presented with two congruent speech segments and two incongruent speech segments for each language, thus each video contained four congruent and four incongruent speech segments. Incongruent speech conditions were collapsed across participants for analyses. Each trial type was a total of three minutes and 35 seconds in length.

In order to center the infant before each speech segment, a 10-second silent 'attention-getter' video (composed of brightly colored moving shapes across colored backgrounds) was presented before each speech segment, for a total of eight attention-getters. Several different attention-getter videos were used (different colored backgrounds and shapes, and different shapes) in order to maintain infants' interest, but the order of these 'attention-getter' videos were kept consistent across conditions.

*Procedure:*

After the infants' caregiver arrived at the lab, the experimenter briefly explained the goals and procedure of the study and obtained consent for their infants' participation. Next, the caregiver and infant were led into the testing room and seated in front of the screen, the cameras were turned on, and headphones were given to the caregiver to put on. The participant number, condition number and date were recorded on a whiteboard and held in front of the camera before trial onset.

Infants were held in their caregivers' laps (alone in testing room) facing the monitor, while the experimenter watched from the control room via surveillance camera. Infants were presented with one of the eight videos (i.e., trial types one through eight) while being videorecorded. After the video was finished, caregivers were thanked for their participation and given a book or toy for their infant.

*Analyses:*

Infants' looking-times (in seconds) at the monitor for each video segment was measured. Looking-times were organized by the type of video segment: Congruent English, Congruent Spanish, and Incongruent (both auditory Spanish paired with visual English, and auditory English paired with visual Spanish). Looking-times towards attention-getters were recorded to ensure that infants were paying attention to the stimulus. Looking-times for each type of video segment were recorded and totaled (Total Towards Overall Congruent, Total Towards Congruent English, Total Towards Congruent Spanish, Total Towards Incongruent, Total Toward Attention-Getters). These totals were divided by the total length (in seconds) of each respective type of video segment to obtain the proportion of time that infants looked towards each type of video segment compared to the total lengths of each video segment type. Thus, the measures used in analyses were the Proportion Time Overall Congruent (i.e., Total Towards Congruent English plus Total Towards Congruent Spanish, then divided by Total English Video plus Total Spanish Video), Proportion Time Congruent English (i.e., Total Towards Congruent English divided by Total English Video), Proportion Time Congruent Spanish (i.e., Total Towards Congruent Spanish divided by Total Spanish Video), and Proportion Time Incongruent (i.e., Total Time Towards Incongruent divided by Total Incongruent Video).

Comparisons were made between Mean Proportion Time Overall Congruent, Mean Proportion Time Congruent English, Mean Proportion Time Congruent Spanish, and Mean Proportion Time Incongruent. The mean proportion looking-time conditions were also compared across the eight trial types as well as across gender (female, male), to ensure that congruency or language effects (or lack of) were not

the result of variability between participant trial types or gender.

### Results and Discussion

The means and standard deviations of infants' proportion of looking-time towards congruency conditions across participants are presented in Table 1, and mean proportion looking-time towards congruency conditions across participants are presented in Figure 1. Scrutiny of the table and figure suggests that infants' mean proportion looking-time towards congruent speech was greater than towards incongruent speech, and likewise that infants' mean proportion looking time towards congruent English was nearly equal to their looking-time towards congruent Spanish. One-way ANOVAs were performed to compare infants' proportion of looking-time towards congruent and incongruent conditions across the trial types and gender to ensure that these did not confound the results. Then, paired-samples t-tests were performed to investigate whether significant differences existed between infants' mean proportion looking-time towards congruent speech (English, Spanish) and incongruent speech.

One-way ANOVAs were performed to compare proportion of looking-time towards congruent conditions (Overall, English, Spanish) and incongruent speech across the eight trial types to ensure that the analyses were not confounded by any differences between each trial type as well as gender. The comparisons between congruent and incongruent speech across the eight trial types were not significant. Thus, there were not significant effects of presentation order for congruent speech nor for incongruent speech across participants. Likewise, none of the comparisons between genders across congruent and incongruent speech were significant.

A paired-samples t-test was conducted to compare infants' mean proportion of looking-time towards overall congruent speech (English + Spanish), and mean proportion looking-time towards incongruent speech (mixed English and Spanish audio and visual). There was a significant difference between mean proportion looking-time towards overall congruent speech ( $M=0.81$ ,  $SD=0.11$ ) and mean proportion looking-time towards incongruent speech ( $M=0.67$ ,  $SD=0.21$ );  $t(10) = 1.72$ ,  $p = .05$ . Thus, there was a

significant effect of congruency, with greater mean proportion looking-time towards congruent speech overall than towards incongruent speech.

A paired-samples t-test was conducted to compare infants' mean proportion of looking-time towards congruent Spanish speech and towards incongruent speech. There was a marginally significant difference between mean proportion looking-time towards matched Spanish speech ( $M=0.80$ ,  $SD=0.14$ ) and incongruent speech ( $M=0.67$ ,  $SD=0.21$ );  $t(10) = 1.58$ ,  $p = .07$ . Thus, there was a marginally significant effect of congruency, with greater mean proportion looking-time towards congruent speech than towards incongruent speech.

A paired-samples t-test was conducted to compare infants' mean proportion of looking-time towards congruent English speech, and towards incongruent speech. There was a significant difference between mean proportion looking-time towards congruent English speech ( $M=0.83$ ,  $SD=0.10$ ) and incongruent speech ( $M=0.67$ ,  $SD=0.21$ );  $t(10) = 1.87$ ,  $p = .04$ . Thus, there was a significant effect of congruency, with greater mean proportion looking-time towards congruent speech than towards incongruent speech.

The paired-samples t-test conducted to compare infants' mean proportion looking-time towards congruent English speech and towards congruent Spanish speech revealed no significant differences between languages. However, whereas mean proportion looking-time towards congruent English speech compared to looking-time towards incongruent speech was significant, infants' mean proportion looking-time towards congruent Spanish compared to looking-time towards incongruent speech reached only marginal significance. This suggests that there was a slightly greater difference between infants' looking-time towards congruent English speech compared to incongruent speech, than between infants' looking-time towards congruent Spanish speech compared to incongruent speech. Thus, despite no significant differences in infants' mean proportion looking-time towards congruent English speech compared to looking-time towards congruent Spanish speech, there seems to be a slight preference in infants to look towards English speech than towards Spanish speech in comparison to incongruent speech.

In order to further explore this slight difference in infants' looking-time towards English compared to incongruent speech than towards Spanish compared to incongruent speech, a correlation was performed across age between each matched condition (Congruent English, Congruent Spanish, Overall Congruent English and Spanish) and the Incongruent condition (see Figure 2 for the correlation between age and the incongruent condition). A significant correlation was found between congruency conditions across ages, specifically, infants' age and mean proportion looking-time towards incongruent speech was negatively correlated ( $r(10) = -0.57, p < .05$ ). Thus, younger infants tended to look longer towards incongruent speech than older infants. From this correlation we can infer that older infants tended to look longer towards congruent speech than younger infants.

In summary, a significant effect of congruency was found in overall congruent speech, as well as in congruent English, whereas looking-time towards congruent Spanish speech only reached marginal significance. In contrast, looking-time toward incongruent speech was not significant. Furthermore, a trend was found between age and looking-time toward incongruent speech, specifically younger infants tended to look longer towards incongruent speech than older infants. These results were not due to either variance between the eight trial types nor due to variance between genders.

Although these results seem to suggest perceptual tuning in infants to a native language (English), the marginally significant effect of congruency on looking-time towards a non-native language (Spanish) may suggest that the perceptual narrowing process is both subtle and flexible in older infants. Given that these infants were never exposed to Spanish speech (according to infants' parents), the fact that they looked longer, on average, towards congruent Spanish speech than towards incongruent speech suggests that the infants were sensitive to the auditory and visual signals of a non-native language. Overall, these results suggest that older infants look longer towards audiovisually congruent speech regardless of language compared to incongruent speech; as well, this tendency is stronger in older infants than in younger infants, and there also may be a slightly greater tendency in infants to look longer towards a native

language compared to a non-native language.

### GENERAL DISCUSSION

Consistent with my hypothesis, infants looked longer towards congruent speech presented in both a native language as well as a non-native language compared to incongruent speech. Likewise, no significant difference was found between infants' mean proportion-looking time towards congruent speech in a native language compared to congruent speech in a non-native language. Infants' looking-time towards congruent native speech was significantly greater than towards incongruent speech, whereas infants' looking-time towards congruent non-native speech reached only marginally greater significance compared to looking-time towards incongruent speech. The greater significance of infants' looking-time towards native speech compared to incongruent speech than towards non-native speech compared to incongruent speech seems to suggest a slight effect of narrowing to infants' native language. However, because there was no significant difference between infants' looking-time towards native speech compared to non-native speech, this effect of narrowing may not be reliable and may be eliminated with a larger sample.

There was also a significant age trend between age (in months) and looking-time towards incongruent speech, such that younger infants looked longer towards incongruent speech than older infants. This trend suggests that older infants tend to pay more attention to the synchronous auditory and visual signals of speech than younger infants, and likewise ignore speech that contains mismatched auditory and visual signals, whereas younger infants tend to look towards both congruent and incongruent speech for similar lengths of time. This suggests that older infants are sensitive to the synchronous auditory and visual information present in speech regardless of language and thus have learned to ignore audiovisual speech that contains conflicting information (i.e., incongruent audio and visual signals), whereas younger infants are perhaps less sensitive to the audio and visual synchrony and have not learned to selectively attend to synchronous speech. Furthermore, the differences found between congruency conditions were not due to order effects or differences in auditory and visual matching in the incongruent

conditions because no differences were found between the eight trial types across all participants. Likewise, no gender effects were found.

According to the perceptual narrowing hypothesis, infants should have looked significantly longer towards native speech compared to non-native speech; as well, they should have looked equally long towards incongruent speech compared to congruent non-native speech. The result that infants' looking-times towards a native language did not significantly differ from looking-times towards a non-native language, along with the marginally significant result that infants looked longer towards congruent non-native speech than towards incongruent speech both provide a challenge to the perceptual narrowing hypothesis. This suggests that infants' perception of native and non-native speech was highly modulated by the fluency of the auditory and visual signals presented, whereas previous studies that used only contrastive syllabic pairings suggested that older infants ignore non-native speech information relative to native speech (Kuhl et al., 1992, as described by Pons et al., 2009; Werker, and Tees, 2005, as described by Scott et al., 2007; Werker, and Tees, 1984, as described by Lewkowicz and Ghazanfar, 2009). This highlights the relevance of audiovisual, fluent speech to infants' language learning, regardless of whether the language is native or non-native. While perceptual narrowing may play a role in infants' perception of isolated speech sounds in a native versus non-native language, this hypothesis understates the vast flexibility infants have in tuning in to a non-native language when it is presented fluently in synchronous auditory and visual modalities.

There were a few disadvantages to the methods used in this study. For one, the visual speech only contained the mouth portion of the face (i.e., not the nose, eyes, or any other part other than the square of face around the mouth, see Figure 3). Originally, we planned to present speech that contained the speaker's entire face, but the speaker's eyes inconsistently focused towards the camera and off to the side of the camera as a result of reading the story passages at the same time. To see whether her eye movements would be an issue, we conducted a pilot trial with a 12-month-old and found that the infant's eyes followed

the eye gaze of the speaker, which meant that coding infants' looking behavior would be unreliable using this stimulus. Thus, we limited the visual information to only mouth and lip-movements. Since it is unusual for an infant to see only a speaker's mouth while speaking (and not the eyes or rest of the face), infants' looking-behavior may have been affected—however, since all infants were presented the same stimulus, the results may still reliably reflect infants' speech perception.

Another issue with our stimulus was that the all incongruent speech pairs (audio and visual) contained both the native and non-native language, whereas the congruent speech was presented in either the native or non-native language. A better design would be to provide incongruent speech in English and in Spanish (e.g., incongruent speech that is auditory Spanish paired with visual Spanish) in order to assess whether the language had an impact on perception of incongruent speech. Similarly, the use of only one speaker may have had an impact on infants' looking behavior, thus future studies should present speakers of different ethnicities in order to control for infants' bias towards same-race faces.

On another note, there is an inherent disadvantage in only using behavioral methods to infer differences or changes in infants' perceptual processing. While looking-behavior is a good indicator of infants' interest and attention towards a stimulus, it is difficult to suggest or infer from this information alone whether or not older infants are processing speech comparably to younger infants. Although in this study subtle differences were shown between infants' proportion looking-times towards a native versus a non-native language, neurophysiological methods may reveal a stronger effect of language. Future studies implementing this paradigm paired with neurophysiological methods may be able to provide complementary understanding of perceptual narrowing processes in infancy by examining corresponding cortical activation or efficiency of processing, which could not be assessed in the present study.

A further disadvantage was the small sample size. With larger power, the marginally significant result that infants looked longer towards congruent non-native speech than towards incongruent speech may have reached greater significance. There may have also been a stronger correlation between infants' ages

and looking-time towards congruent (English, Spanish) and incongruent speech. Furthermore, many infants were distracted by the bright table lamp positioned near their head during presentation, which was needed to be on to reliably code infants' looking-behavior towards the screen. A less bright lamp or more central light may have amended this issue.

In general, infants seem to readily attend to and perceive fluent multisensory speech in both native and non-native languages. Specifically, infants' attention towards speech that is auditorily and visually synchronous compared to speech that contains conflicting auditory and visual information seems to increase with age, perhaps as a result of greater exposure to spoken language (thus greater efficiency in processing), compared to younger infants. Whether or not perceptual narrowing plays a role in older infants' attention towards native versus non-native speech cannot be concluded in this study; however, it seems that infants' acute sensitivity to the congruency of fluent speech regardless of language may outweigh any narrowed preference for native speech.

### References

1. Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303-306.
2. Jusczyk, P. (1997). *The Discovery of Spoken Language*. MIT Press: Cambridge, MA.
3. Jusczyk, P., Rosner, B., Cutting, J., Foard, C., & Smith, L. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception and Psychophysics*, *21*, 50-54.
4. Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences, USA*, *105*, 11442-11445.
5. Lewkowicz, D.J., & Ghazanfar, A.A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences, USA*, *103*, 6771-6774.
6. Lewkowicz, D.J., & Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Science*, *13*, 470-478.
7. Lewkowicz, D. J., Sowinski, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. *Brain Research*, *1242*, 291-302.
8. Polka, L. & Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421-435.
9. Pons, F., Lewkowicz, D.J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences, USA*, *106*, 10598-10602.
10. Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P.K. (2005). Brain potentials to native and non-

- native speech contrasts in 7- and 11-month-old American Infants. *Developmental Science*, 8, 162-172.
11. Sai, F.Z. (2005). The role of a mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, 14, 29-50.
  12. Scott, L.S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. *Neuropsychologia*, 48, 1857-1861.
  13. Scott, L.S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, 20, 676-680.
  14. Scott, L.S., Pascalis, O., & Nelson, C.A. (2007). A Domain-General Theory of the Development of Perceptual Discrimination. *Current Directions in Psychological Science*, 16, 197-201.
  15. Weikum, W.M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316, 1159.
  16. Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, 7, 49-63.

### Tables and Figures

Table 1: Means (and standard deviations) of infants' mean proportion looking-times towards matched English, matched Spanish, total matched, and total non-matched speech.

	Congruent	Incongruent
English	0.83 (0.10)	
Spanish	0.80 (0.14)	
Overall (English + Spanish)	0.81 (0.11)	
Mixed English + Spanish		0.67 (0.21)

Figure 1: Mean Proportion Looking-Times towards congruency condition across participants.

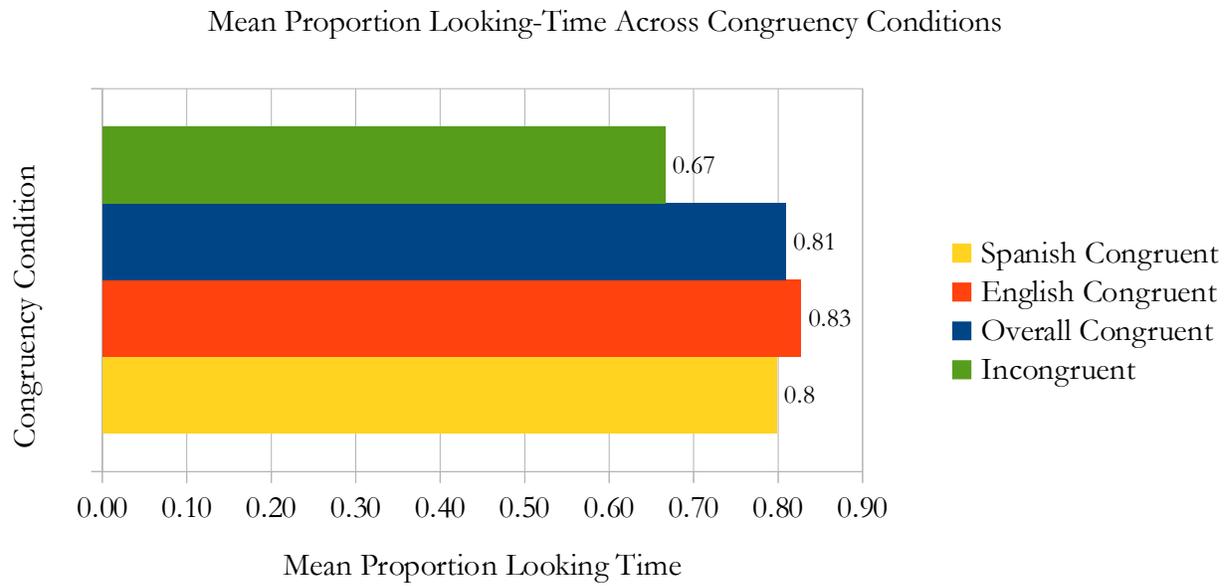


Figure 2: Correlation between Age (in months) and infants' proportion looking-time (as a percentage) towards incongruent speech.

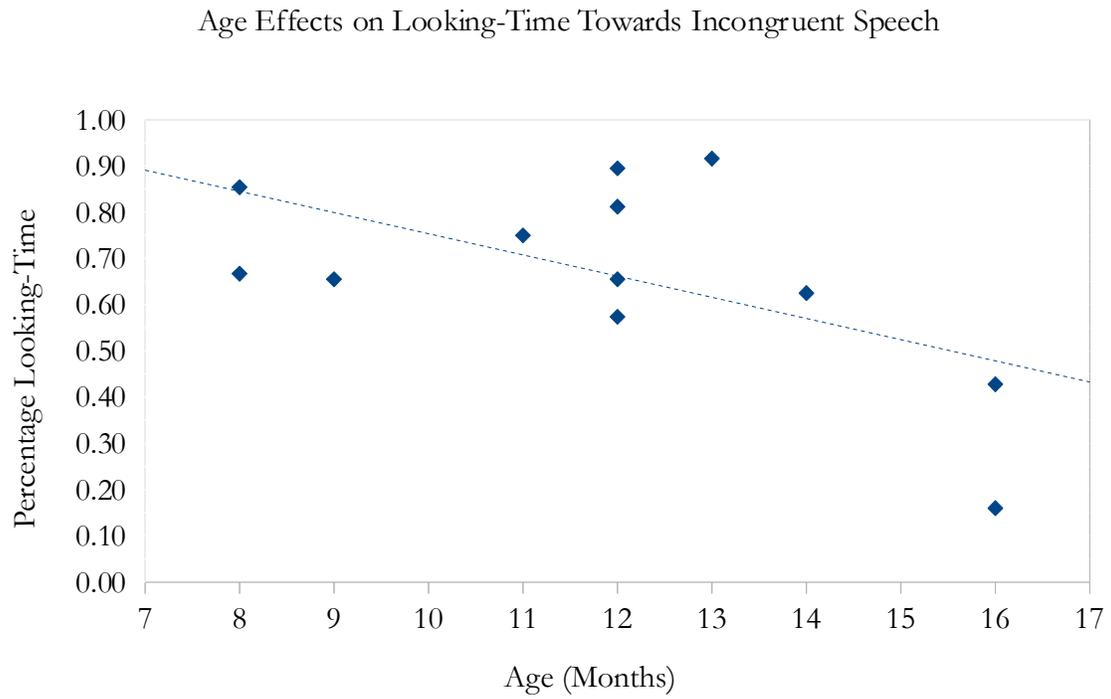


Figure 3: Still-frame from visual speech stimulus.



## Appendix

Summary of trial types 1 - 4 across audio and visual language for each story, where A = Auditory Language, V = Visual Language, E = English, and S = Spanish.

---

		Trial Type	1	2	3	4
		Language	A, V	A, V	A, V	A, V
<i>Story Order, length (sec)</i>						
<i>Dog</i>	13		E, E	S, S	S, E	E, S
<i>Farm</i>	16		E, S	S, E	S, S	E, E
<i>Goldfish</i>	12		S, S	E, E	E, S	S, E
<i>Park</i>	25		S, E	E, S	E, E	S, S
<i>Hippo</i>	9		E, E	S, S	E, S	S, E
<i>Bee</i>	21		E, S	S, E	E, E	S, S
<i>Rabbit</i>	23		S, E	E, S	S, S	E, E
<i>Ladybug</i>	12		S, S	E, E	S, E	E, S

---

Summary of trial types 5 - 8 across audio and visual language for each story, where A = Auditory

Language, V = Visual Language, E = English, and S = Spanish.

---

		Trial Type	5	6	7	8
		Language	A, V	A, V	A, V	A, V
<i>Story Order, length (sec)</i>						
<i>Ladybug</i>	12		S, S	E, E	S, E	E, S
<i>Rabbit</i>	23		S, E	E, S	S, S	E, E
<i>Bee</i>	21		E, S	S, E	E, E	S, S
<i>Hippo</i>	9		E, E	S, S	E, S	S, E
<i>Park</i>	25		S, E	E, S	E, E	S, S
<i>Goldfish</i>	12		S, S	E, E	E, S	S, E
<i>Farm</i>	16		E, S	S, E	S, S	E, E
<i>Dog</i>	13		E, E	S, S	S, E	E, S

---