NIH Public Access
Author Manuscript

The Decline of Cross-Species Intersensory Perception in Human Infants: Underlying Mechanisms & Its Developmental Persistence

David J. Lewkowicz, **Ryan Sowinski**, and **Silvia Place**
Department of Psychology, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431, USA

Abstract

The current study investigated the mechanisms underlying the developmental decline in cross-species intersensory matching first reported by Lewkowicz and Ghazanfar (2006) and whether the decline persists into later development. Experiment 1 investigated whether infants can match monkey vocalizations to asynchronously presented faces and found that neither 4–6 nor 8–10 month-old infants did. Experiment 1 also assessed whether a visual processing deficit may account for the developmental decline in cross-species matching and indicated that it does not because both age groups discriminated silent monkey calls. Experiment 2 investigated whether an auditory processing deficit may account for the decline and indicated that it does not because 8–10 month-old infants discriminated the acoustic versions of the calls. Finally, Experiment 3 asked whether the developmental decline persists into later development by testing cross-species intersensory matching in 12- and 18-month-old infants and showed that it does because neither age group made intersensory matches. Together, these results bolster prior evidence of a decline in cross-species intersensory integration in early human development and shed new light on the mechanisms underlying it.

Keywords

Infants; Audiovisual Perception; Face-Voice Integration; Perceptual Narrowing

Talking faces are ubiquitous and provide a rich array of verbal, affective, and social multisensory information. Extraction and interpretation of this information depends, to a large extent, on the observer's ability to perceive and integrate the multidimensional sources of audiovisual information inherent in talking faces. Indeed, studies have shown that adults can integrate the multidimensional auditory and visual attributes specifying talking faces (Fowler, 2004; Munhall & Vatikiotis-Bateson, 1998, 2004; Yehia, Rubin, & Vatikiotis-Bateson, 1998). When they do, they rely on the fact that the spectral characteristics of audible speech are typically closely linked with the dynamic characteristics of visible speech and that the auditory and visual attributes specifying talking faces are usually in temporal and spatial register and correspond in terms of duration, tempo, and rhythmic patterning.

Research has shown that the ability to integrate auditory and visual information inherent in speech as well as in non-speech events emerges early in infancy and that it improves rapidly once it does so (Lewkowicz, 2000a; Lickliter & Bahrick, 2000; Walker-Andrews, 1997). The

improvement is characterized by the fact that as infants get older they become capable of perceiving increasingly more complex and higher-level types of intersensory relations (Lewkowicz, 2002). For example, the ability to perceive simple audio-visual synchrony relations inherent in non-speech events has been found at two months of age (Bahrick, 1983; Lewkowicz, 1992b, 1996), the ability to perceive rhythm-based intersensory relations has been found at four months (Mendelson & Ferland, 1982), the ability to perceive duration-based intersensory relations has been found at six months (Lewkowicz, 1986), and the ability to perceive intersensory dynamic distance relations has been found between five and seven months of age (Walker-Andrews & Lennon, 1985). Studies of infant response to multimodal speech have found that infants as young as two months of age can match audible and visible phonetic information but that only 8-month-old infants can perform intersensory gender matching (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2002, 2003; Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994; Walker-Andrews, 1986; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991; Walker, 1982).

Perceptual Narrowing in Early Development

The early emergence of intersensory perceptual abilities and their rapid improvement is consistent with the basic assumptions underlying all extant theories of intersensory perceptual development. Although these theories differ in their predictions regarding the specific timing of the emergence of intersensory perceptual abilities - some assume that intersensory perception emerges at or shortly after birth (Gibson, 1969; Thelen & Smith, 1994; Werner, 1973) whereas others assume that it emerges slowly over the first few years of life (Birch & Lefford, 1963, 1967; Piaget, 1952) - all assume that intersensory perceptual abilities improve and broaden in scope as development progresses. This general assumption is consistent with conventional wisdom in the developmental sciences that structure and function broaden in scope as development progresses. Although this conventional wisdom has been the dominant view, there is evidence that some aspects of development also may be characterized by a parallel process of narrowing of structure and function.

The evidence in support of narrowing comes from research on neural as well as behavioral development. At the neural level, studies have shown that the nervous system initially produces many more neurons and neural connections and that over time many of these are selectively pruned as a function of experience (Greenough, Black, & Wallace, 1987). At the behavioral level, studies have shown that some functions that are initially present in early development actually decline as a function of early canalizing experiences (Gottlieb, 1991b; Kuo, 1976). For example, Gottlieb (1991a) showed that early canalizing experiences in birds lead to a narrowing of call recognition. He found that when mallard duck embryos are prevented from hearing the contact call that they produce before hatching, subsequently they prefer the call of another species. In other words, prenatal exposure to self-produced calls buffers the mallard duckling from developing a preference for the calls of the wrong species.

In human infants, studies have shown that the perception of speech, face, and music narrows with development and experience and that this leads to a decline in responsiveness to nonnative information. For example, in speech perception, Werker and Tees (1984) found that 6–8 month-old English-learning infants can discriminate native speech contrasts (/ba/vs./ga/) as well as nonnative speech contrasts (i.e., the Hindi retroflex/Ta/vs. the dental/Ta/) but that infants older than 11 months of age no longer discriminate the nonnative contrasts. Subsequent studies have indicated that this decline in phonetic perception is general in that it also occurs in Zulu (Best, McRoberts, & Sithole, 1988) and in Swedish (Kuhl, Williams, Lacerda, Stevens, & et al., 1992) and that it is shaped by experience with native speech (Kuhl, Tsao, & Liu, 2003). In face perception, studies have shown that perception of nonnative faces as well as of the faces of other races declines. For example, Pascalis, deHaan, and Nelson (2002) found that 6-month-

old infants can discriminate pictures of native (i.e., human) as well as nonnative (i.e., monkey) faces but that 9-month-old infants only discriminate pictures of native faces and Pascalis et. al. (2005) showed that this decline is the result of specific experience. Weikum et. al. (2007) have found that 4- and 6-month-old infants demonstrate the remarkable ability to discriminate the visual-only attributes of their native language from the visual-only attributes of an unfamiliar language, that monolingual 8-month-old infants only discriminate the visual attributes of their own language, and that bilingual 8-month-olds maintain the ability to discriminate between their native language and a nonnative language. Studies of infant response to the faces from other races have shown that infants younger than three months of age can discriminate same-race faces as well as other-race faces but that older infants only discriminate same-race faces (Kelly et al., 2007; Kelly et al., 2005; Sangrigoli & de Schonen, 2004). Finally, studies of infant response to nonnative musical rhythms have found that 6-month-old infants can discriminate complex non-Western musical rhythms, that 12-month-old infants do not (Hannon & Trehub, 2005a), and that this decline is a function of specific experience with Western musical rhythms (Hannon & Trehub, 2005b).

Perceptual Narrowing of Intersensory Perception in Early Development

The fact that the ability to perceive nonnative speech, faces, and musical rhythms declines in human infancy recently prompted Lewkowicz & Ghazanfar (2006) to propose that this decline may reflect the operation of a developmental process that operates not only in individual sensory modalities but across them as well. Based on this hypothesis, these researchers reasoned that a similar decline may occur in infant intersensory integration of nonnative faces and their accompanying vocalizations. To investigate this possibility, Lewkowicz & Ghazanfar (2006) tested 4-, 6-, 8-, and 10-month-old infants' ability to match the visible and audible attributes of monkey calls. To do so, they presented side-by-side faces of a rhesus monkey producing a coo call on one side and a grunt call on the other side and assessed infants' response to the two calls in silence as well as in the presence of one of the corresponding audible vocalizations.

Results showed that the 4- and 6-month-old infants exhibited greater looking at the visible call when the corresponding vocalization was presented than when it was not but that the 8- and 10-month-old infants did not. This developmental pattern of successful integration of the visible and audible attributes of a vocalizing face at a younger but not at an older age was in sharp contrast to all previous findings from studies of face-voice integration in infancy. These studies have all shown that the ability to integrate the audible and visible attributes of human talking faces does not decline (Kahana-Kalman & Walker-Andrews, 2001; Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2002, 2003; Poulin-Dubois, Serbin, & Derbyshire, 1998; Poulin-Dubois et al., 1994; Walker-Andrews, 1986; Walker-Andrews et al., 1991) and that infants as old as 18 months of age can match human faces and voices (Poulin-Dubois et al., 1998). Thus, the Lewkowicz & Ghazanfar (2006) study was the first to provide empirical evidence that the ability to integrate the faces and voices of another species is present early in infancy and then declines as infants get older.

Although the Lewkowicz & Ghazanfar (2006) findings provided the first evidence of intersensory perceptual narrowing, they also raised some important questions. First, what was the basis for the younger infants' successful intersensory matching? The two monkey calls had different durations and, as a result, the onset of the accompanying vocalization was synchronized with the onset of both visible calls but only with the offset of the matching visible call. Thus, it is possible that the younger infants made intersensory matches on the basis of temporal synchrony cues, duration equivalence, and/or both. This possibility is made all the more likely by the fact that infants as young as two months of age are sensitive to intersensory temporal synchrony relations (Bahrick, 1983; Dodd, 1979; Lewkowicz, 1986, 1992a, 1992b,

1996b, 2000b) and that they become sensitive to audio-visual duration relations between three and six months of age (Lewkowicz, 1986). Second, might the older infants' failure to make intersensory matches have been due to unisensory processing deficits rather than to an intersensory integration deficit? This is likely in the visual modality because, as indicated earlier, responsiveness to nonnative faces declines during infancy (Kelly et al., 2005; Pascalis et al., 2002; Weikum et al., 2007). It is less clear whether this is likely in the auditory modality because, so far, no evidence has been reported that responsiveness to nonnative vocalizations declines. Finally, does the decline in cross-species intersensory perception reflect the waning of the early dominance of synchrony detection mechanisms and, thus, a temporary developmental phenomenon, or does it represent an enduring aspect of early perceptual development that reflects the operation of an experiential canalization process?

The purpose of the current study was to answer the three questions outlined above in three separate experiments. Experiment 1 addressed the first question by investigating whether audio-visual temporal synchrony and/or duration equivalence might underlie successful cross-species intersensory perception in young infants. Experiment 1 also addressed the first part of the second question by investigating whether a visual discrimination deficit might have contributed to older infants' failure to make cross-species intersensory matches. Experiment 2 addressed the second part of the second question by asking whether an auditory discrimination deficit might have contributed to older infants' failure to make cross-species intersensory matches. Finally, Experiment 3 addressed the third question by investigating whether the decline in cross-species intersensory perception persists into the second year of life.

Experiment 1

Lewkowicz and Ghazanfar (2006) presented a set of initial silent test trials designed to assess infant response to the calls in silence and then a set of in-sound test trials designed to assess infant response to the calls in the presence of one of the matching vocalizations. The data from the silent trials make it possible to determine whether the decline in cross-species intersensory matching observed by eight months of age was due to a visual processing deficit. Lewkowicz and Ghazanfar (2006) reported that their infants looked longer at the coo but they did not report whether this difference was statistically reliable. Thus, the current experiment had two aims. The first was to investigate whether infants can perceive the difference between the silent calls. To this end, we tested a group of 4–6 and 8–10 month-old infants' response to the identical silent calls presented by Lewkowicz and Ghazanfar (2006). The second aim was to investigate the perceptual basis for the successful matching obtained in the younger infants in the Lewkowicz and Ghazanfar (2006) study. Because the coo and the grunt calls differed in duration, the correspondence between a given vocalization and the matching call was based on onset and offset synchrony as well as duration. To determine which of these perceptual cues mediated successful intersensory matching, we presented two additional trials during which we presented the identical visible calls accompanied by one of the vocalizations presented out of synchrony with respect to the matching call. If synchrony alone mediated responsiveness then infants were not expected to make intersensory matches. If, however, duration mediated matching then the younger infants were expected to make successful intersensory matches.

Method

Participants—The participants in this experiment, as well as in the subsequent ones, were fullterm, healthy infants, with birth weights above 2500 g, and were recruited from local birth records. Because Lewkowicz and Ghazanfar (2006) found that 4- and 6-month-old infants' response was the same and that the 8- and 10-month-old infants' response also was the same, in the current study we tested only a younger (4–6 month-old) and an older (8–10 month-old) age group. The 4–6 month-old group had a mean age of 23.7 weeks (SD = 4.7 weeks) and

consisted of 25 boys and 22 girls while the 8–10 month-old group had a mean age of 41.8 weeks (SD = 4.3 weeks) and consisted of 18 boys and 21 girls. We tested an additional 21 infants but they did not contribute usable data because of equipment failure (four infants) and fussing (17 infants; five younger infants and 12 older ones).

Stimuli and Apparatus—The stimuli were identical to those presented in the Lewkowicz and Ghazanfar (2006) study and consisted of pairs of digitized videos of one of two vocalizing rhesus monkeys (*Macaca mulatta*). As in that study, we presented two different stimulus sets consisting of two different vocalizing monkeys in order to permit generalization and to rule out the possibility that responsiveness was based on some idiosyncratic property of the animal. For one of these monkeys, the duration of the coo was 735 ms and the duration of the grunt was 180 ms, whereas for the other monkey, the duration of the coo was 760 ms and the duration of the grunt was 275 ms. The videos were 2-s digital clips of the calls looped continuously for 1 minute. The videos were presented on two, side-by-side, 17-inch LCD display monitors inside 19.05 cm (H) × 25.40 cm (W) windows on each monitor (the monkey's head filled most of the window). The sound pressure level (SPL) of the coo vocalization was 77 dB for one monkey and 72 dB for the other monkey and the SPL of the grunt vocalization was 81 dB for one monkey and 77 dB for the other.

The space between the two monitors (6.4 cm) was just large enough to accommodate a small video camera together with a circuit-board on top of it containing several light-emitting-diodes (LEDs). The LEDs were used to center the infant's gaze before the start of each trial. The audible portion of the video was presented through a speaker that was placed behind the camera and, thus, was located in-between the two monitors. Infants were tested individually in a quiet room (ambient SPL was 58 dB, A level) while seated either in an infant seat (n = 67) or in the parent's lap (n = 19) 50 cm in front of the camera. The parents of the infants holding the infant on the lap were not informed about the specific hypothesis under test.

Procedure—The experiment consisted of four 1-min test trials. Each trial began with the LEDs flashing on and off. As soon as the infant looked at the LEDs they were turned off and the videos began to play on each monitor. As the videos began, the face of the same monkey appeared at the same time on each monitor and began to produce each respective call at the same time. During the first two silent test trials, the calls were presented in silence (their lateral position was reversed on the second of these trials). During the second two in-sound trials the same calls were presented except that this time one group of infants heard the coo vocalization whereas another group heard the grunt vocalization.

In contrast to the Lewkowicz & Ghazanfar (2006) study, here the onset of the audible call was presented 666 ms prior to the initiation of lip movement in both videos (see Fig. 1). This specific degree of temporal asynchrony was chosen on the basis of prior studies indicating that infants can detect an audio-visual asynchrony of 350 ms produced by a simple moving/sounding object (Lewkowicz, 1996) and an asynchrony of 666 ms produced by a talking face (Lewkowicz, 2000b). Given that the stimuli presented here were most similar to a talking face, we chose to present the audible vocalization 666 ms prior to the onset of the visible vocalization to maximize the possibility that the auditory and visual attributes would be perceptually asynchronous.

The entire test session was recorded onto videotape. Subsequently, coders who were blind with regard to the testing conditions and the stimuli presented on each trial, scored the duration of visual fixations. Inter-coder reliability, calculated on a sample of 16 infants representing the two age groups tested, was 97%.

Results and Discussion

Overall, the younger infants spent 71% of their allotted time looking at the two visual stimuli during the silent test trials and 74% of their allotted time looking at the two visual stimuli during the in-sound test trials. The older infants spent 68% and 71% of their allotted time looking at the stimuli during the two types of test trials, respectively. Thus, both age groups found the stimuli to be equally interesting.

To determine whether infants distinguished between the visible coo and grunt, we examined the proportion of total looking time that infants devoted to each call during the silent test trials. As can be seen in Fig. 2a, 4–6 month-old infants looked longer at the silent coo than at the silent grunt, t (47) = 3.85, $p < .001$, as did the 8–10 month-old infants, t (38) = 2.25, $p < .05$. In addition, we performed statistical comparisons of the data from the Lewkowicz and Ghazanfar (2006) study. As can be seen in Fig. 2b, the pattern of looking in the silent test trials in that study was the same as in the current experiment. Two-tailed t-tests indicated that the 4–6 month-old infants looked longer at the silent coo, t (90) = 4.93, $p < .001$, as did the 8–10 month-old infants, t (85) = 4.96, $p < .001$.

To determine whether infants made intersensory matches, we compared looking time to the matching visible call across the silent and in-sound conditions. To make this comparison, we computed two proportion scores for each vocalization group, respectively. Thus, for the coo vocalization group, first we computed the proportion of time these infants looked at the coo in the silent condition by dividing the amount of time they spent looking at the silent coo by the total amount of looking they spent looking at the silent coo and grunt. Then, we computed the proportion of time these infants looked at the coo during the in-sound condition by dividing the total amount of time they looked at the in-sound coo by the total amount of time they spent looking at the in-sound coo and grunt. For the grunt vocalization group, first we computed the proportion of time these infants looked at the grunt in the silent condition by dividing the amount of time they spent looking at the silent grunt by the total amount of looking they spent looking at the silent coo and grunt. Then, we computed the proportion of time this group of infants looked at the grunt during the in-sound condition by dividing the total amount of time they looked at the in-sound grunt by the total amount of time they spent looking at the in-sound coo and grunt. We expected infants to exhibit greater looking at the matching call during the in-sound test trials than during the silent trials if they perceived the intersensory relations.

First, we submitted the proportion-of-looking scores to an overall repeated analysis of variance (ANOVA), with vocalization group (2), animal (2) and age (2) as the between-subjects factors and sound condition (2; silent, in-sound) as the within-subjects factor. The ANOVA did not yield any significant interactions, indicating that looking was not affected by the particular vocalization, the particular animal producing it, nor by the infants' age. Thus, we collapsed the data across vocalization group and animal and then compared the proportion of looking in the silent vs. in-sound test trials at each age, respectively. As Fig. 3 shows, neither the younger, t (46) = −0.46, $p = 0.65$, nor the older, t (38) = −1.52, $p = 0.14$, infants looked more at the matching call in the in-sound trials than in the silent trials. In other words, neither age group exhibited evidence of cross-species intersensory matching.

The finding that the infants in, both, the current study and the Lewkowicz and Ghazanfar (2006) study preferred to look at the coo indicates that the failure of 8–10 month-old infants to make cross-species intersensory matches is not due to their inability to distinguish between the visible calls. Thus, the older infants' failure to make matches either may have been due to their failure to perceive some aspect of the vocalizations or to their failure to perceive the audio-visual relationship *per se*. The finding that the 4–6 month-old infants in the current experiment did not exhibit cross-species intersensory matching indicates that it was the audio-visual synchrony cues that mediated the younger infants' successful matching in the Lewkowicz and

Ghazanfar (2006) study and not duration. Had the infants perceived duration as an amodal invariant, they would still have performed intersensory matching. Their failure to perceive the amodal invariance of duration parallels similar findings from other studies showing that 6- and 8-month-old infants do not match auditory and visual information on the basis of duration unless the onset and offset of the information in the two modalities is also synchronous (Lewkowicz, 1986).

Experiment 2

The intersensory correspondence between a given vocalization and the matching visible call in the Lewkowicz & Ghazanfar (2006) study depended on their having synchronous onsets and offsets. This, in turn, depended critically on the fact that the vocalization and the matching call had equal durations but that the vocalization and the non-matching call did not. In other words, although the findings from Experiment 1 indicated that infants did not match the vocalizations and the calls on the basis of duration alone, duration still played a critical role in that it determined which call corresponded to a given vocalization. The results from Experiment 1 and from the silent trials in Lewkowicz and Ghazanfar (2006) also indicated that both age groups perceived the difference between the visible calls. Thus, there is no doubt that both age groups perceived the difference between the visible calls and that they most likely did so on the basis of duration. Given that the calls were produced by the same monkey's face, the most salient difference between them was duration.

If older infants perceived the difference between the visible calls on the basis of their duration then this raises the question of whether these infants, who failed to make cross-species intersensory matches in the Lewkowicz and Ghazanfar (2006) study, may be insensitive to the duration-based difference in the vocalizations. If they are then this would suggest that this perceptual deficit accounts for their failure to make intersensory matches. If, however, they are sensitive to the duration difference in the auditory modality then this would help rule out an auditory perceptual deficit as being a possible cause of their failure to perform cross-species intersensory matching. Thus, to determine whether older infants can discriminate between the vocalizations on the basis of their different durations, we habituated 8–10 month-old infants to one vocalization and then tested them for discrimination by presenting the other vocalization.

One potential interpretive problem in a study such as this is that the two vocalizations differed not only in terms of their durations but also in terms of their spectral characteristics. Fortunately, it is possible to make *a priori* predictions regarding outcome that can help determine which of these perceptual cues mediated responsiveness. Thus, if infants' response is based primarily on duration differences then they are likely to exhibit an asymmetrical response pattern. Specifically, infants habituated to the short vocalization (the grunt) and then tested with the longer one (the coo) should exhibit greater response recovery than infants habituated to the longer vocalization and then tested with the shorter one. This is because in the former case there is an overall increase in stimulation whereas in the latter case there is an overall decrease in stimulation. If, however, infants' response is based on spectral differences then it is reasonable to expect equivalent magnitudes of response recovery regardless of which vocalizations serve as the habituation and test stimuli.

Method

Participants—Twenty 8–10 month-old infants (mean age = 36.7 weeks, SD = 4.1 weeks; 9 girls, 11 boys) completed the experimental session and contributed data in this experiment. We tested three additional infants but did not include their data because of fussing (two infants) and equipment failure (one infant).

Stimuli and Apparatus—All stimuli in the current experiment were presented in the form of multimedia movies on a single 17-inch monitor. One was the attractor movie and showed a continuously expanding and contracting green disk in the middle of the screen. Its purpose was to bring the infant's gaze back to the stimulus-presentation monitor whenever he or she looked away from it. A second movie, used to test for possible fatigue effects, showed a 1 minute segment of a Winnie-the-Pooh cartoon (70–74 dB SPL, A scale). The remaining four vocalization movies, each one minute in duration, showed a continuously rotating disk in the center of the screen (the disk was composed of alternating beige and brown wedges). As the disk rotated, the infant could hear one of the two vocalizations performed by one of the two monkeys. The vocalization movies were constructed by extracting the auditory portion of the stimulus videos used in Experiment 1 and pairing it with the rotating disk. As in the original movies, the vocalizations could be heard every 2 s.

Infants were seated either in an infant seat or in the parent's lap at a distance of 50 cm from the stimulus-presentation monitor in a dimly lit, sound-attenuated, single-wall, chamber (IAC Corp., ambient sound pressure level was 48 dB, A scale). Most of the infants were seated alone in the infant seat. In those few cases where infants were seated in the parent's lap, the parent wore headphones and listened to white noise, was blind with respect to the hypothesis under test, and was asked to sit as still as possible and not to interact with the baby. The audio part of the movie was presented through two speakers placed on each side of the monitor. The average SPL of the coo vocalization was 77 dB for one animal and 75 dB for the other animal and the corresponding SPL for the grunt calls was 75 dB and 77 dB, respectively. A video camera was located on top of the stimulus-presentation monitor and was used to transmit an image of the infant's face to a monitor outside the chamber.

Procedure—An experimenter, who was located outside the testing chamber, observed the infant on a video monitor and could neither see nor hear the stimuli being presented. An infant-controlled habituation/test procedure was used and, thus, the infant's looking behavior controlled movie presentation. Whenever the infant looked at the stimulus-presentation monitor, the movie began to play and continued to do so until the infant either looked away from the monitor for more than one second or until the movie ended (after 60 seconds). Once either of these two conditions was met, the attractor appeared again and remained there until the infant looked back at the monitor. As soon as the infant looked at the attractor, it was turned off and the next trial was initiated.

The experiment began with a single pre-test trial during which the Winnie-the-Pooh cartoon was played. The purpose of this trial was to measure the infant's initial level of attention. Once this trial ended, the habituation phase began during which one of the two vocalizations produced by one of the two monkeys was presented. The habituation phase continued until the infant reached a pre-defined habituation criterion which required that the infant's total duration of looking during the last three habituation trials decline to 50% of the total duration of looking during the infant's first three habituation trials. Once the infant reached the habituation criterion, the habituation phase ended and the test phase began. Here, each infant was given one familiar and one novel test trial, with the order of these two trials counterbalanced across infants. During the familiar test trial, the same vocalization that was presented during the habituation phase was presented again and was used as a measure of baseline responding following attainment of the habituation criterion. During the novel test trial, the novel vocalization produced by the same monkey was presented. Following completion of these two test trials, infants received a single post-test trial during which the Winnie-the-Pooh cartoon movie was presented again to measure the terminal level of attention. Once this trial was completed, the test session ended.

Results and Discussion

The data from the test trials were submitted to a repeated-measures ANOVA, with habituation vocalization (2) as a between-subjects factor and trials (2) as the within-subjects factor. This analysis yielded a significant trials effect, F (1, 18) = 33.97, $p < .001$, as well as a significant trials x habituation vocalization interaction, F (1, 18) = 14.39, $p < .01$. The interaction is depicted in Figure 4 where it can be seen that the expected pattern of response consistent with discrimination based on duration was obtained. That is, as expected, response recovery was greater to the coo following habituation to the grunt than was recovery to the grunt following habituation to the coo, F (1, 18) = 14.39, $p < .01$. Interestingly, however, and in spite of the interaction, both groups of infants exhibited significant response recovery to the novel vocalization. This was confirmed by a direct comparison of responsiveness in the familiar and novel test trials in each habituation group. Thus, those infants who were habituated to the coo exhibited significant response recovery to the grunt, t (11) = 1.80, $p < .05$, and those infants who were habituated to the grunt exhibited significant response recovery to the coo, t (7) = 5.40, $p < .01$. Finally, as evidence of the consistent nature of the infants' response, 17 out of the 20 infants tested in this experiment exhibited response recovery to a novel vocalization (binomial $p < .01$).

The current results indicate that 8–10 month-old infants perceived the difference between the coo and grunt vocalizations and the pattern of these results suggests that infants distinguished between the two vocalizations on the basis of their duration. When the current findings are considered together with the fact that 8–10 month-old infants can perceive the difference between the two silent calls, it is clear that the failure of the older infants to make cross-species intersensory matches is not due to a decline in unisensory processing in either modality.

Experiment 3

Although Lewkowicz and Ghazanfar (2006) established that cross-species intersensory matching declines by eight months of age, there are theoretical and empirical reasons to expect that this ability returns later in development. From a theoretical standpoint, because the detection of intersensory temporal synchrony only requires low-level perceptual mechanisms, its use as a basis for intersensory integration is likely to be limited to the earliest months of development (Lewkowicz, 2000a). Indeed, studies have shown that whereas young infants rely on audiovisual synchrony in perceptual tasks (Bahrick, Lickliter, & Flom, 2004; Dodd, 1979), older infants no longer do (Walker-Andrews, 1986; Walker, 1982). Thus, the synchrony-based cross-species intersensory matching found in young infants may not reflect a permanent decline in intersensory perceptual abilities but a shift away from reliance on this low-level stimulus features for intersensory integration. While this shift occurs, visual (Kellman & Arterberry, 1998), auditory (Jusczyk, 1997), and intersensory processing skills (Lewkowicz, 2000a; Lickliter & Bahrick, 2000; Walker-Andrews, 1997) continue to improve at a rapid pace during infancy and beyond. As they do, infants acquire a great deal of experience and, in the process, become increasingly more expert at processing higher-level stimulus features. Consistent with this improvement, evidence shows that in intersensory matching tasks involving native faces and vocalizations, older infants begin to exhibit the ability to match them on the basis of higher-order amodal invariant relations such as affect (Walker-Andrews, 1986) or gender (Patterson & Werker, 2002; Walker-Andrews et al., 1991). This shift in responsiveness to higher-level amodal invariant relations in infant response to native faces and vocalizations may also occur in older infants' response to nonnative faces and vocalizations. It may, however, be somewhat delayed because perceptual learning and differentiation of less familiar visual and auditory information may take longer.

Although the prediction of a return of cross-species intersensory matching is a reasonable one, an alternative prediction is that the decline represents the permanent effects of experiential

canalization on the development of intersensory perception. On this view, the decline is an enduring aspect of early perceptual development for two reasons. First, it is an enduring characteristic because older infants no longer integrate multisensory information on the basis of low-level relations such as synchrony. Second, it is an enduring characteristic presumably because older infants can no longer extract the higher-level intersensory relations inherent in nonnative faces and vocalizations due to the fact that the facial and vocal information is unfamiliar to them.

The purpose of the current experiment was to adjudicate between the two predictions outlined above. To do so, we tested 12- and 18-month-old infants with the identical stimuli and procedures used in the Lewkowicz and Ghazanfar (2006).

Method

Participants—The participants were twenty-five 12-month-old infants (11 boys, 14 girls; mean age = 51.9 weeks, SD = .83 weeks) and twenty 18-month-old infants (9 boys, 11 girls; mean age = 78.7 weeks, SD = 0.9 weeks). An additional nine infants were tested but did not contribute data because of fussing (seven) and inability to see infant's eyes (two).

Stimuli & Apparatus—The stimuli and apparatus were the same as in Experiment 1 with the following exceptions. First, like in the Lewkowicz and Ghazanfar (2006) study, here the onset and offset of the vocalization was synchronous with the onset and offset of the matching call and synchronous with the onset but not the offset of the non-matching call. Second, because we were concerned that older infants would not tolerate 1-minute test trials, we reduced each trial to 30 seconds. Finally, rather than use the flashing LEDs to center the infant's gaze prior to the start of each trial, here we presented a modified version of the attractor stimulus that was presented in Experiment 2. This consisted of presenting half the green disk on each of the two monitors such that it was visible at the right edge of the left monitor and the left edge of the right monitor. The disk expanded and contracted continuously and each time it reached its maximum size it made a brief tone.

Procedure—The procedure was the same as in Experiment 1 except that here we tested the infants in the sound-attenuated chamber. Inter-coder reliability, calculated on 12 randomly chosen infants, was 99%. Twenty six infants were tested on the parent's lap and 19 in the infant seat (parents were not aware of the specific hypothesis under test).

Results & Discussion

Overall, the 12-month-old infants spent 80% of their allotted time looking at the two calls during the silent test trials and 76% of their allotted time looking at the two calls during the in-sound test trials. The 18-month-old infants spent 77% and 75% of their allotted time looking at the calls during the two types of test trials, respectively. This level of overall attention to the stimuli is comparable to the level of attention found in the 4–6 month-old and the 8–10 month-old infants in Experiment 1.

To determine whether the infants in this experiment exhibited evidence of intersensory matching, we submitted the proportion-of-looking scores to an overall repeated-measures ANOVA, with vocalization group (2), animal (2), and age (2) as the between-subjects factors and trial type (2; silent, in-sound) as the within-subjects factor. This analysis did not yield a significant trial type effect, $F(1, 37) = 0.18$, n.s., nor any significant interactions between trial type and the other factors. The absence of a trial type effect indicates that relative to their looking at the matching call in silence, infants did not increase their looking to the matching call in the presence of the matching vocalization.

Despite the absence of a significant main effect of age, or of an interaction between age and trial type, it was still theoretically possible that the decline in intersensory matching might occur sometime between 12 and 18 months of age. To determine if this was the case, we examined the data separately at each age and, as can be seen in Fig. 5, neither the 12-month-old infants, t (24) = 0.60, n.s., nor the 18-month-old infants, t (19) = −1.10, n.s., exhibited greater looking at the matching call in the in-sound test trials than in the silent trials.

To determine whether the absence of intersensory matching in the current experiment might be attributable to infants' inability to distinguish between the visible calls, we examined the proportion of looking at the silent calls. This analysis indicated that, overall, the infants exhibited a robust preference for the coo, t (45) = 4.10, p < .001. Separate analyses at each age indicated that the 12-month-old infants exhibited a preference for the coo, t (24) = 4.70, p < .001 while the 18-month-old infants exhibited a marginal preference for it, t (19) = 1.53, p < .10. The marginal preference in the older infants is most likely due to the relatively small sample size.

The results from this experiment indicated that despite the fact that the 12- and 18-month-old infants found the stimuli to be as interesting as did the 4–6 and the 8–10 month-old infants, they still did not exhibit cross-species intersensory matching. Moreover, this failure to match was found despite the fact these infants perceived the difference between the visible calls. Overall, the current findings are consistent with Lewkowicz and Ghazanfar's (2006) finding of a decline in cross-species intersensory matching and show clearly that this decline persists into the second year of life.

General Discussion

The current study investigated the mechanisms underlying the decline in cross-species intersensory perception in human infants reported by Lewkowicz and Ghazanfar (2006) and whether this decline persists beyond 10 months of age. Specifically, the study asked whether synchrony and/or duration cues contribute to cross-species intersensory matching, whether unisensory processing deficits are responsible for the observed decline in cross-species intersensory matching, and whether the decline persists into later development. To permit direct comparisons between the results from the current study and those from the Lewkowicz and Ghazanfar (2006) study, here we presented the same dynamic faces and accompanying vocalizations and employed the identical testing procedures (the only exception was that the faces and voices were asynchronous in Experiment 1). Experiment 1 showed that neither 4–6 nor 8–10 month-old infants matched monkey calls and their vocalizations when the onset and offset of these two attributes was not in temporal register. This indicated that synchrony plays a critical role in early cross-species intersensory matching. Moreover, the results from the silent test trials in Experiment 1 and from the silent test trials in the Lewkowicz and Ghazanfar (2006) study indicated that infants of both ages successfully discriminated the two silent monkey calls and Experiment 2 indicated that 8–10 month-old infants successfully discriminated the two monkey vocalizations and that they did so on the basis of their duration. Together, these findings indicate that unisensory processing deficits do not account for the developmental decline. Finally, Experiment 3 showed that neither 12- nor 18-month-old infants performed cross-species intersensory matching of synchronous faces and vocalizations indicating that the decline in cross-species intersensory matching persists into later development.

Role of Intersensory Mechanisms

The finding that intersensory temporal synchrony mediates cross-species intersensory matching early in development is consistent with the general theoretical notion that infants begin life by integrating multisensory attributes primarily on the basis of whether or not they

are in close temporal register (Lewkowicz, 2000a; Thelen & Smith, 1994). This finding is also in line with a body of evidence indicating that young infants are sensitive to audio-visual temporal synchrony relations in speech and in non-speech events (Lewkowicz, 1992b, 1996, 2000b, 2003), that they prefer synchronous over asynchronous audiovisual speech (Dodd, 1979), that they rely on intersensory temporal synchrony when learning multisensory events (Bahrick & Lickliter, 2000), and that they rely on intersensory temporal synchrony when they respond to multimodal emotional expressions (Walker-Andrews, 1986).

Although intersensory temporal synchrony appears to mediate integration of nonnative faces and voices, it is possible that infants in the Lewkowicz and Ghazanfar (2006) study also relied on duration equivalence to perform intersensory matching. This is because, in addition to being synchronous, the corresponding visible and audible calls in that study corresponded in terms of their duration. Despite this possibility, the overall pattern of results shows clearly that duration did not mediate matching. Neither the younger infants in the current study nor the older infants in the Lewkowicz and Ghazanfar (2006) study made intersensory matches despite the fact that the duration of one of the visible calls was the same as the duration of the vocalization. It is interesting to note that the failure of the 4–6 month-old infants in Experiment 1 to match, despite the fact that the visual and auditory stimuli corresponded in terms of duration, is consistent with results from previous investigations of infant intersensory response to duration (Lewkowicz, 1986). In these investigations, Lewkowicz found that 6- and 8-month-old infants can match equal-duration flashing checkerboards and tones but only when they are also synchronous. In other words, like in the current study, when the visual and auditory stimuli were not synchronous, infants did not match them on the basis of duration equivalence.

It is also interesting to note that the 12- and 18-month-old infants in the current study, as well as the older infants in the Lewkowicz and Ghazanfar (2006) study, also failed to match the visible and audible calls and that they did so despite the fact that these stimuli corresponded in terms of duration and synchrony. This failure suggests that duration/synchrony-based intersensory integration depends, in part, on the nature of the information to be matched. Moreover, it may reflect a shift in the mechanisms of intersensory integration in early human development. That is, even though intersensory synchrony plays a key role in the integration of auditory and visual information early in infancy, basing one's integration of multisensory information on this cue alone does not permit the perceiver to detect the higher-level intersensory relations that are typically available in dynamic faces and the vocalizations that they produce.

Fortunately, as infants grow and acquire perceptual experience, they shift their attention away from simple synchrony-based intersensory relations to higher-level ones (Gibson, 1969; Lewkowicz, 2000a, 2002; Lickliter & Bahrick, 2000; Walker-Andrews, 1997). For example, whereas 5-month-olds can only perceive multimodally specified affect when its audible and visible attributes are synchronous, 7-month-olds can perceive it even when these attributes are asynchronous (Walker-Andrews, 1986). What makes the current findings from the 12- and 18-month-old infants and those from the older infants in the Lewkowicz and Ghazanfar (2006) study interesting is that they suggest that the presumed shift away from responsiveness to synchrony-based audio-visual relations and towards responsiveness to higher-level intermodally invariant relations does not seem to occur in infants' cross-species intersensory responsiveness. This suggests that older infants no longer rely on synchrony to integrate faces and their vocalizations because they are now presumably searching for more complex intermodally invariant stimulus features. The problem is that when they search for such features, they are unable to detect them because they have not had any experience with nonnative faces and their vocalizations.

Role of Unisensory Mechanisms

In addition to demonstrating that younger infants rely on temporal synchrony cues for successful intersensory matching, the current study provides evidence that older infants' failure to make intersensory matches in the Lewkowicz and Ghazanfar (2006) study was not due to unisensory processing deficits. Specifically, Experiment 1 and findings from the Lewkowicz and Ghazanfar (2006) study indicated that 8–10 month-old infants can discriminate between the visible coo and grunt and the findings from Experiment 3 showed that 12- and 18-month-old infants can do so as well. The findings from Experiment 2 showed that 8–10 month-old infants also can discriminate between the two audible calls. This latter ability contributes in an indirect way to synchrony-based intersensory matching because infants must be able to perceive the specific durations of the visible and audible information in order to detect their synchronous onsets and offsets. Overall, the visual and auditory discrimination results demonstrate that the older infants' failure to make intersensory matches was not due to unisensory perceptual processing deficits but rather to a specific failure to integrate the visual and acoustic attributes of the monkey calls.

In addition to ruling out unisensory processing deficits, the fact that infants performed unisensory discrimination has important implications for our understanding of the nature and specific timing of the developmental decline in infant responsiveness to nonnative vocalizations and nonnative faces. With regard to vocalizations, the 8–10 month-old infants' discrimination of the audible calls suggests that a decline in infant responsiveness to the vocalizations of other species either does not occur, occurs later in development, or that older infants do not assimilate such vocalizations to their native vocalization categories and, as a result, respond to such vocalizations in terms of low-level acoustic features. With regard to the reported decline in responsiveness to nonnative faces, the current study showed that the ability to discriminate between different dynamic monkey facial expressions is present as late as 18 months of age. This differs from the fact that 9-month-old infants no longer discriminate between static representations of monkey faces (Pascalis et al., 2002) and suggests that the decline in infant discrimination of the faces of other species may be restricted to static faces. This conclusion is consistent with the finding that attributes that specify static faces provide less information than do attributes that specify dynamic faces (O'Toole, Roark, & Abdi, 2002). This, in turn, suggests that dynamic attributes may maintain discriminability of nonnative faces into later development. At the same time, however, it should be noted that the positive effects of dynamic cues on discriminability may be restricted to the faces of other species. This is suggested by a recent study by Weikum and colleagues (Weikum et al., 2007) reporting that the ability to discriminate between visually specified native and nonnative speech declines by eight months of age despite the fact that the faces are dynamic. Thus, the facilitative effects of dynamic cues on discrimination most likely interact with the degree of experience with a particular type of face.

In conclusion, the present findings demonstrate that intersensory integration of the faces and vocalizations of other species is based primarily on their usual synchronous occurrence, that this ability declines by the latter part of the first year of life, that the decline is not due to unisensory processing deficits, and that the decline persists into later development. From a conventional developmental perspective, it is reasonable to expect that increasing experience with vocalizing faces may actually promote the development of better, rather than worse, intersensory integration skills. If so, older infants can be reasonably expected to be able to overcome their initial inability to integrate the auditory and visual attributes of nonnative vocalizing faces. The fact that they don't suggests that experience after 10 months of age either maintains the already narrowed perceptual processing characteristics of the human perceptual system and/or continues to narrow them more. In either case, this developmental pattern is similar to the irreversible type of experience-dependent perceptual narrowing that occurs in

avian species (Gottlieb, 1991a). The parallels between avian and human canalization effects make it tempting to speculate that these types of effects reflect a general characteristic of perceptual development and that this characteristic has been conserved in evolution. If that is the case then future studies might ask whether this decline occurs in other domains of perceptual functioning as, for example, in the perception of multisensory affect, and whether this phenomenon is widespread across other primate as well as non-primate species.

Acknowledgements

References

Bahrick LE. Infants' perception of substance and temporal synchrony in multimodal events. Infant Behavior & Development 1983;6(4):429–451.

Bahrick LE, Lickliter R. Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. Developmental Psychology 2000;36(2):190–201. [PubMed: 10749076]

Bahrick LE, Lickliter R, Flom R. Intersensory Redundancy Guides the Development of Selective Attention, Perception, and Cognition in Infancy. Current Directions in Psychological Science 2004;13 (3):99–102.

Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of Experimental Psychology: Human Perception and Performance 1988;14(3):345–360. [PubMed: 2971765]

Birch HG, Lefford A. Intersensory development in children. Monographs of the Society for Research in Child Development 1963;25(5)

Birch HG, Lefford A. Visual Differentiation, Intersensory Integration, and Voluntary Motor Control. Monographs of the Society for Research in Child Development 1967;32(2):1–87. [PubMed: 5339915]

Dodd B. Lip reading in infants: Attention to speech presented in- and out-of-synchrony. Cognitive Psychology 1979;11(4):478–484. [PubMed: 487747]

Fowler CA. Speech as a Supramodal or Amodal Phenomenon. Calvert, Gemma A. 2004

Gibson, EJ. Principles of perceptual learning and development. New York: Appleton; 1969.

Gottlieb G. Experiential canalization of behavioral development: Results. Developmental Psychology 1991a;27(1):35–39.

Gottlieb G. Experiential canalization of behavioral development: Theory. Developmental Psychology 1991b;27(1):4–13.

Greenough WT, Black JE, Wallace CS. Experience and brain development. Child Development 1987;58 (3):539–559. [PubMed: 3038480]

Hannon EE, Trehub SE. Metrical Categories in Infancy and Adulthood. Psychological Science 2005a; 16(1):48–55. [PubMed: 15660851]

Hannon EE, Trehub SE. Tuning in to musical rhythms: infants learn more readily than adults. Proc Natl Acad Sci U S A 2005b;102(35):12639–12643. [PubMed: 16105946]

Jusczyk, PW. The discovery of spoken language. Cambridge, MA: MIT Press; 1997.

Kahana-Kalman R, Walker-Andrews AS. The role of person familiarity in young infants' perception of emotional expressions. Child Development 2001;72(2):352–369. [PubMed: 11333071]

Kellman, PJ.; Arterberry, ME. The cradle of knowledge: Development of perception in infancy. Cambridge, MA: MIT Press; 1998.

Kelly DJ, Quinn PC, Slater AM, Lee K, Ge L, Pascalis O. The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. Psychological Science 2007;18(12):1084–1089. [PubMed: 18031416]

Kelly DJ, Quinn PC, Slater AM, Lee K, Gibson A, Smith M, et al. Three-month-olds, but not newborns, prefer own-race faces. Developmental Science 2005;8(6):F31–F36. [PubMed: 16246233]

Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. Science 1982;218(4577):1138–1141. [PubMed: 7146899]

Kuhl PK, Meltzoff AN. The intermodal representation of speech in infants. Infant Behavior & Development 1984;7(3):361–381.

Kuhl PK, Tsao FM, Liu HM. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc Natl Acad Sci U S A 2003;100(15):9096–9101. [PubMed: 12861072]

Kuhl PK, Williams KA, Lacerda F, Stevens KN, et al. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992;255(5044):606–608. [PubMed: 1736364]

Kuo, ZY. The dynamics of behavior development: An epigenetic view. New York: Plenum; 1976.

Lewkowicz DJ. Developmental changes in infants' bisensory response to synchronous durations. Infant Behavior & Development 1986;9(3):335–353.

Lewkowicz DJ. Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. Infant Behavior & Development 1992a;15(3):297–324.

Lewkowicz DJ. Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. Perception & Psychophysics 1992b;52(5):519–528. [PubMed: 1437484]

Lewkowicz DJ. Perception of auditory-visual temporal synchrony in human infants. Journal of Experimental Psychology: Human Perception & Performance 1996;22(5):1094–1106. [PubMed: 8865617]

Lewkowicz DJ. The development of intersensory temporal perception: An epigenetic systems/limitations view. Psychological Bulletin 2000a;126(2):281–308. [PubMed: 10748644]

Lewkowicz DJ. Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. Child Development 2000b;71(5):1241–1257. [PubMed: 11108094]

Lewkowicz DJ. Heterogeneity and heterochrony in the development of intersensory perception. Cognitive Brain Research 2002;14:41–63. [PubMed: 12063129]

Lewkowicz DJ. Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. Developmental Psychology 2003;39(5):795–804. [PubMed: 12952394]

Lewkowicz DJ, Ghazanfar AA. The decline of cross-species intersensory perception in human infants. Proc Natl Acad Sci U S A 2006;103(17):6771–6774. [PubMed: 16618919]

Lickliter R, Bahrick LE. The development of infant intersensory perception: advantages of a comparative convergent-operations approach. Psychological Bulletin 2000;126(2):260–280. [PubMed: 10748643]

Mendelson MJ, Ferland MB. Auditory-visual transfer in four-month-old infants. Child Development 1982;53(4):1022–1027. [PubMed: 7128253]

Munhall, KG.; Vatikiotis-Bateson, E. The moving face during speech communication. In: Campbell, R.; Dodd, B.; Burnham, D., editors. Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis; 1998. p. 123-139.

Munhall, KG.; Vatikiotis-Bateson, E. Spatial and Temporal Constraints on Audiovisual Speech Perception. In: Calvert, GA.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. Cambridge, MA: MIT Press; 2004. p. 177-188.

O'Toole AJ, Roark DA, Abdi H. Recognizing moving faces: A psychological and neural synthesis. Trends in Cognitive Sciences 2002;6(6):261–266. [PubMed: 12039608]

Pascalis O, Haan Md, Nelson CA. Is face processing species-specific during the first year of life? Science 2002;296(5571):1321–1323. [PubMed: 12016317]

Pascalis O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, et al. Plasticity of face processing in infancy. Proc Natl Acad Sci U S A 2005;102(14):5297–5300. [PubMed: 15790676]

Patterson ML, Werker JF. Matching phonetic information in lips and voice is robust in 4.5-month-old infants. Infant Behavior & Development 1999;22(2):237–247.

Patterson ML, Werker JF. Infants' ability to match dynamic phonetic and gender information in the face and voice. Journal of Experimental Child Psychology 2002;81(1):93–115. [PubMed: 11741376]

Patterson ML, Werker JF. Two-month-old infants match phonetic information in lips and voice. Developmental Science 2003;6(2):191–196.

Piaget, J. The origins of intelligence in children. New York: International Universities Press; 1952.

Poulin-Dubois D, Serbin LA, Derbyshire A. Toddlers' intermodal and verbal knowledge about gender. Merrill-Palmer Quarterly 1998;44(3):338–354.

Poulin-Dubois D, Serbin LA, Kenyon B, Derbyshire A. Infants' intermodal knowledge about gender. Developmental Psychology 1994;30(3):436–442.

Sangrigoli S, de Schonen S. Recognition of own-race and other-race faces by three-month-old infants. Journal of Child Psychology and Psychiatry 2004;45(7):1219–1227. [PubMed: 15335342]

Thelen, E.; Smith, LB. A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press; 1994.

Walker-Andrews AS. Intermodal perception of expressive behaviors: Relation of eye and voice? Developmental Psychology 1986;22:373–377.

Walker-Andrews AS. Infants' perception of expressive behaviors: Differentiation of multimodal information. Psychological Bulletin 1997;121(3):437–456. [PubMed: 9136644]

Walker-Andrews AS, Bahrick LE, Raglioni SS, Diaz I. Infants' bimodal perception of gender. Ecological Psychology 1991;3(2):55–75.

Walker-Andrews AS, Lennon EM. Auditory-visual perception of changing distance by human infants. Child Development 1985;56(3):544–548. [PubMed: 4006566]

Walker AS. Intermodal perception of expressive behaviors by human infants. Journal of Experimental Child Psychology 1982;33:514–535. [PubMed: 7097157]

Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. Visual language discrimination in infancy. Science 2007;316(5828):1159. [PubMed: 17525331]

Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior & Development 1984;7(1):49–63.

Werner, H. Comparative psychology of mental development. New York: International Universities Press; 1973.

Yehia H, Rubin P, Vatikiotis-Bateson E. Quantitative Association of Vocal-Tract and Facial Behavior. Speech Communication 1998;26(1–2):23–43.
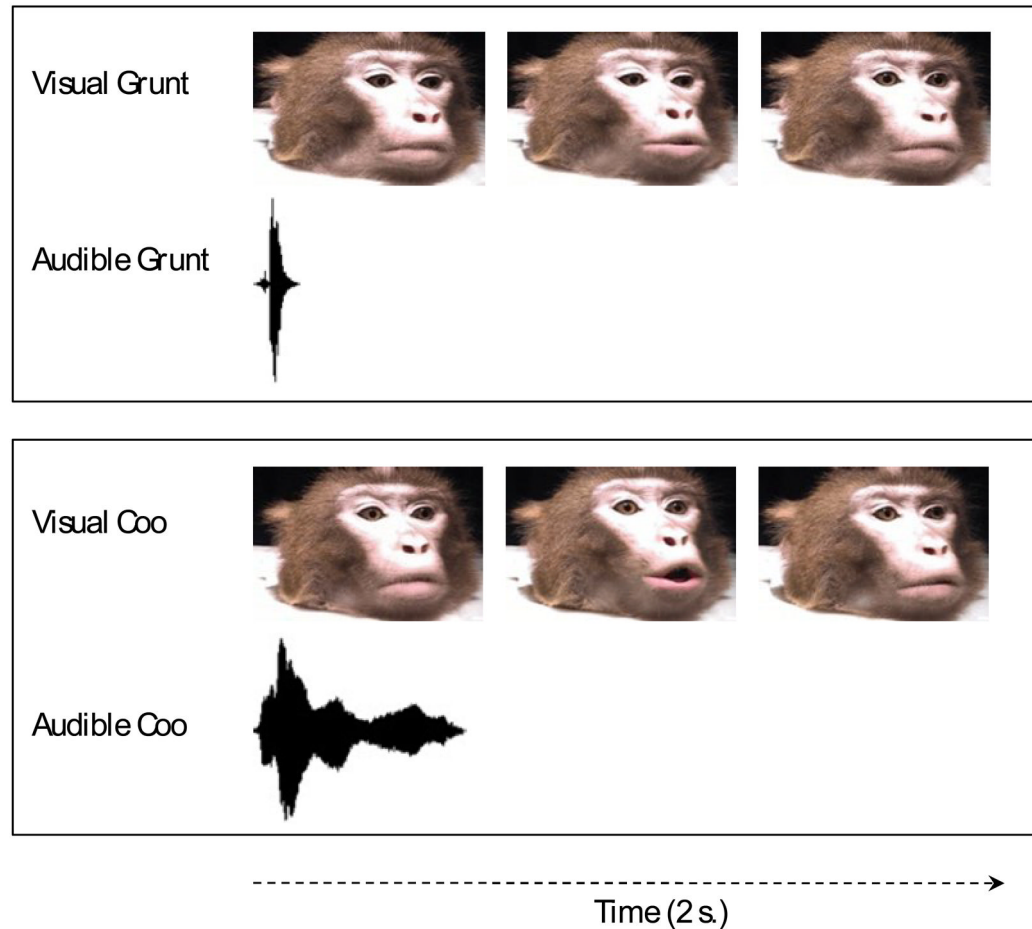
Figure 1.
The asynchronous temporal relation between the audible and visible calls during a single 2-s video clip presented during the in-sound test trials (as indicated in the Methods section, during the actual test trial, the two different visible calls depicted here were shown side-by-side on two different monitors and were shown repeatedly for the duration of the trial and either the audible coo or grunt were presented at the same time). The first facial expression seen on the left shows that the animal's lips were closed during the presentation of the audible call, the second facial expression shows that no audible call was presented while the lips were open and the animal could be seen making the call, and the third facial expression shows that no audible call was presented when the lips were closed following termination of the visible call.
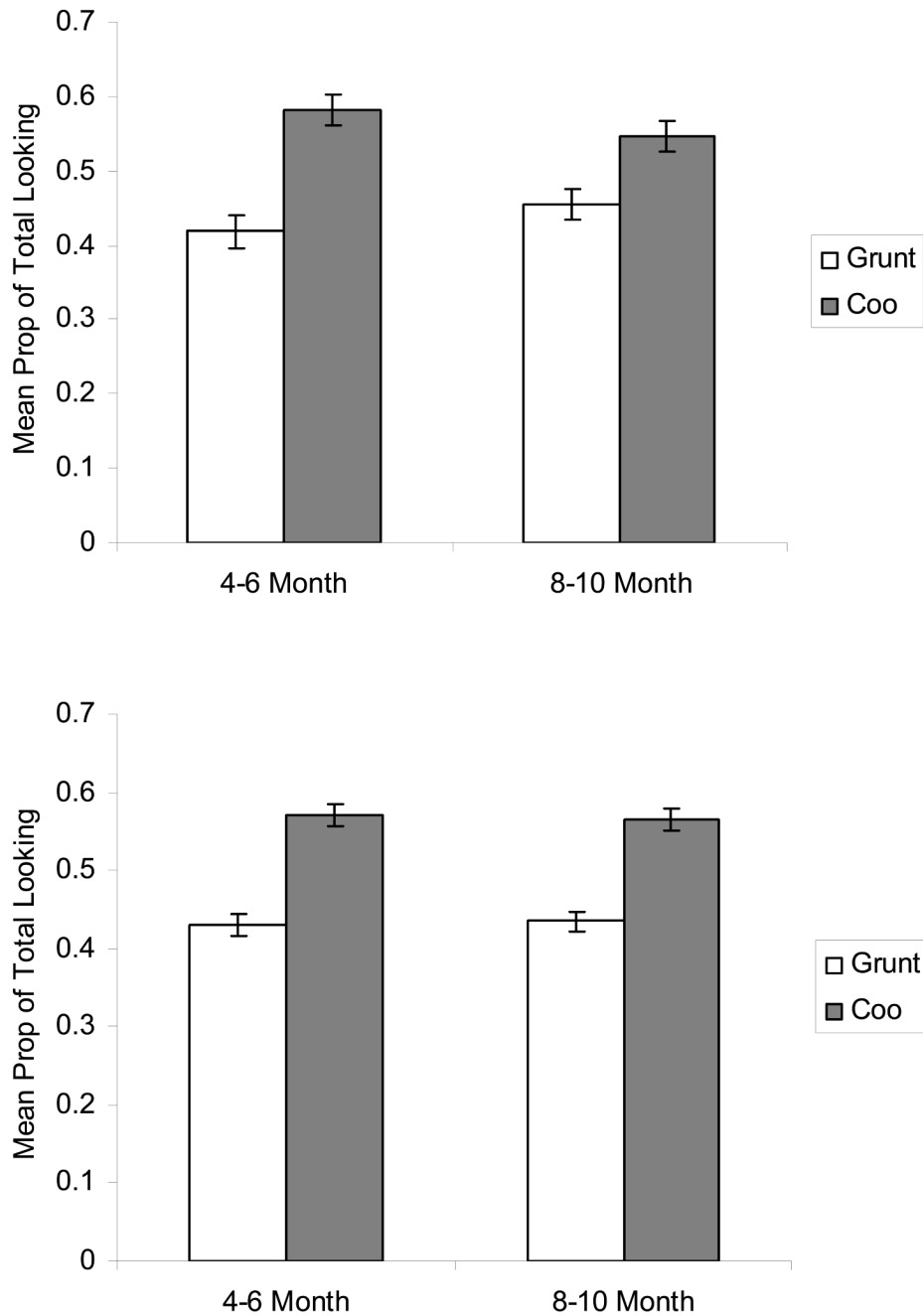
Figure 2.
a. Proportion of total looking directed at the grunt and at the coo in the silent condition in Experiment 1. b. Proportion of total looking directed at the grunt and at the coo in the silent condition in the Lewkowicz & Ghazanfar (2006) study.
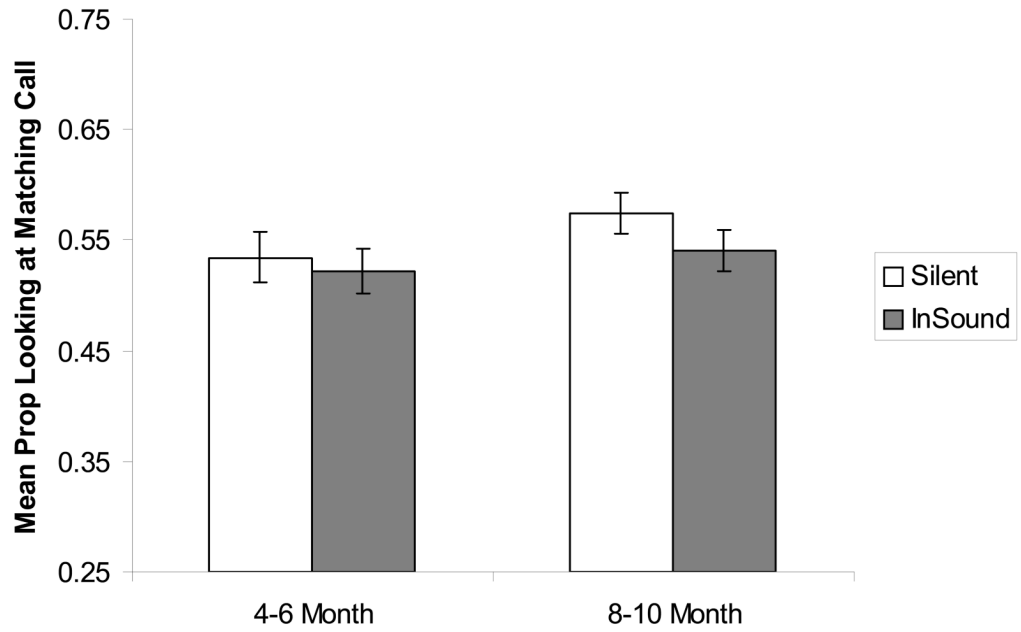
Figure 3.
Proportion of looking at the matching call in the silent and in-sound conditions in 4–6 and 8–10 month-old infants in Experiment 1.
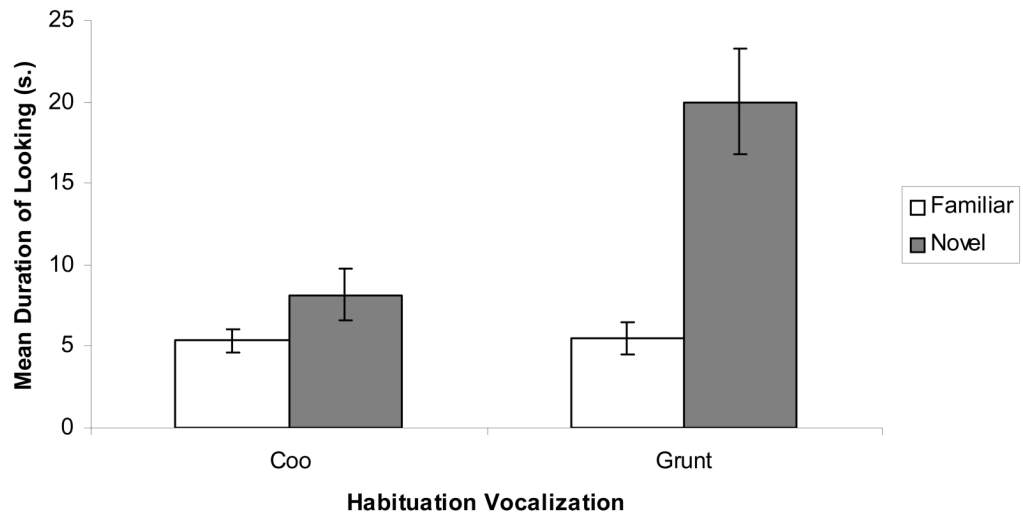
Figure 4.
Discrimination of the coo and grunt vocalizations in Experiment 2.
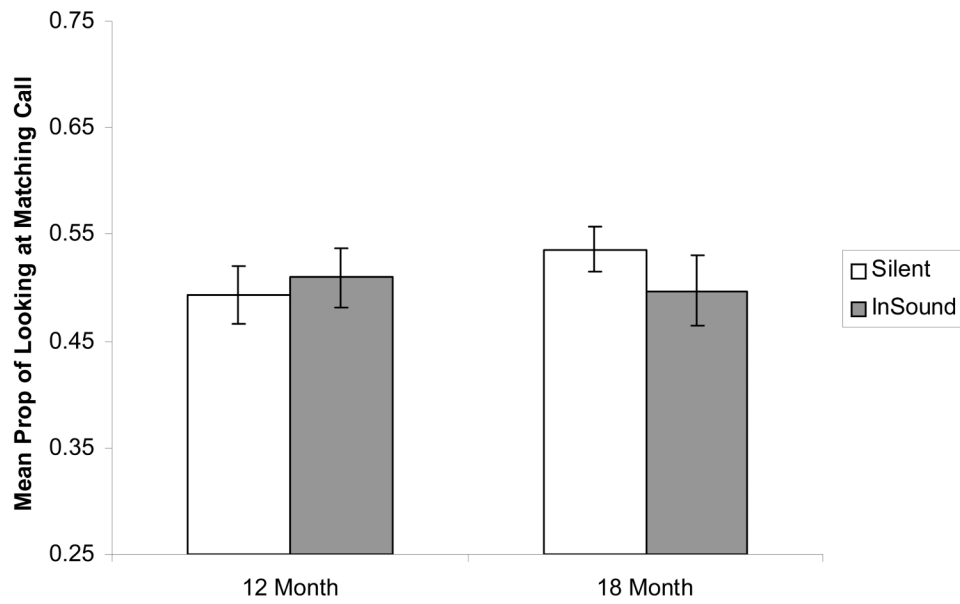
Figure 5.
Proportion of total looking at the matching call in the silent and the in-sound conditions in 12- and 18-month-old infants in Experiment 3.