⊛WILEY-BLACKWELL

# Intersensory Perception at Birth: Newborns Match Nonhuman Primate Faces and Voices

David J. Lewkowicz
*Department of Psychology*
*Florida Atlantic University*

Irene Leo and Francesca Simion
*Dipartimento di Psicologia dello Sviluppo e della Socializzazione*
*Università degli Studi di Padova*

Previous studies have shown that infants, including newborns, can match previously unseen and unheard human faces and vocalizations. More recently, it has been reported that infants as young as 4 months of age also can match the faces and vocalizations of other species raising the possibility that such broad multisensory perceptual tuning is present at birth. To test this possiblity, we investigated whether newborns can match monkey facial and vocal gestures. Using a paired preference procedure, in Experiment 1 we presented pairs of different visible monkey calls in silence and then in the presence of one or the other corresponding audible call and compared preferences across the silent and in-sound conditions. In Experiment 2, we presented the same monkey visible calls but this time together with a tone analog of the natural calls in the in-sound trials. We found that newborns looked longer at the matching visible call in the in-sound condition than in the silent condition in both experiments. These findings indicate that multisensory perceptual tuning is so broad at birth that it enables newborns to integrate the facial and vocal gestures of other primates and that integration is based on newborns' detection of audio-visual temporal synchrony relations.

Correspondence should be sent to David J. Lewkowicz, Department of Psychology, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431. E-mail: lewkowic@fau.edu

Our everyday perceptual world is usually specified by concurrent multisensory information (Gibson, 1966; Lewkowicz, 2000; Stein & Meredith, 1993). For example, whenever we interact with other people, we can see their facial gestures as well as hear their concurrent vocalizations. The fact that most of our daily perceptual experiences are usually multisensory is highly advantageous because multisensory inputs provide perceivers with redundant information that increases stimulus salience and enhances detection, discrimination, and learning (Bahrick, Lickliter, & Flom, 2004; Ernst & Bülthoff, 2004; Lewkowicz & Kraebel, 2004; Partan & Marler, 1999; Rowe, 1999; Stein & Meredith, 1993; Sumby & Pollack, 1954; Summerfield, 1979).

Despite the fact that multisensory redundancy is highly advantageous, the benefits arising from it can only be realized if the diverse information specified in the different modalities can be perceived as belonging to the same object or event. This requires that the perceiver be capable of detecting the various types of relations that often specify multisensory sources of information. For example, audiovisual speech is represented by a hierarchy of audio-visual (AV) relations (Munhall & Vatikiotis-Bateson, 2004). At the bottom of this hierarchy are relatively simple AV relations that specify the synchronous onsets and offsets of facial gestures and vocalizations and their spatial co-location. At the next level of the hierarchy are AV relations that derive from the correlation between the dynamics of vocal tract motion and the dynamics of accompanying vocalizations and can be specified by such amodal attributes as duration, tempo, and rhythmical patterning. Finally, at the highest level of the hierarchy are abstract AV relations that specify amodal categorical attributes, such as gender, affect, and identity.

Although the ability to perceive AV relations emerges early in human development, initially infants are only capable of detecting the relatively low-level types of AV relations. For example, infants as young as 3 weeks of age can perceive the equivalent intensity of auditory and visual stimulation and infants as young as 2 months of age can perceive the synchronous onsets and offsets of auditory and visual stimulation (Lewkowicz, 2000; Lewkowicz & Turkewitz, 1980). The fact that young infants are capable of detecting these two types of intersensory relations is not surprising because in the case of intensity infants only have to detect the equivalence of energy magnitude across modalities and in the case of intersensory temporal synchrony they only have to detect energy onsets and offsets. There is no doubt that the relatively simple AV relations, such as intensity and temporal synchrony, provide infants with only very basic relational information. Nonetheless, they make it possible for young infants to begin to discover the coherent nature of their multisensory world. As they do so, through the continuous process of perceptual learning and differentiation, they then begin to discover the higher level types of AV relations (Lewkowicz, 2000, 2002;

Walker-Andrews, 1997). For example, even though newborn infants can match simple facial and vocal speech gestures (Aldridge, Braga, Walton, & Bower, 1999; Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003; Sai, 2005), only older infants (6 months of age and older) can perceive amodal gender and affect and they can do that even in the absence of temporal synchrony cues (Patterson & Werker, 2002; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991; Walker-Andrews & Gibson, 1986).

The developmental improvement in infant perception of vocalizing faces is consistent with the conventional theoretical view that structure and function broadens and improves as development progresses (Gibson, 1969; Piaget, 1952; Thelen & Smith, 1994). Despite that, a number of studies of responsiveness to unisensory perceptual features have provided evidence that runs contrary to this developmental broadening view. In essence, these studies have indicated that a process of perceptual narrowing also operates during infancy and that it leads to a decline in responsiveness to some key perceptual attributes. For example, it has been found that young infants can discriminate between different monkey faces (Pascalis, Haan, & Nelson, 2002), other-race faces (Kelly et al., 2007), non-native audible speech contrasts (Best, McRoberts, & Sithole, 1988; Kuhl, Williams, Lacerda, Stevens, & Lindblom,1992; Werker & Tees, 1984), non-native visible speech contrasts (Weikum et al., 2007), and non-native musical rhythms (Hannon & Trehub, 2005), but that older infants no longer do. In addition, and most pertinent to the issue at hand, recent studies have demonstrated for the first time that intersensory responsiveness also narrows during infancy. For example, Pons, Lewkowicz, Soto-Faraco, and Sebastián-Gallés (2009) have reported that infant intersensory response to a non-native phonetic contrast narrows between 6 and 11 months of age. In a similar vein, Lewkowicz and Ghazanfar (2006) found that young (4- to 6-month-old) infants can match vocalizing monkey faces with the corresponding monkey vocalizations but that older (8- to 10-month-old) infants no longer do. Moreover, in a follow-up study, Lewkowicz, Sowinski, and Place (2008) found that the ability to perform cross-species AV integration is mediated by the younger infants' response to the synchronous onsets and offsets of the matching facial and vocal gestures.

Earlier it was suggested that AV integration during the first months of life is based primarily on responsiveness to low-level intersensory relations (e.g., temporal synchrony). If that is the case, then it is actually not surprising that young infants should exhibit the kind of cross-species AV integration found by Lewkowicz and Ghazanfar (2006). In other words, if young infants attend primarily to the onsets and offsets of corresponding facial and vocal gestures and if, as a result, they cannot as yet extract the higher level perceptual features inherent in the gestures, then it is not surprising that younger

infants' multisensory tuning is much broader than older infants' tuning. Given that this is the case at 4–6 months of age, it is highly likely that it is at birth as well. So far, this possibility has not been tested but if it turns out to be the case then this would have important theoretical implications for our understanding of the development of intersensory perception and of the processes involved in the developmental construction of multisensory coherence. For example, Gibson's (1969) highly influential theoretical view of the development of intersensory perception holds that humans are ready to pick up the intermodal structure of the multisensory perceptual array at birth. Interestingly, however, Gibson never considered the possibility that multisensory tuning may be so broad at birth that newborns may even be capable of perceiving the faces and voices of other related species in a unitary fashion. In other words, Gibson's theory of intersensory development underestimates the breadth of initial global perceptual organization.

Based on the fact that young infants rely primarily on low-level AV relations for intersensory integration, here we tested the likelihood that newborn infants may be capable of matching the faces and vocalizations of another primate species, the rhesus monkey (*Macacca mulatta*). In brief, we used an intersensory matching procedure to measure visual preferences for each of two side-by-side visual monkey calls. First, we measured preferences in silence and then in the presence of the audible call that corresponded to one of the visible calls. If infants perceived the visible calls and the accompanying audible vocalizations as belonging to the same event then we expected that they would look longer at a given visible call in the presence of the matching audible call than in its absence (Experiment 1). If infants also perceived the unitary nature of the visible and audible calls on the basis of temporal synchrony—and not the higher level AV relations—then we expected that they would continue to make intersensory matches even when the higher level relations were disrupted (Experiment 2).

## EXPERIMENT 1

### Method

#### Participants

We tested sixteen 1- to 3-day-old (mean age = 42 hr) newborn infants (3 girls, 13 boys) who were predominantly Caucasian, middle-class, healthy, full-term, had a normal delivery, had birthweights of 2,015–4,110 g, and a 5-min Apgar score of 9 or 10. We tested an additional eight infants but did not include them in the final sample because of fussing ($n = 2$), a strong position preference ($n = 3$; defined as looking in one direction more than 80% of time), distraction ($n = 1$), and technical error ($n = 2$).

### Apparatus and stimuli

During testing, infants were seated on an experimenter's lap in front of a 30-inch monitor at a distance of 30 cm in a quiet room (ambient sound pressure level (SPL) = 59 dB, A scale) and watched pairs of side-by-side videos of one of two monkeys producing a coo on one side and a grunt on the other. The experimenter holding the infants was blinded to the stimuli by being required to fix his/her gaze on a camera located on the ceiling. Each video was presented on each side of the display monitor and measured 22.5 × 17 cm. A red, blinking, light-emitting diode (LED), located in the middle of the monitor, was used to center the infant's attention between trials. A video camera, located above the monitor, transmitted an image of the infant's face. This image was used by a second experimenter to determine whether the infant was looking at the LED.

### Procedure

As soon as the infant fixated the LED, the second experimenter turned it off, started the videos, and then observed and coded the infant's visual fixations. This experimenter also was blind with respect to the stimuli and could not see them in the infant's corneal reflections. Following testing, an independent observer watched the video recordings of all the infants and coded their visual fixations. Interobserver agreement between this observer and the second experimenter, calculated on the total duration of looking per trial (Pearson correlation), was .92.

We presented two calls that monkeys typically produce during social interactions: a coo and a grunt. The coo is a long tonal signal that is accompanied by a small mouth opening and lip protrusion, while the grunt is a short, pulsatile signal that is accompanied by a larger mouth opening and no lip protrusion. The experimental stimuli consisted of pairs of digitized videos of a rhesus monkey (*M. mulatta*). One of the videos showed the monkey producing a coo while the other video showed the same monkey producing a grunt. To increase generalizability, we presented the faces and vocalizations of two different animals across different infants. The videos were 2-sec digital recordings of the facial gestures associated with the coo and grunt and were presented either in silence or in the presence of one of the corresponding vocalizations and were looped continuously for 1 min. During each 2-sec recording, the visible coos and grunts appeared at the same time and then disappeared at different times because they had different durations. Specifically, the durations of the coo and the grunt were 735 and 180 msec, respectively, for one of the monkeys, and 760 and 275 msec, respectively, for the other monkey. The SPL of the coo and the grunt was 76 and 68 dB, respectively, for one of the monkeys, and 80 and

76 dB, respectively, for the other monkey. The audible calls were presented through two speakers positioned on either side of the monitor.

The experiment consisted of a total of four 1-min test trials, divided into an initial silent condition (two trials) followed by an in-sound condition (two trials). During the silent condition we presented pairs of silent side-by-side faces of the same monkey uttering the two different calls, with side of presentation counterbalanced across the two silent test trials. The data from the silent condition provided a baseline against which we assessed the effects of concurrent audible call presentation in the subsequent in-sound test condition. The in-sound test condition differed from the silent condition only in that this time infants also heard the audible call that corresponded to one of the two visible calls. Half the infants heard the coo vocalization, while the other half of the infants heard the grunt vocalization. The onset of the audible call was synchronized with the onset of both visible calls but its offset was only synchronized with the offset of the corresponding visible call. Figure 1a shows the visible gestures associated with the visible calls of one of the animals and the spectrograms of the corresponding audible calls.
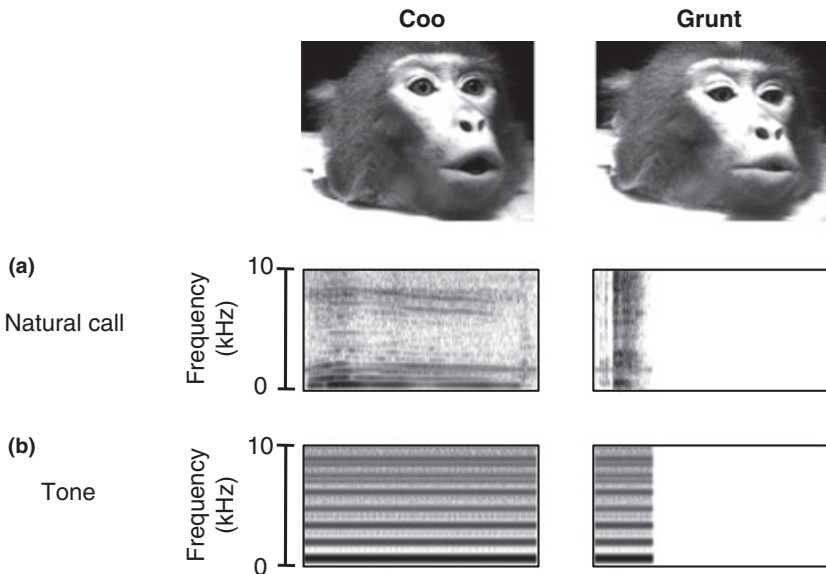


**Figure 1** Single video frames depicting the facial gestures made by one of the monkeys when producing the coo and the grunt. The gestures shown are at the point of maximum mouth opening. Below the facial gestures are the corresponding spectrograms of the natural audible call (a) and the corresponding complex tone (b).

## Results and Discussion

To determine whether infants matched the visible and audible calls, we computed separate proportion of looking scores for the silent and the in-sound conditions. These scores consisted of the proportion of looking directed at the matching visible call out of the total amount of looking directed at both calls. Specifically, for those infants who heard the coo vocalization, we first calculated the total amount of looking at the silent coo face and divided it by the total amount of looking at the silent coo and grunt faces. Then, we calculated the total amount of looking at the in-sound coo face and divided it by the total amount of looking at the in-sound coo and grunt faces. We followed the same procedure for calculating the proportions for the infants who heard the grunt vocalization except that for these infants the number that was entered in the numerator was the total amount of looking at the grunt face. The resulting proportion of looking scores allowed us to ask whether infants matched the visible calls to the corresponding audible calls regardless of the specific vocalization that they heard during the in-sound test trials. If they matched then we expected that they would look longer at the visible call in the presence of the corresponding audible call than in its absence.

A preliminary repeated measures analysis of variance (ANOVA), with call (2) and animal (2) as the between-subjects factors and condition (silent, in-sound) as the within-subjects factor yielded no significant interactions. Consequently, we collapsed the data across call and animal and compared the proportion of looking at the matching call in the presence versus the absence of the audible call by way of a paired, two-tailed, $t$-test. As can be seen in Figure 2a, infants looked significantly longer at the matching monkey face when they heard the audible call than in silence, $t(15) = 2.17$, d.f. $= 15$, $p < .05$, two-tailed, partial $\eta^2 = .25$. These findings show that newborn infants can match the facial and vocal signals of a nonhuman primate species and, thus, confirm our prediction that the newborn's perceptual system is broadly tuned.

The successful intersensory matching that we obtained in this experiment could have been mediated either by the equal duration of the corresponding visible and audible calls and/or by their synchronous onsets and offsets because both cues were available. Previous findings have shown, however, that even older and more perceptually sophisticated infants (4- to 6-month-olds) do not match visible and audible monkey calls when they only correspond in terms of their durations (Lewkowicz et al., 2008). Moreover, other studies have shown that even 3-month-old infants do not match auditory and visual stimuli on the basis of their duration (Lewkowicz, 1986). Thus, the most reasonable conclusion is that the newborns matched on the basis of synchrony.
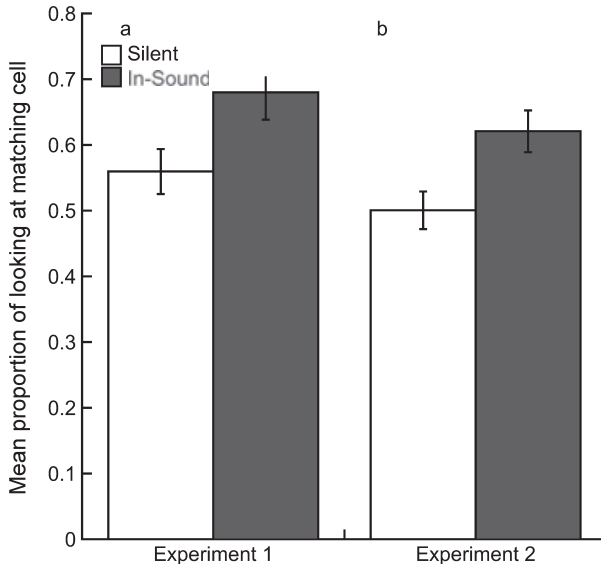
**Figure 2**  Visual preference for matching visible calls in the absence and presence of the matching audible call or tone. (a) Mean proportion of looking at the matching visible call when it was presented in the silent and in-sound trials, respectively, in Experiment 1 where the natural audible call was presented. (b) Mean proportion of looking at the matching visible call in the silent and in-sound test trials, respectively, in Experiment 2 where the complex tone was presented. Error bars represent the standard errors of the mean.

## EXPERIMENT 2

When the successful matching obtained in Experiment 1 is considered together with the fact that young infants can perceive AV synchrony relations (Lewkowicz, 1992, 1996, in press; Scheier, Lewkowicz, & Shimojo, 2003) and that detection of AV temporal synchrony is mediated by low-level, subcortical, tecto-thalamo-insular mechanisms (Bushara, Grafman, & Hallett, 2001), it is clear that the neural mechanisms needed for the detection of AV synchrony are functional at birth. This is especially interesting because newborn infants have a highly immature nervous system and, relatively speaking, are perceptually naïve. Thus, it is unlikely that the newborns in Experiment 1 relied on the extraction of higher level AV relations to match the monkey faces and vocalizations. In particular, it is unlikely that the newborns in Experiment 1 relied on the correlation between the dynamic facial and acoustic cues to match the visible and audible calls. If this conclusion is correct, and if our hypothesis that successful intersensory matching early in life is based primarily on low-level AV relations is also correct then newborns should be

able to match the visible and audible monkey calls even when the correlation between the dynamic facial and acoustic cues is disrupted.

We tested this possibility in Experiment 2 by replacing the natural audible calls presented in Experiment 1 with spectrally rich tones. The tones had a constant intensity and a linear spectral profile and, as a result, no longer had the temporally modulated amplitude envelope and formants that are characteristic of the natural coo and grunt vocalizations. This, in turn, meant that the fine-grained correlation between the amplitude and spectral fluctuations of the audible calls and dynamic facial cues was now disrupted, while the overall onset and offset synchrony cues linking the audible and visible calls were still available. If the fine-grained correlation did not mediate successful intersensory matching in Experiment 1 then newborns should still exhibit intersensory matching.

## Method

### Participants

The participants were 16 (5 girls, 11 boys) healthy, full-term, 1- to 3-day-old infants (mean age = 41 hr). The infants were predominantly Caucasian, middle-class, had a normal delivery, birth weight between 2,300 and 4,280 g, and a 5-min Apgar score of 9 or 10. Four additional infants were excluded from the final sample because three were fussy and one had a strong position preference.

### Apparatus, stimuli, and procedure

The apparatus, stimuli and procedure were identical to those in Experiment 1 with one exception. The naturalistic audible calls were replaced by broadband complex tones (triangular waveform, Adobe Audition 1.5) that matched each call's duration but that eliminated the temporal modulation in the envelope of the signal. The fundamental frequency ($F_0$) of the complex tone was based on an average between the fundamental frequencies of the coo and the grunt and it had a constant intensity and a linear spectral profile. As can be seen in Figure 1b, this effectively removed the amplitude envelope and the formants that make the coo and the grunt so acoustically distinct. The SPL of the coo tone and the grunt tone was 72 and 68 dB, respectively, for one of the monkeys, and 82 and 72 dB, respectively, for the other monkey. Half the subjects heard the complex tone version of the coo, while the other half heard the complex tone version of the grunt. Interobserver agreement calculated on the total amount of looking (Pearson correlation) was .94 for Experiment 2.

Results and Discussion

The results from Experiment 2 supported our hypothesis that the successful cross-species AV matching obtained in Experiment 1 was not based on higher level AV relations. That is, despite the fact that the auditory stimuli no longer resembled the natural calls and that the correlation between the dynamic facial and acoustic cues was disrupted, the newborns still matched the faces and vocalizations. As can be seen in Figure 2b, infants looked longer at the matching visible call in the presence of the corresponding tone than in its absence. To determine if this difference was statistically reliable, we first performed a preliminary ANOVA on the proportion scores from the silent and the in-sound conditions and found that there were no significant interactions between animal, type of call, and condition. As a result, we collapsed the data across animal and call and compared the proportion of looking scores across the two conditions. This comparison indicated that infants looked longer at the matching call when they heard the tone than when they did not hear it, $t(15) = 2.23$, $p < .05$, two-tailed, partial $\eta^2 = .24$.

As expected, we found that the newborns once again successfully performed AV matching in that they matched the naturalistic facial call gestures with the corresponding tones. This finding is particularly interesting for two reasons. First, the newborns made successful AV matches despite the fact that the audible signals no longer carried the critical identity information that normally characterizes audible monkey calls. This suggests that this feature of monkey audible calls does not contribute to newborn infants' ability to match monkey facial and vocal calls. Second, the newborns made the matches, even though the correlation between the dynamic properties of the visible and audible calls was disrupted. This finding is consistent with the findings from Experiment 1 and suggests that the newborns performed the AV matches on the basis of temporal synchrony.

# GENERAL DISCUSSION

The current study investigated the theoretically intriguing possibility that multisensory tuning is so broad at birth that it enables newborn infants to perceive the AV unity of non-native faces and vocalizations. The results from Experiment 1 were consistent with this prediction by demonstrating that newborn infants can match dynamic monkey facial gestures and corresponding vocalizations. Specifically, Experiment 1 showed that newborns looked longer at a monkey face producing a particular visible call when that call was accompanied by the corresponding audible call than when it was not accompanied by it. Experiment 2 investigated one possible basis

for the successful AV matching obtained in Experiment 1 by substituting a tone for the natural audible call and, thus, removing auditory identity information and disrupting the correlation between the dynamic attributes of the visible and audible calls. Despite the absence of identity information and dynamic AV relations, newborns still integrated the auditory and visual information.

Overall, the current results demonstrate for the first time that human newborns can perform cross-species AV matching. This appears to be due to the fact that the newborn perceptual system is so broadly tuned that newborns are not only sensitive to the correlations of multisensory signals of their own species, but that they are sensitive to the correlations of multisensory signals of other primate species as well. The current results also suggest that newborns accomplished the AV matching task by attending to the synchronous onsets and offsets of the matching facial and vocal gestures (recall that the nonmatching facial and vocal gestures had synchronous onsets but asynchronous offsets). The finding that newborns relied on the temporal synchrony of the matching visible and audible signals is of particular interest because it indicates that newborns matched by relying on a low-level AV relation. This is consistent with recent work showing that older infants (4–10 months of age) respond to the temporal relationship between the audible and visible attributes of talking human faces by attending to AV energy onsets and offsets (Lewkowicz, in press). The current findings add to the data from older infants by showing that newborns also rely on energy onsets and offsets for their perception of AV synchrony relations and, like 4- to 6-month-old infants (Lewkowicz & Ghazanfar, 2006), also use this ability to integrate non-native auditory and visual inputs.

Previous studies have shown that newborns can perceive AV intensity relations suggesting that newborns possess a relatively primitive but useful way of bootstrapping their perception of an integrated multisensory world (Lewkowicz & Turkewitz, 1980). The findings from the current study show that newborns' sensitivity to AV temporal synchrony relations provides them with another useful, and particularly powerful, way of linking the visible and audible attributes of everyday multisensory events. As our findings indicate, this sensitivity is so broad that it even enables newborns to integrate non-native facial and vocal gestures.

Our finding that a sensitivity to intersensory synchrony at birth mediates AV matching is consistent with the theoretical view that infants are selectively tuned to intersensory temporal coherence from birth on and that this tuning enables them to build a multimodally unified conception of their world (Gibson, 1969; Piaget, 1954; Thelen & Smith, 1994). As indicated earlier, our findings demonstrate that a relatively low-level perceptual ability provides a critical adaptation that helps bootstrap intersensory perception

at birth and that this, in turn, helps launch infants on their path to the development of a unitary conception of their largely multisensory world. This, in turn, is consistent with the classic theoretical view that the development of intersensory perception proceeds from a state of global unity of the senses to a state where the sensory modalities become differentiated from one another (Gibson, 1969; Werner, 1973).

Critically, however, our findings demonstrate that the classic view underestimates the breadth of initial global perceptual organization and, in so doing, ignores humans' evolutionary history. In our view, the broad perceptual tuning that we have uncovered reflects humans' evolutionary past as well as the evolution of an adaptation that probably arose to confer maximum perceptual plasticity at the beginning of life. This adaptation makes it possible for developing infants to take full advantage of all the multisensory stimulation available in their natural ecology and, at the same time, allows them to tune themselves to the features of their own species through species-specific experience (Gottlieb, 1991b). Once they begin to do that they not only begin to discover the higher level modality-specific perceptual attributes associated with inputs in different modalties, but they also begin to discover higher level intersensory relations. As this occurs, their need to rely solely on low-level multisensory relational cues begins to wane.

Our findings provide a glimpse into the postnatal beginnings of the perceptual narrowing process. When they are considered together with available evidence of intersensory perceptual narrowing (Lewkowicz & Ghazanfar, 2006; Lewkowicz et al., 2008; Pons et al., 2009), the current findings suggest that as infants accumulate selective experience with the faces and vocalizations of their own species, they cease to integrate the facial and vocal gestures of other species because such gestures do not represent their everyday perceptual experience. It is interesting to note, however, that the perceptual narrowing that has so far been observed in early human development may not be a universal feature of development. Recent studies of perceptual narrowing in young vervet monkeys have shown that this species does not exhibit narrowing of intersensory responsiveness to the facial and vocal gestures of rhesus monkeys (Zangenehpour, Ghazanfar, Lewkowicz, & Zatorre, 2009). When this finding is considered together with the fact that birds exhibit perceptual narrowing in their response to heterospecific vocalizations (Gottlieb, 1991a), the picture that emerges is a complex one. First, it is possible that unisensory and intersensory perceptual narrowing processes differ across different species. Second, it is possible that vervet monkeys may exhibit narrowing of intersensory integration later in life. Finally, it may be that vervets may not undergo narrowing at all because of their more precocious neural state at birth. That is, they may not be capable of incorporating the effects of selective early experience due to the relative lack of neural

plasticity. If this is the case, then higher primates (e.g., apes) may exhibit perceptual narrowing because their fetal and neural development is more similar to that of humans. Future studies should investigate this possibility. In the meantime, the current findings indicate that humans begin postnatal life with a broadly tuned perceptual system that makes it possible for them to integrate the audible and visible signals of other species by relying on a low-level but powerful perceptual processing mechanism that is sensitive to the onsets and offsets of stimulus energy in different sensory modalities.

## ACKNOWLEDGMENTS

## REFERENCES

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, *2*, 42–46.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). 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*, *14*, 345–360.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience*, *21*, 300–304.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin.

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

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

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, *16*, 48–55.

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

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

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

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

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

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

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

Lewkowicz, D. J. (2002). Heterogeneity and heterochrony in the development of intersensory perception. *Cognitive Brain Research*, *14*, 41–63.

Lewkowicz, D. J. (in press). Infant perception of audio-visual speech synchrony. *Developmental Psychology*.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6771–6774.

Lewkowicz, D. J., & Kraebel, K. (2004). The value of multimodal redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing* (pp. 655–678). Cambridge, MA: MIT Press.

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.

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

Munhall, K. G., & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 177–188). Cambridge, MA: MIT Press.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*, 1272–1273.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321–1323.

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

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

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

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

Piaget, J. (1954). *The construction of reality in the child*. London: Routledge & Kegan.

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 of the United States of America*, *106*(26), 10598–10602.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*, 921–931.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, *14*, 29–50.

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

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: The MIT Press.

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

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

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456.

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

Walker-Andrews, A. S., & Gibson, E. J. (1986). What develops in bimodal perception? *Advances in Infancy Research*, *4*, 171–181.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, *316*, 1159.

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.

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

Zangenehpour, S., Ghazanfar, A. A., Lewkowicz, D. J., & Zatorre, R. J. (2009). Heterochrony and cross-species intersensory matching by infant vervet monkeys. *Public Library of Science One*, *4*, e4302.