REPORT

Development of multisensory spatial integration and perception in humans

Patricia A. Neil,[1] Christine Chee-Ruiter,[1] Christian Scheier,[1] David J. Lewkowicz[2] and Shinsuke Shimojo[3]

1. Computation and Neural Systems Department, California Institute of Technology, USA
2. Department of Psychology, Florida Atlantic University, USA
3. JST ERATO Shimojo 'Implicit Brain Function' Project, Japan

Abstract

Previous studies have shown that adults respond faster and more reliably to bimodal compared to unimodal localization cues. The current study investigated for the first time the development of audiovisual (A-V) integration in spatial localization behavior in infants between 1 and 10 months of age. We observed infants' head and eye movements in response to auditory, visual, or both kinds of stimuli presented either 25° or 45° to the right or left of midline. Infants under 8 months of age intermittently showed response latencies significantly faster toward audiovisual targets than toward either auditory or visual targets alone They did so, however, without exhibiting a reliable violation of the Race Model, suggesting that probability summation alone could explain the faster bimodal response. In contrast, infants between 8 and 10 months of age exhibited bimodal response latencies significantly faster than unimodal latencies for both eccentricity conditions and their latencies violated the Race Model at 25° eccentricity. In addition to this main finding, we found age-dependent eccentricity and modality effects on response latencies. Together, these findings suggest that audiovisual integration emerges late in the first year of life and are consistent with neurophysiological findings from multisensory sites in the superior colliculus of infant monkeys showing that multisensory enhancement of responsiveness is not present at birth but emerges later in life.

Introduction

The location of objects and events in our environment is often specified by concurrent auditory and visual inputs. Adults of many species, including humans, take advantage of such multisensory redundancy in spatial localization. Detection, discrimination, and localization are often performed more quickly and more reliably when bimodal as opposed to unimodal cues are available (Miller, 1982). For example, the spatial localization of simple A-V targets is significantly faster than the localization of the same auditory or visual targets alone (Hughes, Reuter-Lorenz, Nozawa & Fendrich, 1994). Despite the fact that adults profit from multimodal source specification when performing spatial localization tasks, it is not known when multisensory facilitation of localization behavior first emerges in human development. Ascertaining when it does is important because the ability to integrate multisensory inputs is critical to the development of a unified perceptual world and ultimately to the acquisition of veridical knowledge (E.J. Gibson, 1982; J. Gibson, 1979; Lewkowicz, 2000a; Piaget, 1952). Research on multisensory perceptual development suggests that multisensory facilitation may develop early in life, although it is not clear at what specific age. To investigate the possibility that facilitation processes might emerge in infancy, we investigated responsiveness to unimodal (auditory and visual) and bimodal localization cues in infants across the first year of life.

Research to date has shown that infants possess a variety of multisensory perceptual abilities (Lewkowicz, 2000b, 2002). For example, it has been shown that infants can perform A-V intensity matching (Lewkowicz, 1980), detect A-V synchrony relations (Dodd, 1979), and even perceive illusions based on A-V interactions (Scheier, Lewkowicz & Shimojo, 2003). Research also has shown that some multisensory abilities differ across early development. Thus, the ability to make duration-based

multisensory matches emerges by 6 months of age (Lewkowicz, 1986) but the ability to perceive synchrony relations embedded in a rhythmic pattern does not emerge until 10 months of age (Lewkowicz, 2003). Developmental differences such as these are not surprising given the rapid changes in basic sensory/perceptual abilities that occur during the first year of life.

So far, investigations of infant multisensory perception have asked whether infants can integrate the auditory and visual attributes of speech, affect, or shape and whether infants can use various forms of temporal information to perceive intermodally unified events (see Lewkowicz, 2000b, for a review of much of this literature). Only a handful of studies have investigated the development of spatial multisensory integration. Some previous work has found that young infants are sensitive to the spatial co-location and temporal contiguity of auditory and visual inputs (Lawson, 1980; Lyons-Ruth, 1977) while other studies indicate that infants as young as 3.5 months of age are insensitive to the spatial separation of auditory and visual inputs when temporally coincident (Bahrick, 1988; Lewkowicz, 1992; Spelke, 1979). However, none of these studies have systematically investigated spatial intersensory integration skills across a wide enough age range during infancy to capture possible developmental changes.

From birth, infants orient their eyes and head toward novel visual and auditory stimuli (Fantz, 1963; Wertheimer, 1961). Voluntary saccades in adults serve to center the fovea on an event or object in visual space, and their generation is controlled through the superior colliculus, a midbrain structure containing spatially aligned maps of visual, auditory, and somatosensory inputs. Animal studies have shown that an examination of neurons in the deeper layers of the superior colliculus reveals a large subset responsive to multiple sensory stimuli from coincident spatial locations. Two characteristic features of these multisensory neurons are an enhanced activation response to spatially co-located multimodal stimuli compared to unimodal activation (and a depressed response with spatially dislocated multimodal stimuli) and lower response thresholds for the single modality components of a multimodal stimulus than for unimodal stimuli (Meredith & Stein, 1983, 1986; Wallace, Wilkinson & Stein, 1996). Several developmental studies in both cats and monkeys suggest that these multisensory neurons, abundant in the adult, are either lacking or incapable of this integrative function in newborns (Wallace & Stein, 1997, 2001). In addition, there is some indirect, behavioral evidence for postnatal changes in the superior colliculus in humans (Harman, Posner, Rothbart & Thomasthrapp, 1994). Finally, multisensory neurons of the superior colliculus receive projections from many cortical and subcortical regions (Wallace, Meredith & Stein, 1993), some of which appear to mediate the multisensory integration observed in the superior colliculus (Jiang, Wallace, Jiang, Vaughan & Stein, 2001). It has been suggested that the projections from cortical association areas to the superior colliculus develop postnatally and may be the final stage necessary for multisensory integration to occur (Wallace & Stein, 2000).

The neurophysiological findings suggest that multisensory facilitation of spatial localization may not be present during human infancy. To investigate whether it is or not, we examined the development of unimodal and multimodal spatial localization in human infants between 1 and 10 months of age and compared it to that in adults. Our measure of responsiveness was the latency of head and eye movements toward auditory, visual, or spatially congruent audiovisual targets located at 25° or 45° to the right or left of midline. Adults are known to respond more rapidly to A-V targets co-localized in space than to targets specified by the auditory or visual component alone. The probability distribution of the response latency to a bimodal target is less than the minimum probability that would be possible by a simple probability summation of the two unimodal conditions (Hughes *et al.*, 1994). This result violates the 'race model' (Miller, 1982) which presumes that response latency is determined by the faster of two sensory inputs rather than by non-linear summation of the auditory and visual inputs. Thus, violations of the race model constitute evidence of multisensory integration/facilitation. The purpose of the current study was, in part, to determine whether response latencies in early human development exhibit violations of the race model and, thus, provide evidence of multisensory spatial integration.

We examined the response latency of infants and adults orienting toward unimodal (visual- or auditory-only) or bimodal (co-localized, temporally coincident audiovisual) targets in the horizontal plane at two possible eccentricities: ±25° and ±45°. As has been shown in many previous studies, response performance and latency can often be manipulated by a target's relative intensity as well as other task-dependent conditions. In order to simplify the comparison of latencies over the wide range in subject ages (approximately 1 month to adult) the intensity levels of the auditory, visual, and audiovisual stimuli were kept constant for all subjects. Visual stimuli were a vertical line of red LEDs and the auditory stimuli a sustained burst of white noise (55–65 dB). The intensity levels of the auditory and visual stimuli were selected to produce comparable response latencies (in pilot studies with adults). To accommodate the short attention span of young infants, the unimodal and bimodal target conditions were separated into separate test sessions

utilizing the same stimulus presentation method. During the first session (Experiment 1) infants were tested with unimodal stimuli and during the second session (Experiment 2) infants were tested with bimodal stimuli.

Experimental design

Method

Participants

Five age groups were defined, *a priori*, and consisted of infants aged: 0–2, 2–4, 4–6, 6–8, and 8–10 months of age. Infants were recruited at various ages from the local community and tested at monthly intervals. During each visit, infants were tested first with the unimodal stimuli (Experiment 1) and then with the bimodal stimuli (Experiment 2). Their performance was given a rating of 'Good', 'Okay', or 'Bad' based upon their attentiveness, fussiness, and completion of each experiment. Only infants with a rating of 'Good' or 'Okay', and only those who had completed both experiments (with the exception of two infants in the 0–2-month group included due to low *N*) were selected for further analysis. From this pool of candidates, 12 subjects per age group (only 11 for the 0–2-month group) were randomly selected. Based on these criteria, a total of 33 full-term infants (16 female, 17 male), with no known medical conditions and ranging from 1.18 to 9.49 months old, participated in our study (Table 1). In addition, seven adults (four female, three male) participated in the study. Parents were encouraged to bring their infant back until he or she reached 10 months of age. As a result, some infants are represented at more than one age. A minimum of 1 month passed between visits for those infants who were tested at more than one age, with repeat participation as follows: one visit (20 infants: 9 at 0–2 mo, 3 at 2–4 mo, 4 at 4–6 mo, 1 at 6–8 mo, 3 at 8–10 mo), two visits (4 infants: 1 at 0–2 and 2–4 mo, 1 at 2–4 and 4–6 mo, 1 at 0–2 and 6–8 mo, 1 at 2–4 and 6–8 mo), three visits (5 infants: 2 at 2–4, 6–8 and 8–10 mo, 3 at 4–6, 6–8, and 8–10 mo), and four visits (4 infants: 4 at 2–4, 4–6, 6–8, and 8–10 mo). The data from 17 other infants were not used due to fussiness or inattentiveness, equipment failure, or failure to complete one or both experiments.

Apparatus and stimuli

The experimental apparatus had five independently controlled stimulus delivery modules positioned at 0°, ±25°, and ±45° on a level semicircular hoop. Each module had nine clusters of four variably colored LEDs (red, yellow, white and green) in a 3″ × 3″ grid. Behind the LED plate was positioned a small speaker. The entire apparatus was encircled by a ceiling-to-floor length black curtain. Two video cameras were mounted, one overhead and one hidden just above the 0° module, with the video signal fed real-time to two monitors outside the curtain where the experimenter sat. Three types of target stimuli were used, auditory-only and visual-only (Experiment 1) and concordant audiovisual (Experiment 2) targets. The visual stimulus consisted of a vertical line of three standard, red LEDs and the auditory stimulus was a sustained burst of white noise (55–65 dB). For the concordant condition, the same visual and auditory stimuli were presented synchronously at the same position. A fixation stimulus of alternating red and green Xs and +s with short bursts of white noise was presented at the center module. The duration of each stimulus was controlled online by the experimenter.

Procedure

Infants were seated 22.5″ from the hoop in a car seat, or on the parent's lap on a stiff foam pillow, with their head centered and level with the modules. When on the lap, parents were instructed to stabilize their child's body but to not move or cue them in any way during the experiment. A piece of adhesive tape (one-quarter inch wide) was affixed to the subject's head, visible from both video cameras, to serve as a spatial calibration measure. The room lights were turned off several minutes prior to starting the experiment to allow the subjects to become dark adapted, and remained off throughout. Padding on the room walls minimized acoustic reflection.

Each trial began with the presentation of the fixation stimulus. When the experimenter determined that the

Table 1 *Total number of subjects used in each age group, their mean age, and the total percentage of valid trials (see Methods) completed. In the youngest age group (0–2 months), one baby completed only Experiment 1, and one completed only Experiment 2*

Subject info\Group	0–2 mo	2–4 mo	4–6 mo	6–8 mo	8–10 mo	adult
N	11*	12	12	12	12	7
Mean age (mo)	1.8	3.1	5.1	6.8	8.7	*Not collected*
% trials good	36.1%	54.9%	62.9%	73.1%	59.4%	87.6%

Table 2 *The group mean was calculated for each modality*eccentricity condition for each age*

| Age (mo) | Visual (Experiment 1) | | | | Auditory (Experiment 1) | | | | Audiovisual (Experiment 2) | | | |
| | 25° | | 45° | | 25° | | 45° | | 25° | | 45° | |
	Mean (ms)	SE	Mean (ms)	SE	Mean (ms)	SE	Mean (ms)	SE	Mean (ms)	SE	Mean (ms)	SE
0–2	849	59	954	66	922	74	1078	63	738	58	694	56
2–4	611	31	675	40	723	34	714	41	493	33	556	42
4–6	387	21	474	21	615	24	525	21	356	23	377	23
6–8	307	17	384	10	535	20	374	10	288	19	347	11
8–10	313	23	363	15	645	31	423	16	278	25	304	16
adult	209	6	212	6	220	6	210	6	173	6	169	6

subject's eyes were fixed on the center module, the fixation stimulus was turned off and the target stimulus was presented at one of the other four positions after a short delay ranging between 300 and 500 ms (this variability was due to the physical constraints of our equipment). After the subject responded either by performing either an eye and/or a head movement, the target stimulus was turned off and the fixation stimulus was presented again. There were eight possible test conditions in Experiment 1 (two modalities × four positions) and four possible test conditions in Experiment 2 (four positions where the bimodal stimulus was presented). In each experiment, each condition was presented five times resulting in a total of 40 trials for Experiment 1 and a total of 20 trials for Experiment 2. The order of target presentation was randomized within each block.

The video, filmed at 29.97 frames per second, was digitized using a Sony DVMC-DA2 Media Converter and captured on a standard PC with a Pinnacle DV500 video capture card and the Adobe Premiere software package. A custom-written software program named Media-Analyzer was used to manually mark the left pupil for each frame of video. These markers provided the horizontal and vertical position displacement of the eye (yoked to the head) which was then converted into velocity data. Response onset was defined as the frame where the horizontal component of the subject's response velocity exceeded 3.1 standard deviations from its baseline value (mean velocity during last 300 ms prior to each target onset) with the vertical velocity below 41.2 mm/s (to distinguish valid responses from random head/body motions). Response latency was then calculated as the time between target onset and the response onset. Trials where the baseline horizontal velocity exceeded 30 mm/s were excluded (i.e. where the infant was moving during fixation and presumably inattentive to the stimulus onset). Trials also were excluded if the subject blinked or lost fixation at stimulus onset, if the response onset was in the wrong direction, or if the response latency was less than 100 ms or greater than 2000 ms, this latter condition not likely being stimulus driven. Outliers were defined as those trials outside the 5th and 95th percentiles (within each modality/eccentricity condition for each age group) and were also removed. All valid trials were pooled within each age group to calculate the grand mean (Table 2).

Results

A three-way univariate ANOVA ($3 \times 2 \times 5$), with modality (auditory-only, visual-only, and audiovisual) and eccentricity (25° or 45°) as the within-subjects factors, and age group (0–2, 2–4, 4–6, 6–8, or 8–10 months) as the between-subjects factor, was used to analyze the response latency data. We adopted the more conservative p value of .01 for significance testing to reduce the possibility of committing a Type I error. Results of the analysis showed that there was a significant main effect of modality, $F(2, 2010) = 124.3$, $p < .001$, and age, $F(4, 2010) = 242.9$, $p < .001$, as well as a significant modality × eccentricity interaction, $F(2, 2010) = 14.1$, $p < .001$, a modality × age interaction, $F(4, 2010) = 3.4$, $p = .001$, and a modality × eccentricity × age interaction, $F(8, 2010) = 3.8$, $p < .001$. Finally, we found a marginally significant eccentricity × age interaction, $F(4, 2010) = 3.2$, $p = .012$.
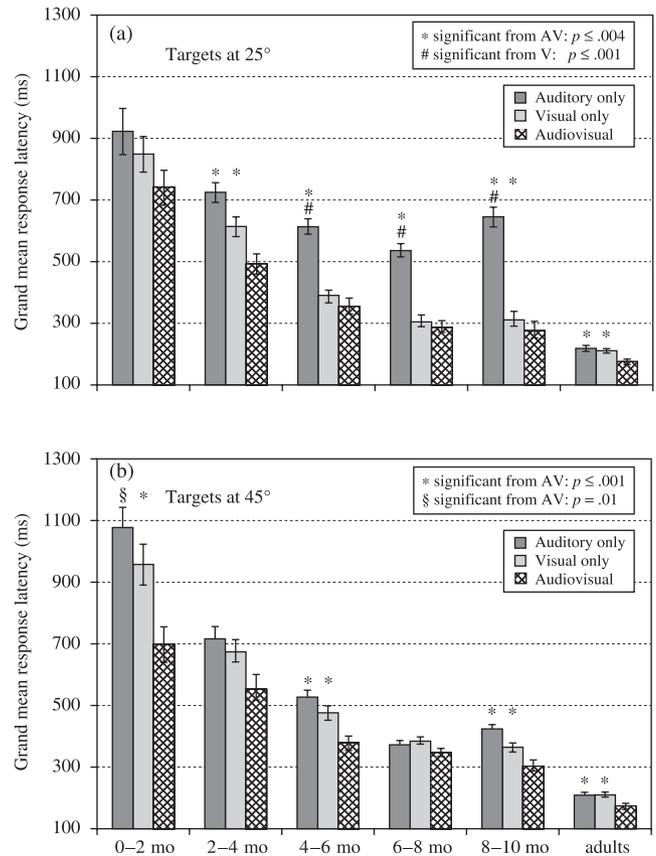
Planned contrast analyses were performed to examine the main effect of age. This was done by comparing the response latencies of infants in one age group with the latencies in next oldest. In addition, we compared the response latencies of the oldest group of infants (8–10 months) and those of adults. These comparisons indicated that there was a significant difference in response latency between each infant age group ($p < .001$), except for that between the two oldest age groups (6–8- versus 8–10-month-olds, $p = .32$). Although there was a steady decrease in the response latency with age (see Figure 1a), the response latency of even the oldest infants was still longer than that of the adults ($p < .001$).
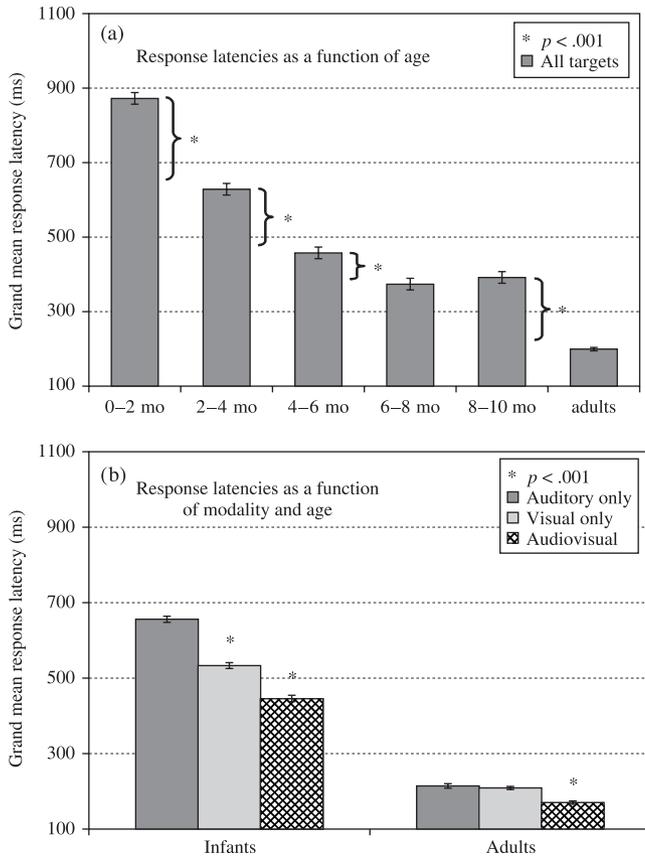
Figure 1   *Main effects of (a) age and (b) modality on response latency. Error bars represent the standard error of the mean.*

Figure 2   *Response latencies for auditory, visual, and audiovisual targets at (a) 25° and (b) 45° eccentricities across all ages. Error bars represent the standard error of the mean.*

Because our sample sizes were unequal and had non-homogeneous variances, we examined the main effect of modality by using multiple post-hoc comparisons using the *Games-Howell* procedure (Field, 2000). Response latency to auditory targets was slower than to both visual and audiovisual targets ($p < .001$), and response latency to visual targets was slower than to audiovisual targets ($p = .001$). This response pattern was different from the adult response pattern where response latency to the auditory and visual targets did not differ and where response latency in both unimodal conditions was slower than in the audiovisual condition ($p = .7$, $p < .001$) (Figure 1b).

To explore the eccentricity × modality interaction, follow-up, one-way univariate ANOVAs were performed, separately at each age and each eccentricity, with modality (A, V, and AV) as the factor in each of these analyses. These analyses were then followed up with post-hoc multiple comparison tests (summarized below). Figure 2 (a, b) shows the results for 25° and 45° eccentricities, respectively, for responses to auditory, visual, and audio-

visual stimuli. As can be seen, the results indicate a general trend toward decreasing response latency with age for all modalities at both eccentricities. The only exception to this general trend is the response latency to auditory targets at 25° (see below).

Unimodal responses

At both eccentricities, response latencies to visual targets decreased linearly with age for infants less than 8 months of age and leveled off between 8 and 10 months of age. The response pattern to auditory targets was somewhat different. Similar to the pattern observed in response to visual targets, the response latencies to auditory targets at 45° decreased as a function of age. In contrast, the response pattern to auditory targets at 25° of eccentricity decreased more slowly between 0–2 and 6–8 months of age and then increased between 6–8 and 8–10 months of age to a level seen in 4–6-month-old infants. This was confirmed with post-hoc multiple comparison tests performed separately between age groups

for each modality/eccentricity condition. Results of these tests yielded significant differences for all comparisons except those for the auditory targets presented at 25°. For the latter, there was no difference in response latency across the age groups from 0–2 to 6–8 months of age. All unimodal response latencies in infants were significantly slower than those observed in adults.

Bimodal responses

Mean response latency to audiovisual stimuli decreased steadily over the entire age range tested, though the most dramatic decrease in response latency occurred over the first 6 months before leveling off between 6 and 10 months. Similar to the visual-only conditions, audiovisual response latencies decreased at a comparable rate as a function of age at both eccentricities. As was the case for the unimodal conditions, response latency was significantly slower in infants than in adults, regardless of condition.

Comparison of visual-only and auditory-only responses

Comparisons between visual-only and auditory-only responses across age groups for 25° and for 45° (Figure 2) showed no difference between the two unimodal conditions in infants under 4 months of age for either eccentricity. In infants over 4 months, there was no difference at 45°, but the latencies diverged significantly at 25°, with shorter latencies found in the visual-only condition. No differences were found between visual and auditory response latencies at either eccentricity for adult subjects.

Comparison of unimodal and bimodal responses

As can be seen in Figure 2, response latencies to audiovisual targets were, in general, significantly faster than to auditory-only and visual-only targets in adults. In infants, on the other hand, it was not the case that response to audiovisual targets was always faster than to unimodal targets. Thus, infants younger than 8 months of age exhibited faster response to audiovisual targets than to auditory-only and visual-only targets, respectively at 2–4 months at 25° eccentricity, at 0–2 and 4–6 months at 45° eccentricity and at 8–10 months at both eccentricities. In contrast, we found no differences at 6–8 months of age.

Race Model

To verify that the faster response times to audiovisual targets in adults reflected the kind of nonlinear multisensory integration reported in previous studies (Hughes

et al., 1994; Miller, 1982; Molholm, Ritter, Murray, Javitt, Schroeder & Foxe, 2002), we analyzed the data to determine whether they violated the Race Model inequality: $P(RT_{AV}) \leq P(RT_A) + P(RT_V)$. Specifically, we compared the cumulative distribution for audiovisual (RT_{AV}) reaction times with the sum of the visual (RT_V) and auditory (RT_A) cumulative distributions. The statistical significance of the violation of the Race Model inequality (CDF of RT_{AV} greater than CDF of $RT_V + RT_A$) was tested using a one-sided Kolmogorov-Smirnov goodness-of-fit test. Adults showed a violation of the Race Model Inequality for all combined eccentricities (Figure 3d, $p < .0001$, $D = 0.34$) and for 25° ($p < .0001$, $D = 0.35$) and 45° ($p < .0001$, $D = 0.32$) separately, for reaction times less than 200 ms (approximately 75% of trials). A test of those unimodal versus bimodal differences found above to be significantly different in infants (i.e. at 0–2 and 4–6 months at 45°, 2–4 months at 25°, and 8–10 months at both eccentricities) showed that the Race Model was violated at 0–2 months at 45° (Figure 3a, $p < .001$, $D = 0.43$) and at 8–10 months of age at 25° (Figure 3b, $p < .001$, $D = 0.43$). For the 8–10-month-olds, the violation at 45° was borderline (Figure 3c, $p = .013$, $D = 0.20$). Additionally, it should be noted that a closer examination of the data from individual subjects in the 0–2-month group for 45° indicated that although the violation of the Race Model found is seemingly substantial, it is mainly due to a disparity between bimodal and unimodal responses for a subset of the infants and is not representative for the group as a whole. Audiovisual response latencies of 400 ms and less were only found in 6 out 10 subjects in the 0–2-month group, but with all 10 subjects contributing to the CDF by 500 ms. By comparison, the shape of the unimodal distributions (and hence the shape of the Race Model) at 400 ms was determined by the responses of only four out of the 10 subjects, with not all subjects being represented in the curve until nearly 1400 ms. This suggests that the shapes of the cumulative distributions are strongly influenced by individual subject differences and the reliability of this violation should be questioned. The Race Model was not reliably violated at any other age less than 8 months, indicating that non-linear multisensory integration cannot be reliably found in infants under 8 months, but is present – at least within a limited spatial range – between 8 and 10 months of age.

Discussion

It has been known for some time that adults take advantage of multisensory redundancy in spatial localization tasks, as evidenced by shorter head and/or eye movement
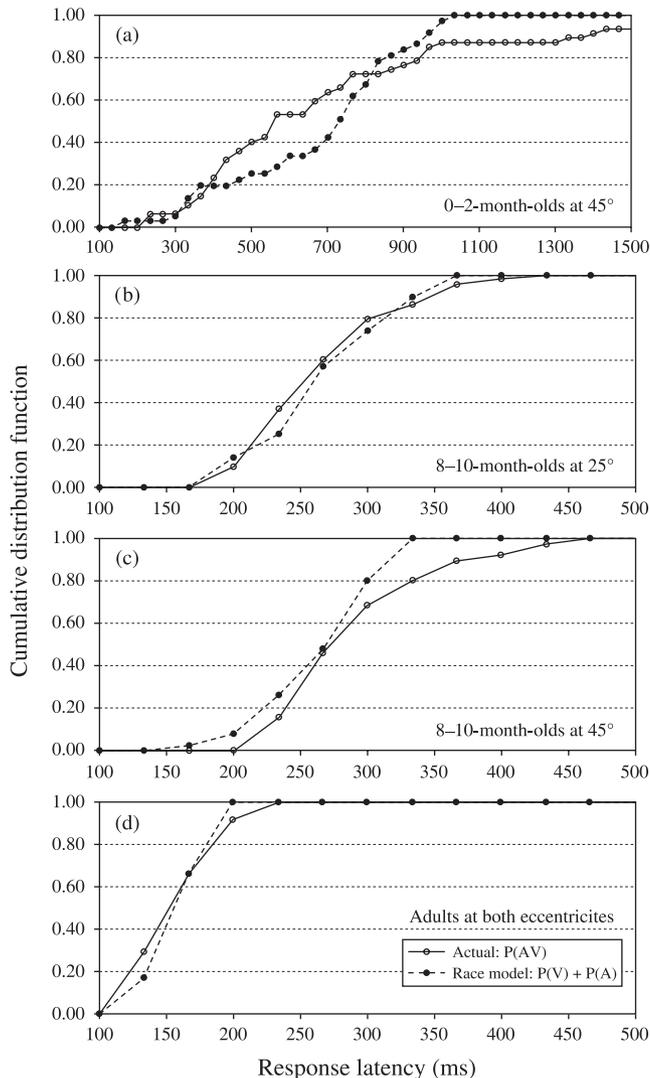
Figure 3 *Cumulative distribution of response latencies to audiovisual targets compared to cumulative distribution of response latencies to both auditory and visual targets at (a) 0–2 months at 45°, (b) 8–10 months at 25°, (c) 8–10 months at 45°, and (d) adults at both eccentricities. Dotted line shows upper boundary for Race Model.*

latencies to bimodally specified co-localized targets than to unimodally specified ones. What has not been known until now, however, is when in development multisensory facilitation of localization behavior emerges. The current study is the first to examine the development of spatial localization behavior in infancy across a broad age range, to compare localization of auditory, visual, and audiovisual targets and to compare infants' response to such targets with those of adults. Overall, the results of the current study indicate that response latencies to all types of targets generally decrease over

the first year of life. In addition, the results show that response latencies are generally fastest to audiovisual targets, somewhat slower to visual targets, and slowest to auditory targets. In contrast to adults, who exhibited nonlinear multisensory facilitation of localization behavior, only older infants exhibited some evidence of such facilitation.

This study revealed an interesting spatial dependence in the development of visual versus auditory localization. This might not be unexpected because the computational demands placed on vision and audition in spatial localization tasks are different. Whereas vision is specialized primarily for spatial perception, hearing is specialized for temporal perception. As a result, spatial location must be computed by the auditory system integrating multiple types of auditory cues (e.g. Hofman, Van Riswick & Van Opstal, 1998; King, Schnupp & Doubell, 2001). Our findings suggest that visual localization skills mature at a uniform rate and that they do so regardless of spatial eccentricity. These results are consistent with studies showing rapid development in the human visual system within the first 6 months of age. For example, stereopsis emerges between 10 and 20 weeks of age and stereoacuity improves immediately following the onset of stereo vision (Braddick, 1996; Held & Birch, 1980). Likewise, visual acuity develops and improves rapidly over the first 6 months of age (Dobson & Teller, 1978; Gwiazda, Bauer, Thorn & Held, 1986; Maurer & Lewis, 2001).

In marked contrast, the ability to localize auditory targets in different regions of auditory space appears to mature at different rates. Thus, at an eccentricity of 45°, response latencies to auditory targets decrease with age in a fashion similar to the decrease found in response to visual targets. In contrast, at 25° eccentricity with its obvious requirement of greater sensitivity and finer discrimination, response latencies decrease much more slowly with age and, in addition, there is a plateau between 4 and 10 months of age. In general, the developmental course of auditory sensory functions is less well known. Studies have shown that spatial localization abilities are rather course at birth (Morrongiello, 1988) and that binaural response capabilities are present – if still underdeveloped compared to adults – by 12 months of age (Schneider, Bull & Trehub, 1988). Animal studies in the wallaby have shown that brain regions mediating binaural processing (the superior olivary complex and inferior colliculus) exhibit adult-like responses only after postnatal day 160 (Liu, 2003). If the findings from animal studies such as these are projected to the human case, they suggest that adult-like auditory functionality in humans might not emerge until several years following birth. This is consistent with other such estimates (J.K. Moore, 2002).

Our finding that developmental changes in auditory response latencies differ at different eccentricities may reflect the fact that the developmental course of binaural versus monaural responses is not the same. In general, sound localization in humans is based on binaural (interaural time and intensity differences) and monaural (spectral) cues (Moore, 1991). Interaural differences are mainly used to localize a source's azimuth whereas spectral cues are used for determining elevation and front–back position and the neural mechanisms underlying responsiveness to these two cues appear to be independent (Hofman *et al.*, 1998). It seems reasonable to assume that as a sound source moves off-center there is a switch in the weighting assigned to binaural versus monaural cues. This is supported by psychophysical studies examining the accuracy of monaural and binaural sound localization in blind and sighted adults. When subjects have one ear artificially blocked (monaural condition), their ability to localize sound sources on the unblocked side is only slightly diminished for peripheral eccentricities between 40° and 80°, more degraded for smaller, pericentral eccentricities, and severely degraded on the blocked side (Lessard, Pare, Lepore & Lassonde, 1998). Surprisingly, in totally blind subjects, monaural sound localization to sources on either side (blocked and unblocked) is unchanged from their binaural localization ability, indicating a capability for more efficient use of monaural cues in spatial regions thought to require interaural differences to localize. While pointing to a large degree of plasticity in blind humans, these findings basically suggest that in normal adults, horizontal sound localization in the periphery is less reliant on binaural cues than more centrally located sources, or at least that monaural cues are sufficient for the task when binaural cues are not available. If young infants have a less mature binaural processing system and have to rely more on monaural cues to localize auditory targets, they would perform more poorly where binaural cues are more critical (i.e. areas close to the midline) and better where monaural cues are sufficient. This is consistent with our findings in that at 25° eccentricity there was no effective improvement in response latency to auditory targets up through 10 months but at 45° eccentricity responsiveness improved over age. This may reflect differences in the maturity of neural systems processing binaural versus monaural cues, with the development of monaural regions of auditory perception maturing before binaural regions.

A second area of interest in the response latencies of infants toward auditory targets at 25° was the slower grand mean latency in 8–10-month-olds compared with 6–8-month-olds. A closer examination of these responses in 8–10-month-olds indicated that the number of valid trials dropped dramatically (approximately 39% of presented trials) compared to the number of valid trials for the other five target conditions (58%–68% valid). Also, there was an interesting bimodal distribution of reaction times that was not seen for any other modality/eccentricity conditions in that age group nor for any conditions in younger infants. The distribution of response latencies for each of the five infant groups for all six target condition types (Figure 4a) shows how the distributions change from a broad distribution at the youngest age group to a single, narrow, positively skewed peak for the oldest infants, with the distinctive exception of 8–10-month-olds for auditory-only targets at 25° (Figure 4b). Unlike the other conditions, where the mean response latencies generally align well with the distribution peaks, this secondary peak pulls the mean latency value off to a point in between. The first peak (representing 72% of the total valid trials for this condition) is centered around 350 ms and is in line with a slow but steady improvement in response latencies with age. The second peak (28% of total valid trials) is centered around 1250 ms with a span of several hundred milliseconds (700 to 1000 ms) between the two peaks where no reaction times were found. This indicates a more complicated development profile in the auditory domain, perhaps with factors such as attention and stimulus saliency involved, particularly when orienting toward targets more centrally located.

It should be noted that the stimulus intensity levels used for all subjects were selected from matching studies performed with adults. As a result, the stimuli presented to the infants were not tailored for the fastest response at each age group. Considering the fact that the auditory and visual systems change in a major way during early development, there is no reason to expect that intensity matches that are appropriate for adults are appropriate for infants. On the other hand, considering the very limited time with each infant subject and their short attention span, as well as the huge individual differences between infants even at the same age, we were forced to stick to our simpler design. This underlies the general difficulty involved in performing infant–adult comparisons. Given that there are no clear and direct ways to choose the 'correct' stimulus intensity values, the best approach is a simple, fixed set of parameters held constant across the age range tested. Thus, it is conceivable that the slower auditory response found across all infants at 25° might have been improved by a more intense or salient auditory stimulus. Evidence to the contrary, however, is the fact that the auditory response latency at 45° was no different from the corresponding visual response latency at any age and that both latencies decreased steadily with age. Issues of stimulus intensity notwithstanding,
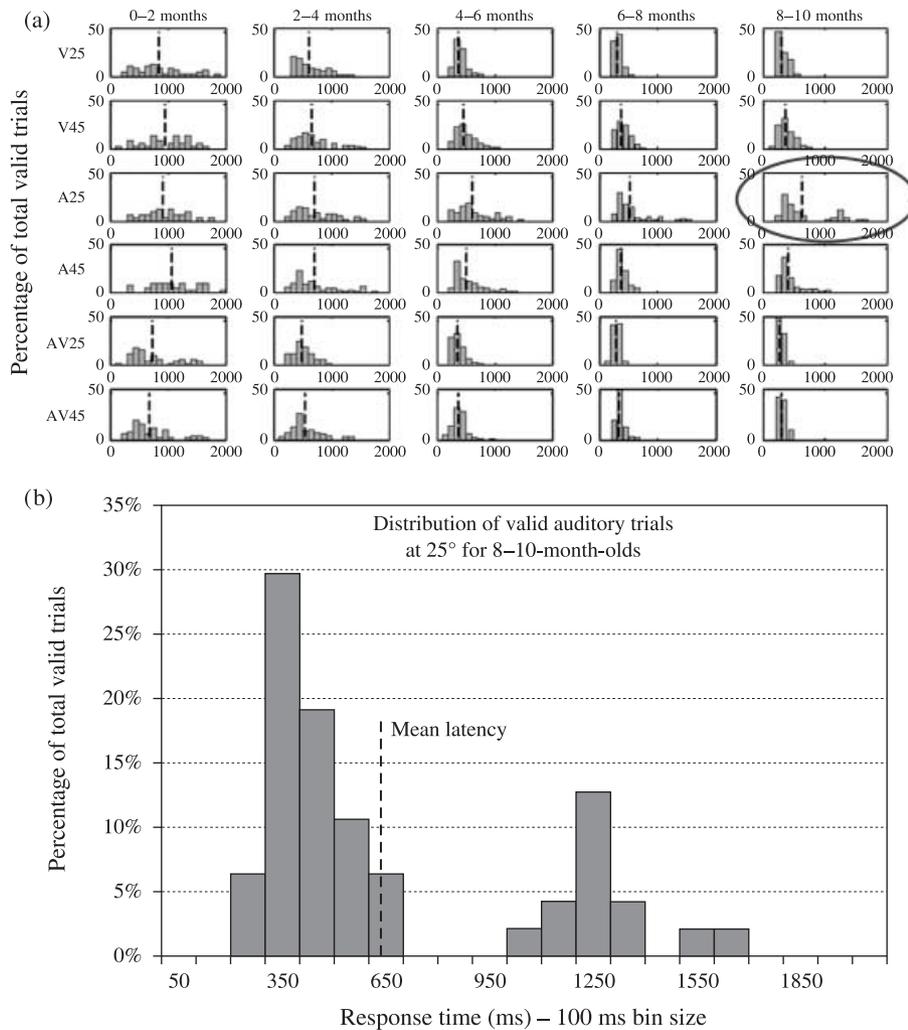
Figure 4   *(a) Histograms for all five infant age groups (columns 1–5) at each of the six target conditions (rows 1–6); dotted line represents mean response latency. (b) Auditory trials at 25° for 8–10-month-olds showed a distinct bimodal distribution only hinted at in the 6–8 month group.*

the conclusion that the slower response to more centrally located auditory targets may reflect immature binaural processing certainly calls for further investigations of monaural and binaural response capabilities in infancy.

The present findings permit the first opportunity to determine whether infants exhibit adult-like nonlinear multisensory facilitation. The findings showed that even though the younger infants (less than 6 months of age) had faster response latencies toward synchronous, co-located audiovisual stimuli under certain situations (25° for 2–4-month-olds, 45° for 0–2- and 4–6-month-olds), these responses cannot be reliably distinguished from a faster response time due to probability summation of independent sensory systems. In addition, response to audiovisual targets was not faster than to visual targets

or to auditory targets at 45° at 6–8 months. Together, these findings suggest that adult-like multisensory facilitation of localization is not present at birth and is consistent with developmental electrophysiological data from newborn cats and monkeys (Wallace & Stein, 1997, 2001). It should be remembered, however, that the data were very noisy in the youngest infants (0–2 months) and that there were large individual response differences. As a result, the conclusion that young infants do not exhibit true multisensory integration should be treated as a tentative one until additional studies are conducted. At the same time, however, it should also be noted that the preponderance of the data from the current study suggests that true integration does not occur until later in infancy. First, there was a steady decrease in response

latency to audiovisual targets and it was only at 8–10 months of age that this response was sufficiently faster than the unimodal responses to result in a violation of the Race Model. This suggests that until 8 months of age responsiveness to audiovisual stimuli reflects the faster of the two unimodal responses. At the same time, however, in 8–10-month-old infants, a violation of the Race Model was found only at 25°, which is different from what is observed in adults. This finding, coupled with the poor response to auditory targets at 25°, suggests that multisensory integration of auditory and visual localization signals is still immature by 10 months of age. This conclusion is consistent with anatomical findings in neonatal monkeys showing that the multisensory neurons that are already present and active in the superior colliculus at birth do not show the multisensory enhancement found in comparable sites in adults (Wallace & Stein, 2001). Furthermore, the finding that responses to auditory targets at 25° are significantly slower than responses to visual or audiovisual targets suggests that multisensory pathways activated during development may facilitate the development of auditory localization for certain eccentricities. This idea receives support from findings that the auditory system of juvenile barn owls is 'tutored' by visual experience (Knudsen, 2002; Knudsen & Knudsen, 1989). Likewise, partially-blind human subjects show very poor auditory localization compared with early-blind and sighted subjects (Zwiers, Van Opstal & Cruysberg, 2001).

In conclusion, the current results indicate that integration of multisensory localization signals emerges slowly over the first year of life and that it has not reached its mature, adult-like level even as late as 8–10 months of age. This developmental scenario may be due to one of two developmental processes. On the one hand, it may be that the development of bimodal responsiveness depends on the prior development of unimodal responsiveness. On the other hand, it may be that the development of unimodal and multisensory responsiveness proceed in parallel. The current findings also raise some interesting questions for future investigations. For example, at what age do adult-like multisensory integration abilities first emerge? What neural mechanisms underlie the developmental differences observed here? Answers to these questions will go a long way in helping better define how in early development we come to construct a unified and coherent representation of our multimodal world.

Acknowledgements

References

Bahrick, L.E. (1988). Intermodal learning in infancy – learning on the basis of 2 kinds of invariant relations in audible and visible events. *Child Development*, **59** (1), 197–209.

Braddick, O. (1996). Binocularity in infancy. *Eye*, **10**, 182–188.

Dobson, V., & Teller, D.Y. (1978). Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vision Research*, **18**, 1469–1483.

Dodd, B. (1979). Lip reading in infants – attention to speech presented in-synchrony and out-of-synchrony. *Cognitive Psychology*, **11** (4), 478–484.

Fantz, R.L. (1963). Pattern vision in newborn infants. *Science*, **140** (356), 296–297.

Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage Publications.

Gibson, E.J. (1982). The concept of affordances in development: the renascence of functionalism. In W.A. Collins (Ed.), *The concept of development: The Minnesota symposia on child psychology*. Hillsdale, NJ: L. Erlbaum.

Gibson, J. (1979). *The ecological approach to visual perception* (vol. 15, pp. 55–82). Boston, MA: Houghton Mifflin Company.

Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1986). Meridional amblyopia does result from astigmatism in early-childhood. *Clinical Vision Sciences*, **1** (2), 145–152.

Harman, C., Posner, M.I., Rothbart, M.K., & Thomasthrapp, L. (1994). Development of orienting to locations and objects in human infants. *Canadian Journal of Experimental Psychology–Revue Canadienne De Psychologie Experimentale*, **48** (2), 301–318.

Held, R., & Birch, E.E. (1980). Stereoacuity of human infants. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 5572–5574.

Hofman, P.M., Van Riswick, J.G.A., & Van Opstal, A.J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, **1** (5), 417–421.

Hughes, H.C., Reuter-Lorenz, P.A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing – saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, **20** (1), 131–153.

Jiang, W., Wallace, M.T., Jiang, H., Vaughan, J.W., & Stein, B.E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology*, **85** (2), 506–522.

King, A.J., Schnupp, J.W.H., & Doubell, T.P. (2001). The shape of ears to come: dynamic coding of auditory space. *Trends in Cognitive Sciences*, **5** (6), 261–270.

Knudsen, E.I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature*, **417** (6886), 322–328.

Knudsen, E.I., & Knudsen, P.F. (1989). Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience*, **9** (9), 3306–3313.

Lawson, K.R. (1980). Spatial and temporal congruity and auditory-visual integration in infants. *Developmental Psychology*, **21**, 185–192.

Lessard, N., Pare, M., Lepore, F., & Lassonde, W. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, **395** (6699), 278–280.

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

Lewkowicz, D.J. (1980). A method for adapting headphones for use with infants. *Infant Behavior and Development*, **3** (3), 283–284.

Lewkowicz, D.J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development*, **9** (3), 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 and Development*, **15** (3), 297–324.

Lewkowicz, D.J. (2000a). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological Bulletin*, **126** (2), 281–308.

Lewkowicz, D.J. (2000b). Perceptual development: visual, auditory, and speech perception in infancy. *American Journal of Psychology*, **113** (3), 488–500.

Lewkowicz, D.J. (2002). Heterogeneity and heterochrony in the development of intersensory perception. *Cognitive Brain Research*, **14** (1), 41–63.

Liu, G.B. (2003). Functional development of the auditory brainstem in the tammar wallaby (*Macropus eugenii*): the superior olivary complex and its relationship with the auditory brainstem response (ABR). *Hearing Research*, **175** (1–2), 152–164.

Lyons-Ruth, K. (1977). Bimodal perception in infancy: response to auditory-visual incongruity. *Child Development*, **48**, 820–827.

Maurer, D., & Lewis, T.L. (2001). Visual acuity: the role of visual input in inducing postnatal change. *Clinical Neuroscience Research*, **1** (4), 239–247.

Meredith, M.A., & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, **221** (4608), 389–391.

Meredith, M.A., & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, **56** (3), 640–662.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, **14** (2), 247–279.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, **14** (1), 115–128.

Moore, D.R. (1991). Anatomy and physiology of binaural hearing. *Audiology*, **30** (3), 125–134.

Moore, J.K. (2002). Maturation of human auditory cortex: implications for speech perception. *Annals of Otology, Rhinology and Laryngology*, **111** (5), 7–10.

Morrongiello, B.A. (1988). Infants' localization of sounds along the horizontal axis – estimates of minimum audible angle. *Developmental Psychology*, **24** (1), 8–13.

Muir, D.W., Clifton, R.K., & Clarkson, M.G. (1989). The development of a human auditory localization response – a u-shaped function. *Canadian Journal of Psychology–Revue Canadienne De Psychologie*, **43** (2), 199–216.

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

Scheier, C., Lewkowicz, D.J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, **6** (3), 233–241.

Schneider, B.A., Bull, D., & Trehub, S.E. (1988). Binaural unmasking in infants. *The Journal of the Acoustical Society of America*, **83** (3), 1124–1132.

Spelke, E.S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, **15** (6), 626–636.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology*, **69** (6), 1797–1809.

Wallace, M.T., & Stein, B.E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience*, **17** (7), 2429–2444.

Wallace, M.T., & Stein, B.E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *Journal of Neurophysiology*, **83** (6), 3578–3582.

Wallace, M.T., & Stein, B.E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, **21** (22), 8886–8894.

Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, **76** (2), 1246–1266.

Wertheimer, M. (1961). Psychomotor coordination of auditory and visual space at birth. *Science*, **134** (349), 1692.

Zwiers, M.P., Van Opstal, A.J., & Cruysberg, J.R.M. (2001). A spatial hearing deficit in early-blind humans. *Journal of Neuroscience*, **21** (9), art. no.-RC142.