Chapter 2

Study 1: Spatially and Temporally Congruent Audiovisual Stimuli

2.1 Overview

Previous studies have found that adults respond faster and more reliably to bimodal compared to unimodal localization cues. The current study (Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, in press) investigated for the first time the development of audioviusal integration in spatial localization behavior in infants between one and ten months of age. We observed infants' head and eye movements in response to auditory, visual, or both kinds of stimuli presented at either 25° or 45° to the right or left of midline. Infants under eight 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 eight and ten 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 is 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.

2.2 Experimental Design and Methods

2.2.1 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 2.1). In addition, seven adults (4 female, 3 male) participated in the study. Parents were encouraged to bring their infant back until he or she reached ten months of age. As a result, some infants are represented at more than one age. A minimum of one month passed between visits for those infants who were tested at more than one age, with repeat participation as follows: one age group (20 infants), two age groups (4 infants), three age groups (5 infants), and four age groups (4 infants).

2.2.2 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” x

Table 2.1: Total number of subjects used in each age group, their mean age, and the total percentage of good 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%

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 spatially and temporally congruent 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 congruent 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.

2.2.3 Procedure

Infants were seated 22.5” from the hoop in a car seat, or on the parents 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 childs body but not to move or cue them in any way during the experiment. A piece of adhesive tape (one quarter inch wide) was affixed to the subjects 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 subject 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 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 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 x 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 MediaAnalyzer was used to mark manually 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 it's 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 was 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

Table 2.2: The group mean was calculated for each modality/eccentricity condition for each age.

	Visual (Experiment 1)				Auditory (Experiment 1)				Audiovisual (Experiment 2)			
Age (mo)	25°		45°		25°		45°		25°		45°	
	Mean (ms)	s.e.	Mean (ms)	s.e.	Mean (ms)	s.e.	Mean (ms)	s.e.	Mean (ms)	s.e.	Mean (ms)	s.e.
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

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 also removed. All good trials were pooled within each age group to calculate the grand mean (Table 2.2).

2.3 Results

A 3-way univariate ANOVA (3 x 2 x 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 0.01 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, $F_{(2,2010)} = 124.3, p < 0.001$) and age, $F_{(4,2010)} = 242.9, p < 0.001$), as well as significant modality x eccentricity, $F_{(2,2010)} = 14.1, p < 0.001$), modality x age, $F_{(4,2010)} = 3.4, p = 0.001$), and modality x eccentricity x age, $F_{(8,2010)} = 3.8, p < 0.001$) interactions. Finally, we found a marginally significant eccentricity x age interaction,

$F_{(4,2010)} = 3.2, p = 0.012$).

2.3.1 Main Effect of Age

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 the 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 ($p < 0.001$), except for that between the two oldest age groups (6–8 versus 8–10 month olds, $p = 0.32$). Although there was a steady decrease in the response latency with age (Figure 2.1a), the response latency of even the oldest infants was still longer than that of adults ($p < 0.001$).

2.3.2 Main Effect of Modality

Because our sample sizes were unequal and had non-homogenous variances, we examined the main effect of modality by using multiple *post-hoc* comparisons using the Games-Howell procedure (A. Field, 2000). Response latency to auditory targets was slower than to both visual and audiovisual targets ($p < 0.001$), and response latency to visual targets was slower than to audiovisual targets ($p = 0.001$). This response pattern was different from the adult response pattern where response latencies to the auditory and visual targets did not differ and where response latencies in both unimodal conditions were slower than in the audiovisual condition ($p = 0.7, p < 0.001$) (Figure 2.1b).

2.3.3 Eccentricity and Modality Interactions

To explore the eccentricity x modality interaction, follow-up 1-way univariate ANOVAs were performed, separately at each age and each eccentricity, with modality (A, V, and AV) as the factor in each. These analyses were then followed up with *post-hoc* multiple comparison tests (summarized below). Figures 2.2a, b show the results for

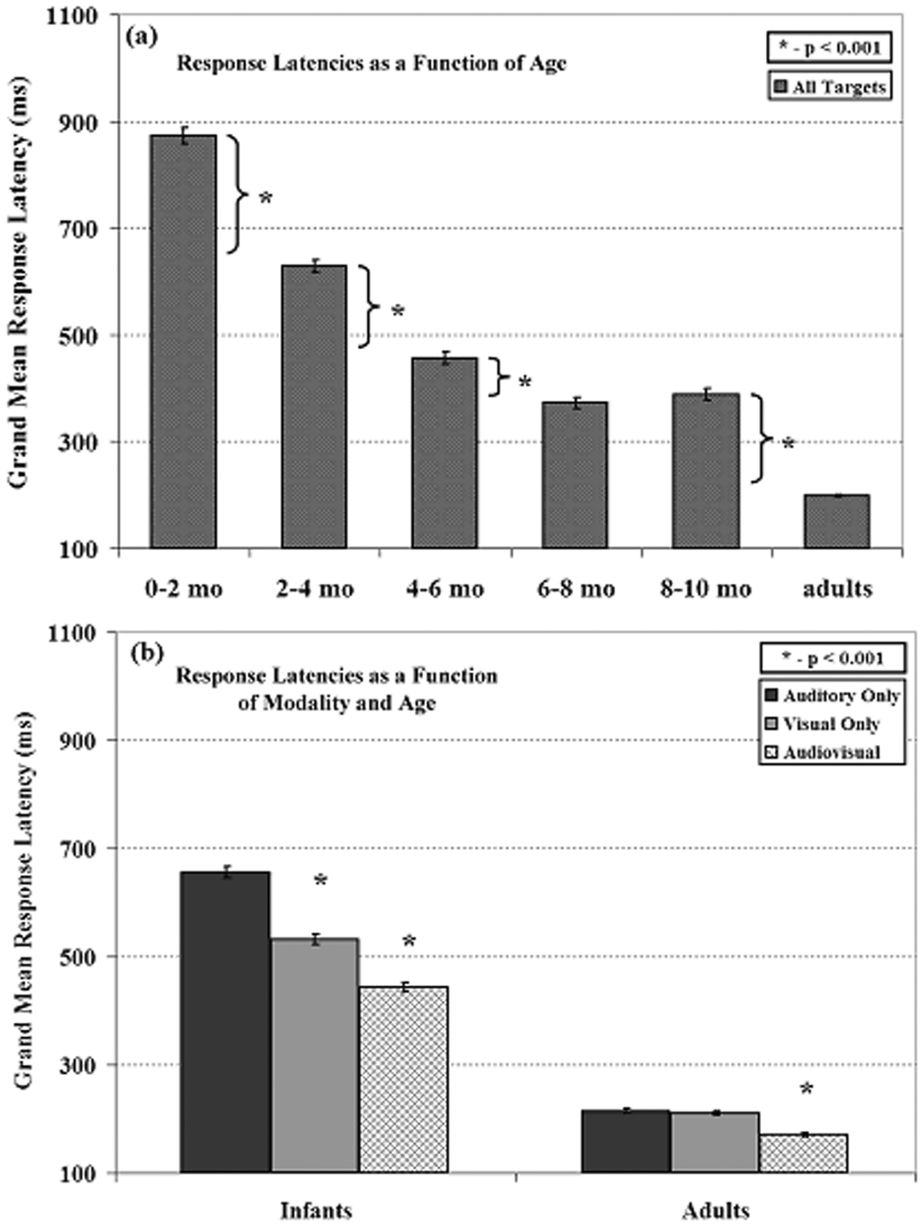

Figure 2.1: Main effects of (a) age and (b) modality on response latency. Error bars represent the standard error of the mean.

25° and 45° eccentricities, respectively, for responses to auditory, visual, and audiovisual 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).

2.3.4 Unimodal Responses

At both eccentricities, response latencies to visual targets decreased linearly with age for infants less than eight months and leveled off between eight and ten months. 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° decreased more slowly between two and eight months but then increased for the oldest infants (8–10 months) to a level last seen in 4–6 month olds. 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 auditory targets at this eccentricity, there was no difference in response latency for ages under eight months of age. All unimodal response latencies in infants were significantly slower than those observed in adults.

2.3.5 Bimodal Responses

Mean response latencies to audiovisual stimuli decreased steadily over the entire age range tested, though the most dramatic decrease in response latency occurred over the first six months before leveling off between six and ten 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.

2.3.6 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 2.2) showed no difference between the two unimodal conditions in infants under four months of age for either eccentricity. In infants over four months, there was no difference at 45°, but the latencies diverged significantly at 25°, as the visual latencies quickened with age but auditory latencies did not. No differences were found between visual and auditory response latencies at either eccentricity for adult subjects.

2.3.7 Comparison of Unimodal and Bimodal Responses

As can be seen in Figure 2.2, 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. Infants younger than eight months exhibited a limited, faster response to audiovisual targets than to both unimodal targets at 2–4 months (25°), 0–2 and 4–6 months (45°), and 8–10 months (both). In contrast, we found no differences at 6–8 months of age.

2.3.8 Race Model

A standard measurement for the non-linear facilitation of response latency provided by a spatially and temporally congruent audiovisual target in adults is looking for a violation the Race Model, in order to rule out statistical facilitation as an explanation for the improved response—the upper boundary of the Race Model defined by the summed cumulative distribution functions of the unimodal response latencies. If the cumulative distribution of the measured bimodal response latencies falls below this boundary, the enhanced response could be explained as just statistical facilitation. Adult psychophysical studies routinely find violations of the Race Model when presented with a congruent audiovisual target. To verify that the faster response times to audiovisual targets of certain infant groups reflected the kind of non-linear mul-

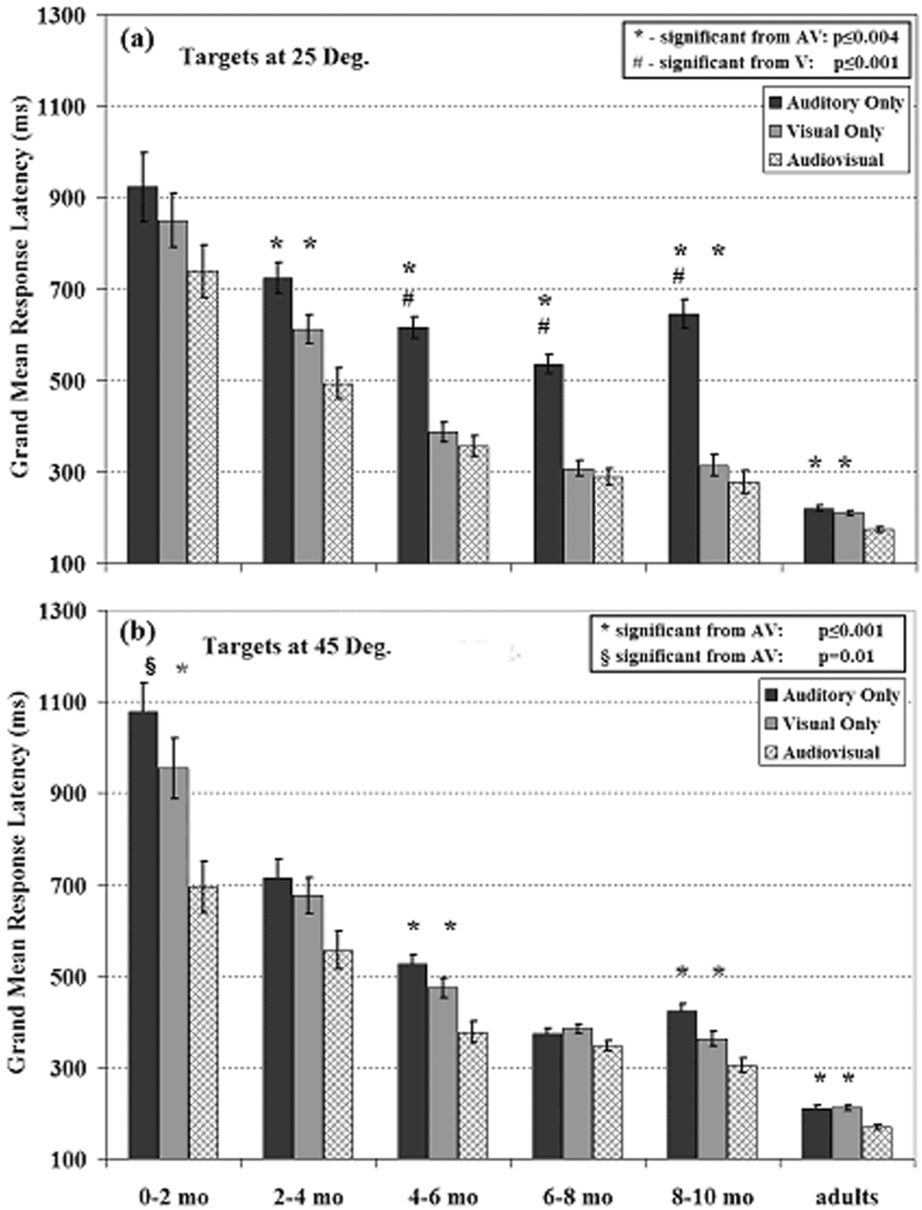

Figure 2.2: Response latencies for auditory, visual, and audiovisual targets at (a) 25° and (b) 45° eccentricities across all ages. Error bars represent the standard error of the mean.

tisensory integration reported in previous adult studies (Hughes et al., 1994; Miller, 1982; Molholm et al., 2002), we looked for violations in the Race Model Inequality:

$$P(RT_{AV}) \leq P(RT_A) + P(RT_V)$$

where $P(RT_i)$ represents the cumulative distributions of the measured response latencies for audiovisual, auditory-only, and visual-only targets, respectively. The statistical significance of any violation of the Race Model inequality, $P(RT_{AV} > P(RT_V) + P(RT_A)$, 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 2.3.8d, $p < 0.0001, D = 0.34$) and for 25°($p < 0.0001, D = 0.35$) and 45°($p < 0.0001, D = 0.32$) separately, for reaction times less than 200 ms (approximately 75% of trials). A test of those conditions in infants where the bimodal response significantly difference from both unimodal responses (i.e., 0–2 and 4–6 months (45°), 2–4 months (25°), and 8–10 months (both)) showed that the Race Model was violated for 0–2 month olds at 45° (Figure 2.3.8a, $p < 0.001, D = 0.43$) and 8–10 month olds at 25° (Figure 2.3.8b, $p < 0.001, D = 0.43$), with a borderline difference for 8–10 month olds at 45° (Figure 2.3.8c, $p = 0.013, D = 0.20$). 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 of 10 subjects in this group, but with all 10 subjects contributing to the CDF at $\geq$ 500 ms. By comparison, the shape of the unimodal distributions (and hence the shape of the Race Model boundary) at 400 ms was determined by the responses of only four out of the ten 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 eight months, indicating that non-linear multisensory integration can not be reliably found in infants under eight months, but is present—at least within a limited spatial range—between eight and ten months of age.

2.4 Discussion

This study revealed an interesting spatial dependence in the development of visual versus auditory localization. This is not unexpected given 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., (King, Schnupp, & Doubell, 2001; Hofman, Van Riswick, & Van Opstal, 1998)). 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 six months of age. For example, stereopsis emerges between ten and 20 weeks of age and stereoacuity improves immediately following the onset of stereo vision (Held, Birch, & Gwiazda, 1980; Braddick, 1996). Likewise, visual acuity develops and improves rapidly over the first six months of age (Maurer & Lewis, 2001; Dobson & Teller, 1978; Gwiazda, Bauer, Thorn, & Held, 1986).

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°—with it's obvious requirement of greater sensitivity and finer discrimination—response latencies decreased much more slowly with age and remained unchanged between four and ten 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 coarse at birth (Morrongiello, 1988b) and that binaural response capabil-

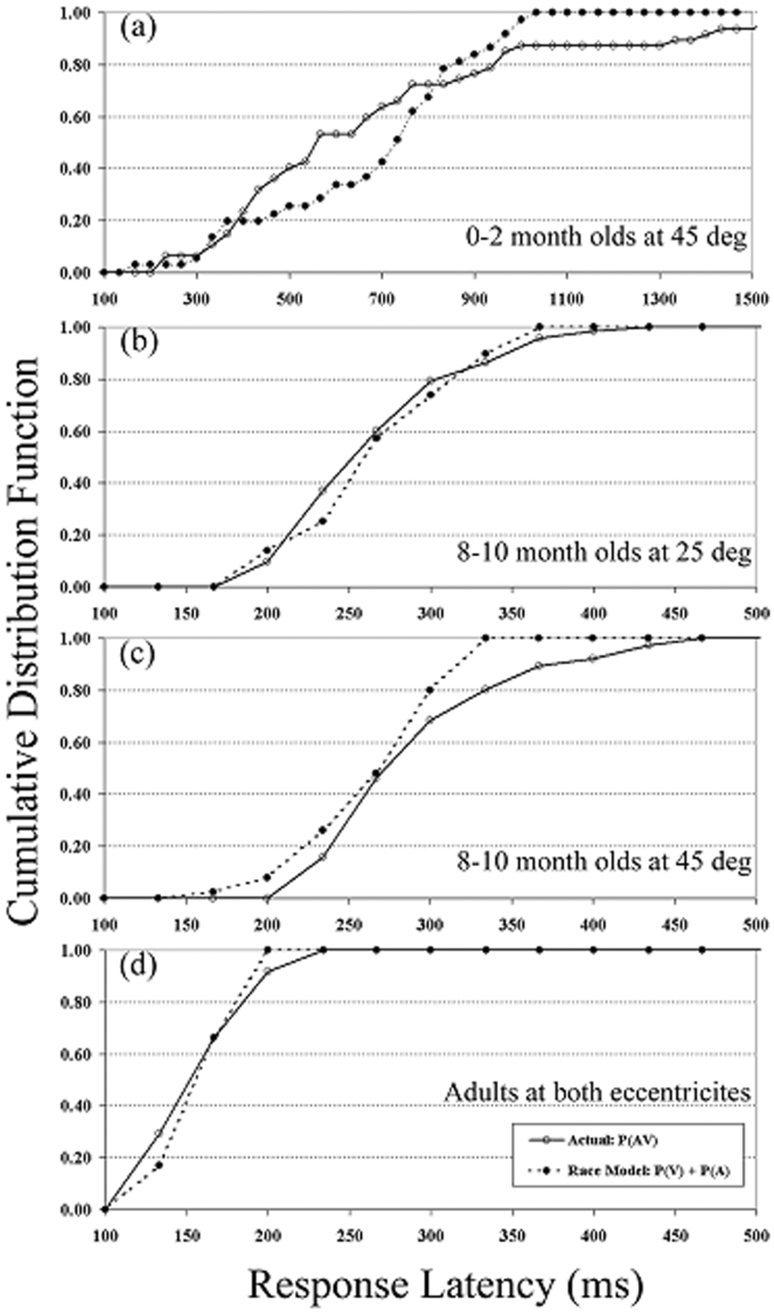

Figure 2.3: Cumulative distribution of response latencies to audiovisual targets compared to cumulative distribution of response latencies to both auditory and visual targets for (a) 0–2 months at 45°, (b) 8–10 months at 25° and (c) 45°, and (d) adults at both eccentricities. Dotted line shows upper boundary for Race Model.

ities 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 (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 sources 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 far 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 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 little effective improvement in response latency to auditory targets up through ten months but at 45° eccentricity responsiveness improved with 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. One motivation for multisensory interaction in this instance is that the poor auditory spatial localization network may require spatial feedback from the more capable visual system. 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).

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 good trials dropped dramatically (approximately 39% of presented trials) compared to the number of good trials for the other five target conditions (58% to 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 2.4a) 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 2.4b). 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 due to their short attention spans, 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, 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.

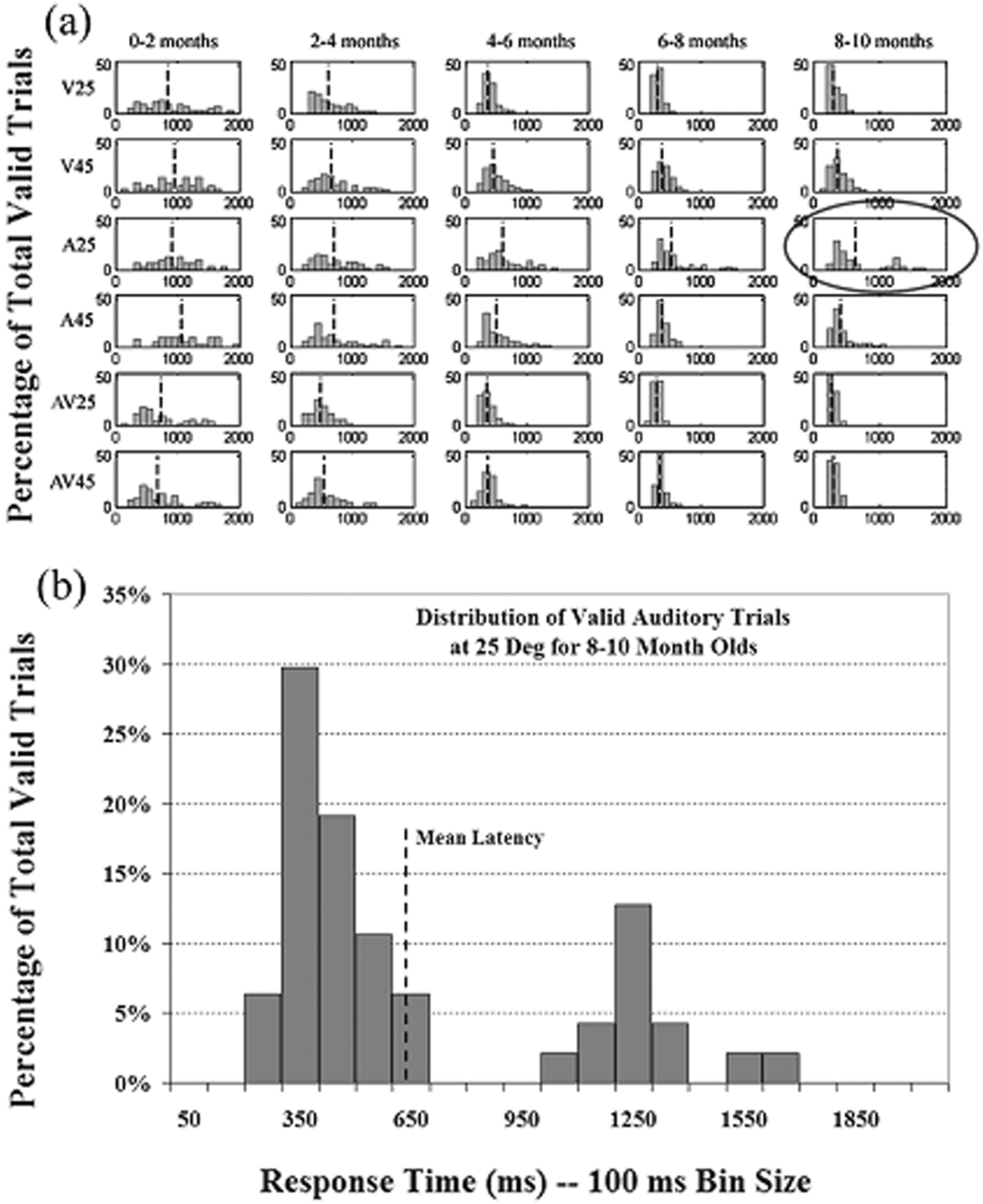

Figure 2.4: (a) Histograms for all five infant age groups (columns 1–5) at each of the six target conditions (rows 1 - 6); dotted line represents mean response latency. (b) Auditory trials at 25° for 8–10 month olds showed a distinct bimodal distribution only hinted at in the 6–8 month group.

The present findings permit the first opportunity to determine whether infants exhibit adult-like non-linear multisensory facilitation. The findings showed that even though the younger infants (less than six months of age) had significantly 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 months old), these responses cannot be reliably distinguished from a faster response time due to probability summation of independent sensory systems. In addition, responses to audiovisual targets were comparable to unimodal targets in 6–8 month olds. 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 eight months of age responsiveness to audiovisual stimuli reflects the faster of the two unimodal responses, most often vision. At the same time, however, in 8–10 month old infants, a violation of the Race Model was only found 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 ten months of age. Together, these findings are consistent with anatomical findings in neonatal monkeys. Multisensory neurons, while present and active in the superior colliculus, are immature at birth and adult-like multisensory facilitation is not yet possible (Wallace & Stein, 1997, 2001).