NFANT BEHAVIOR AND DEVELOPMENT 2, 13–27 (1979)

Intermodal Exploration and Knowledge in Infancy*

ELIZABETH S. SPELKE
University of Pennsylvania
CYNTHIA J. OWSLEY
Cornell University

By the age of 15 weeks, human infants search visually for a parent who is heard to speak, even when no spatial or temporal information unites the face and voice. They can apparently use knowledge about a parent's audible and visible characteristics to guide intermodal exploration. Perceivers evidently learn at an early age about some of the optic and acoustic information that specifies certain objects.

Adults can experience objects with manifold properties from optic, acoustic, haptic, or chemical information alone. They can perceive rigid, solid objects by looking (Gibson, 1969), and objects of definite size, color, and texture by listening (James, 1890). *Intermodal knowledge*—knowledge of the visible, audible, tangible, and olfactory properties of objects—underlies these common perceptual experiences. Adults typically know a good deal about the visible appearance of events they hear, the sound and scent of objects they see, and the tactual impressions that seen or heard objects create.

Intermodal knowledge is related to *intermodal exploration:* the search for information through multiple perceptual systems. Mature perceivers are most apt to acquire intermodal knowledge of an object if they seek and attend to information in several stimulus modalities at once. The knowledge they acquire, in turn,

*The research was supported by a Training Grant from the National Institute of Child Health and Human Development (5TO1 HD00381-05) and by a grant to Eleanor J. Gibson from the National Science Foundation (BNS76-14942). A version of this paper was presented at the meetings of the Society for Research in Child Development, New Orleans, March, 1977. Requests for reprints should be sent to Elizabeth Spelke at the University of Pennsylvania, Department of Psychology, 3813-15 Walnut Street T3, Philadelphia, Pa. 19104.

makes possible further investigations of the object. Perceivers who detect an object through one perceptual system become able to anticipate information that will specify the object to other systems, and so they can seek that information directedly. Their searching may yield still further knowledge, and thus intermodal exploration and knowledge will develop in a cycle of perceptual activity (cf. Neisser, 1976).

Debate over the nature and origin of intermodal knowledge has continued for centuries (e.g., Berkeley, 1709/1910; Mill, 1829; Titchener, 1902) and has, in recent years, stimulated research with human infants. The intermodal knowledge of infants has been inferred either from observations of conflict behavior or from observations of exploratory activity. In a conflict experiment, stimulation to two perceptual systems is rearranged so as to give discrepant information about an event. If an infant shows signs of conflict when confronting a discrepant display, he or she is inferred to know something about the normal arrangement of bimodal information. Infants as young as 30 days have been reported to show conflict, indexed by "tonguing," when faced with an object that was not heard in the direction from which it was seen (Aronson & Rosenbloom, 1971). Babies as young as 6 days reacted with "frustration and tears" to an object that could not be felt in its visible location (Bower, Broughton, & Moore, 1970). These findings must be questioned, however, owing to failures of replication (Condry, Haltom, & Neisser, 1977; Field, 1977; McGurk & Lewis, 1974), and to the inherent ambiguity of these behavioral indexes of conflict.

In a study of exploratory behavior, information about an object is presented in one stimulus modality, and the infant's subsequent investigations are observed. Infants below four months reach, grasp, and engage in activities precursory of skilled manipulation when they see an object of graspable size (Bower, 1972; Bruner & Koslowski, 1972). Newborn infants also look in the direction of a sound (Mendelson & Haith, 1976; Wertheimer, 1961), and four-month-old infants look preferentially to objects that move in temporal synchrony with an auditory accompaniment (Spelke, 1978a). These observations indicate that young infants know something about the multimodally specified properties of objects, and use such knowledge when they explore.

Investigations of the exploratory activity of newborns do not support an associationist position. Perceivers need not always learn to relate sensory impressions arising from stimulation in each of the modalities. Empiricist arguments should not, however, be completely dismissed; most intermodal knowledge is surely acquired. A voice, for example, specifies a person with distinctive visible properties only to the perceiver who has become acquainted with that person. Research on infancy might felicitously begin to investigate the nature of the human capacity to acquire intermodal knowledge.

Only two experiments have probed the infant's capacity to learn about bimodally specified objects, and both have used conflict behavior to index intermodal knowledge. Lyons-Ruth (1977) familiarized infants with an inanimate

object undergoing movements in temporal synchrony with a ringing or squeaking sound. Sound and object were presented in the same spatial direction. During a test that immediately followed, the sound was played from a new direction, in which the original object or a different object was visible. Four-month-old infants were reported to show "gaze aversion" when they encountered an unfamiliar object in the direction of the familiar sound. Infants apparently expected to see, in the direction of the sound, the object that had been paired with the sound in the past. Cohen (1973) tested knowledge of an already familiar object, the mother. She presented infants at five and eight months with the tape-recorded voice of the mother or of an unfamiliar woman, as both adults sat in view of the infant. A voice was heard in the direction of the appropriate or inappropriate adult. Although there was no effect of an auditory–visual mismatch at five months, the eight-month-olds showed signs of conflict (again, gaze aversion) in the mismatch conditions. These experiments reveal that infants respond differentially to congruent and to discrepant optic and acoustic information. If infants' responses reflect conflict, then one may conclude that infants expect optic and acoustic information about an object to be available at the same time and place. It would be desirable to collect converging evidence, however, that gaze aversion is an index of conflict.

Our own investigations of acquired intermodal knowledge focused on the infant's exploratory activity. Rather than violating infants' expectations, we gave them an opportunity to use expectations to guide looking for objects that they heard. Infants were presented with two objects and a sound specifying one of them. Sounds and visible objects were not presented in the same spatial direction and were not temporally synchronized. We observed whether infants would search visually for the object that was acoustically specified, in a situation where only knowledge could guide their looking.

The research was undertaken to determine if infants can use knowledge about the audible and visible characteristics of objects to guide exploration. Since the parents served as stimulus objects, the experiment focused on intermodal knowledge of the mother and father. Experiment I examined auditory–visual search for the mother and the father by infants at 3½, 5½, and 7½ months of age. Experiment II examined four-month-old infants' search for the mother and father or for the mother and a second adult woman.

EXPERIMENT I

Infants participated in one experimental session consisting of two episodes. In both episodes, they sat with the two parents in view and heard each parent speaking in turn. We observed their visual search. In the first, tape-recorded voice episode, neither adult actually spoke. The parents remained motionless while their tape-recorded voices were played through speakers that centered the

sounds between them. Thus, no temporal or spatial relationship united the sight of a parent with the sound of his or her voice. Looking to the "speaking" parent could only have been guided by knowledge of the visible object which that voice specified. The parents actually spoke during the live voice episode that followed. Optic and acoustic information for the speech of each parent were spatially coincident and temporally synchronized. This condition was included to ensure that infants were capable of turning to look to a speaking parent when these added sources of information were available.

Method

Subjects. Twenty-seven infants participated in this experiment: nine at each of three age levels. The youngest group of infants ranged in age from 3 months, 8 days to 4 months, 7 days (mean age = 3 months, 20 days). The second group ranged from 4 months, 28 days to 6 months, 1 day (mean age = 5 months, 16 days). The eldest group ranged from 7 months, 1 day to 7 months, 29 days (mean age = 7 months, 18 days). Infants were healthy and full term and lived in or near Ithaca, New York. Race, sex, and socioeconomic status were not controlled. Parents were recruited by letter and by telephone from the birth records of local newspapers, and they were reimbursed for their transportation. One of the babies in the sample, a 3½-month-old, failed to complete the live voice episode due to fussiness. Three additional infants participated in the experiment, but were not included in the final sample because crying brought an early termination to the tape-recorded voice episode.

Procedure. Each baby was placed in a reclinable infant seat facing midway between the parents, who sat facing their child and looked at his or her eyes. The lateral positions of the mother and father randomly varied across infants. The distance from the infant's face to either parent's face was about 80 cm, and the parents' faces were about 80 cm apart. Thus, when the infant looked straight ahead, each parent's face was located about 30° in the periphery. An experimenter crouched below the infant, out of sight, with a collection of small toys. One second before a voice was heard, she attempted to draw the baby's gaze to a position midway between the parents by waving a toy. The experimenter did not know which parent was about to speak.

During each episode, the voice of each parent was presented eight times. These voices occurred in a different order for every infant; orders were random, with the restrictions that neither parent speak more than three times in succession and that each parent speak four times within the first eight trials. Eight seconds after one utterance, the next voice began. A soft beep indicated when a parent was to begin speaking. Within these constraints, we attempted to make the speech as natural, and as familiar to the baby, as possible. Parents were asked to speak as if they were talking to their baby to elicit his or her attention. They were

encouraged to use their own words, and the duration of each utterance was permitted to vary between one and four seconds. Parents were encouraged, but not required, to say different things on different trials. Most parents varied their utterances considerably.

The voices for the tape-recorded episode were recorded prior to the session by a second experimenter. Parents spoke into a microphone as if talking to their baby. If a parent spoke too long, spoke out of turn, or failed to speak at the correct time, we began the recording again; this occurred in two cases. Except for the rare baby who would not tolerate a parental separation, infants were not present while the voices were recorded. Parents spoke directly to their infant during the live voice episode. They were otherwise given the same instructions for speaking as they had received for making the tape recording. Again, they chose their own words—often, but not always, words that had been spoken during the previous episode. It was never necessary to recommence the live voice episode because a parent spoke too long or spoke out of turn.

During the tape-recorded voice episode, the voices were played in stereo through speakers located behind each parent. To an adult, they were most easily localized between the parents. The parents were asked to remain quiet and motionless during the session and to refrain from initiating interactions with the baby. During the live voice episode, which followed a five-minute break, a parent's voice was localized in the place where he or she could be seen, and the parent's face was seen to move as his or her speech was heard.

The infant's face and the parents' voices were videorecorded throughout the two experimental episodes. Infants' visual behavior was scored during the seven seconds that followed the onset of a voice, regardless of the length of the utterance.

Dependent Measures. The video record was scored first by two observers, working cooperatively. The observers were unaware of the lateral positions of the mother and father; they recorded looks to the "left" and to the "right." On each trial, the observers judged the direction of the infant's first look after the onset of a voice, provided that any look to a parent occurred within 7 sec. They also judged whether the infant subsequently looked to the other parent within this 7-sec period. Finally, they judged whether, despite our efforts to attract his or her attention, the infant was already looking at a parent when the voice began. The observers viewed each trial repeatedly until they agreed on each judgment. Their reliability with a second pair of observers was calculated on two infants at each age. Of the 32 judgments given at each age, the pairs of observers agreed on the direction of the first look on 73%, 70%, and 95% of the trials for the 3½-, 5½-, and 7½-month-olds, respectively. They gave 64 judgments at each age concerning the existence of any look, first or second, to each parent, and they agreed on 81%, 79%, and 89% of the trials. They had to judge whether the infant looked before the onset of a voice on 32 trials at each age, and they agreed on 90%, 75%, and 90% of these.

The video record was scored next by a third observer and a coder, both of whom were blind to the lateral positions of the mother and father. The observer continually recorded looking to the left and to the right onto two channels of a polygraph record. The moment of onset of each voice was recorded onto a separate channel. The coder worked from this polygraph record. He scored looking during a 5-sec period, beginning with the onset of a voice, for each trial. He derived two further measures of visual search. First, on each trial the coder measured the latency, from the onset of a voice, of the infant's first look to the parent on the left and to the parent on the right. Latency of looking to a parent was considered to be 5 sec if no look to that parent occurred within the 5-sec time period. Second, the coder judged the duration of looking to each parent during the 5-sec time period. All looks to a parent that occurred during that period were counted in deriving a duration score. The duration of looking to a parent was scored as 0 sec if no look to that parent occurred within 5 sec after the voice onset. A second observer and coder independently derived latency and duration scores for two infants at each age. The two pairs of scorers thus gave 64 judgments of latency and duration at each age, 32 per infant. The correlations between their judgments were .71, .90, and .79 (latency) and .72, .90, and .81 (duration) for the 3½-, 5½-, and 7½-month-olds, respectively.

Trials were excluded from further analysis if an infant was judged to be looking at either parent when a voice began. An average of three trials were excluded for each infant. Tables 1 and 2 present the number of trials that remained for analysis in each episode and for infants at each age. From these remaining trials, four measures of visual search were calculated. For the *First Look* measure, we counted the number of mother-voice trials on which the infant looked first to the mother (the "speaking" parent) and to the father (the "silent" parent), and the number of father-voice trials on which the infant looked first to the father (the speaking parent) and to the mother (the silent parent). For the *Eventual Look* measure, we counted the number of mother-voice and father-voice trials on which the infant looked at all, first or second, to the speaking and to the silent parent. For the *Latency* measure, we calculated the mean duration of time that elapsed between the onset of a voice and the infant's first look to the speaking and to the silent parent on mother-voice and on father-voice trials. For the *Duration* measure, we calculated the mean duration of looking to the speaking and silent parent on mother-voice and father-voice trials. First Looks and Eventual Looks were derived from the record of looking made by the first pair of observers, while Latency and Duration scores were derived from the data recorded and reduced by the second pair of assistants.

Separate 3 × 2 × 2 mixed factor analyses of variance (ANOVAs) were performed on each measure for each of the two episodes. The between-subjects factor was Age, and the within-subjects factors were Adult (looking to the mother vs. father) and Speaker (looking to the parent whose voice was heard vs. to the other parent).

Results

Tape-Recorded Voice Episode. Infants looked primarily to the parent who was heard to speak (Table 1). On the First Look measure, there was a main effect of Speaker: infants looked first to the parent whose voice was played on more trials than to the other parent, $F(1, 24) = 10.22, p < .005$. This effect did not interact with that of Age, $F(2, 24) < 1$. A similar effect of Speaker emerged in analysis of Eventual Looks, $F(1, 24) = 21.24, p < .001$. It was not qualified by an Age by Speaker interaction, $F(2, 24) < 1$. No other effects were significant in either analysis.

TABLE 1
Visual Search for the Parents: Experiment I, Tape-Recorded Voice Episode

	Age (months)		
	3½	5½	7½
Number of scorable trials			
Mean	12.33	13.00	13.22
Range	10–16	9–16	11–15
Number of trials with a first look			
Speaking parent	6.67	6.56	7.44
Silent parent	4.33	5.67	4.89
$F(1, 24)$	5.48[a]	1.36	5.98[a]
Number of trials with an eventual look			
Speaking parent	8.11	9.78	11.22
Silent parent	6.22	8.22	8.33
$F(1, 24)$	5.67[a]	3.80[b]	13.25[c]
Mean latency of looking			
Speaking parent	2.96	2.38	2.09
Silent parent	2.98	2.83	2.92
$F(1, 24)$	<1	2.15	7.48[a]
Mean duration of looking			
Speaking parent	1.15	1.63	1.79
Silent parent	1.26	1.32	0.82
$F(1, 24)$	<1	1.47	16.44[c]

[a] $p < .05$.
[b] $p < .10$.
[c] $p < .01$.

Analysis of the Latency measure revealed that infants looked faster to the speaking parent, $F(1, 24) = 6.08, p < .025$, an effect that did not interact with Age, $F(2, 24) = 1.77, p > .10$. There was an interaction of Age and Adult in the Latency analysis, $F(2, 24) = 3.93, p < .05$. Tests for simple effects revealed that 3½-month-olds looked somewhat faster to the father, $F(1, 24) = 3.28, p < .10$,

5½-month-olds looked faster to the mother, $F(1, 24) = 4.23$, $p < .10$, and 7½-month-olds showed no preference $F(1, 24) < 1$. Analysis of the Duration measure revealed that infants looked longer to the speaking parent, $F(1, 24) = 7.12$, $p < .025$. This effect increased with age, producing an Age by Speaker interaction, $F(2, 24) = 4.65$, $p < .025$. No other significant effects emerged from the Latency or the Duration analyses.

The results of tests for simple effects of Speaker at each age are given in Table 1. The effect of Speaker was most pronounced at 7½ months, but it was also evident at the youngest age on the First and Eventual Looks measures.

Live Voice Episode. Infants showed a strong tendency to search visually for parents who actually spoke (Table 2). They looked first and eventually more often to the speaking than to the silent parent, $F(1, 23) = 82.12$ and 89.79, respectively, $p < .001$. Infants also looked faster and longer to the speaking parent, $F(1, 23) = 60.68$ and 105.85, $p < .001$. No other effects were significant in these analyses. In particular, there were no hints of an Age by Speaker interaction, all Fs$(2, 23) < 1.5$. The simple effects of Speaker were significant at each age for each measure (see Table 2).

TABLE 2
Visual Search for the Parents:
Experiment I Live Voice Episode

	Age (months)		
	3½	5½	7½
Number of scorable trials			
Mean	12.75	13.33	13.78
Range	10–16	10–16	10–16
Number of trials with a first look			
Speaking parent	8.67	8.33	8.33
Silent parent	2.35	2.22	2.00
$F(1, 23)$	23.40[a]	22.01[a]	23.63[a]
Number of trials with an eventual look			
Speaking parent	9.33	9.33	8.78
Silent parent	5.89	3.78	4.67
$F(1, 23)$	9.61[a]	24.01[a]	13.69[a]
Mean latency of looking			
Speaking parent	1.84	1.77	2.27
Silent parent	3.75	3.94	4.14
$F(1, 24)$	34.02[a]	46.28[a]	34.39[a]
Mean duration of looking			
Speaking parent	1.97	1.88	1.29
Silent parent	0.76	0.43	0.38
$F(1, 24)$	10.89[a]	30.31[a]	13.10[a]

[a] $p < .01$.

Discussion

Infants explored the mother and father by listening and looking. They looked systematically for each parent when he or she was heard to speak. This pattern of search was especially pronounced when parents actually spoke to their infant,[1] but it was also observed when their speech was tape-recorded. In the tape-recorded voice episode, looking could not have been guided by the spatial direction of the voice or by visible movements of the speaker. Infants appeared to use knowledge of a parent's audible and visible characteristics to direct their search.

During the tape-recorded voice episode, optic and acoustic information about a parent were both spatially separated and temporally asynchronized. These discrepancies did not, however, induce in infants any apparent state of conflict. The subjects displayed no distress or obvious surprise during the tape-recorded voice episode. Although more subtle conflict responses may have escaped our attention, strong reactions to audiovisual discrepancy were not observed in this situation. This experiment offers no clear support for the claim that infants respond negatively to discrepancies between auditory and visual information about familiar objects.

Some important questions remain from Experiment I. First, the study did not control for a possible source of bias in the tape-recorded voice episode of the visual search procedure. Searching for the speaking parent may have been guided by vicarious visible signals from parent to baby. In the tape-recorded voice episode, the parents, like the baby, could hear each voice that was played. Parents were asked not to react to the onset of a voice, and they appeared to follow these instructions. Furthermore, parents' faces appeared rather far in the infant's peripheral vision at the time of voice onset. Their subtle gestures would not likely be detectable to a baby. Nevertheless, we did not monitor the parents' behavior closely and cannot be certain that they inhibited all gestures to which a baby might respond.

If intermodal knowledge truly guided auditory–visual exploration, a second question arises concerning the specificity of that knowledge. Did an infant, hearing a parent's voice, know to search for a person with the particular visible characteristics of that parent, or did he or she know only to search for an adult of a particular sex? Conversely, would an infant look to a parent only after hearing that parent's voice, or would the voice of any adult man or woman initiate a

[1]The superiority of search for a speaking parent in the live voice episode may indicate that looking can be guided by the spatial direction of a parent's voice or by the visible movements of his or her face. There are, however, at least three alternative explanations for this effect. First, acoustically guided search may have been attenuated in the tape-recorded voice episode because of the reduction in fidelity caused by the tape recording, despite the fact that the quality of the recordings was high. Second, searching may have improved due to a warm-up effect, since the live voice episode always occurred after the tape recorder voice episode. Finally, searching in the tape-recorded voice episode was conceivably attenuated by a reaction of surprise or conflict aroused by the spatial separation or temporal asynchrony of faces and voices.

search for the parent of the same sex? Further experiments could address this question by presenting infants with other visually and aurally specified objects.

A final question concerns the age at which infants first engaged in visual search in the tape-recorded episode. On two measures of searching, the tendency to look for an acoustically specified parent increased with age and was not reliable at the youngest ages. Interobserver reliabilities, however, also tended to be low at the youngest age. Observer errors may have introduced extra noise into the data from the younger infants and magnified the apparent age effect. One account of our results is that the youngest infants genuinely searched for the parents, but that only the First and Eventual Looks measures were sensitive enough to reveal their searching. To check this account, it seemed wise to replicate the experiment with young infants in order to determine if the First and Eventual Looks measures would again prove to be sensitive indexes of visual search.

Experiment II explored the intermodal knowledge of four-month-old infants. Eighteen infants were observed in a replication of the tape-recorded voice episode of Experiment I. Eighteen additional infants were observed in a search experiment with two adults of the same sex to assess the specificity of the infants' knowledge of the mother. The procedure for this latter group was the same as in the tape-recorded voice episode of Experiment I, except that each baby watched and listened to the mother and a second adult woman. To investigate the possibility of parental signaling to the child, one parent's face was videotaped during some of the sessions. This tape was monitored for the presence of visible gestures to the baby.

EXPERIMENT II

Method

Subjects. Eighteen infants participated in each condition of this experiment. Those in the Mother–Father Condition ranged in age from 3 months, 9 days to 4 months, 22 days (mean age = 3 months, 26 days). Those in the Mother–Other Woman Condition ranged in age from 3 months, 7 days to 4 months, 26 days (mean age = 3 months, 23 days). Infants were recruited and selected as in Experiment I. Two additional babies were eliminated from the Mother–Father Condition because of crying.

Procedure. The procedure was identical to the tape-recorded voice episode of Experiment I, except as follows. Infants in the Mother–Other Woman Condition participated in the experiment with the mother and a second adult woman. For six of the infants, the woman was chosen by the mother and was familiar to the baby. The remaining twelve babies participated in the study with

one of several unfamiliar research assistants. The mother and the second woman were instructed to speak as if they were speaking to that particular baby, again using their own words. The research assistants spoke as they spoke normally to unfamiliar infants in the laboratory. In general, we noticed no pronounced differences in the intonation or phrasing of speech by mothers, other familiar women, and research assistants.

For thirteen subjects, nine from the Mother–Father Condition, a video recording of one parent's face was made during the episode. The voices were also recorded on videotape. These parents were given the same instructions as in Experiment I: to face the baby quietly and to refrain from reacting to a voice. The video recordings were played silently to each of six college undergraduates to whom the nature and purpose of the experiment had been explained. They were told when a voice began, and were asked to guess whether that voice belonged to the adult they watched. The observers knew that the voice would belong to that adult on half the trials, and that a parent might reveal when his or her voice was played by attempting to elicit the attention of the baby in some manner. The proportion of correct guesses was tested against the chance value of .50 for each baby, for each observer, and over all babies and observers.

The video record of the baby's face was scored and reduced as in Experiment I. Table 3 gives the number of usable trials in each condition. The data for each condition were analyzed by 2 × 2 repeated measures ANOVAs with Adult (looking to the mother vs. father or other woman) and Speaker as the within-subjects factors.

Results

Mother–Father Condition. Infants tended to search for the speaking parent, as indicated by analyses of the measures that had proved most sensitive in Experiment I (Table 3). There was a marginally significant tendency to look first more often to the speaking parent, $F(1, 17) = 3.56, p < .08$, and a significant tendency to look eventually to that parent on more trials, $F(1, 17) = 7.37, p < .02$. No other effects emerged in the analyses, and no significant effects emerged in the analysis of latency or duration of looking.

Searching for the speaking parent did not appear to be guided by visual detection of vicarious gestures from the parent. The six observers who monitored the recordings of one parent's face for such gestures guessed correctly on 53% of the trials, a nonsignificant departure from chance, $t(77) = 0.21$. No observer achieved a level of accuracy above that expected by chance; no adult on videotape was scored by the observers with greater than chance accuracy (for each analysis, $p > .05$, one-tailed).

The introduction of video monitoring equipment might, conceivably, have reduced the tendency to gesture by parents who were monitored. Other parents, not inhibited by the video recorder, might have gestured to their infants. Those

TABLE 3
Visual Search: Experiment II

	Condition	
	Mother–Father	*Mother–Other Woman*
Number of scorable trials		
Mean	13	14
Range	7–16	6–16
Number of trials with a first look		
Speaking adult	5.66	3.99
Silent adult	4.66	5.39
Number of trials with an eventual look		
Speaking adult	7.33	4.61
Silent adult	6.16	6.33
Mean latency of looking		
Speaking adult	3.04	3.52
Silent adult	3.19	3.01
Mean duration of looking		
Speaking adult	1.10	0.77
Silent adult	1.17	1.14

infants alone might have tended to look first and eventually more often to the "speaking" parent. To check this possibility, searching by infants was tabulated separately for the 9 infants for whom a parent was videotaped (Monitored Group) and for the 9 infants for whom no parent was videotaped (Nonmonitored Group). These tabulations appear in Table 4a. For infants in the Monitored Group, searching for the videotaped parent and for the other parent is tabulated separately in Table 4b. Searching by infants in the two groups was about equally consistent,

TABLE 4
The Effect of Video Monitoring on
Visual Search for the Parents, Experiment II

a. Search by infants in the monitored and nonmonitored groups

	Monitored group			*Nonmonitored group*		
	Speaking	*Silent*	*F(1, 16)*	*Speaking*	*Silent*	*F(1, 16)*
First looks	5.22	4.56	<1	6.11	4.78	3.04
Eventual looks	7.00	6.11	2.06	7.67	6.22	5.45[a]

b. Search by infants in the monitored group in response to the voice of the monitored and nonmonitored parent

	Videotaped parent's voice			*Other parent's voice*		
	Speaking	*Silent*	*F(1, 8)*	*Speaking*	*Silent*	*F(1, 8)*
First looks	2.56	2.33	<1	2.67	2.22	<1
Eventual looks	3.56	2.56	5.36[a]	3.44	3.56	<1

[a] $p < .05$.

and infants in the Monitored Group searched as strongly for the videotaped as for the nonvideotaped parent. Two by two mixed factor ANOVAs on the First Looks and Eventual Looks measures, with Group (Monitored vs. Nonmonitored) between subjects and Speaker within subjects, revealed the expected main effects of Speaker, $F(1, 16) = 3.43$ and 7.11, respectively, $p < .10$ and .05, but no interaction of Speaker with Group, both Fs < 1. Two by two repeated measures ANOVAs on the results for the Monitored Group alone, testing the effects of Videotaping (looking to the videotaped parent vs. the other parent) and Speaker, revealed no significant effects of Speaker, $F(1, 8) < 1$ for First Looks and $F(1, 8) = 2.23$ for Eventual Looks, and no Speaker by Videotaping interaction, both Fs < 1. The results of tests for the simple effects of Speaker, in each of the four analyses, appear in Table 4. The only significant effects appeared in analyses of the Eventual Looks measure. Infants in the Nonmonitored Group tended to search for the speaking parent. Infants in the Monitored Group searched only for the Speaking parent who was monitored on videotape.

Mother–Other Woman Condition. Preliminary analyses of searching revealed no effect of the familiarity of the second woman, so the 18 infants were treated as a single group. They showed a surprising tendency to look more often to the silent adult (Table 3). This main effect of Speaker appeared in the analysis of First Looks, $F(1, 17) = 7.95, p < .02$, but not in the analysis of Eventual Looks, $F(1, 17) = 1.74, p > .10$. Infants looked more quickly and for a longer time to the silent woman $F(1, 17) = 8.63$ and $8.28, p < .01$. No other effects were significant in these analyses.

Discussion

The results of the Mother–Father Condition provide evidence that a four-month-old infant can search visually for a parent whose tape-recorded voice is played. The measures that had provided evidence for auditory—visual search by 3½-month-olds in Experiment I also provided comparable evidence for search by four-month-olds in Experiment II. This searching does not appear to be guided by vicarious gestures from the parents. The chance performance of the six observers who judged parental signaling from the video record, contrasted with the visual activity of the infants, strengthens our conviction that searching was guided by intermodal knowledge. It seems highly unlikely that a baby would respond to signals, 30° in the periphery, that a prepared adult observer could not detect in central vision.

The results of the Mother–Other Woman Condition were unexpected and are difficult to interpret. Visual activity in this condition might conceivably reflect gaze avoidance of a speaking adult who is not the infant's parent. If these results are replicable, they indicate that infants do not invariably use intermodal knowledge of the mother in order to search for her. There appear to be circumstances in which infants show the very opposite tendency.

Once again, infants in these conditions were not obviously surprised or distressed by the discrepant optic and acoustic information.

GENERAL DISCUSSION

Infants patterns of auditory–visual exploration provided evidence for knowledge of the audible and visible characteristics of their parents. Infants were able to look for a parent whose voice was presented, even when the parent's face and voice were spatially and temporally unrelated. Searching in this situation became more consistent between 3½ and 7½ months, but it was discernible even at the youngest age. Infants in the fourth month appeared to use knowledge to guide visual search for a parent whose voice was heard.

This finding indicates that for a young infant, as for an adult, acoustic information may specify an object with definite visible characteristics. Young perceivers evidently possess some of the knowledge that underlies the adult's ability to experience objects with manifold properties from stimulus information in a single modality. Although these studies do not reveal how infants acquire intermodal knowledge, they suggest an experimental approach to that question. The auditory–visual search method might be used to study the acquisition of knowledge of visible and audible objects. Indeed, one investigation indicates that four-month old infants can acquire intermodal knowledge of some inanimate objects in minutes (Spelke, 1978b).

This research underscores the interrelation of perceptual exploration and knowledge in the very young. Infants are able and persistent seekers of optic and acoustic information. Their investigations can be guided by the intermodal relationships that they have come to know. Intermodal exploration may yield further information about audible and visible objects, increasing the infant's knowledge of their properties. Young perceivers thus may engage in a cycle of exploring and learning about bimodally specified objects. The adult's perceptual knowledge may ultimately derive from this cycle.

ACKNOWLEDGMENTS

We thank Richard Evans, Dale Klopfer, Kate Loveland, Celia Reaves, Michael Slivka, Nancy Van Derveer, and especially Eleanor J. Gibson for their advice and assistance.

REFERENCES

Aronson, E., & Rosenbloom, S. Space perception in early infancy: Perception within a common auditory–visual space. *Science*, 1971, *172*, 1161–1163.

Berkeley, G. Essay towards a new theory of vision, 1709. In A. C. Fraser (Ed.), *Selections from Berkeley*. Oxford: Clarendon Press, 1910.

Bower, T. G. R. Object perception in infants. *Perception*, 1972, *1*, 15–30.

Bower, T. G. R., Broughton, J. M., & Moore, M. K. The coordination of visual and tactual input in infants. *Perception and Psychophysics*, 1970, *8*, 51–53.

Bruner, J. S., & Koslowski, B. Visually preadapted constituents of manipulatory action. *Perception*, 1972, *1*, 3–14.

Cohen, S. E. Infant attentional behavior to face-voice incongruity. Paper presented at the meetings of the Society for Research in Child Development, Philadelphia, March, 1973.

Condry, S. M., Haltom, M. K., & Neisser, U. Infant sensitivity to audio-visual discrepancy: A failure to replicate. *Bulletin of the Psychonomic Society*, 1977, *9*, 431–432.

Field, J. Coordination of vision and prehension in young infants. *Child Development*, 1977, *48*, 97–103.

Gibson, E. J. *Principles of perceptual learning and development*. New York: Appleton–Century–Crofts, 1969.

James, W. *The principles of psychology*. New York: Holt, 1890.

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

McGurk, H., & Lewis, M. Space perception in early infancy: Perception within a common auditory–visual space? *Science*, 1974, *186*, 649–651.

Mendelson, M. J., & Haith, M. M. The relation between audition and vision in the human newborn. *Monographs of the Society for Research in Child Development*, 1976, *41*, Whole No. 4.

Mill, J. *Analysis of the phenomena of the human mind*. London: Longmans & Dyer, 1829.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Spelke, E. Perceiving bimodally specified events. Paper presented at the meeting of the International Conference on Infant Studies, Providence, March, 1978. (a)

Spelke, E. Acquiring intermodal knowledge. Paper presented at the meeting of the International Conference on Infant Studies, Providence, March, 1978. (b)

Titchener, E. B. *An outline of psychology*. New York: Macmillan, 1902.

Wertheimer, M. Psychomotor coordination of auditory–visual space at birth. *Science*, 1961, *134*, 1692.