



Learning Stimulus-Location Associations in 8- and 11-Month-Old Infants: Multimodal Versus Unimodal Information

Sophie ter Schure and Dorothy J. Mandell
*Cognitive Science Center Amsterdam
University of Amsterdam*

Paola Escudero
*Cognitive Science Center Amsterdam
University of Amsterdam
MARCS Institute
University of Western Sydney*

Maartje E. J. Raijmakers
*Cognitive Science Center Amsterdam
University of Amsterdam*

Scott P. Johnson
*Department of Psychology
University of California Los Angeles*

Research on the influence of multimodal information on infants' learning is inconclusive. While one line of research finds that multimodal input has a negative effect on learning, another finds positive effects. The present study aims to shed some new light on this discussion by studying the influence of

multimodal information and accompanying stimulus complexity on the learning process. We assessed the influence of multimodal input on the trial-by-trial learning of 8- and 11-month-old infants. Using an anticipatory eye movement paradigm, we measured how infants learn to anticipate the correct stimulus–location associations when exposed to visual-only, auditory-only (unimodal), or auditory and visual (multimodal) information. Our results show that infants in both the multimodal and visual-only conditions learned the stimulus–location associations. Although infants in the visual-only condition appeared to learn in fewer trials, infants in the multimodal condition showed better anticipating behavior: as a group, they had a higher chance of anticipating correctly on more consecutive trials than infants in the visual-only condition. These findings suggest that effects of multimodal information on infant learning operate chiefly through effects on infants' attention.

INTRODUCTION

Infants are able to integrate auditory and visual information from a very early age (for a review, see Lewkowicz, 2000). For instance, they look longer at a matching speaking face when hearing a syllable at 2 months (Patterson & Werker, 2003), and discriminate a tempo change when habituated to both the sound and the movement of a tapping hammer but not in unimodal conditions at 3 months of age (Bahrick, Flom, & Lickliter, 2002). However, when auditory and visual information are arbitrarily connected, the literature is equivocal (Plunkett, Hu, & Cohen, 2008; Robinson & Sloutsky, 2004, 2010; Stager & Werker, 1997; Waxman & Braun, 2005).

Previous literature has mostly studied whether children could process and represent critical auditory or visual features that are shared across the stimuli after habituation in multimodal versus unimodal contexts. However, in these habituation studies, criterion effects may play an important role in infants' behavior (McMurray & Aslin, 2004). That is, outcomes rely on an individual judgement of whether a new exemplar is dissimilar from the familiarized exemplars. The research question in those studies is whether the way infants process information varies between contexts, but thresholds determining when stimuli are judged to be different may vary as well. There is little research studying how multimodal versus unimodal information affects learning in a two-alternative forced choice task, which does not depend on such a threshold. The current study aims to study infants' learning process when they are presented with unimodal versus multimodal information in an anticipatory eye-movement task (McMurray

& Aslin, 2004). This paradigm allows for testing the variability of the speed and consistency with which infants are able to associate a discriminating feature with a location during the learning process.

There are a number of explanations for why multimodal (auditory and visual) information may impair learning. One explanation focuses on the earlier development of the auditory system over the visual system, which results in auditory information being dominant over visual information. The Auditory Dominance Hypothesis was introduced by Lewkowicz (1988a,b) and expanded by Robinson and Sloutsky (2004). Robinson and Sloutsky have shown that infants who are trained with a multimodal stimulus attend to an auditory change more than to a visual change (Robinson & Sloutsky, 2004), and that while infants trained with a unimodal visual stimulus do succeed at noticing a visual change, infants trained with the same visual stimulus but combined with auditory input fail to notice the change (Robinson & Sloutsky, 2010). They concluded that auditory input overshadows visual processing in infants younger than 14 months (Robinson & Sloutsky, 2004, 2010). Between 14 and 24 months this dominance abates, resulting in more efficient formation of arbitrary auditory-visual associations.

Werker and colleagues (Stager & Werker, 1997; Werker, Cohen, Lloyd, Casasola, & Stager, 1998) have shown that linguistic input specifically seems to lead to this difficulty. Casasola and Cohen (2000) showed that linguistic labels (but not nonlinguistic sounds) impaired 14-month-old children's ability to discriminate between observed actions. Further, when the difference in the linguistic information is minimal, object-word associations can be formed by 17-month olds but not by younger children, who fail to pay attention to a switch in the object-word pair that they were trained with (Stager & Werker, 1997), even though they can discriminate the words in the absence of a possible visual referent. These results have led to the conclusion that the linguistic information either overshadows the processing of visual information or directs infants' attention towards irrelevant features.

In contrast, Waxman and colleagues (Ferry, Hespos, & Waxman, 2010; Waxman & Booth, 2003; Waxman & Braun, 2005) suggested that adding linguistic information provides a label that infants can use to group visual stimuli. They consistently showed that a word, but not an attention-getting phrase, facilitates processing visual information. Interestingly, Plunkett et al. (2008) showed that adding a word helped infants only if the visual information could be easily divided into multiple categories, but not if this grouping was difficult to make. Thus, this line of studies suggests that auditory information seems to aid, but not create, the discrimination of visual information.

More recently, Plunkett (2010) attempted to bring these lines of research together by proposing that the ease with which infants can process

multimodal information depends on the familiarity and the complexity of the information in each modality. Linguistic labels might have special salience for infants, but if the visual information is novel and complex, they will not benefit from the presence of an auditory stimulus. Further, Plunkett's computational model of infant categorization predicts that auditory-visual compound stimuli will result in longer looking times than unimodal stimuli, because they have a higher complexity or higher cognitive load. In this study, as well as in the previous studies, the relation between auditory and visual information was arbitrary.

Bahrick, Lickliter, and Flom (2004) proposed that auditory-visual compound stimuli can be easier to process than unimodal stimuli under particular circumstances. Their Intersensory Redundancy Hypothesis (Bahrick & Lickliter, 2000; for reviews, see Bahrick et al., 2004; Bahrick & Lickliter, 2012) postulates that when information from auditory and visual modalities is linked by an amodal property such as synchrony, infants will process the amodal information before and more easily than modality-specific information, even when the auditory and visual content is not related (Hollich, Newman, & Jusczyk, 2005; Hyde, Jones, Porter, & Flom, 2010). According to this hypothesis, intersensory redundancy directs infants' attention to amodal properties, while under unimodal stimulation—or multimodal stimulation without synchrony—attention is focused on modality-specific information.

The hypotheses of both Plunkett (2010) and Bahrick and Lickliter (2012; Bahrick et al., 2004) focus on how properties of the stimuli influence infants' attention during the task and hence also how the information is processed. In doing so, they address the apparent discrepancies in the previous literature: infants will benefit from multimodal input under optimal conditions of complexity and synchrony of the auditory and visual components of the stimuli. Both hypotheses suggest that infants will first focus on the most salient features of the stimuli, but where Plunkett (2010) proposes that this might be the auditory component if it is a linguistic label, Bahrick and Lickliter (2012) suggest that it will be an amodal feature (e.g., the synchronicity of visual and auditory information).

Previously published studies mainly focused on the *outcome* of learning and not on the learning *process*. Specifically, they presented infants with unimodal or multimodal information and subsequently tested how the type of information presented during training affected infants' performance during testing. That is, the learning phase itself was not subject of study. However, it seems that differences in methods specifically affected the learning process (e.g., fixed duration trials as in the Ferry et al., 2010 study versus habituation in the Plunkett et al., 2008 study), which possibly affected infants' behavior during the test phase as well. Given that criterion effects play a role in habituation paradigms (McMurray & Aslin,

2004), these aspects of habituation based paradigms cloud unambiguous interpretation of looking times during test. Therefore, it is difficult to determine whether looking time differences at test are due to higher stimulus complexity in the multimodal condition or failure to process information from one of the two modalities.

To our knowledge, no previous study has examined how infants' learning and attention *unfolds* across trials during presentation of multimodal versus unimodal stimuli. To address this question, we employed the anticipatory eye movement paradigm (AEM; McMurray & Aslin, 2004), which allows the learning process to be assessed through changes in overt behavior. Specifically, the AEM tests whether an infant will anticipate where a moving stimulus will reappear on the basis of its features. Infants see an object appear on the screen, move upwards until it is completely hidden behind an occluder, and reemerge on either the left or right side of the occluder. Only after infants have attended to the discriminating features of the two stimuli will they be able to process the trajectory of the object and the association between the stimulus features and the reappearance location (Markman & Ross, 2003). Thus, infants have a learning curve that characterizes how long it takes them to use discriminating features for making associations with a reappearance location and how well they can apply these associations (Mandell & Raijmakers, 2012).

Using the AEM paradigm, Albareda-Castellot, Pons, and Sebastián-Gallés (2011) showed that bilingual 8-month-old infants could successfully learn to associate words that were distinguished by a single speech sound (i.e.,/dedi/versus/dædi/) with the reemergence of an attractive visual object (an Elmo face) at two screen locations. Similarly, Mandell and Raijmakers (2012) demonstrated that 11-month olds associated two visual objects with different sides of the screen and generalized this association to visual objects with similar features. Thus, infants are able to learn discriminating stimulus features and associate these with a reappearance location in both the auditory and visual modalities.

Based on the success of these previous unimodal (auditory-only or visual-only) studies, we employed the AEM paradigm to compare the learning process of infants that are presented with auditory-only, visual-only, or auditory and visual (multimodal) distinctive information. It is important to note that all discriminating features of the two stimuli are modality-specific. In the multimodal and auditory-only conditions, the presentation of auditory and visual components of the stimulus were synchronized, resulting in an amodal cue, which might drive attention to the object itself, but not specifically to the discriminating features. Our aim was to study how multimodal (arbitrarily related) cues versus unimodal cues affect the learning of object–location associations. Does the type of information

only affect the learning speed or also how well associations can be learned? We would expect that due to complexity differences, stimuli in unimodal conditions are processed faster than in multimodal conditions. However, the literature does not provide us with expectations towards the strength of the associations between conditions. We tested two age groups, 8- and 11-month olds, because the previous studies suggest a developmental change in how multimodal input would affect learning (Bahrick & Lickliter, 2012; Casasola & Cohen, 2000; Robinson & Sloutsky, 2004, 2010).

METHOD

Participants

Sixty-three infants, 31 8-month olds (age range: 7.5–8.5 months) and 32 11-month olds (age range: 10.5–11.5 months), were included in the analysis. All infants were full term and had no known developmental difficulties or hearing or visual impairments. They were randomly assigned to three conditions: multimodal ($n = 19$); auditory-only ($n = 22$) and visual-only ($n = 22$). An additional 40 infants participated but were excluded from further analysis due to fussiness (multimodal: $n = 7$, auditory-only: $n = 5$, visual-only: $n = 11$) or anticipating on fewer than 50% of the trials¹ (multimodal: $n = 8$, auditory-only: $n = 4$, visual-only: $n = 5$).

Apparatus

Infants' fixations were captured with a Tobii 1750 eye tracker with a 50 Hz sampling frequency (20 ms per sample). Point of gaze was calibrated through the native Clearview software, and E-prime (Psychology Software Tools, Sharpsburg, PA, USA) was used for task control and data collection. The trials were shown on the Tobii monitor and sound was played through two speakers located at the infant's eye level. Trial number, x and y coordinates of the upper left corner of the stimulus, x and y coordinates of the infant's gaze and timing were collected.

Stimuli

The auditory stimuli consisted of two nonsense words, *feep* (/fip/) and *fap* (/fap/), recorded by a female native speaker of American English. The

¹All tests were also run with these low-anticipating infants resulting in no change to the overall model effects. However, including low-anticipating infants attenuated the magnitude of the parameter estimates for the differences on specific trials.

vowels of these words differ mainly on their first and second formant (F1 and F2), and infants are able to distinguish these vowels from an early age, because their formant frequencies are maximally distinct (Polka & Bohn, 1996). The auditory stimuli were matched on length (585 ms) and amplitude (75 dB). Pitch for *feep* increased from 150 to 275 Hz and for *fap* it increased from 150 to 300 Hz. The main formant frequencies of the vowels, measured at the midpoint of each vowel, were an F1 of 350 Hz and an F2 of 2950 Hz for the /i/ in *feep*, and an F1 of 975 Hz and an F2 of 1820 for the /a/ in *fap*.

The two visual stimuli, a circle and a triangle, were drawn with Adobe Illustrator. They were equal in color (light purple) and size (150 × 150 pixels). Shape was used as the visual dimension because it has been shown that infants as young as 2 months discriminate between shapes and view this dimension as an invariant property of an object even across occlusion (Wilcox, 1999; cf. Bremner, Slater, Mason, Spring, & Johnson, 2013).

Procedure

Infants sat on their parent's lap approximately 60 cm away from the display. Parents were instructed not to interact with the child during the trials. Prior to the experiment, infant's point of gaze was calibrated with a standard 5-point calibration procedure, where gaze is directed to a sequence of five coordinates on the screen. Calibration was deemed successful for an infant when it resulted in at least four acceptable points.

The occluder, a bright purple tube with a center 'opening' and 'openings' on both sides, was shown at the middle of the screen and was present throughout each trial. A trial started with the appearance of the visual stimulus at the bottom center of the screen. It loomed twice, shrinking to 80% of its size, and moved up with a constant velocity until it was completely hidden behind the occluder. The visual stimulus remained hidden for 3 sec, which was the time it needed to move through the occluder at the same velocity. It then reemerged from the left or right of the occluder, made a rapid figure-eight movement, and disappeared horizontally off the screen. Figure 1 shows an example trial.

For the multimodal and auditory-only conditions, the auditory stimulus (*feep* or *fap*) was played twice when the visual stimulus first appeared and loomed prior to its upward movement. The onset of the first utterance of the word was synchronous with the onset of the appearance of the object. The offset of the second utterance of the word was synchronous with the end of the looming. The auditory stimulus was played twice again concurrent with the reemergence of the object and its figure-eight movement, again with synchronous onset and offset. In the multimodal condition,

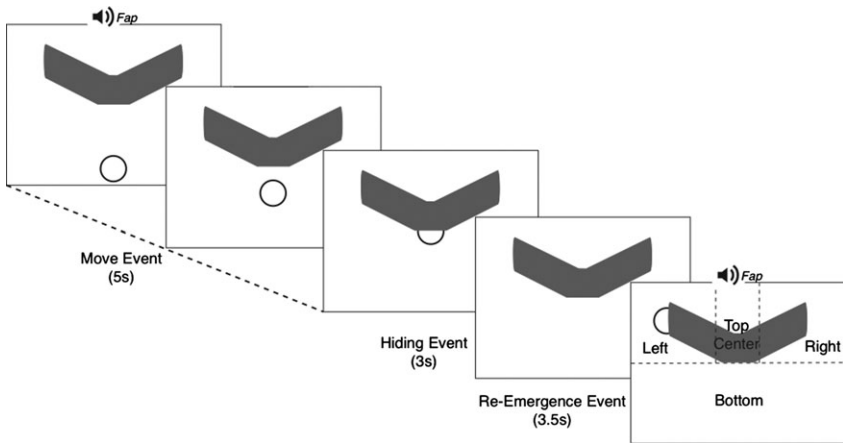


Figure 1 Illustration of an example trial. In the visual-only and multimodal conditions, infants also saw a triangle emerge on the right together with the auditory stimulus *feep*. In the auditory-only condition, the circle also emerged on the right together with the auditory stimulus *feep*. The auditory stimulus was played four times each trial: twice during appearance and twice during reemergence. The visual-only condition was silent.

feep was always paired with the triangle and *fap* with the circle. In the auditory-only condition, the infant saw a circle as the visual stimulus and only the auditory stimulus cued the reemergence location. In the visual-only condition, infants saw either a circle or a triangle without any auditory stimulus.

An attention getter consisting of both auditory (but not linguistic) and visual information was played before each trial to center the infant's gaze. Testing proceeded until infants disengaged or became fussy. The test session lasted about 5 min.

Data analysis

Raw gaze data were assigned to one of four possible areas of interest (AOIs) that corresponded to the bottom half, the upper middle, the upper right and the upper left portions of the screen.² The AOI was identified as

²Possible calibration error was checked by plotting the infant's gaze data against the actual location of the object during the move event, as this is when infants tracked the object. If their mean tracking was more than 150 pixels plus one standard deviation away from the x -axis center of the object, the gaze data were corrected along the x -axis to prevent incorrectly assigning these infants' looks as anticipations. Only three infants needed data correction in this way.

missing if there were no x and y gaze coordinates for the sample. The gaze data were then aggregated into look sequences in each AOI, maintaining the sequential order and duration of each look. If the duration of a missing AOI was shorter than 500 ms it was reassigned to the last valid AOI. Missing AOIs with duration longer than 500 ms were coded as a 'look away' from the screen.

The crucial measure of anticipation in each trial was where the infant looked between 150–0 ms before the stimulus reemerged from the occluder. If they looked at either side of the reappearance area (i.e., upper left or upper right regions) within that time window, the fixation was counted as an *anticipation* (Gredebäck, Johnson, & von Hofsten, 2010; Johnson, Amso, & Slemmer, 2003). Importantly, a fixation on the reappearance area was considered an anticipation only if the infant had looked at the object when it first appeared on the bottom of the screen for at least 250 ms. Anticipations were coded as being 'correct' or 'incorrect' based on whether the object would reemerge on that side of the screen. If the infant did not have an anticipation on the trial but looked at the center of the screen instead, it was coded as 'no anticipation.'

We coded for not anticipating because previous work with this paradigm has shown that looking at the center of the occluder while waiting for the object to reappear is an important and meaningful behavior during learning (Mandell & Raijmakers, 2012). Mandell and Raijmakers' trial-by-trial analysis shows that there is a progression from not anticipating to anticipating correctly when infants learn in this paradigm, rather than having a gradual increase in correct versus incorrect anticipations. The chance of having a correct anticipation on a trial is consequently 33%. Trials in which the infant looked away for more than 90% of the anticipation phase were treated as missing trials. Trial number was then resequenced to represent the 1st, 2nd, 3rd, etc. valid trial for each infant. Because previous studies found that infants attended to up to 40 trials (Albareda-Castellot et al., 2011; McMurray & Aslin, 2004), we set up the experiment similarly. In our study, very few infants attended to the screen for the full course of the experiment. To limit the number of missing trials, we cut off the trial sequence at 12 trials.³ Figure 2 shows the number of infants for which there was data per trial.

The outcome measure used in this study was a categorical variable scoring whether the last anticipatory look before the object reemerged was correct, incorrect, or whether there was no anticipation. By making the outcome measure categorical and using only one anticipation per trial, we controlled for any differences between infants arising from longer or

³Clusters with missing values are not used in a GEE-model.

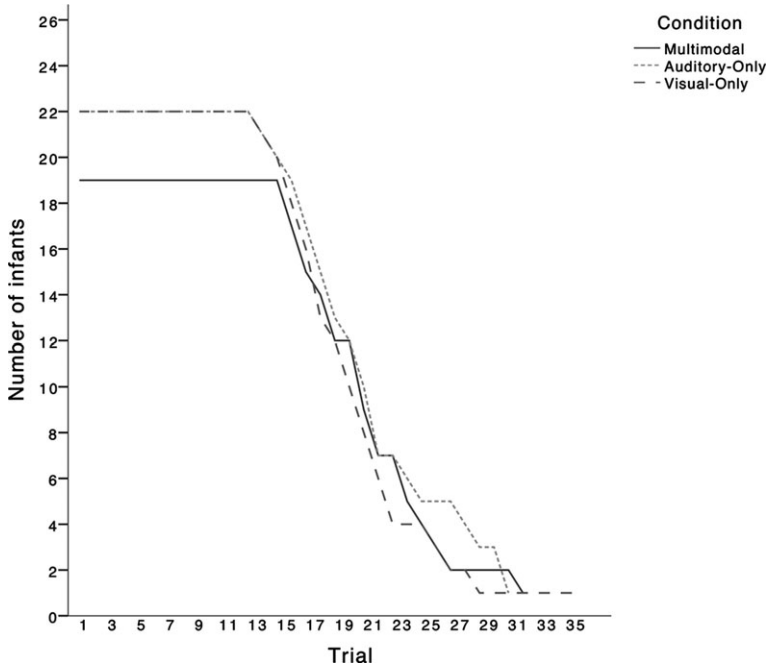


Figure 2 Plot of the number of infants for which there was data on each trial. Each experimental condition is depicted by a separate line.

shorter looking times during trials. Trial length was fixed. We used only the last anticipation instead of total looking times because, in the majority of cases, infants only made one anticipation per trial, which was immediately before the object reappeared.⁴

In keeping with previous findings using the AEM paradigm, two attention measures from the anticipation phase were calculated for each trial: (1) the duration of time that the infant spent looking away from the screen, and (2) the duration they spent looking at the center. These measures were analyzed separately to assess whether there were differences

⁴On average, infants had one anticipation on 80% of the trials. On trials where there was more than one anticipation, the first look was to the reemergence location of the previous trial in 26% of the cases regardless of whether this was the 'correct' location in the current trial. This was not a common behavior, however; it occurred 0.67 times per infant on average (SD 1.05, median 0, range 0–5), with only three infants who did this more than twice and no differences between conditions (F [2, 60] = 0.130, p = .878). In a previous study using the same method (Mandell & Raijmakers, 2012), the accuracy of the first versus last look was calculated which yielded significantly better scores for the last look.

between the conditions on the level of attention that infants in each condition allocated to the task.

All data were analyzed with Generalized Estimating Equations (GEE; Zeger & Liang, 1986; Zeger, Liang, & Albert, 1988). GEE are used to estimate the parameters of general linear models of repeated measures that do not assume that all measurements are independent, but allow for correlated repeated measures. Hence, multinomial data of learning trials are suited to be analyzed with GEE to test the difference between experimental conditions and the interactions with trial number.

We used GEE to model the repeated measures data with predictor variables that were treated as fixed effects. The anticipation data were analyzed using a multinomial cumulative-log linking function and a first-order autoregressive correlation structure to represent the learning nature of the data. This correlation structure assumes that trials that are consecutive are more correlated than trials that are further apart. The two attention measures were also analyzed with GEE using a first-order autoregressive correlation structure and an identity linking function, because these measures were normally distributed. For all analyses, condition, trial number, and infants' age were entered as factors. For the anticipation analysis, the duration of time the infant spent looking away was also included in the GEE as a covariate nested in trial, as Mandell and Raijmakers (2012) showed that looking-away is an important covariate in assessing an infant's learning process.

A full factorial model was fit to the data. The condition by trial and the condition by age interactions were always kept in the analysis as they tested our research questions: whether there were differences in the learning curves between conditions and whether the effect of modality of information varied across this age range.

RESULTS

Our research question was how multimodal arbitrarily related cues affect the learning of associations as compared to unimodal cues. To this aim, we assessed infants' trial-by-trial anticipatory behavior and their attention during the task in three different conditions: trials with auditory-only cues, visual-only cues, and multimodal cues. We first discuss the effect of multimodal versus unimodal cues on infants' general task attention. We measured whether infants looked at the appearing object on the center of the screen at the start of each trial and the amount of time that infants looked away during each trial.

Table 1 shows the results of the final GEE model for the two attention measures. For duration looking at the center, there was a significant main effect of trial ($\chi^2 [11] = 51.45, p < .001$), showing a general decrease in this behavior over trials. There was also a main effect of condition ($\chi^2 [2] = 11.79, p = .003$), with infants in the auditory-only condition looking at the center less than infants in the visual-only ($M_{\text{diff}} = -514.1, p = .001$) and ultimodal ($M_{\text{diff}} = -307.4, p = .05$) conditions. For looking-away, a significant main effect of trial was found ($\chi^2 [11] = 49.88, p < .001$), showing that infants had a general increase in the duration of time they spent looking away over trials. Additionally, a significant main effect of condition was found ($\chi^2 [2] = 11.37, p = .003$), with infants in the visual-only condition looking away significantly less than infants in the auditory-only ($M_{\text{diff}} = -540.0, p = .001$) and marginally less than infants in the multimodal condition ($M_{\text{diff}} = -322.5, p = .054$). In short, infants in the auditory-only condition had the lowest attention to the task, with more time spent looking away and a shorter duration of looking at the center than the other two conditions. Neither of the attention measures revealed main effects for or interactions with infants' age.

Because emergence location of the objects was not counterbalanced between infants, we tested whether emergence at one of the two sides was easier to learn. An ANOVA on the number of correct anticipations with stimulus location as a repeated measure and condition as a factor did not result in a significant main effect of stimulus location ($F [1, 75] = 0.147, p = .864$) nor in a significant interaction with condition ($F [2, 75] = 0.108, p = .744$).

Our measure of learning was whether infants anticipated correctly, incorrectly, or not at all. For this anticipation measure, the GEE model revealed a significant condition by trial interaction ($\chi^2 [22] = 34.73, p = .04$; see Table 2 and Figure 2). The auditory-only group did not significantly differ from either group. Analysis of the observed and the

TABLE 1
Full Model Effects for the Attention Measures

	<i>Looking to center</i>			<i>Looking away from screen</i>		
	<i>Wald-χ^2</i>	<i>df</i>	<i>p-value</i>	<i>Wald-χ^2</i>	<i>df</i>	<i>p-value</i>
Intercept	4437.67	1	<.001	207.02	1	<.001
Trial	51.45	11	<.001	49.89	11	<.001
Age	0.643	1	.42	0.002	1	.96
Condition	11.79	2	.003	11.37	2	.003
Trial*Condition	29.23	22	.14	29.23	22	.14
Age*Condition	3.17	2	.21	0.77	2	.68

TABLE 2
Full Model Effects for the Last Anticipation Measure

	<i>Last anticipation</i>		
	<i>Generalized-χ^2</i>	<i>df</i>	<i>p-Value</i>
Looking-away (nested in trial)	27.72	12	.006
Trial	18.32	11	.07
Age	0.05	1	.82
Condition	0.51	2	.77
Trial*Condition	34.73	22	.04
Age*Condition	2.06	2	.36

predicted response probabilities from the auditory-only condition showed that these infants were generally random in their behavior. Therefore, the differences between the visual-only and multimodal group were explored further. When only these groups were included, there was a condition by trial interaction (χ^2 [11] = 28.08, p = .003) with infants in the visual-only condition slightly more likely to anticipate correctly than infants in the multimodal condition.

Figure 3 shows the raw response probabilities for each trial which were also analyzed to identify if infants made a correct anticipation above-chance (= .33). Response probabilities that were more than 2.5 *SEs* from .33 were considered different from chance. Because the previous analysis did not reveal systematic or significant differences between age groups, the ages were collapsed in this analysis. The visual-only group showed above-chance correct anticipations on trials 3 and 4, then again on 8, 9, 10 and 12. On the other trials, the below-chance correct responding was complemented by an above-chance probability of not anticipating. This indicates that when they were not anticipating correctly, infants were off task instead of anticipating incorrectly. The multimodal infants showed above-chance responding on trial 5 through 9 and 12. As with the visual-only group, the below-chance probability of making a correct prediction on the other trials was complemented by an above-chance probability of not anticipating (Figure 3).

We also explored individual differences by looking at each infant's anticipations during the second half of the experiment. Within these six trials, in the auditory group, six infants did not once anticipate the reemergence of the stimulus, while in the multimodal and visual-only groups, all infants made at least one correct anticipation. Further, in the auditory-only group, six of 22 infants anticipated correctly more often than incorrectly, three had an equal number of correct and incorrect anticipations and 13 anticipated incorrectly most of the trials. In the

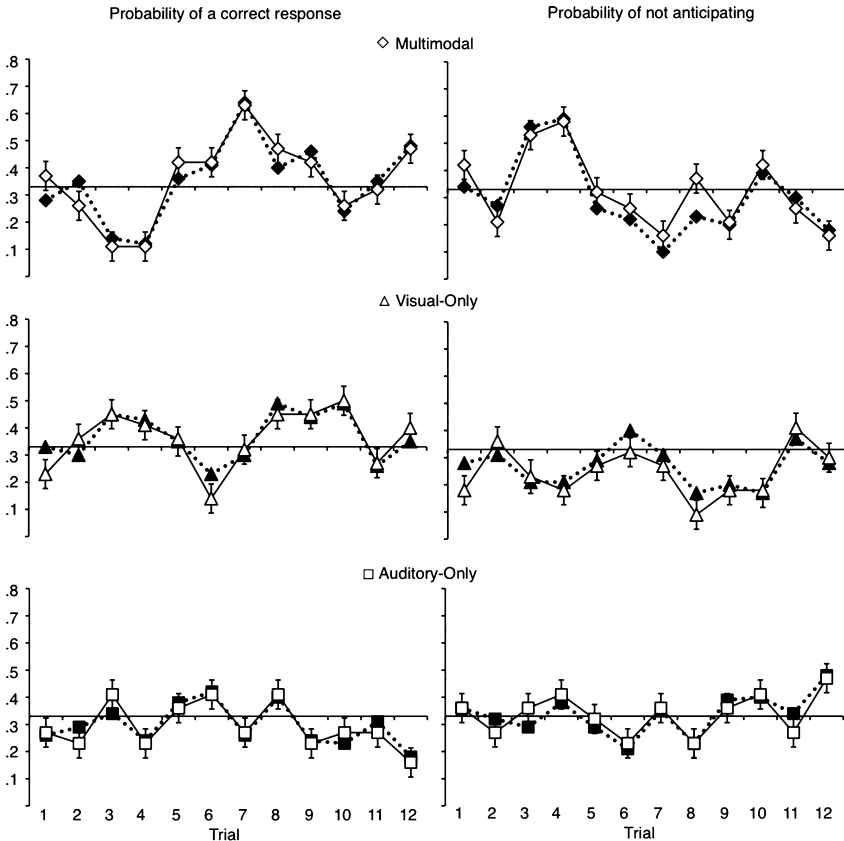


Figure 3 Comparison between the groups collapsed across age for the probability of a correct response and the probability of not anticipating. The solid symbols with the dashed lines show the predicted results from the final GEE model. These results control for individuals, age, and for the amount of time the infant looked away from the screen on the trial. The open symbols with the solid lines show the observed data. Error bars on the observed data are ± 2.5 standard errors of the multinomial distribution. The horizontal line represents chance level responding. GEE, Generalized Estimating Equations.

multimodal group, 11 of 19 infants made more correct than incorrect anticipations, while six had an equal number and two had more incorrect anticipations. In the visual-only group, 10 of 22 infants anticipated correctly during most trials, four had an equal number and eight had more incorrect anticipations. A chi-square test on these distributions yielded a significant difference between conditions ($\chi^2 [4] = 10.560, p = .032$), but no

differences between multimodal and visual-only groups ($\chi^2 [2] = 3.849$, $p = .146$). A chi-square test on the same measures during the first half of the experiment revealed no differences between conditions ($\chi^2 [4] = 0.063$, $p = 1$). Taken together, these results suggest that our test reveals evidence of differences in learning as a function of input modality: Multimodal and visual information (but not auditory information) were effective in facilitating learning of the object's emergence location, in particular during the latter half of experimental trials.

DISCUSSION

An important skill infants need to acquire is to predict the behavior of a stimulus on the basis of its features so that they can quickly react upon potential danger or allocate cognitive resources to what is most relevant in a particular situation. The present study set out to investigate how infants learn stimulus–location associations depending on whether they are exposed to unimodal (auditory-only or visual-only) or multimodal (auditory and visual) information. The formation of these associations was tested with the AEM paradigm (McMurray & Aslin, 2004), which is an ideal paradigm to measure how learning unfolds on a trial-by-trial basis (Mandell & Raijmakers, 2012). The synchronized presentation of visual and auditory information linked the information from two modalities to one multimodal compound stimulus. However, the amodal component of the multimodal stimulus did not cue the reappearance location of the stimulus. Hence, the stimulus information that infants could use for learning the stimulus-side association was modality-specific in all three conditions (Bahrick & Lickliter, 2000).

Using the AEM-paradigm, we were able to assess learning in a two-choice context. Instead of longer looking to a novel item as compared to a prefamiliarized item, all infants were exposed to stimuli that would either reappear left or right on the basis of their visual and/or auditory features. The relevant behavior, anticipating to the right or the left, is equally difficult between conditions and does not suffer from a criterion effect. That is, the relevant behavior does not depend on the judgement whether stimuli are different from each other. Infants who simply look at the screen and attend to the most dynamic components on the screen have not been included in our measure of learning; only infants who choose to look at the relevant portion of the screen at the relevant time window—when there is no dynamic event happening at that location at that moment—provide data on a trial. In this way, we can be relatively sure that infants included in our analyses provide meaningful data, although of

course it is possible that infants sometimes look at one of the anticipation locations by chance. The way learning behavior could be different between conditions is twofold: associations can be learned earlier or associations can be consistent over a larger number of trials.

We found clear signs of learning for infants in the visual-only and multimodal conditions, with both groups able to anticipate the reemergence of a visual object correctly within 12 trials regardless of age. In contrast, no such learning was observed in the auditory-only group. The attention measures showed that the auditory-only group was significantly less attentive on the task than the other two groups. For this group, the same visual stimulus (a circle) was used throughout the whole experiment, which may have rendered the visual component of the task too simple. If stimuli are too simple or too complex, infants have a high probability of looking away (Kidd, Piantadosi, & Aslin, 2012). Not learning the location association in this condition might therefore be explained by saying that infants only learn the association if they attend to the screen for a sufficient amount of time. Our results for the auditory-only condition did not replicate Albareda-Castellot et al.'s (2011) successful discrimination of auditory stimuli. The visual stimuli that were used in that study were attractive faces of cartoon figures (e.g., Elmo faces) that occasionally changed, while we used the same simple circle stimulus for all auditory-only trials. Infants in Albareda-Castellot et al.'s (2011) study looked at a minimum of 18 trials instead of our cut-off point of 12. Thus, in our study, the invariant visual stimulus could have resulted in low task attention, so that infants may have not performed well because they looked at such a small number of trials that they were unable to learn the location-sound association. One might argue that the invariant visual stimulus in our study paired with two auditory stimuli confused the infants, resulting in random behavior. This explanation seems unlikely, however, given the relatively short duration that the infants spent processing the visual stimulus at the beginning of each trial.

Because we assessed learning on a trial-by-trial basis, a detailed evaluation of the differences in learning curves between conditions was possible. The AEM-paradigm revealed divergent learning curves between the visual-only and the multimodal groups. The visual-only group had above-chance correct anticipations within the first three trials, and therefore appeared to learn the associations faster than the multimodal group, who did not show a higher-than-chance probability to anticipate correctly until trials 5 or 6. Yet, infants presented with multimodal information predicted the object reemergence for five consecutive trials, while infants exposed to visual-only information as a group had more sporadic behavior. The discriminating features in the visual-only condition seemed to be processed earlier during

the learning process, such that the location association was also learned earlier.

Multimodal information seems to have sustained infants' correct anticipations for longer intervals than visual-only information, which also suggests that the multimodal information heightened attention or engagement in the task and consequently improved task behavior. This is compatible with the ideas from Plunkett (2010) and the Intersensory Redundancy Hypothesis (Bahrick & Lickliter, 2000, 2012; Bahrick et al., 2004): the synchrony between auditory and visual components in the multimodal condition captured infants' attention. Neurological evidence supports the idea that multimodal information enhances processing: In an EEG-study, Hyde et al. (2010) find increased auditory processing under multimodal compared to unimodal stimulus presentation when the visual component was factored out. In a similar set-up, Reynolds et al. (2014) report enhanced processing of synchronous multimodal stimulation as compared to asynchronous or unimodal stimulation in 5-month-old infants. In our study, multimodal presentation slowed down learning, however, probably because reappearance location in this task is inherently a modality-specific, namely visual, feature.

Our findings are not compatible with the auditory dominance hypothesis raised by Robinson and Sloutsky (2004, 2010). The multimodal group's higher consistency suggests that multimodal information did have a positive influence on infants' learning. However, our results provide no evidence for the hypothesis that auditory labels facilitate learning, in the sense that associations are learned more easily (Ferry et al., 2010; Plunkett et al., 2008; Waxman & Booth, 2003; Waxman & Braun, 2005). Instead, multimodal (relatively complex) information seems to have helped capture infants' attention, resulting in the greater behavioral consistency for this group. In the present study, infants in the multimodal group paid attention to the stimuli longer than infants in the other groups, and therefore had more stimulus exposure, which could have led to their more consistent anticipatory behavior.

The findings of Reynolds et al. (2014) support this idea: their EEG-study with 5-month olds found that the Nc-component associated with attentional salience was largest in infants presented with multimodal synchronous information as compared to infants presented with the same events without intersensory redundancy. Further work is required to test whether increased attention to the stimuli is indeed the crucial factor in learning the associations. It is expected that a more complex or varying visual stimulus would improve attention for infants in our auditory-only condition (Kidd et al., 2012; Reynolds, Bahrick, Lickliter, & Guy, 2014; Reynolds, Zhang, & Guy, 2013), and consequently would result in more anticipations to the

correct reappearance location. A more complex visual stimulus might also result in a better learning environment for infants in the visual-only condition, keeping them interested for longer consecutive trials.

We set out to study the influence of multimodal versus unimodal information on infants' attention and learning of stimulus–location associations. Our combination of the AEM paradigm and GEE analysis revealed that unimodal visual information was simplest to discriminate, which led to fast learning of associations, but also gave rise to lower attention than multimodal information. Multimodal information took longer to process, but led to sustained task engagement, which had a positive effect on the consecutive number of correctly anticipated stimuli of infants in this group as compared with infants in the visual-only group. These findings suggest that multimodal synchronous stimuli are interpreted as a more reliable source of information for orienting behavior than unimodal stimuli.

ACKNOWLEDGMENTS

This research was funded by a grant from the priority program Brain & Cognition of the University of Amsterdam. PE's work was also supported by NWO grant 277-70-008 awarded to Paul Boersma. MEJR's and DJM's work was supported by NWO-VIDI grant 452-06-008 to MEJR. Infant testing was conducted at SPJ's baby lab at UCLA, funded by NIH grants R01-HD40432 and R01-HD73535 awarded to SPJ.

REFERENCES

- Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *2*, 395–401.
- Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41*, 352–363.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190–201.
- Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–205). Oxford, UK: Oxford University Press.
- Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102.
- Bremner, J. G., Slater, A. M., Mason, U. C., Spring, J., & Johnson, S. P. (2013). Trajectory perception and object continuity: Effects of shape and color change on 4-month-olds' perception of trajectory identity. *Developmental Psychology*, *49*, 1021–1026.

- Casasola, M., & Cohen, L. B. (2000). Infants' association of linguistic labels with causal actions. *Developmental Psychology, 36*, 155–168.
- Ferry, A. L., Hespous, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development, 81*, 472–479.
- Gredebäck, G., Johnson, S. P., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*, 1–19.
- Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development, 76*, 598–613.
- Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology, 52*, 181–189.
- Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences USA, 100*, 10568–10573.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7*, e3699, 1–8.
- Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 155–171.
- Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 172–182.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*, 281–308.
- Mandell, D. J., & Raijmakers, M. E. J. (2012). Using a single feature to discriminate and form categories: the interaction between color, form and exemplar number. *Infant Behavior and Development, 35*, 348–359.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.
- McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*, 203–229.
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*, 191–196.
- Plunkett, K. (2010). The role of auditory stimuli in infant categorization. In L. M. Oakes, C. H. Cahson, M. Casasola, & D. H. Rakison (Eds.), *Infant perception and cognition: Recent advances, emerging theories, and future directions* (pp. 203–221). New York, NY: OUP.
- Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665–681.
- Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America, 100*, 577–592.
- Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5-month-old infants. *Developmental Psychobiology, 56*, 355–72.
- Reynolds, G. D., Zhang, D., & Guy, M. W. (2013). Infant attention to dynamic audiovisual stimuli: Look duration from 3 to 9 months of age. *Infancy, 18*, 554–577.
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.
- Robinson, C. W., & Sloutsky, V. M. (2010). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology, 107*, 351–358.

- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.
- Waxman, S. R., & Booth, A. E. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, *6*, 128–135.
- Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, *95*, B59–68.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*, 1289–1309.
- Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*, 125–166.
- Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data-analysis for discrete and continuous outcomes. *Biometrika*, *42*, 121–130.
- Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrika*, *44*, 1145–1156.