



Published in final edited form as:

Lang Learn Dev. 2009 July 1; 5(3): 172–190. doi:10.1080/15475440902754086.

Infants' recognition of meaningful verbal and nonverbal sounds

Alycia Cummings^{1,2}, Ayse Pinar Saygin^{2,3}, Elizabeth Bates², and Frederic Dick^{2,4}

¹ San Diego State University

² University of California – San Diego

³ University College London

⁴ Birkbeck College, University of London

Abstract

To examine how young children recognize the association between two different types of meaningful sounds and their visual referents, we compared 15-, 20-, and 25-month-old infants' looking time responses to familiar naturalistic environmental sounds, (e.g., the sound of a dog barking) and their empirically matched verbal descriptions (e.g., “*Dog barking*”) in an intermodal preferential looking paradigm. Across all three age groups, performance was indistinguishable over the two domains. Infants with the largest vocabularies were more accurate in processing the verbal phrases than the environmental sounds. However, after taking into account each child's verbal comprehension/production and the onomatopoeic test items, all cross-domain differences disappeared. Correlational analyses revealed that the processing of environmental sounds was tied to chronological age, while the processing of speech was linked to verbal proficiency. Overall, while infants' ability to recognize the two types of sounds did not differ behaviorally, the underlying processes may differ depending on the type of auditory input.

INTRODUCTION

What is the relationship between the development of speech and language comprehension and the comprehension of other complex auditory signals? Previous developmental studies have addressed this intriguing question in a variety of manners and modalities (e.g., Hollich, Hirsh-Pasek, Golinkoff, Brand, Brown, Chung, Hennon, and Rocroi, 2000; Gogate & Bahrack, 1998; Gogate, Bolzani, & Betancourt, 2006; Gogate, Walker-Andrews, & Bahrack, 2000), though few have directly compared how infants recognize sound-object associations for nonverbal sounds vs. language. Rather, most have contrasted infants' preference for speech and non-speech sounds. For instance, even at a young age (2.5–6.5 months) infants show a preference for natural speech sounds over synthetic sine-wave speech (Vouloumanos & Werker, 2004); nine-month-old infants also appear to show a preference for the human voice, as opposed to similar sounds conveyed by musical instruments (Glenn, Cunningham, & Joyce, 1981). However, these studies do not directly address the infant's use and understanding of speech as a means of conveying semantic information. This is important to understand since the ability to recognize the association between a sound and/or word and an object is a likely precursor to word comprehension, as knowing that a word can “stand for” an object is the first step towards developing language (Golinkoff, Mervis, Hirsh-Pasek 1994).

Indeed, studies examining the development of intermodal semantic associations suggest that for much of the second year of life, infants will accept gestures, spoken words, and nonverbal sounds as possible object associates when these are presented in an appropriate context. Early in development, perceptual saliency can assist young infants' (7- to 12-month-olds) associative abilities (Bahrick, 1994; Gogate & Bahrick, 1998; Gogate, Bolzani & Betancourt, 2006; Hollich et al., 2000). However, sometime after their first birthday, infants start to attend to and use social and referential cues for symbolic associations. For instance, Hollich and colleagues (2000) and Roberts (1995) found that 15-month-old infants developed a categorization bias for novel objects when previous object presentations were predictably paired with either instrumental music or speech sounds. Similarly, results of Namy (2001), Namy and Waxman (1998, 2002), and Namy, Acredolo, and Goodwyn (2000) have shown that pictograms, gestures, and nonverbal sounds were equally likely to be learned as labels for object categories at 17 or 18 months. In particular, Campbell and Namy (2003) showed that both at 13 and 18 months, toddlers were equally likely to recognize associations between novel objects and either novel words or novel, relatively complex artificial sounds (e.g., a series of tone pips), when either type of sound was linked to social and referential cues during learning. However, neither words nor nonverbal sounds were associated with the objects when such social or referential cues were not present during learning – see also Balaban and Waxman (1997) and Xu (2002) for instances of infants not categorizing objects in the absence of referential cues.

As children get older, they become less likely to accept a non-linguistic label for an object than they are a word. For instance, Woodward and Hoyne (1999) found that 13 month-old infants would learn a novel association between a nonverbal label (an artificially produced sound) and an object, but 20-month-old infants did not; Namy & Waxman (1998) found similar trends in gestural comprehension between 18 and 26 months. Likewise, in gesture production, Iverson, Capirci, and Caselli (1994) found a greater acceptance of verbal rather than gestural labels between 16–20 months (see also Volterra, Iverson, & Emmorey, 1995).

Undoubtedly, there are multiple factors driving the emergence of this greater acceptance of spoken words as labels (for discussion and review, see Volterra, Caselli, Capirci, & Pizzuto, 2005). However, it is unclear whether this greater acceptance of spoken words holds when words are compared with natural, complex, and frequently occurring sounds that convey meaningful information, such as environmental sounds. Both word-object and environmental sound-object associations occur frequently in everyday situations (e.g., Ballas and Howard, 1987), and this frequency of occurrence may be important for learning meaningful sound-object relationships. However, the relationship between objects and their associated verbal and environmental sound labels are somewhat different. Environmental sounds can be defined as sounds generated by real events – for example, a dog barking, or a drill boring through wood - that gain sense or meaning by their association with those events (Ballas & Howard, 1987). Thus, individual environmental sounds are typically causally bound to the sound source or referent, unlike the arbitrary linkage between a word's pronunciation and its referent. Environmental sounds might be considered *icons* or *indexes*, rather than true *symbols*.

There is wide variation in how environmental sounds are generated. They can be produced by live things (e.g., dogs, cats), they can be produced when animate things act upon them (e.g., playing an instrument, using a tool), and inanimate objects are also able to produce a sound on their own without the intervention of an animate being (e.g., water running in a river, alarm clock ringing). With the advent of electronic toys, books, and other media, the linking of environmental sounds to their source has become somewhat more abstract; for instance, it is much more likely for a child in San Diego to hear the mooing of a cow from a picture-book or video than on a farm.

Studies with adult subjects have shown that, like words, processing of individual environmental sounds is modulated by contextual cues (Ballas & Howard, 1987), item familiarity, and frequency (Ballas, 1993; Cycowicz & Friedman, 1998). Environmental sounds can prime semantically related words and vice versa (Van Petten & Rheinfelder, 1995) and may also prime other semantically related sounds (Stuart & Jones, 1995; but cf. Chiu and Schacter, 1995 and Friedman, Cycowicz, & Dziobek, 2003, who showed priming from environmental sounds to language stimuli, but no priming in the reverse direction). Spoken words and environmental sounds share many spectral and temporal characteristics, and the recognition of both classes of sounds breaks down in similar ways under acoustical degradation (Gygi, 2001). Finally, Saygin, Dick, Wilson, Dronkers, and Bates (2003) have shown that in adult patients with aphasia, the severity of patients' speech comprehension deficits strongly predicts the severity of their environmental sounds comprehension deficits.

Environmental sounds also differ from speech in several fundamental ways. The 'lexicon' of environmental sounds is small and semantically stereotyped; these sounds are also not easily recombined into novel sound phrases (Ballas, 1993). There is wide individual variation in exposure to different sounds (Gygi, 2001), and correspondingly, healthy adults show much variability in their ability to recognize and identify these sounds (Saygin, Dick, & Bates, 2005). Finally, the human vocal tract does not produce most environmental sounds. In fact, the neural mechanisms of non-linguistic environmental sounds that can and cannot be produced by the human body may differ significantly (Aziz-Zadeh, Iacoboni, Zaidel, Wilson, & Mazziotta, 2004; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005; Pizzamiglio, Aprile, Spitoni, Pitzalis, Bates, D'Amico, & Di Russo, 2005).

Recognizing the association between environmental sounds and their paired objects or events appears to utilize many of the same cognitive and/or neural mechanisms as those associated with a verbal label, especially when task and stimulus demands are closely matched (reviewed in Ballas, 1993; Saygin, et al., 2003; Saygin et al., 2005, Cummings, Čeponienė, Dick, Saygin, & Townsend, 2008; Cummings, Čeponienė, Koyama, Saygin, Townsend, & Dick, 2006). Indeed, spoken language and environmental sounds comprehension appear to show similar trajectories in typically developing school-age children (Dick, Saygin, Paulsen, Trauner, & Bates, 2004; Cummings et al., 2008), as well as in children with language impairment and perinatal focal lesions (Cummings, Čeponienė, Williams, Townsend, & Wulfeck, 2006; Borovsky, Saygin, Cummings & Dick, in preparation).

The acoustical and semantic similarities between environmental sounds and speech sounds discussed above, and the findings of previous studies have led us to explore the relationship between the early development of speech processing and that of environmental sounds. Previous studies have shown an emerging preference for speech compared to other complex sounds, but these sounds were not naturally meaningful or referential. While it has been shown that adults and school age children may use similar resources for processing meaningful environmental and verbal sounds, it is still possible that distinct mechanisms may direct the recognition of the sound-object associations between these sounds early in development.

There is some evidence from early development that recognizing sound-object associations may rely on domain-general perceptual mechanisms. For example, Bahrack (1983, 1994) demonstrated how the perception of amodal relations in auditory-visual events could serve as a building block for the detection of arbitrary auditory-visual mappings. Infants at 3 months of age were able to identify color/shape changes in objects and pitch changes in sound, but it was not until 7 months of age that the infants were able to identify the arbitrary relationship between objects and sounds. Thus it is possible that by first attending to invariant amodal relations, infants learn about arbitrary intermodal relationships (Bahrack, 1992). Alternatively, there has also been recent evidence that infants are better able to learn non-linguistic

information by first learning about speech (Marcus, Fernandes, & Johnson, 2007, but see Dawson & Gerken, 2006; Saffran, Pollack, Seibel and Shkolnik (2007)).

Thus, it is also important not only to examine the links between the *receptive* processing of verbal and environmental sounds and overall (chronological) development, but between receptive abilities and the development of the child's *production* of verbal sounds. Although no previous study has examined the relationship of environmental sounds and productive vocabulary size, previous looking-while-listening studies have reported positive relationships between looking accuracy and productive vocabulary (e.g., Conboy & Thal, 2006; Zangl, Klarman, Thal, Fernald, Bates, 2005). Following a large sample of children longitudinally, Fernald, Perfors, and Marchman (2006) observed that accuracy in spoken word recognition at 25 months was correlated with measures of lexical and grammatical development gathered from 12 to 25 months. Moreover, Marchman and Fernald (2008) observed that infants' vocabularies at 25 months are strongly related to expressive language, IQ, and working memory skills tested at 8 years of age. Thus, a link between vocabulary size and environmental sound-object associations might suggest that an understanding of sound-object relationships, verbal or nonverbal, is linked to more general language and cognitive development. Alternatively, we may see that productive vocabulary growth is only linked to increases in language comprehension, suggesting that verbal and nonverbal sound processing might be differentiated to some degree over early development.

Here, by testing infants before, during and after the vocabulary spurt, we examined whether infants are able to recognize associations to an object for verbal descriptions and environmental sounds similarly, and whether this recognition would shift in response to dramatic gains in language abilities due to the vocabulary spurt. In particular, we designed our study to address the following questions, motivated by the literature summarized above: a) Are infants and toddlers better able to recognize the association between pictures of objects and speech labels, as compared to pictures of objects and non-speech sounds, even when the sounds are naturalistic and meaningful?; b) If there is no initial difference in the matching of meaningful speech sounds and environmental sounds to associated pictures, will one emerge around 18–20 months of age (or later), as suggested by the literature discussed above?; and c) Will a child's recognition of environmental sounds and spoken words be predicted by chronological age, productive vocabulary size, or some combination of the two?

To test infants' ability to recognize sound-object associations, we used an intermodal preferential looking paradigm. We played either an environmental sound (e.g., the sound of a dog barking), or the spoken language gloss of the sound (the phrase *Dog Barking*) and measured infants' looking to matching and mismatching pictures (in this case, a matching picture would be a dog, a mismatching picture would be a piano). Based on the results of previous studies of early linguistic and non-linguistic acquisition and development briefly presented above, we might expect younger (15- and 20-month-old) infants to process environmental sounds and spoken language similarly, whereas the oldest and/or most linguistically advanced children (e.g., 25-month-olds or infants with very high productive vocabularies) would process the two types of sounds differently, more readily recognizing language sounds than environmental sounds. Alternatively, given the results of environmental sounds/language comparisons in older children and adults, the trajectories of auditory linguistic and non-linguistic processing might go hand-in-hand across this age span.

METHODS

Participants

Participants were 60 typically developing, English-learning infants of three ages: 20 15-month-old children (9 female), 20 20-month-old children (11 female), and 20 25-month-old children

(12 female). Parents of the infants were asked to fill out the MacArthur-Bates Communicative Development Inventory (CDI; Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). The Words and Gestures version was used for the 15-month-olds, while the Words and Sentences form was given to the parents of the 20- and 25-month-olds. The Words and Gestures form tracks both infants' comprehension and production of words, but that form is only valid up to the age of 17 months. The Words and Sentences form is valid from 16 to 30 months, but it only tracks productive vocabulary (due to the difficulty parents would have in keeping track of the hundreds and thousands of words older infants know). While it would have been ideal to know all our infant subjects' *receptive* vocabularies, our vocabulary assessment instruments limited us to only using productive vocabulary measures. Verbal proficiency groups were defined by dividing the entire infant sample into thirds based on their productive vocabulary: Low (< 50 words), Mid (51–261 words) and High (> 262 words)¹.

Experimental Design and Materials

There was a single within-subjects factor, Domain (Verbal or Environmental Sounds); this was crossed with the between-subjects factor Age Group (15-, 20-, 25-month-old) or CDI Verbal Proficiency (Low, Medium, High). The experiment contained three types of stimuli: full-color photographs, nonverbal environmental sounds, and speech sounds. All of the sound tokens, nonverbal and verbal, were presented after the attention-grabbing carrier phrase, "Look". The duration of a single trial - composed of carrier phrase, sound, and intervening silences - was ~4000 msec (see Supplemental Web Materials -<http://crl.ucsd.edu/~saygin/soundsdev.html> for individual sound durations and sample stimuli). All auditory stimuli were digitized at 44.1 kHz with a 16-bit sampling rate. The average intensity of all auditory stimuli was normalized to 65 dB SPL using the Praat 4.4.30 computer program (Boersma, 2001). All sound-picture pairs were presented for the same amount of time.

Nonverbal Environmental Sounds—The sounds came from many different nonlinguistic sources: animal cries (n=8, e.g. cow mooing), human nonverbal vocalizations (n=4, e.g. sneezing), vehicle noises (n=2, e.g. airplane flying), alarms/alerts (n=3, e.g. phone ringing), water sounds (n=1, toilet flushing), and music (n=2, e.g. piano playing) – see Appendix for a complete list of stimuli. All environmental sound trials were edited to 4000ms in length, including the carrier phrase, "Look!". The average length of the environmental sound labels was 3149 ms (Range 1397–4000 ms), including 964 ms for the "Look!" and its subsequent silence prior to the onset of the environmental sound. In order to extend the files to the required 4000 ms, 11 of the environmental sound labels had "silence" appended to the end of the sound files. The addition of the silence standardized the amount of time the infants were exposed to each picture set (2000 ms prior to the onset of the sound, followed by 4000 ms of auditory input), even if the amount of audible auditory information varied across trials.

Sounds were selected to be familiar and easily recognizable based on extensive norming studies (detailed in Saygin et al., 2005), and a novel 'sound exposure' questionnaire that was completed by the children's parents. This questionnaire asked parents to rate how often their child heard the environmental sounds presented in the study (e.g., every day, 3–4 times/week, once a week, twice a month, once a month, every 3–4 months, twice a year, yearly, or never). Because there was little consistency in parents' responses (perhaps due to problems with the survey itself, as

¹The parents of four of the infants (2 15-month-olds and 2 20-month-olds) did not return the CDIs. One 20-month-old's mother did keep track of his vocabulary and stated that he was producing more than 400 words at the time of testing, so he was informally placed in the High vocabulary group. The parents of the two 15-month-olds reported that their infants only spoke a few words, so those two children were informally grouped into the Low vocabulary group. The last 20-month-old child was reported to speak more than 50 words, but was not overly verbal, thus she was informally placed into the Mid vocabulary group. These four children were not included in the analyses that examined the production of specific words (see Results for more detail).

well as with differences in kind and degree of exposure to sounds), we did not have enough confidence in these data to use them as covariates.

Verbal Labels—Speech labels were derived from a previous norming study, Saygin et al. (2005), and were of the form ‘*noun verb-ing*’ (see Appendix). Labels were read by an adult male North American English speaker, digitally recorded, and edited to 4000ms in length, including the carrier phrase, “Look!”. The average duration of the verbal labels was 2431 ms (Range 2230–2913 ms), including 964 ms for the “Look!” and its subsequent silence prior to the onset of the verbal label. In order to extend the files to the required 4000 ms, all of the verbal labels had “silence” appended to the end of the sound files. Fifteen (of 20) of these labels were matched to items on the MacArthur-Bates CDI, allowing us to calculate whether each label was in a child’s receptive or productive vocabulary.

Visual stimuli—Pictures were full-color, digitized photos (280 × 320 pixels) of common objects corresponding to the auditory stimuli. Pictures were paired together so that visually their images did not overlap, thus lessening the potential confusion infants would have with two perceptually similar items (e.g., two animals with four feet: dog vs. cat). These pairings remained the same throughout the experiment, and were the same for all infants. Trials were quasi-randomized, where all twenty targets were presented once (either with a word or an environmental sound) before being repeated (presented with the other auditory stimulus). Additionally, each picture was presented twice on each side; as the “target” picture, it was the target once on the left side and once on the right side to control for looking side bias. Over the course of the experiment, each picture appeared four times in a counterbalanced design: picture type (target/distracter) × domain (verbal/nonverbal).

Procedure

This experiment was run on a PC using the Babylab Preferential Looking software, custom-developed by the Oxford Babylab (Oxford, England). The participants sat on their mothers’ laps in a darkened room in front of a computer booth display (Figure 1). The two computer monitors, 90cm apart, and the loudspeaker were placed in the front panel of the booth for stimuli presentation. The infants sat 72cm away from the monitors, which were placed at infant eye level. A small video camera placed between the monitors recorded subjects’ eye movements. The camera was connected to a VCR in the adjacent observation area, where the computer running the study was located.

The experiment consisted of 47 experimenter-advanced trials (Figure 2). In each trial, infants were presented with two pictures on the two computer screens. After 2000ms, the sound stimulus (either verbal or nonverbal) was presented through the loudspeaker. The infants’ looking task was to visually match each verbal or environmental sound to one of two pictures presented on computer screens in front of them. For example the target picture “dog” appeared twice: once with a verbal sound (*‘dog barking*), and once with a nonlinguistic sound (the sound of a dog barking). Overall, twenty pictures were targets, for a total of 40 test trials.² An additional 7 attention-getting trials (*Hey, you’re doing great!*, paired with “fun” pictures, such as Cookie Monster and Thomas the Train) were interspersed throughout the experiment to keep infants on task.

²The present study used many more different stimuli pairs than most preferential looking studies. However, the large number of items did not necessarily decrease the infants’ ability to semantically map the auditory stimuli onto the visual stimuli. For instance, Killing and Bishop (2008) used a bimodal preferential looking paradigm, similar to that used in the present study. They presented their infants (20 to 24 months) with 16 different pairs of items, which were also semantically and visually distinct (e.g., comb vs. star, tongue vs. boot, clock vs. bee, etc.), and demonstrated interpretable results, as in the present study.

During the session a digital time code accurate to 33ms was superimposed on the videotape. Videos were digitized by Apple iMovie 2.1.1, converted to Quicktime 6.2 movie files, and coded using custom-designed software (from the Stanford University Center for Infant Studies) on a Macintosh G3 by a highly trained observer, blind to the side of the target. Each session was analyzed frame-by-frame, assessing whether the child was looking at the target picture, distracter picture, or neither picture. Accuracy was measured by calculating the amount of time in each time window an infant looked at the target picture divided by the total amount of time the child fixated on both the target and distracter pictures. Time windows were defined as Early (533–1500msec) and Late (1533–2500 msec), with time measured from the onset of each phrase or environmental sound (e.g. Fernald, Perfors, & Marchman, 2006; Houston-Price, Plunkett, & Harris, 2005; see Figure 3). In other words, the analysis windows did not include the carrier phrase, “Look!”. Rather, they began at the *onset* of the verbal label (e.g., “dog barking”) or the environmental sound (e.g., the sound of a dog barking). Data from the first 500 ms of each environmental sound and verbal label are not reported as this time window did not show significant differences between modalities or age groups. (It is standard in infant preferential looking paradigms to disregard at least the initial 367 ms of an analysis window due to the amount of time infants need to initiate eye movements (e.g., Fernald, Pinto, Swingley, Weinberg, McRoberts, 1998; Fernald et al., 2006)).

Two different sets of mixed-effects analyses of variance (ANOVA) were performed to examine looking time accuracy: 1) a 2-within-subjects (Domain (Verbal/Nonverbal) and Window (Early/Late)) by 1-between-subjects (Age (15/20/25)) ANOVA; and 2) a 2-within-subjects (Domain (Verbal/Nonverbal) and Window (Early/Late)) by 1-between-subjects (CDI Verbal Proficiency (Low/Mid/High)) ANOVA. Due to the small number of varied stimulus items, ANOVAs were carried out with both subjects (F1) and items (F2) as random factors. When appropriate, Bonferroni corrections were applied.

The predictability of infants’ performance on the task was also tested with multiple regression using age and productive vocabulary as predictors of looking accuracy. All estimates of unique variance were derived from adjusted r^2 .

RESULTS

Looking Time Results

Age Group Analysis—Accuracy significantly improved with age ($F(2,57) = 12.84, p < .0001$; $F(2,38) = 19.50, p < .0001$; Table 1). Pre-planned group comparisons using ANOVA revealed that both the 20- and 25-month-old infants performed significantly better than the 15-month-olds (20m: $F(1,38) = 9.71, p < .004$; 25m: $F(1,38) = 25.62, p < .0001$), while not being statistically different from each other. Infants’ accuracy performance also improved significantly from the Early to Late Window ($F(1,57) = 94.88, p < .0001$; $F(1,19) = 42.25, p < .0001$). No main effect of Domain (i.e., verbal or environmental sound labels) was observed. No other accuracy effects were significant.

Verbal Proficiency Group Analysis—Infants’ accuracy improved as their verbal proficiency grew ($F(2,57) = 17.53, p < .0001$; $F(2,38) = 18.6, p < .0001$; Table 1). Pre-planned group comparisons using ANOVA revealed that infants in both the Mid and High verbal proficiency groups were more accurate than those in the Low group (Mid: $F(1,38) = 27.45, p < .0001$; High: $F(1,38) = 27.20, p < .0001$), while not differing from each other. Infants’ accuracy also improved from the Early to Late time window ($F(1,57) = 93.76, p < .0001$; $F(1,19) = 42.44, p < .0001$). As with the Age group analyses, no main effect of Domain was observed.

Unlike age, there were differential effects of verbal proficiency on Verbal and Environmental Sounds looking accuracy, as shown by the Domain \times Verbal Proficiency interaction ($F(2,57) = 6.59, p < .003$; $F(2,38) = 3.38, p < .06$; Figure 4). Pre-planned group comparisons using ANOVA (Table 1) demonstrated that for Environmental Sound trials, the Low proficiency group's performance was significantly worse than the Mid group ($F(1,38) = 15.69, p < .0002$), and showed a strong trend for being less accurate than infants in the High group; the Mid and High groups did not differ from each other. Similarly in the Verbal trials, infants in the Low group were less accurate than those in both the Mid ($F(1,38) = 10.84, p < .002$) and High ($F(1,38) = 33.56, p < .0001$) groups, who did not differ from each other. Interestingly, when the individual proficiency groups were examined, only the High proficiency group showed significant processing differences for speech and environmental sounds ($F(1,19) = 21.55, p < .0002$), being more accurate in response to the Verbal labels. No other effects or interactions were significant.

A lack of item repetition across trials may have potentially influenced infants' performance in the task.³ Thus, a comparison to chance performance (i.e., 50% responding rate) is a prudent check to establish a difference between systematic and random responding (i.e., whether or not infants were recognizing the sound-object associations in the task), especially when comparatively young and nonverbal infants are involved. As expected, for the 25-month-old group, t-tests showed that group accuracy was significantly greater than chance in each time window and domain (Verbal Early: $t(1,19) = 4.35, p < .0004$; Verbal Late: $t(1,19) = 8.10, p < .0001$; Sounds Early: $t(1,19) = 4.89, p < .0002$; Sounds Late: $t(1,19) = 9.64, p < .0001$). The 20-month-old infants responded significantly greater than chance to the environmental sounds in both windows (Early: $t(1,19) = 5.21, p < .0001$; Late: $t(1,19) = 6.12, p < .0001$), as well as to the words in the Late window ($t(1,19) = 5.48, p < .0001$). The 15-month-olds' accuracy was also greater than chance for environmental sounds in the Late window ($t(1,19) = 4.76, p < .0002$) but not for the words in either window. Thus, the word stimuli, particularly in the Early time window, did not appear to give the younger infants (15-month-olds, 20-month-olds) enough information to begin systematically processing the presented pictures. However, as more of the verbal information was presented (i.e., the verb was spoken) and/or more time was allotted for the infants to process the information, their performance increased above chance levels. Indeed, while nouns are the more prevalent and salient word class in young infants' vocabularies (Fenson et al., 1994), the verbs in the phrases might have provided at least some of the infants with additional context to help them recognize the sound-object associations (e.g., *kiss* is comprehended by 52% of infants by 10 months of age; Dale & Fenson, 1996). In regards to the environmental sound stimuli, the youngest infants did perform at chance levels during the Early window. However, no other age group for either analysis window was at or below chance in their sound-object mapping of the environmental sounds onto the pictures.

In order to ascertain that these results were not simply due to the children not knowing the words in the linguistic stimuli, we also analyzed the data after re-calculating each child's mean looking accuracy based only on items whose verbal labels were in their individual receptive (15-month) and/or productive (15, 20, 25-month) vocabulary according to the MacArthur-Bates CDI (see Appendix for list of corresponding items). As in the analyses where mean accuracy was calculated over all items, there were no significant interactions of Domain with

³Relative to other preferential looking studies (e.g. Swingley, Pinto, & Fernald, 1999; Fernald et al., 1998) overall looking accuracy in our study was relatively low; this may have been driven by several factors. First, while many preferential looking studies test on a small set of items presented repeatedly, our experiment had forty stimuli, each presented once. Thus, higher accuracy observed in other studies may be in part driven by opportunities for stimulus learning over multiple trials. Additionally, because the two modalities were not separately blocked within the experiment, infants were not able to build up expectancies for the upcoming trial. The mixed trial design also may have decreased accuracy, since young children's performance can be modulated by stimulus predictability (Fernald & McRoberts, 1999).

Age Group at either time window, but there was an interaction of Domain by Verbal Proficiency Group ($F(2,50)=3.90, p < .03$)⁴.

Since onomatopoeic sounds are arguably closer to linguistic sounds than sounds that are not typically verbalized, this subset of sounds could make infants' looking to environmental sounds seem more similar to verbal sounds than it truly is. Therefore, we wanted to ensure that the pattern of results in mean verbal and nonverbal looking accuracy, in particular the similarities in the patterns, was not driven by the subset of animal-related environmental sounds that were paired with labels containing onomatopoeic verbs, like 'cow mooing', 'rooster crowing', etc. (8/20 experimental items – see Appendix for list). Thus we re-ran each ANOVA listed above, including only the 12 non-onomatopoeic items in the analyses. The only difference was in the CDI-based analyses: when the 8 onomatopoeic items were excluded, the interaction between Verbal Proficiency Group and Domain was no longer significant. Otherwise, the exclusion of these items did not change the direction or significance of any analyses, aside from an approximately 2–3% overall drop in accuracy.

Finally, because 8 of the 10 picture pairs pitted an animate versus an inanimate object (e.g. with the target being 'dog' and foil being 'piano', and vice versa), we wanted to make sure that simple animacy cues were not allowing children to distinguish between environmental sounds on the basis of animacy alone, as this would have confounded a proper comparison of environmental sound and verbal label comprehension. If indeed accuracy on the environmental sounds trials was being driven (even in part) by a difference in animacy in 8 of 10 target/foil pairs, then the infants' looking accuracy would have been correlated over the environmental sounds target/distracter pairs. For example, accuracy for the target 'piano' should positively predict target accuracy for 'dog', if infants are using animacy differences rather than, or in addition to, item knowledge to guide their sound/picture matching. Instead, we found no hint of a positive correlation between paired items. For each age group (20/25 months), and time window (Early/Late), correlations between environmental sounds items in a pair were all *negative* with Spearman's Rho values of: -0.7619 (20-month-old, Window 2); -0.4072 (20 month-old, Window 3); -0.3810 (25-month-old, Window 2); and -0.1557 (25-month-old, Window 3). Thus, these analyses give no indication that animacy differences were affecting infants' looking accuracy. (Note that we did not include the 15-month-old data here, as they were on average near chance levels – see Discussion).

Regression Analyses

Chronological age and productive vocabulary—We carried out a set of regressions to examine the predictive value of age and CDI productive vocabulary for verbal and nonverbal accuracy in Early and Late time windows. One might expect chronological age and productive vocabulary size to have roughly equivalent predictive value, given that children's lexicons typically grow as they get older. Indeed, the two measures showed a very strong positive association (adjusted $r^2 = .527, p < .0001$). However, as suggested by the ANOVAs above, young children's accuracy with verbal and environmental sounds was differentially predicted by these two variables when both age and productive vocabulary were entered into the regression model. Increases in environmental sounds accuracy were associated with greater chronological age in both Early (adjusted $r^2 = .140, p < .0026$) and Late (adjusted $r^2 = .1804, p < .0006$) time windows, whereas productive vocabulary scores had no significant predictive value by itself, or when it was added to chronological age in the regression model. In contrast to environmental sounds accuracy, increases in verbal accuracy were associated both with chronological age in Early (adjusted $r^2 = .1495, p < .0019$) and Late (adjusted $r^2 = .1963, p < .$

⁴The change in degrees of freedom from the original analyses is due to the fact that 4 subjects' parents did not return a completed CDI questionnaire, and the lack of sufficient items in some cells for 3 subjects.

0004) windows and with productive vocabulary (Early Window: adjusted $r^2 = .2401$, $p < .0001$; Late Window: adjusted $r^2 = .2944$, $p < .0001$). However, whereas productive vocabulary accounted for significant additional variance when entered after chronological age in the regression model in both Early (increase in adjusted $r^2 = .0789$, $p < .0136$) and Late (increase in adjusted $r^2 = .0905$, $p < .0071$) time windows, chronological age accounted for no significant additional variance when entered after productive vocabulary.

SUMMARY AND DISCUSSION

This study compared 15-, 20-, and 25-month-old infants' eye-gaze responses associated with the recognition of intermodal sound-object associations for verbal and environmental sound stimuli. Infants' looking time accuracy increased from 15- to 20-months of age, while no measurable improvement was noted between the 20- and 25-month-old infants. Looking time accuracy also improved with infants' reported verbal proficiency levels. Infants with very small productive vocabularies (under 50 words) were less accurate than children with larger vocabularies.

Based on earlier research and as described in the introduction, two different outcomes could have been expected when comparing verbal and nonverbal processing. There could be an early advantage for the nonverbal sounds, with a gradual preference for the verbal modality as children get older, as has been found in studies comparing gesture and language development in the first two years of life. Alternatively, we could observe similar patterns of processing, as evidenced by studies with older children and adults.

We found that the recognition of sound-object associations for environmental sounds and verbal sounds is quite similar for all infants, especially when examined in the late time window (1500–2500ms post-stimulus onset). We also observed that infants' ability to recognize these sound-object associations improves with increasing chronological age, especially for environmental sounds. Changes in performance for verbal sounds on the other hand are more closely tied to infants' productive vocabulary sizes (consistent with Conboy & Thal, 2006; Fernald et al., 2006). Thus, although the infants' overt behavior (i.e., directed eye gaze) might be indistinguishable as they process verbal and nonverbal sounds, it is likely that there are some differences in the underlying processes driving the infants' responses to the two types of sounds. The design of the present study, and the data collected, cannot shed further light on this issue but neuroimaging studies using methods such as event-related potentials (ERPs) or near-infrared spectroscopy (NIRS) may be useful in exploring potential underlying differences or similarities.

In the present study, we did not observe a pronounced shift towards better recognizing sound-object associations for verbal stimuli as children get older, although as noted before, children with the largest productive vocabularies did show greater accuracy for verbal labels than for environmental sounds. Most studies that observed such a shift were carried out in the context of learning new associations. For example, studies have compared infants' ability to associate novel objects with either new words (an experience quite familiar to children of this age group) or with artificial, arbitrary sounds (a relatively unnatural and unfamiliar experience; e.g., Woodward & Hoyne, 1999). The relationship between the nonverbal environmental sounds and visual stimuli in the present study was quite different: Infants were already familiar with most of the audiovisual associations in our stimuli. Thus, these associations were not arbitrary but were based on infants' prior real-world knowledge, with sensitivity to such meaningful associations increasing with age and presumably experience. Future research could examine whether infants learn new non-arbitrary associations between objects and novel/unfamiliar environmental sounds as quickly as verbal labels, or whether there are differences in the

processing of verbal and nonverbal stimuli in the context of learning new associations vs. using already learned associations.

While our results do not necessarily bear on the question of how infants initially acquire and learn words and the environmental sounds, they suggest that at this stage in development, children are capable of recognizing correct sound-object associations, regardless of whether it involves a word or meaningful nonverbal sound. More generally, infants can detect mismatches between heard and seen stimuli, and their ability to do so improves with age, as has been reported in previous studies on the perception of object-sound and face-voice relationships (e.g., Bahrick, 1983; Dodd, 1979; Hollich, Newman, & Jusczyk, 2005; Kuhl & Meltzoff, 1982; Kuhl & Meltzoff, 1984; Lewkowicz, 1996; Patterson & Werker, 2003; Walker-Andrews & Lennon, 1991).

It is possible that the frequency with which infants are exposed to verbal and environmental sounds, as well as to intermodal associations involving each sound type, may have affected their responses in the present study. Obtaining specific measurements of infants' exposure to sounds and intermodal associative situations is a vast and complicated undertaking (see Ballas & Howard (1987) who conducted a study of adults' exposure to different environmental sounds) and beyond the scope of this study. However, given the influence frequency of occurrence has on language learning and processing (Dick, Bates, Wulfeck, Aydelott Utman, Dronkers, & Gernsbacher, 2001; Matthews, Lieven, Theakston, & Tomasello, 2005; Roland, Dick, & Elman, 2007) it is very likely that frequency of occurrence does play a role in establishing both nonverbal and verbal sound-object associations.

As mentioned above, it appears that the children who are more sensitive to intermodal sound-object associations (and thus are presumably more attentive to the important distinguishing characteristics in their world) also have larger vocabularies. As would be predicted by innumerable studies, we found a strong correlation between chronological age and productive vocabulary. Presumably, infants who attend to all types of meaningful input, verbal, visual, or otherwise, are more likely to benefit and learn from their environmental interactions. More prosaically, productive vocabulary may also reflect more basic individual differences, such as 'speed of processing' that would influence children's ability to learn new words quickly and efficiently. We cannot rule out this possibility here, as we have no independent measure of intellectual and sensorimotor ability for this cohort.

The advantage the most verbally proficient infants showed for the verbal stimuli was no longer significant when the sounds that could be verbally produced (i.e., onomatopoeias) were excluded from the analyses. This suggests that onomatopoeic items are a significant, and potentially important, component of toddlers' (15–25 months) vocabularies. This is interesting since, except for onomatopoeic items, the recognition of verbal sound-object associations is thought to be a completely arbitrary process and varies across languages. However, there is evidence that young children are able to recognize, and prefer, the association between coordinating sounds and objects: Maurer, Pathman, and Mondloch (2006) found that 2½-year-old children appear to have a bias for a sound-shape correspondence (i.e., mouth shape and vowel sound) during the production of nonsense words, a preference that may serve to potentiate language development. Even very young infants can match sounds and their sources by localizing sound sources in space (e.g., Morrongiello, Fenwick, & Chance, 1998; Morrongiello, Lasenby, & Lee, 2003; Muir & Field, 1979), as well as identify transparent dynamic relations between mouth movement and sound (e.g., Bahrick, 1994; Gogate & Bahrick, 1998; Hollich et al., 2005). Thus, these early-acquired biases for bi-modal correspondence could structure how young infants interact during potential language learning situations (e.g., focusing on mouths speaking or producing noise, as opposed to feet). Onomatopoeic items provide young children with information regarding intermodal

associations, potentially providing a basis from which to learn and recognize more arbitrary sound-object associations, such as those involving words.

CONCLUSIONS

To our knowledge, this study was the first of its kind to examine young children's recognition of sound-object associations for different types of complex, meaningful sounds. Overall, across the age span studied, infants were just as accurate in recognizing the association between a familiar object and its related environmental sound, as when the object was verbally named. Additionally, we observed that the processing of environmental sounds is closely tied to chronological age, while the processing of speech was linked to both chronological age and verbal proficiency, thus demonstrating the subtle differences in the mechanisms underlying recognition in verbal and nonverbal domains early in development. The consistency with which nonverbal sounds and onomatopoeias are associated with their object referents may play an important role in teaching young children about intermodal associations, which may bootstrap the learning of more arbitrary word-object relationships.

Acknowledgments

AC was supported by NIH training grants #DC00041 and #DC007361, APS was supported by European Commission Marie Curie award FP6-025044, and FD was supported by MRC New Investigator Award G0400341. We would like to thank Leslie Carver and Karen Dobkins for allowing us to use their subject pool at UCSD and the Project in Cognitive in Neural Development at UCSD for the use of their facility, supported by NIH P50 NS22343. We would also like to thank Arielle Borovsky and Rob Leech for their comments on previous versions of this paper.

References

- Aziz-Zadeh L, Iacoboni M, Zaidel E, Wilson S, Mazziotta J. Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience* 2004;19:2609–2612. [PubMed: 15128415]
- Bahrack LE. Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development* 1983;6:429–451.
- Bahrack LE. Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology* 1992;53:180–199. [PubMed: 1578197]
- Bahrack LE. The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology* 1994;6(2):111–123.
- Balaban M, Waxman S. Do words facilitate object categorization in 9-month-old infants. *Journal of Experimental Child Psychology* 1997;64:3–26. [PubMed: 9126625]
- Ballas J. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance* 1993;19(2):250–267. [PubMed: 8473838]
- Ballas J, Howard J. Interpreting the language of environmental sounds. *Environment and Behavior* 1987;19(1):91–114.
- Boersma, Paul. Praat, a system for doing phonetics by computer. *Glott International* 2001;5(910):341–345.
- Borovsky A, Saygin AP, Cummings A, Dick F. Emerging paths to sound understanding: Language and environmental sound comprehension in specific language impairment and typically developing children. in preparation. Manuscript in preparation.
- Campbell A, Namy L. The role of social-referential context in verbal and nonverbal symbol learning. *Child Development* 2003;74(2):549–563. [PubMed: 12705572]
- Chiu C, Schacter D. Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds. *Consciousness and Cognition* 1995;4(4):440–458. [PubMed: 8750418]
- Conboy B, Thal D. Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development* 2006;77(3):712–735. [PubMed: 16686797]

- Cummings A, Čeponienė R, Dick F, Saygin AP, Townsend J. A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research* 2008;1208:137–149. [PubMed: 18387601]
- Cummings A, Čeponienė R, Koyama A, Saygin AP, Townsend J, Dick F. Auditory semantic networks for words and natural sounds. *Brain Research* 2006;1115:92–107. [PubMed: 16962567]
- Cummings, A.; Čeponienė, R.; Williams, C.; Townsend, J.; Wulfeck, B. Auditory word versus environmental sound processing in children with Specific Language Impairment: An event-related potential study. Poster presented at the 2006 Symposium on Research in Child Language Disorders; Madison, WI. 2006.
- Cycowicz Y, Friedman D. Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain and Cognition* 1998;36:30–51. [PubMed: 9500881]
- Dawson, C.; Gerken, LA. Proceedings of the Twenty-ninth annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum; 2006. p. 1198-1203.
- Dick F, Bates E, Wulfeck B, Aydelott Utman J, Dronkers N, Gernsbacher MA. Language deficits, localization and grammar: evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review* 2001;108(4):759–788. [PubMed: 11699116]
- Dick, F.; Saygin, AP.; Paulsen, C.; Trauner, D.; Bates, E. The co-development of environmental sounds and language comprehension in school-age children. Poster presented at the meeting for Attention and Performance; Colorado. 2004.
- Dodd B. Lip reading in infants: Attention to speech presented in- and out-of synchrony. 1979
- Fenson L, Dale P, Reznick J, Bates E, Thal D, Pethick S. Variability in the early communicative development. *Monographs of the Society for Research in Child Development* 1994;59(5):1–173. Serial No. 242. [PubMed: 7845413]
- Fenson L, Dale P. Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers* 1996;28:125–127.
- Fernald, A.; McRoberts, GW. Listening ahead: how repetition enhances infants' ability to recognize words in fluent speech. 24th Annual Boston University Conference on Language Development; Boston, MA. 1999.
- Fernald A, Perfors A, Marchman V. Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology* 2006;42(1):98–116. [PubMed: 16420121]
- Fernald A, Pinto JP, Swingle D, Weinberg A, McRoberts GW. Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science* 1998;9:72–75.
- Friedman D, Cycowicz Y, Dziobek I. Cross-form conceptual relations between sounds and words: Effects on the novelty P3. *Cognitive Brain Research* 2003;18(1):58–64. [PubMed: 14659497]
- Glenn S, Cunningham C, Joyce P. A study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Child Development* 1981;52(4):1303–1307. [PubMed: 6459215]
- Gogate LJ, Bahrack LE. Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology* 1998;69:133–149. [PubMed: 9637756]
- Gogate LJ, Bahrack LE, Watson JD. A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Development* 2000;71(4):878–894. [PubMed: 11016554]
- Gogate LJ, Bolzani LH, Betancourt EA. Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy* 2006;9(3):259–288.
- Golinkoff R, Mervis C, Hirsh-Pasek K. Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language* 1994;21:125–155. [PubMed: 8006089]
- Gygi, B. Factors in the identification of environmental sounds. Unpublished Doctoral Dissertation, Department of Psychology and Cognitive Science. Bloomington: Indiana University; 2001. p. 187
- Hollich G, Hirsh-Pasek K, Golinkoff R, Brand R, Brown E, Chung H, Hennon E, Rocroi C. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development* 2000;65(3):v-123.
- Hollich G, Newman RS, Jusczyk PW. Infants' use of synchronized visual information to separate streams of speech. *Child Development* 2005;76(3):598–613. [PubMed: 15892781]

- Houston-Price C, Plunkett K, Harris P. Word-learning wizardry at 1;6. *Journal of Child Language* 2005;32:175–189. [PubMed: 15779882]
- Iverson J, Capirci O, Caselli MC. From communication to language in two modalities. *Cognitive Development* 1994;9:23–43.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe E. Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience* 2005;25(21):5148–5158. [PubMed: 15917455]
- Lewkowicz DJ. Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance* 1996;22(5):1094–1106. [PubMed: 8865617]
- Killing S, Bishop D. Move it! Visual feedback enhances validity of preferential looking as a measure of individual differences in vocabulary in toddlers. *Developmental Science* 2008;11(4):525–530. [PubMed: 18576960]
- Kuhl P, Meltzoff AN. Bimodal perception of speech in infancy. *Science* 1982;218(4577):1138–1141. [PubMed: 7146899]
- Kuhl P, Meltzoff AN. The intermodal representation of speech in infants. *Infant Behavior and Development* 1984;7:361–381.
- Marchman V, Fernald A. Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science* 2008;11(3):F9–F16. [PubMed: 18466367]
- Marcus GF, Fernandes KJ, Johnson SP. Infant rule learning facilitated by speech. *Psychological Science* 2007;18(5):387–391. [PubMed: 17576276]
- Matthews D, Lieven E, Theakston A, Tomasello M. The role of frequency in the acquisition of English word order. *Cognitive Development* 2005;20:121–136.
- Maurer D, Pathman T, Mondloch CJ. The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science* 2006;9(3):316–322. [PubMed: 16669803]
- Morrongiello B, Fenwick K, Chance G. Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior & Development* 1998;21(4):543–554.
- Morrongiello B, Lasenby J, Lee N. Infants' learning, memory, and generalization of learning for bimodal events. *Journal of Experimental Child Psychology* 2003;84:1–19. [PubMed: 12553915]
- Muir D, Field J. Infants orient to sounds. *Child Development* 1979;50(2):431–436. [PubMed: 487883]
- Namy L. What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy* 2001;2:73–86.
- Namy L, Acredolo L, Goodwyn S. Verbal labels and gestural routines in parental communication with young children. *Journal of Nonverbal Behavior* 2000;24:63–79.
- Namy L, Waxman S. Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development* 1998;69(2):295–308. [PubMed: 9586206]
- Namy L, Waxman S. Patterns of spontaneous production of novel words and gestures within an experimental setting in children ages 1;6 and 2;2. *Journal of Child Language* 2002;24(2):911–921. [PubMed: 12471979]
- Patterson ML, Werker JF. Two-month-old infants match phonetic information in lips and voice. *Developmental Science* 2003;6(2):191–196.
- Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D'Amico S, Di Russo F. Separate neural systems for processing action- or non-action-related sounds. *NeuroImage* 2005;24(3):852–861. [PubMed: 15652320]
- Roberts K. Categorical responding in 15-month-olds: Influence of the noun-category bias and the covariation between visual fixation and auditory input. *Cognitive Development* 1995;10(1):21–41.
- Roland D, Dick F, Elman JL. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 2007;57:348–379. [PubMed: 19668599]
- Saffran J, Pollak S, Seibel R, Shkolnik A. Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition* 2007;105(3):669–680. [PubMed: 17188676]
- Saygin AP, Dick F, Bates E. An online task for contrasting auditory processing in the verbal and nonverbal domains and norms for younger and older adults. *Behavior Research Methods, Instruments, & Computers* 2005;37(1):99–110.

- Saygin AP, Dick F, Wilson S, Dronkers N, Bates E. Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain* 2003;126:928–945. [PubMed: 12615649]
- Stuart G, Jones D. Priming the identification of environmental sounds. *Quarterly Journal of Experimental Psychology* 1995;48A(3):741–761. [PubMed: 7568996]
- Swingley D, Pinto J, Fernald A. Continuous processing in word recognition at 24 months. *Cognition* 1999;71:73–108. [PubMed: 10444905]
- Van Petten C, Rieffers H. Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia* 1995;33:485–508. [PubMed: 7617157]
- Volterra, V.; Caselli, MC.; Capirci, O.; Pizzuto, E. Gesture and the emergence and development of language. In: Tomasello, M.; Slobin, D., editors. Chapter in *Beyond Nature/Nurture: A festschrift for Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005. p. 3-40.
- Volterra, V.; Iverson, J.; Emmorey, K. When do modality factors affect the course of language acquisition?. In: Reilly, J., editor. Chapter in *Language, gesture, and space*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1995. p. 371-390.
- Vouloumanos A, Werker J. Tuned to the signal: the privileged status of speech for young infants. *Developmental Science* 2004;7(3):270–276. [PubMed: 15595367]
- Walker-Andrews AS, Lennon E. Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development* 1991;14:131–142.
- Woodward A, Hoynes K. Infants' Learning about Words and Sounds in Relation to Objects. *Child Development* 1999;70:65–77. [PubMed: 10191515]
- Xu F. The role of language in acquiring object kind concepts in infancy. *Cognition* 2002;85(3):223–250. [PubMed: 12169410]

Appendix

Target Picture	Distracter Picture	Verbal Phrase
<i>Cow</i>	Piano	<i>Cow Mooing *</i>
<i>Lion</i>	Alarm Clock	<i>Lion Growling *</i>
Trumpet	Kissing Couple	Trumpet Playing
<i>Horse</i>	Telephone	<i>Horse Neighing *</i>
<i>Dog</i>	Sneezing Person	<i>Dog Barking *</i>
Toilet	Cat	Toilet Flushing
Baby	Grandfather Clock	Baby Crying *
Sneezing Person	Dog	Someone Sneezing
Airplane	Singing Woman	Airplane Flying *
<i>Fly</i>	Sheep	<i>Fly Buzzing</i>
Alarm Clock	Lion	Alarm Clock Ringing *
Singing Woman	Airplane	Woman Singing *
Kissing Couple	Trumpet	Someone Kissing *
<i>Cat</i>	Toilet	<i>Cat Meowing *</i>
Grandfather Clock	Baby	Grandfather Clock Chiming *
Train	Rooster	Train Going By *
Piano	Cow	Piano Playing
<i>Sheep</i>	Fly	<i>Sheep Baaing *</i>
Telephone	Horse	Telephone Ringing *
<i>Rooster</i>	Train	<i>Rooster Crowing *</i>

Items in *italics* were counted as onomatopoeic; starred verbal phrases had equivalents in the MacArthur Bates CDI, and thus were included in the individually-tailored item analyses (see Results).

Front View of Infant Testing Booth

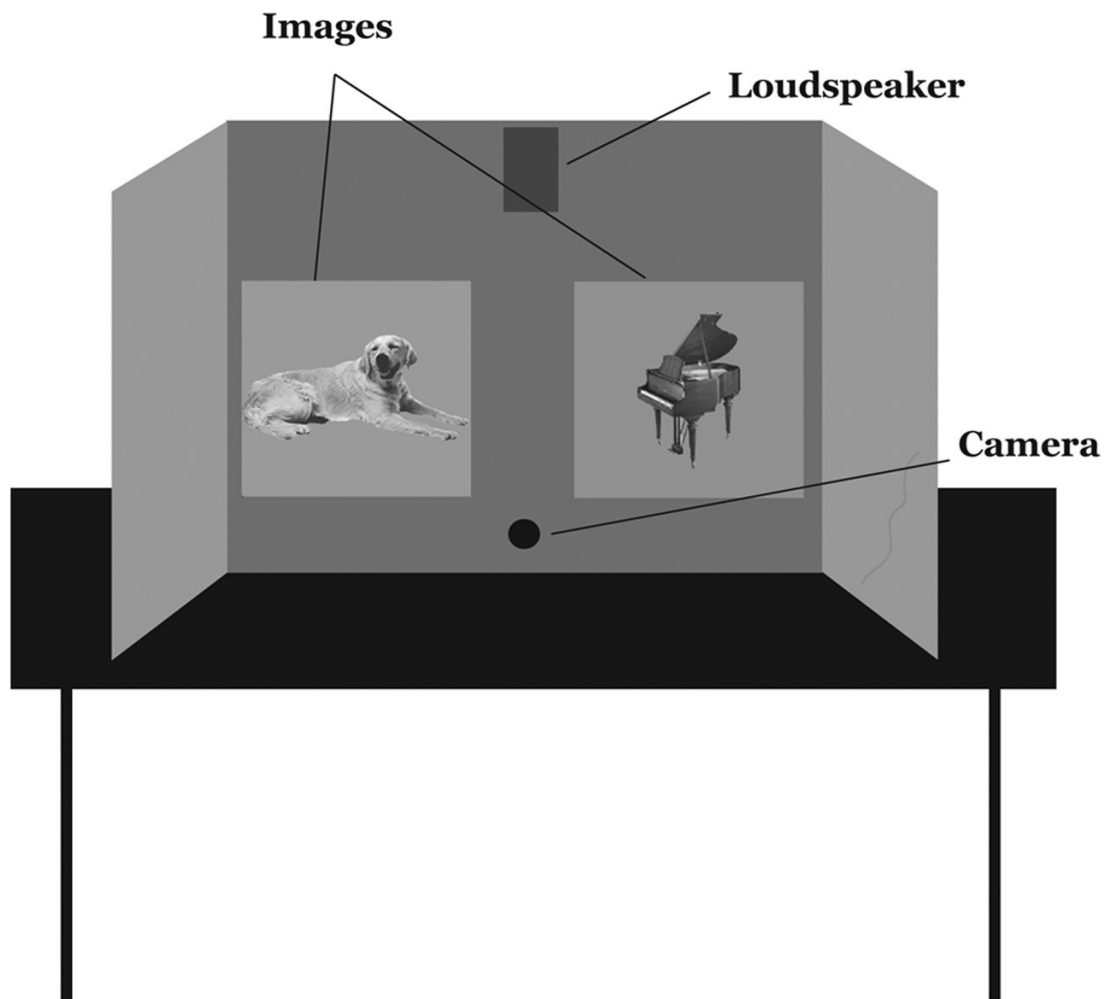


Figure 1. Infants visually matched each verbal or environmental sound to one of two pictures presented on computer screens in front of them. A small video camera placed between the monitors recorded subjects' eye movements during the experiment.

Schematic of a Single Preferential Looking Trial

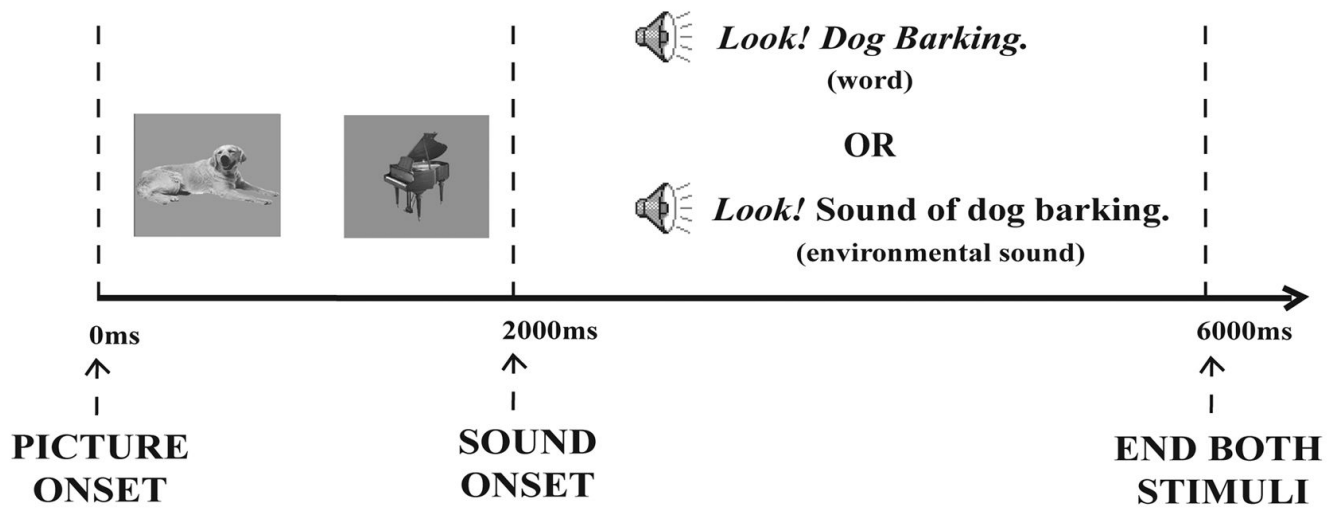


Figure 2.

Two pictures first appeared on two computer screens for 2000ms. The pictures remained visible while the auditory stimulus (verbal or environmental sound label) was presented for 4000ms. At the conclusion of the auditory stimulus, both pictures disappeared together from the computer screens. The experimenter manually started the next trial as long as the child was attending to the computer screens.

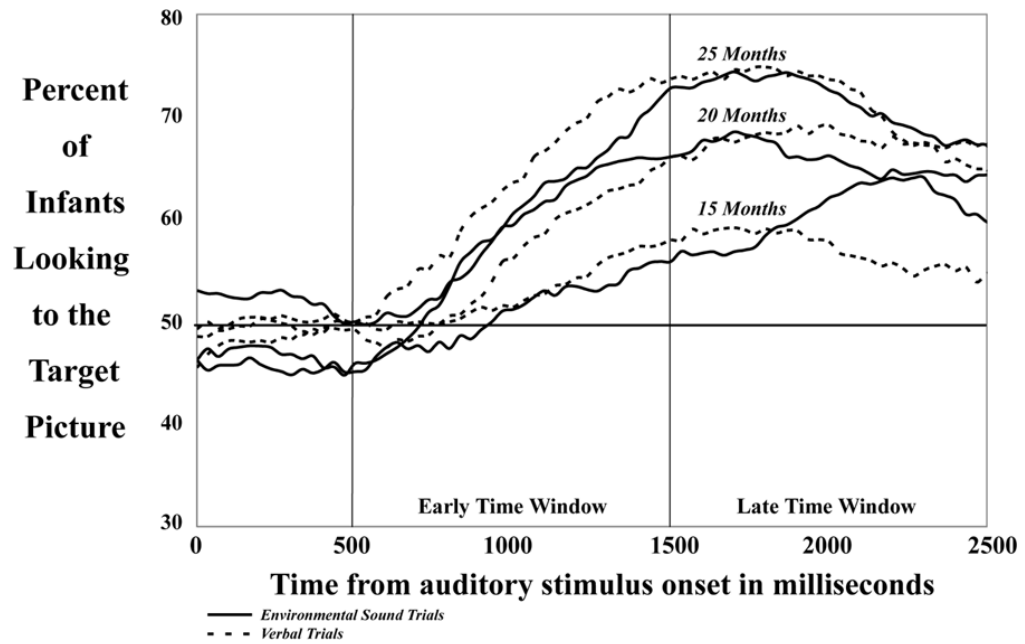


Figure 3. The Early Time Window ran from 533ms after utterance/sound onset to 1500ms after onset. The Late Time Window ran from 1533–2500ms after sentence onset. Note that the analyses did not include the carrier phrase, “Look!”. Here, the 15-, 20-, and 25-month-old infants’ responses to the Verbal trials are represented by dashed lines and the Environmental Sound trials by solid lines.

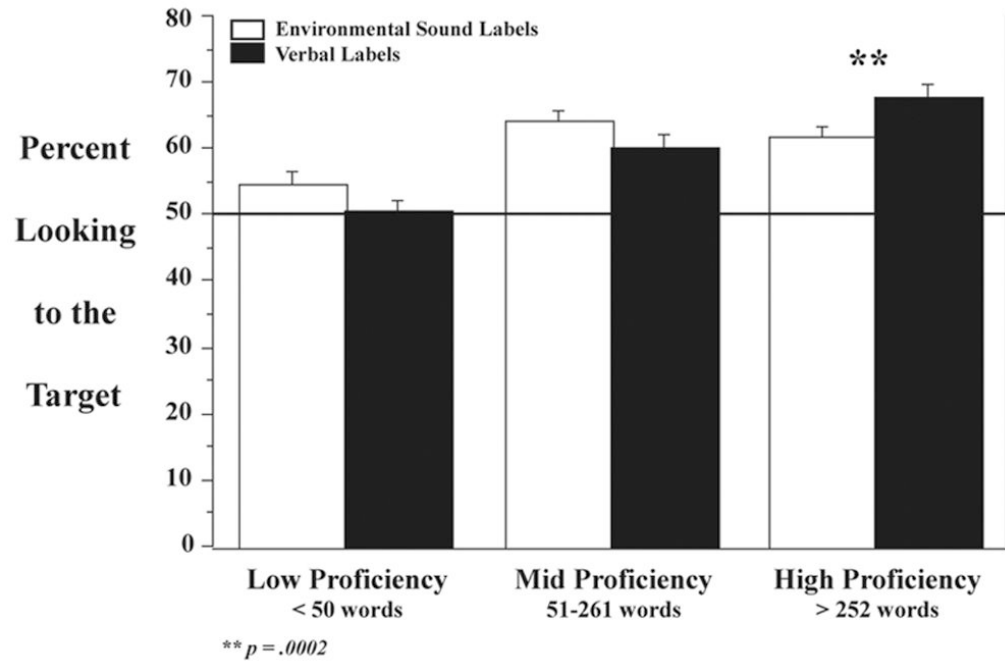


Figure 4.

While the Low and Mid proficiency groups did not differ in their responses to Verbal and Environmental Sound labels, infants in the High proficiency group responded significantly more accurately to Verbal labels. Note that after taking into account each child's verbal comprehension/production and the onomatopoeic test items, all cross-domain differences did disappear (see text for details).

Accuracy was calculated by dividing the proportion of time infants looked at the target picture by the total amount of time infants looked at both the target and distracter pictures. All accuracy measures represent the percent of time infants attended to the target during the designated analysis window. Looking Accuracy Results.

Table 1

	Overall Accuracy	Domain		Early Window		Late Window	
		Sounds	Verbal	Sounds	Verbal	Sounds	Verbal
15 months	53.69	54.87	52.51	50.18	59.56	54.86	
20 months	60.71 ¹	61.25	60.17	57.73	64.77	64.75	
25 months	65.10 ²	64.38	65.84	58.35	70.39	70.3	
Low Proficiency	52.60	54.72	50.49	49.73	59.72	53.01	
Mid Proficiency	62.26 ³	64.22 ⁴	60.29 ⁵	58.7	69.75	65.28	
High Proficiency	64.64 ³	61.54	67.74 ^{6,7}	57.83	62.50	72.99	

Significant as compared to:

¹ 15 months, $p < .003$,

² 15 months, $p < .0001$,

³ Low Proficiency, $p < .0001$,

⁴ Low Proficiency Sounds, $p < .0003$,

⁵ Low Proficiency Verbal, $p < .003$,

⁶ Low Proficiency Verbal, $p < .0001$,

⁷ High Proficiency Sounds, $p < .0003$. Note that F-values are cited in the text.