Toddlers' ability to map the meaning of new words in multi-talker environments

Justine Dombroski and Rochelle S. Newman[a)]
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 24 March 2013; revised 29 June 2014; accepted 2 October 2014)

Whether in a noisy daycare center, home, or classroom, many of the environments children are exposed to are, undoubtedly, not acoustically ideal for speech processing. Yet, somehow, these toddlers are still able to acquire vocabularies consisting of hundreds of words. The current study explores the effect of background speech noise on children's early word learning (specifically, their ability to map a label onto an object). Three groups of children aged 32–36 months were taught two new words either in quiet, or in the presence of multi-talker babble at a +5 or 0 dB signal-to-noise ratio (SNR). They were then tested on their learning of these new word-to-object mappings. Children showed similar accuracy in all three conditions, suggesting that even at a 0 dB SNR, children were successfully able to learn new words. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4898051]

PACS number(s): 43.71.Ft, 43.66.Dc [BRM] Pages: 2807–2815

I. INTRODUCTION

The majority of toddlers today are exposed to a great deal of noise. Although typical noise levels for young children's homes have not been studied, many children find themselves spending some time in day care or classroom-like settings. The acoustic environments for most infant/toddler centers and classrooms are far from ideal for speech perception and word learning (Frank and Golden, 1999; Manlove *et al.*, 2001), and the levels of noise and reverberation typical of classroom settings has the potential to interfere with children's understanding of speech and hence their learning (e.g., Hétu *et al.*, 1990; Mills, 1976; Neuman *et al.*, 2010; Smyth, 1979). Golden and Frank (personal communication) measured signal-to-noise ratios (SNR) in five occupied toddler classrooms and found that, during book-reading time, the SNR for different teachers averaged only 5–6 dB. As a comparison, SNR measures in primary schools should ideally be +15 dB (Crandell and Smaldino, 2000), suggesting that these toddler classrooms are quite loud.

Thus, noise is an all-too-common presence in young children's lives, and this presence could potentially have detrimental effects on their language development (Frank and Golden, 1999; Manlove *et al.*, 2001). These effects can take multiple forms, each of which can adversely affect children's perceptual abilities and thus be an obstacle to word learning.

First, noise has the potential to mask the target signal, making the target signal less audible or less discriminable. In general, children require greater SNRs in order to detect and discriminate speech sounds or words than do adults (cf. Fallon *et al.*, 2000). Noise has been shown to adversely affect the performance of young children on speech tasks such as word discrimination and consonant identification (Finitzo-Hieber and Tillman, 1978; Nábelek and Robinson, 1982). Because young children's auditory and cognitive systems are not fully developed, they are especially vulnerable to the effects of noise and need more exaggerated cues to discriminate syllables and recognize words, particularly in noise (Elliott *et al.*, 1981; Fallon *et al.*, 2000). Children under the age of 5 yr also appear to rely more heavily on bottom-up processing, meaning that they rely more on the speech signal itself than on their knowledge of language structures, and thus have trouble perceptually restoring missing input (Newman, 2006). This may make it more difficult for them to recover the intended signal when portions are masked by background noise.

Noise and background speech can also cause distraction, attracting attention away from the target signal. Infants and children are generally less adept at focusing their attention selectively to a given signal, at least through 5 yr of age (Allen and Wightman, 1995; Bargones and Werner, 1994); (see Gomes *et al.*, 2000 for a review). This suggests it would be harder for them to separate the relevant auditory information from other background noise and pay attention to what someone is saying to them while in a noisy environment. This, too, has the potential to limit their word-learning opportunities.

Closely related to the prior point, both preschoolers and school-aged children are more sensitive to informational masking than are adults (see, for example, Leibold and Neff, 2007; Lufti *et al.*, 2003; Oh *et al.*, 2001; Polka *et al.*, 2008; Wightman *et al.*, 2003). This implies that they may have difficulty recognizing which sources of sound should be grouped together or segregated, further decreasing their ability to listen in noise. This will be particularly the case when the background "noise" consists of other speech, or of sounds that vary in a speech-like manner. However, Polka *et al.* (2008) reported that even high-frequency bird songs (a signal very different from typical speech) can cause informational masking in young infants; it is not clear whether such would also be the case for toddlers.

Yet despite these myriad difficulties, Newman (2011) reported that 24-month-olds can recognize well-known

[a)]Author to whom correspondence should be addressed. Electronic mail: Rnewman1@umd.edu

words in the presence of multi-talker babble at a SNR as low as negative 5 dB. While adults can recognize words at still lower SNRs (Bronkhorst and Plomp, 1992), this is nonetheless a significant achievement. Indeed, the 24-month-olds succeeded at recognizing words at far worse SNRs than do infants in their first year of life, suggesting that, as they age, children become better able to distinguish target speech streams from background speech, and to recognize words with which they are familiar.

Despite this general success, noise still has noticeable effects on word recognition, even at relatively high SNRs and for well-known words. For example, Grieco-Calub *et al.* (2009) found that both normal-hearing children and children with cochlear implants were less accurate at recognizing known words in the presence of two-talker babble at a 10 dB SNR than in quiet, and Newman (2011) found a gradient effect of noise level on accuracy, with poorer performance as the SNR decreased.

To date, most studies of children's ability to recognize speech in noise or other auditory distractions has utilized words that children were expected to already know; i.e., words they presumably had learned in quiet settings. In this situation, noise merely limits the ability to recognize the signal. But background noise, particularly background speech, may also be a factor in situations in which children are attempting to learn new words. Moreover, word learning in the presence of background speech is a much more complex task than word recognition or speech perception, as it involves memory, selective attention, and other cognitive functions on top of the perceptual skills required for recognition; thus, children's ability to recognize known words in noise is likely to be greater than their ability to learn new words in such settings. Along these lines, Mattys *et al.* (2012) reported that noise can result in reduced memory and attentional capacity in general. To the extent that these skills are required for learning words, this suggests that noise could lead to poorer learning even in the face of adequate recognition.

Word learning outside of the laboratory typically happens incidentally as the child engages in activities of daily life. Such learning is a multi-step process, requiring (at minimum) that children segment the target word from the speech stream, identify a likely referent for the word in their surroundings, and make a mapping between the two. This information must then be encoded in memory in such a way that the word form and the referent will be accessed together in the future (Horst and Samuelson, 2008).

In the laboratory setting, word learning is often assessed through an examination of children's ability to make an association between an object and a label during a short testing session. Such mapping is thought to be a critical step involved in true word learning, but may be distinct from symbolic comprehension, when a child understands not only that the sound pattern is associated with a specific referent, but that the sound pattern stands for or refers to it (Oviatt, 1980). Indeed, some research has suggested that initially discovering the referent of a new word in a lab setting may be quite different than actually learning the word (Horst and Samuelson, 2008), which requires the ability to store this information in a way that can be accessed later. However, this initial mapping ability has the advantage that it can be studied in a single visit in a laboratory setting, allowing complete control of the input the child receives (as compared to studies assessing children's recognition of already-known words) while not requiring long-term (multi-session) training regimens. Moreover, this initial mapping is likely to depend critically on the ability to hear the signal; factors that hinder children's ability to make object-label associations are likely to also affect their ability to fully learn and store the meaning of a word.

Thus, the current study examines the extent to which background speech influences children's ability to successfully map a label onto an object. While many studies have looked at word learning generally, few have looked at how noise affects this process. What research has been conducted suggests that noise in general affects receptive and expressive vocabulary growth in children. Riley and McGregor (2012) taught children aged 9–11 yr a set of 16 novel words: half in the context of speech-shaped noise at an 8 dB SNR, and the other half in quiet. The children were then tested on their learning, both through a measure of comprehension and through a production measure. They showed ceiling performance on comprehension, suggesting that the mild noise level used was not sufficient to disrupt word learning. However, children were better able to produce the words accurately when they had been taught in quiet as opposed to noise, suggesting that the presence of noise might result in weaker (or less complete) expressive vocabulary representations.

Blaiser (2010) tested children ranging from 2.5 to 6 yr on a word-learning task. Half of her participants had hearing loss (and used either a hearing aid or a cochlear implant); the other half were matched controls. Children were taught eight new words in each of two conditions: a quiet setting and in the presence of multi-talker babble at a +7 dB SNR. Children were first tested on their fast-mapping abilities, or the ability to link a label with an object after only a small number of examples; they were introduced to the new objects and were subsequently asked to repeat back the names. The children were then taught the words during three more training sessions over a one-week period, and were assessed a second time. The first assessment was considered an evaluation of their fast-mapping abilities, while the second assessment was considered an evaluation of true word learning. Children without hearing loss performed quite similarly in the quiet and noise conditions, both in terms of their fast-mapping and word-learning abilities, suggesting that this level of noise did not interfere with their ability to learn words. Children with hearing loss, who might be expected to be more susceptible to noise, showed marked deficits in their fast mapping abilities in the noise condition, but comparable performance in quiet. However, they showed similar word-learning results, suggesting that repeated exposure could compensate for poorer input at any given time. Children's age correlated strongly with both fast-mapping and word-learning skills, suggesting that there may be developmental differences across the age range she tested.

While these studies as a group indicate that noise can make object-label mapping more difficult (at least at some

SNR levels and for some children), they do not reveal the limits of preschoolers' ability to learn new words in noise. To begin to explore this issue, the present study directly compares the performance of 32- to 36-month-old toddlers on a fast-mapping task in the presence of multi-talker babble at three different SNRs: no noise, +5 dB SNR (where the target is more intense than the background speech), and 0 dB SNR (where the target and background speech are at equivalent intensity levels). We expect excellent performance when the words are taught in quiet, but decreasing performance as the signal-to-noise level becomes more challenging. Although Blaiser (2010) and Riley and McGregor (2012) found no effects of noise on receptive assessments in typically hearing children in their studies, the current study uses both more difficult SNR levels (+5 and 0 dB SNR compared to +7 or +8 dB) and younger children (32–36 months vs 28–72 months and 9–11 yr), as well as a different task and testing method. In addition, both studies did find effects of noise when they used either a more difficult assessment (production in Riley and McGregor) or individuals with poorer hearing (Blaiser), suggesting that their receptive tasks may simply have been too easy to show effects in a typically developing population. We also examined if there was a relationship between children's existing vocabulary size (as indicated by parent-report measures) and their ability to learn new words in noise.

We taught children novel words in a split-screen preferential-looking paradigm (Golinkoff *et al.*, 1987; Hollich, 2006). During the initial portion of the experiment, children were presented with a single object at a time and heard a voice saying, "It's a [novel name]!" At the same time, participants in the background-speech conditions heard nine-talker multi-talker babble at either a +5 dB or 0 dB SNR. Newman (2011) reported that 24-month-olds were able to recognize the names of familiar items at a SNR as low as −5 dB, but only for one word out of six tested, all of which were highly familiar. Moreover, the effect was quite small, and shown by only a subset of children tested (17 of 24 children looked longer at the appropriate object, where 12 would be chance performance). Since learning a word is thought to be more cognitively demanding task than simply recognizing it (Bloom, 2000), it seemed logical to test a slightly easier noise level here.

After children were taught the two words, they were tested on their learning of those word-object mappings. Testing occurred without any background distractors; children were shown both novel objects on the screen at the same time, as a voice said, "Find the [novel name]." We presumed that if children looked significantly longer at the appropriate image than they looked at that image on baseline trials (when not told explicitly where to look), this would be an indication that they had been able to learn the new words during the training stage, despite the presence of any background speech.

We predicted that toddlers would have difficulty creating word-object mappings in the presence of speech distractors. As noted above, the presence of speech distractors not only can reduce signal quality, but it can result in reduced memory and attentional capacity (Mattys *et al.*, 2012) which would be expected to reduce the ability to form lexical representations. We therefore expect poorer word learning in the presence of such distractors.

II. METHODS

A. Participants

In total, 72 monolingual toddlers participated in this study (41 male; 32 female; range 31.6 months to 36.2 months; mean 34 months); 24 heard each of the three SNR levels (no background speech, +5 dB SNR, and 0 dB SNR). The three groups were similar in age (34.4, 33.4, 34.1 month, respectively) and gender breakdown (13 or 14 males each). All participants had at least 80% daily exposure to American English and had no history of visual, language, hearing, or neurological impairment or disorder. Data from an additional 21 participants were excluded for the following reasons: fussiness/crying ($n = 5$), not meeting the language criterion ($n = 5$), side bias during silent looking trial ($n = 1$), parental interference ($n = 1$); experimenter error ($n = 2$), or failure to attend/distractedness ($n = 7$). Children who were excluded for fussiness and failure to attend were distributed equally across conditions (four each in control, 5 and 0 dB conditions).

Participants were 14% African American, 4% Hispanic, 4% Asian, and 78% Caucasian; all parents had at least a high school education, with 15% having some college experience, 28% having a bachelor's degree, 34% having a master's degree, and 19% having a doctoral degree. Education information was not available for two families.

B. Materials

Children were taught two new words, "coopa" and "needoke." These word forms were selected because they are both easily discriminable (having highly distinct vowels, consonants that differ in voicing, and different syllable structures) and multi-syllabic; the latter was based on the assumption that longer words might be easier to make out in the context of an amplitude-varying noise source. These choices were designed to make the task easier for the child, but admittedly run the risk of making the laboratory task less realistic, in that the children might not need to form detailed lexical representations in order to succeed. We come back to this point in the discussion.

The visual stimuli were used by Hollich (2006), and consisted of a green multi-limbed creature and a spiky object on a pedestal; both objects rotated in three-dimensional space. Assignment of name to object was counterbalanced across participants.

Audio stimuli were recorded in a slight infant-directed style, or "happy speech," in a noise-reducing sound booth, using CoolEdit 2000 software and a Shure SM81 microphone at a 44.1 kHz sampling rate and 16-bit precision. The target speaker was a female speaker of English without a particular regional accent. Several tokens of each stimulus sentence were recorded, and the best tokens were used in the study. The same token of each carrier phrase (e.g., "Find the ______!", "It's a _____!") was cross-spliced with the two target words to ensure that the only differences were in the words themselves. The target words were edited to be the same average amplitude, and the test trial stimuli were edited to be the same duration. Training stimuli were 11 s long, and repeated the target word three times ("*Look! It's a needoke! Wow, it's a needoke. Do you see it? A needoke.*"). Test stimuli were 7.8 s long, and were of two types: On labeling trials, the voice instructed the participant to look at one or the other object ("*Look at the coopa! Where is the coopa? Do you see the coopa? Find the coopa! Coopa!*"). On baseline trials, the voice did not instruct children to look at a particular object ("*Wow, look at that! Do you see that? Ooooo! look!*"). This provides a measure of baseline preference to look at one object vs the other. On labeling trials, the first instance of the target word occurred at 528 ms for both target items. The instructions in the two target trial types (coopa and needoke) were identical, so the sentence frame could not cue looking to the appropriate object. Nor did the difference in instructions result in generally greater looking on test trials, as reported in the results section below.

The background babble was the same as that used in Newman (2011). It consists of a combination of nine different women, all native speakers of English, reading passages aloud from a variety of books; these were adjusted to be the same RMS amplitude, and were blended together at equal ratios. This combined distractor passage was then edited to be the same length as the target passages, adjusted to be the appropriate intensity level (either 5 dB less intense than the target items or the same intensity as the target items), and blended with the target recordings.

C. Procedure

Participants were tested using a variant of the split-screen preferential looking paradigm and heard only one of the three SNR conditions. They sat on their caregiver's lap 2 ft away from a large 58 in. LCD monitor, on which the stimuli were presented from DVD.

During the training phase, children saw a single object appear on the screen, and the speaker labeled that object. The two objects were labeled in alternation for eight trials (four trials per object), with some children receiving this training in quiet, and others in the presence of multi-talker babble. This was followed by a single silent trial, in which both objects appeared together, intended to introduce the idea that objects would now occur on the left and right; one child had a strong side bias (>85% looking to one side) and was replaced, as noted in the participants section. Following this was a 10-trial test phase, in which both objects appeared on each trial; which visual object appeared on the left (vs right) was counterbalanced across participants. Two of the trials were baseline trials; the other eight trials consisted of four labeling each object; all names were presented in quiet so as to avoid reduced audibility at the time of test. It is worth noting, however, that this change also requires that children generalize between a situation with noise (during training) and one without (during test), which itself might cause some task difficulty (see Creel *et al.*, 2012 for work with adults). Four different random orders were created for the different trial types, and infants were randomly assigned to one of these orders.

It is worth noting that our method of baseline trials is different from some studies that have used an initial baseline window within each trial as a comparison base. Our method of explicit baseline trials has the advantage that it measures preference for one object over another across the full length of a trial, rather than over a short initial portion thereof. This can be important if children demonstrate different patterns of looking at the start of a trial (when images first appear) vs later in a trial. One potential disadvantage of our method would be if children's looking behavior was different across different trials in the study; we account for this, at least partially, by having the baseline trials interspersed randomly among the 10 test trials. We also analyze data two ways: (1) by comparing looking time to a named object vs to that object when unnamed (baseline trials), and (2) by comparing looking time to a named object vs to the other object on the same trial (ignoring baseline trials; see coding, below).

Although we taught two new objects, our intent was not to compare them; rather, teaching names for two objects allows us to test children's learning of the correct object-word association without having to introduce an unknown object at the test phase. Teaching only a single object during training can cause a confound, in that children might be biased to look preferentially at the object that had received a name. Training with two objects allows us to test that children actually recognize the particular mapping between object and label.

In addition to testing children experimentally, we also gathered information on their language skills more generally. Parents completed the speech language assessment scale (SLAS; Hadley and Rice, 1993), a questionnaire in which they provide a comparison rating of their child's language skills relative to other children of the same age. The SLAS consists of 19 questions, each answered on a seven-point rating scale, which were then summed for each child. A larger number indicates that parents consider their child's language skills relatively advanced compared to their peers.

D. Coding

A digital camera recorded each child's eye gaze throughout the study at a rate of 30 frames per second. Two assistants, blind to condition, individually coded each child's looking behaviors offline, on a frame-by-frame basis using Supercoder coding software (Hollich, 2005). From this, the infants' total duration of looking at each of the two images on each trial was calculated.

On any trial in which the coders disagreed by more than 15 frames (0.5 s), a third coder was added; the average of the two closest codings was used as the final data. These final data were extremely reliable; correlations on the percentage of left (vs right) looking for each individual participant ranged from 0.9833 to 0.9999 with a mean correlation of 0.9971. On average, the coders agreed on looking coding on 95.0% of frames (range per participant: 88.0%–98.7%).

These coding results were used to calculate the percentage of time the child spent looking at the appropriate (named) object on each trial (e.g., looking at the coopa when the coopa was named, and looking at the needoke when the needoke was named), starting from the onset of the first repetition of the target word, minus the time spent looking at that object on the baseline trials. We collapsed across the two objects for several reasons: (1) there was no reason to expect the two objects to be learned differentially, as noted above; (2) even if a child had learned only one object, they would nonetheless be able to look appropriately on all trials; that is, if a child knew what a "coopa" was, he or she would presumably look appropriately when asked to look at it, but also could look appropriately via disambiguation/mutual exclusivity when asked for the "needoke" (see Markman and Wachtel, 1988). Thus, there is no way to test knowledge for the two objects separately from one another with this testing method, and we collapsed across this distinction. The critical issue is thus whether, across trials, children look to the correct object when instructed to do, and whether the size of this effect differs across noise conditions.

As an alternative measure, we also compared looking time behavior during the initial portion of the trial (from 367 ms after word onset to 1800 ms after word onset; this window size has previously been used by Fernald and Hurtado, 2006); here we compared the proportion of time infants spent looking at the named object compared to the alternate object (the inverse value), as has been done by Bergelson and Swingley (2012).

In addition to coding these overall looking measures, we also measured children's reaction times, based on the procedure developed by Fernald and colleagues (Fernald *et al.*, 2006; Fernald *et al.*, 1998). Because all test trials presented the same two objects, in the same positions, children should be able to switch to the correct image as soon as they hear the target word. We examined those test trials in which children were not looking at the target object at word onset, but made a shift toward that object during a window from 300–1800 ms later. For those trials, we recorded the time at which the shift occurred; these were then averaged for the different trials within a participant; we include data only for participants who had reaction time (RT) data from at least two trials. We note, however, that this procedure has typically been used in word recognition experiments; word-learning experiments generally include far fewer test trials (both because part of the session is taken up with learning, and because they cannot include as many different words at test), which makes RT data more variable. This was clearly the case here, as we had, on average, 2.8 RT trials per participant (range: 2–6; SD = 1.02).

III. RESULTS

Looking time can be evaluated as the total amount of time spent attending in general (a general measure of attentiveness to the task), and as the total amount of time spent attending to the named (vs unnamed) object (an indication of actual word learning). We first looked at overall attentiveness, to ensure that children were not looking differentially on target vs baseline trials. In general, children attended on trials asking for the coopa for 198.5 frames, compared to 194.7 frames on trials asking for the needoke, and 201.8 frames on baseline trials. These measures are not significantly different [$F(2,142) = 1.93$, $p > 0.15$, partial eta-squared = 0.027], suggesting that the differences in sentence structure did not cause differential attention. Nor was there a difference between baseline trials and the average of the two target trial types, $t(71) = 0.996$, $p > 0.30$, Cohen's $d = 0.37$). Across baseline trials, children looked slightly longer to the creature-like object (55.6% of the time) than to the spiky object on a pedestal (44.4%), but these differences, while significant [$t(71) = 2.45$, $p = 0.0167$], were not so large as to make changes in looking impossible to detect.

Our primary analyses focused on looking to the correct (named) object, compared to looking towards that object on baseline trials, as a measure of children's learning. We first analyzed data from the control participants. As expected, the children looked significantly longer to the appropriate object than to that object on baseline trials [$t(23) = 3.59$, $p = 0.0015$, Cohen's $d = 0.73$], as shown in the left column of Fig. 1. This demonstrates the appropriateness of our task and stimuli; if children were unable to learn words in the quiet setting, there would be little point in testing them in noise.

We then examined word learning in the two noise conditions. Children looked significantly longer to the appropriate object than to that object on baseline trials in the harder, 0 dB condition [$t(23) = 3.03$, $p = 0.006$, Cohen's $d = 0.62$; right column of Fig. 1], but only marginally so in the easier, +5 dB condition [$t(23) = 1.85$, $p = 0.077$, Cohen's $d = 0.38$; center column of Fig. 1]. Surprisingly, the children appeared to do slightly better in the more difficult noise condition; this could be an indication that noise helps to focus children's attention. However, such a finding could also be the result of the use of different participants in the different conditions, rather than suggesting that children are actually learning better in the presence of greater amounts of noise. An analysis across the conditions supports the latter interpretation, suggesting that there was no effect of the noise on the degree of this learning: A one-way repeated measures analysis of variance (ANOVA) showed no effect of condition,

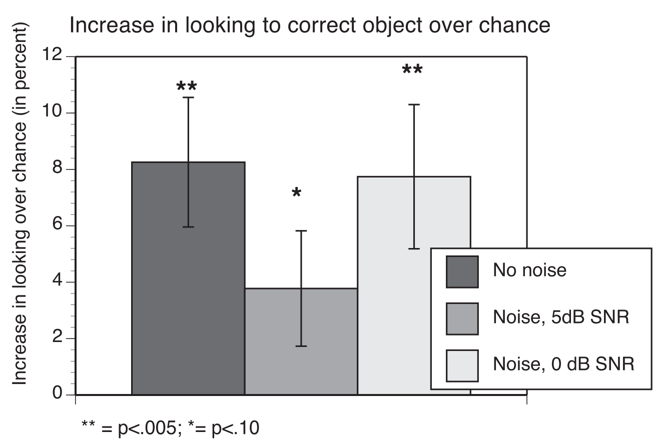

FIG. 1. The increase in children's looking to the correct object compared to their looking on baseline trials (chance), in quiet (left), 5 dB SNR (middle) and 0 dB SNR (right). Error bars indicate standard error.

$F(2,69) = 1.13$, $p > 0.30$, partial eta-squared $= 0.032$. Thus, children apparently performed equally well at test, regardless of the noise level during learning.

One concern may be that using separate baseline trials, where no object was named, was an inappropriate comparison. To assess this, we compared the proportion of time children spent looking to the named object vs the unnamed object on each trial; using this method, and looking across the three conditions, we find that children look appropriately on both the "needoke" trials [$t(71) = 3.19$, $p < 0.005$] and on the "coopa" trials [$t(71) = 3.08$, $p < 0.005$]. A mixed effects ANOVA showed a significant effect of looking (to the correct rather than incorrect object), $F(1,69) = 24.52$, $p < 0.001$, but no effect of noise condition [$F < 1$), target word (needoke vs coopa, $F < 1$), and no interactions [all $F < 1$, except looking x condition, $F(2,69) = 1.06$, $p > 0.35$]. Another concern may be that using the entire trial length may have been too long; perhaps children looked appropriately at the start of trials, but then became bored and looked randomly thereafter; including the full trial lengths may have made it harder to find differences across conditions. To test this, we reanalyzed all data using only the analysis window between 367 and 1800 ms after word onset, again comparing proportion of looking to the named vs unnamed object; although children looked longer to the correct object overall [$F(1,69) = 8.84$, $p = 0.004$], there was no effect of group and no group by object interaction (both $F < 1$). Thus, irrespective of the method of testing, it appears that children performed equally well at test regardless of the noise level during learning.

We also looked at the reaction times for children's eye gaze shifts to the correct object, on the assumption that slower RTs might be indicative of weaker representations. This analysis was based on a total of 48 participants (17 control participants, 15 participants in the 5 dB condition, and 16 participants in the 0 dB condition). While there was a trend toward slower reaction times in the harder condition, this difference was not significant. (Only including children for whom there were at least 3 RTs did not change this pattern.) Children shifted their eye gaze, on average, 886 ms after word onset in the control condition, compared to 841 ms for the 5-dB condition and 1003 ms for the 0 dB condition. A one-way ANOVA showed no significant effect of condition, $F(2,45) = 1.02$, $p > 0.35$, partial eta-squared $= 0.043$. However, it is clearly possible that this null result was the effect of the extreme variability in reaction time data. Still, we again find no clear indication of poorer learning or weaker representations for children who were taught the words in the presence of noise.

Finally, we examined whether children's general language skills (as measured on the SLAS) related to performance on the looking task (the difference score between looking to the correct object when named as compared to the baseline condition). This was calculated for each condition separately, using Spearman rank correlations (Wessa, 2012). There was a significant relationship for the +5 dB SNR, $R_s = 0.41$, $p < 0.05$. However, there was no significant relationship in either the clear condition, $R_s = 0.28$, $p > 0.05$, or for the 0 dB SNR, $R_s = 0.10$, $p > 0.05$. Looking across conditions, this correlation was only marginal, $R_s = 0.22$, $p < 0.10$. Thus, the findings are somewhat unclear; it may be that children whose parents rate them as having better language skills did indeed spend relatively more time looking at the appropriate object during testing, but only for the easier noise condition. The lack of an effect in the other two conditions, however, raises the possibility that this correlation may be simply spurious.

IV. DISCUSSION

The current study suggests that children were no worse at learning words in the presence of equal-amplitude multitalker babble than they were at learning them in quiet. Children showed no significant differences in their looking behavior or reaction times to identify the correct referent of novel words when those words had been taught in quiet, in the presence of background speech at a +5 dB SNR, and in the presence of background speech at a 0 dB SNR. These results are quite surprising; although prior work has shown that children aged 24 months are able to recognize known words in similar noise levels, we predicted that the ability to learn new words would be affected by noise to a greater degree. Apparently that is not the case; children were quite adept at learning new words at even quite difficult noise levels.

It is important to note, however, that we do not actually have a good measure of children's word *recognition* abilities in noise at the age tested here, making it difficult to know what to make of children's successful learning performance. Newman (2011) tested children aged 24 months and found some (minimal) ability to recognize at least one known word at levels as low as −5 dB. We did not opt to test a noise level as difficult as this in the present study, primarily because word recognition performance in Newman's study was quite inconsistent at this level; more consistent recognition required a SNR of at least 0 dB (a value tested here). While children in the present study are significantly older, Grieco-Calub *et al.* tested word recognition in normal-hearing children aged 30 months, closer to the age tested here, and found noticeably poorer performance in noise than in quiet settings even at a relatively easy 10 dB SNR. It seems unlikely that performance would be markedly improved by 34 months, and thus one might have expected to see noticeable effects on word learning at SNRs as high as 10 dB, let alone the 5 and 0 dB tested here. But while there are likely to be only minimal changes in thresholds between 24 and 34 months (Tharpe and Ashmead, 2001), word recognition in noise has not been sufficiently tested to make any strong claims in this regard; perhaps children aged 34 months have developed better ability to understand and learn from speech in noise than the younger children in these earlier studies. Future work should more directly compare across word learning and word recognition in the same ages, or ideally in the same children, so as to examine whether noise affects learning differently than recognition.

The children's word-learning performance in the present study was aided, no doubt, by the fact that the two words were quite dissimilar phonetically. It might well be possible to cause children to fail by using minimal pairs or other extremely difficult words. Indeed, Werker *et al*. (2002) have shown that when words become more similar, more resources are required to discriminate them for word learning even in quiet settings. Yet the point of this study was not to make the hardest possible task for the children, but rather to create something that might be a reasonable analog to the types of situations faced outside the laboratory. While children do have minimal pairs in their vocabulary (cf. Dollaghan, 1994), certainly many of the words children experience are as dissimilar from one another as are the two examples chosen here.

Still, the fact that these words were relatively unlike any other words children were likely to already know undoubtedly helped them succeed at the task, since they could do so with only partial representations. A more realistic task might well include words that were more similar to ones already known. Prior work has suggested that children distinguish known words from mispronounced versions (Swingley, 2003; Swingley and Aslin, 2002), suggesting that they do form detailed representations in their typical word learning. However, such work has also suggested that toddlers may fail to learn new words that are phonologically similar to known words (Swingley and Aslin, 2007), even in quiet settings; this would make their ability to do so in the presence of distraction more difficult to assess. That is, if a child failed to learn a new word, it would be unclear whether this was a result of the presence of noise during learning, or whether it was a continuation of their typical pattern of failing to learn words that compete with known lexical items.

It is also likely that children's performance was abetted by the fact that there were only two choices on the screen during the test phase. This is, in essence, a closed-set recognition task; closed-set tasks are a typical means of testing young children, but this choice of testing methods can have a variety of effects on the results obtained. Not only do closed-set, limited-choice tasks typically result in greater accuracy scores, but they are also less likely to show effects of lexical competition or of talker variability (Clopper *et al*., 2006). Such differences may suggest that "the basic perceptual processes used to recognize spoken words in open-set and closed-set tasks are not equivalent" (Clopper *et al*., 2006, p. 348). With regards to the current design, children may have been able to select the correct choice at test even if their representations for the new words were incomplete or fragmentary. This may be one reason why noise appeared to have so little effect: even if the children received only partial information about the new lexical items during training, that may have been sufficient to distinguish the word from its single competitor in the two-alternative testing format.

Their performance likely also was supported by repeated presentations of the target words. During training, the target words were only repeated three times per trial, but children heard four training trials for each word (for a total of 12 repetitions). These repeated exposures may have been sufficient to compensate for difficulty perceiving the target word on any given occasion. Moreover, the words were repeated five times per trial during the test phase. These repeated presentations resulted in a large analysis window, and may have prevented us from seeing any fine-grained changes in looking behavior across conditions. They may also have allowed children to better access weak representations, preventing us from seeing differences that truly existed in the representations they formed in the presence of noise. Presenting the words at test only a single time would have avoided these potential concerns, but would have left open the possibility that variation in performance was the result of differences in attention or comprehension at the time of test, rather than differences in learning the words originally. It remains unclear how best to satisfy these divergent methodological concerns.

It is also possible (indeed, likely) that effects of noise would have been stronger if we had targeted children with hearing loss (as found by Blaiser, 2010). Children and adults with hearing loss are generally affected by noise to a greater degree than are those with typical hearing (Finitzo-Hieber and Tillman, 1978) and this would likely affect their ability to learn words in noisy environments. Indeed, children with hearing loss generally have smaller vocabularies than their normal-hearing peers (Lederberg and Spencer, 2001), suggesting that their poor perceptual skills make it more difficult to learn words.

Finally, the present study used only one form of distractor, that of multi-talker babble. We selected this type of distractor both because of its prior use in related studies and because it is a type of distractor that children likely encounter quite often. Nonetheless, the results might be quite different with other forms of masking. Multi-talker babble has spectral overlap with the speech targets, and can thus provide both energetic and informational masking, rather than just informational masking alone. It may also attract attention in a manner different from speech-shaped noise or other noise-based distractors. However, babble also tends to have a relatively flat amplitude modulation over time and poses less threat of informational masking than would a single-voice distractor (Newman, 2009), since no one voice in the babble mixture is likely to stand out. Thus, children's learning ability in noise may depend critically on the particular type of distractor present in the environment.

Another possibility is that the presence of noise may have actually encouraged infants' attention, counteracting any decrement caused by the masking itself. Given a more difficult task, infants may have essentially "tried harder" to hear what was being said, and these added cognitive resources could have compensated for any difficulty caused by the noise. The trend towards better recognition for the harder SNR (8% increase over baseline looking) than the easier SNR (4% increase over baseline looking) would support such an argument. Clearly, more research would be needed to explore such a possibility, but the notion that small amounts of noise could actually aid performance is an intriguing direction for future research.

Prior research (Blaiser, 2010; Riley and McGregor, 2012) has likewise found only limited effects of noise on word learning, at least at noise levels in which speech recognition is relatively good. This may suggest that the effect of noise on word learning and word recognition are quite comparable. Such a result would be surprising, since listeners have the opportunity to use top-down knowledge to aid in their recognition of known words, something which they cannot rely on for learning new words. However, a number of studies have suggested that young children do not rely on top-down information to the same extent as do adult listeners (Newman, 2006; Nittrouer and Boothroyd, 1990) and instead seem to weight the bottom-up perceptual aspects of the signal more heavily (Fallon *et al.*, 2000). If toddlers do not rely on their prior knowledge to aid their recognition, they might be affected by noise to similar degrees in the different tasks, in a manner very unlike that from adults.

In sum, these findings suggest that 32- to 36-month-olds are surprisingly adept at learning new words in a multi-talker environment. This bodes well for their ability to acquire language in the varied environments in which they may find themselves. Children's skill at learning words even in noisy environments has clear implications for theories of word learning, which often presume either that the child is hearing language input in ideal listening conditions or that the input has already been separated from background noise. Since the input children receive is a critical factor in the rate and success of language acquisition (e.g., Bornstein *et al.*, 1998; Hart and Risley, 1995; Pan *et al.*, 2005; Pine *et al.*, 1997), understanding the extent to which children are affected by the presence of background noise is important in evaluating the adequacy of such theories. The present findings suggest that it may indeed be appropriate to develop theories of learning without reference to the need to separate target speech from background noise, at least for the moderate levels of noise tested here.

These findings also have implications for child-rearing practices, particularly with regards to noise limits in day cares. The impact of day care vs home child care on children's development has been a topic of much societal debate, and one of the many differences between settings has to do with typical noise levels. Here we found that children are still able to acquire new words in relatively noisy conditions. While this might be taken to suggest that the level of noise in typical day cares is not a major factor in this societal debate, the current study only looked at a single pair of words that were highly distinct; children might still have difficulty distinguishing more similar words, and future work is needed to explore this possibility.

ACKNOWLEDGMENTS

This project was an undergraduate honors thesis for the first author. We particularly thank Dr. Yasmeen Faroqi Shah for serving on the committee and providing helpful feedback, and Giovanna Morini for assistance in experimental setup and many helpful discussions. We also thank Elizabeth Johnson's lab and George Hollich for the various coding and analysis programs used here. We thank the following for assistance in scheduling, testing, and coding participants: Mikayla Abrams, Faraz Ahsan, Jaclyn Albalah, Alison Arnold, Amelie Bail, Catherine Bender, Taryn Bipat, Devon Brunson, Alyssa Cook, Jennifer Coon, Sara Dougherty, Catherine Eaton, Lauren Evans, Andrea Fisher, Arielle Gandee, Sean Hendricks, Mina Javid, Esther Kim, Penina Kozovsky, Stephanie Lee, Perri Lieberman, Rachel Lieberman, Kelly McPherson, Vidda Moussavi, Molly Nasuta, Maura O'Fallon, Amanda Pasquarella, Allie Rodriguez, Rebecca Sherman, Katie Shniderman, Veronica Son, Ashley Thomas, Krista Voelmle, Kimmie Wilson, and Catherine Wu. This work was funded by a Maryland Center for Undergraduate Research Summer Fellowship, and by NSF grant 0642294 to UMD.

Allen, P., and Wightman, F. L. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively; infants do not," Psychol. Sci. **5**(3), 170–174.

Bergelson, E., and Swingley, D. (**2012**). "At 6–9 months, human infants know the meanings of many common nouns," Proc. Nat. Acad. Sci. **109**(9), 3253–3258.

Blaiser, K. M. (**2010**). "Effects of noise on fast mapping and word learning scores in preschool children with and without hearing loss," Diss. Abstr. Int., B **71**(2-B), 966.

Bloom, P. (**2000**). *How Children Learn the Meanings of Words* (MIT Press, Cambridge, MA).

Bornstein, M. H., Haynes, M. O., and Painter, R. M. (**1998**). "Sources of child vocabulary competence: A multivariate model," J. Child Lang. **25**, 367–393.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**(6), 3132–3139.

Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (**2006**). "Effects of open-set and closed-set task demands on spoken word recognition," J. Am. Acad. Audiol. **17**, 331–349.

Crandell, C. C., and Smaldino, J. J. (**2000**). "Classroom acoustics for children with normal hearing and with hearing impairment," Lang., Speech, Hear. Serv. Sch. **31**, 362–370.

Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (**2012**). "Word learning under adverse listening conditions: Context-specific recognition," Lang. Cognit. Process. **27**(7–8), 1021–1038.

Dollaghan, C. A. (**1994**). "Children's phonological neighborhoods: half empty or half full?," J. Child Lang. **21**, 257–271.

Elliott, L. L., Longinotti, C., Clifton, L., and Meyer, D. (**1981**). "Detection and identification thresholds for consonant-vowel syllables," Percept. Psychophys. **30**(5), 411–416.

Fallon, M., Trehub, S. E., and Schneider, B. A. (**2000**). "Children's perception of speech in multitalker babble," J. Acoust. Soc. Am. **108**(6), 3023–3029.

Fernald, A., and Hurtado, N. (**2006**). "Names in frames: Infants interpret words in sentence frames faster than words in isolation." Dev. Sci. **9**(3), F33–F40.

Fernald, A., Perfors, A., and Marchman, V. A. (**2006**). "Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year," Dev. Psychol. **42**(1), 98–116.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., and McRoberts, G. W. (**1998**). "Rapid gains in speed of verbal processing by infants in the 2nd year," Psychol. Sci. **9**(3), 228–231.

Finitzo-Hieber, T., and Tillman, T. W. (**1978**). "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21**, 440–458.

Frank, T., and Golden, M. V. (**1999**). "Acoustical analysis of infant/toddler rooms in daycare centers," J. Acoust. Soc. Am. **106**(4), 2172.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., and Gordon, L. (**1987**). "The eyes have it: Lexical and syntactic comprehension in a new paradigm," J. Child Lang. **14**, 23–45.

Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., and Cowan, N. (**2000**). "The development of auditory attention in children," Front. Biosci. **5**, 108–120.

Grieco-Calub, T. M., Saffran, J. R., and Litovsky, R. Y. (**2009**). "Spoken word recognition in toddlers who use cochlear implants," J. Speech, Lang., Hear. Res. **52**, 1390–1400.

Hadley, P., and Rice, M. L. (**1993**). "Parental judgments of preschoolers's-peech and language development: A resource for assessment and IEP planning," Seminars Speech Lang. **14**, 278–288.

Hart, B., and Risley, T. R. (**1995**). *Meaningful Differences in the Everyday Experience of Young American Children* (Brooks, Baltimore).

Hétu, R., Truchon-Gagnon, C., and Bilodeau, S. A. (**1990**). “Problems of noise in school settings: A review of literature and the results of an exploratory study,” J. Speech-Lang. Pathol. Audiol. **14**(3), 31–39.

Hollich, G. (**2005**). Supercoder: A program for coding preferential looking (Version 1.5) (Computer Software), Purdue University, West Lafayette, IN.

Hollich, G. (**2006**). “Combining techniques to reveal emergent effects in infants’ segmentation, word learning, and grammar,” Lang. Speech **49**(1), 3–19.

Horst, J. S., and Samuelson, L. K. (**2008**). “Fast mapping but poor retention by 24-month-old infants,” Infancy **13**(2), 128–157.

Lederberg, A. R., and Spencer, P. E. (**2001**). “Vocabulary development of deaf and hard of hearing children,” in *Context, Cognition, and Deafness*, edited by M. D. Clark, M. Marschark, and M. Karchmer (Gallaudet University Press, Washington, DC), pp. 88–112.

Leibold, L. J., and Neff, D. L. (**2007**). “Effects of masker-spectral variability and masker fringes in children and adults,” J Acoust. Soc. Am. **121**(6), 3666–3676.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). “One factor underlies individual differences in auditory informational masking within and across age groups,” Percept. Psychophys. **65**, 396–406.

Manlove, E. E., Frank, T., and Vernon-Feagans, L. (**2001**). “Why should we care about noise in classrooms and child care settings?,” Child Youth Care Forum **30**(1), 55–64.

Markman, E. M., and Wachtel, G. F. (**1988**). “Children’s use of mutual exclusivity to constrain the meanings of words,” Cognit. Psychol. **20**, 121–157.

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scot, S. K. (**2012**). “Speech recognition in adverse conditions: A review,” Lang. Cognit. Process. **27**, 953–978.

Mills, J. H. (**1975**). “Noise and children: A review of literature,” J. Acoust. Soc. Am. **58**(4), 767–777.

Nábelek, A. K., and Robinson, P. K. (**1982**). “Monaural and binaural speech perception in reverberation for listeners of various ages,” J. Acoust. Soc. Am. **71**(5), 1242–1248.

Neuman, A., Wroblewski, M., Hajicek, J., and Rubinstein, A. (**2010**). “Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults,” Ear Hear. **31**(3), 336–344.

Newman, R. S. (**2006**). “Perceptual restoration in toddlers,” Percept. Psychophys. **68**, 625–642.

Newman, R. S. (**2009**). “Infants’ listening in multitalker environments: Effect of the number of background talkers,” Atten., Percept., Psychophys. **71**(4), 822–836.

Newman, R. S. (**2011**). “2-year-olds’ speech understanding in multi-talker environments,” Infancy **16**(5), 447–470.

Nittrouer, S., and Boothroyd, A. (**1990**). “Context effects in phoneme and word recognition by young children and older adults,” J. Acoust. Soc. Am. **87**(6), 2705–2715.

Oh, E. L., Wightman, F. L., and Lutfi, R. A. (**2001**). “Children’s detection of pure-tone signals with random multitone maskers,” J. Acoust. Soc. Am. **109**(6), 2888–2895.

Oviatt, S. L. (**1980**). “The emerging ability to comprehend language: An experimental approach,” Child Dev. **51**, 97–106.

Pan, B., Rowe, M., Singer, J., and Snow, C. (**2005**). “Maternal correlates of growth in toddler vocabulary production in low income families,” Child Dev. **76**, 763–782.

Pine, J. M., Lieven, E. V. M., and Rowland, C. F. (**1997**). “Stylistic variation at the ‘Single-Word’ stage: Relations between maternal speech characteristics and children’s vocabulary composition and usage,” Child Dev. **68**(5), 807–819.

Polka, L., Rvachew, S., and Molnar, M. (**2008**). “Speech perception by 6- to 8-month-olds in the presence of distracting sounds,” Infancy **13**(5), 421–439.

Riley, K. G., and McGregor, K. K. (**2012**). “Noise hampers children’s expressive word learning,” Lang., Speech, Hear. Serv. Sch. **43**, 325–336.

Smyth, V. (**1979**). “Speech reception in the presence of classroom noise,” Lang., Speech, Hear. Serv. Sch. **10**(4), 221–230.

Swingley, D. (**2003**). “Phonetic detail in the developing lexicon,” Lang. Speech **46**(2–3), 265–294.

Swingley, D., and Aslin, R. N. (**2002**). “Lexical neighborhoods and the word-form representations of 14-month-olds,” Psychol. Sci. **13**(5), 480–484.

Swingley, D., and Aslin, R. N. (**2007**). “Lexical competition in young children’s word learning,” Cognit. Psychol. **54**, 99–132.

Tharpe, A. M., and Ashmead, D. H. (**2001**). “A longitudinal investigation of infant auditory sensitivity,” Am. J. Audiol. **10**, 104–112.

Werker, J. F., Fennell, C. T., Corcoran, K. M., and Stager, C. L. (**2002**). “Infants’ ability to learn phonetically similar words: Effects of age and vocabulary size,” Infancy **3**(1), 1–30.

Wessa, P. (**2012**). Spearman Rank Correlation (v1.0.1) in Free Statistics Software (Version 1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_spearman.wasp/.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). “Children’s detection of pure-tone signals: Informational masking with contralateral maskers,” J. Acoust. Soc. Am. **113**(5), 3297–3305.