The Official Journal of the International Society on Infant Studies
Infancy

Infancy, ***16***(5), 447–470, 2011

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/j.1532-7078.2010.00062.x

WILEY-BLACKWELL

2-Year-Olds' Speech Understanding in Multitalker Environments

Rochelle S. Newman
Department of Hearing & Speech Sciences
University of Maryland

Infants and toddlers are often spoken to in the presence of background sounds, including speech from other talkers. Prior work has suggested that infants 1 year of age and younger can only recognize speech when it is louder than any distracters in the environment. The present study tests 24-month-olds' ability to understand speech in a multitalker environment. Children were presented with a preferential-looking task in which a target voice told them to find one of two objects. At the same time, multitalker babble was presented as a distracter, at one of four signal-to-noise ratios. Children showed some ability to understand speech and look at the appropriate referent at signal-to-noise ratios as low as −5 dB. These findings suggest that 24-month-olds are better able to selectively attend to an interesting voice in the context of competing distracter voices than are younger infants. There were significant correlations between individual children's performance and their vocabulary size, but only at one of the four noise levels; thus, it does not appear that vocabulary size is the driving factor in children's listening improvement, although it may be a contributing factor to performance in noisy environments.

Understanding spoken language is a remarkable task. No two talkers speak in exactly the same way; thus, understanding speech requires adjusting for differences across individuals (Houston & Jusczyk, 2000; Pisoni, 1992). It also requires adjusting for changes in the environment: speech may be

Correspondence should be sent to Rochelle S. Newman, Department of Hearing & Speech Sciences & Program in Neuroscience & Cognitive Science, University of Maryland, 0100 Lefrak Hall, College Park, MD 20742. E-mail: rnewman@hesp.umd.edu

produced in settings with reverberation or background noise. Listening in noise poses a wide range of problems for the listener. First, if the noise involves the same frequency range as the target, there is the potential for energetic masking, or masking of the acoustic properties at the auditory periphery. Second, background noise can also cause uncertainty as to which components of the signal belong together. This is particularly likely to be the case when the ''noise'' is that of other talkers. As such, listening in background speech is particularly problematic for many groups of individuals, including the elderly, individuals with hearing loss, and individuals listening to a non-native language (Cooke, Garcia Lecumberri, & Barker, 2008; Peters, Moore, & Baer, 1998; Summers & Molis, 2004).

Children also are faced with listening to speech in the presence of background noise. And as with adults, such background noise can consist of low-level machine noise (as in the background hum of an air-conditioner unit), noise from the environment (both the noise of wind and rain, and of manmade objects such as airplanes flying overhead), and noise from other people. The latter is likely to be particularly problematic, as it is for adult listeners.

A great deal of work has examined the noise levels children typically face. Schools and preschools are noisy environments (Frank & Golden, 1999; Golden & Frank, 2000; Manlove, Frank, & Vernon-Feagans, 2001; Picard & Bradley, 2001), and such noise has the potential to interfere with children's learning and understanding (Fallon, Trehub, & Schneider, 2000; Hétu, Truchon-Gagnon, & Bilodeau, 1990; Manlove et al., 2001; Mills, 1976; Smyth, 1979). For example, Golden and Frank (2000; personal communication) measured signal-to-noise ratios (SNRs) in five occupied toddler classrooms and found that the background sound, which often consisted of speech from other children, was typically within 15 dB of the teacher's voice; moreover, during book-reading time, the SNR for different teachers averaged only 5–6 dB. Yet, infants below 1 year of age do not even recognize their name in background speech unless the SNR is at least 10 dB (Newman, 2005), suggesting that the noise levels found in preschool settings may be quite challenging.

Infants and young children are likely to face particular problems listening in the presence of background sound compared to adult listeners. First, infants and children are generally less adept at focusing their attention selectively to a given signal or a given component of a signal, at least through 5 years of age (Allen & Wightman, 1994, 1995; Bargones & Werner, 1994; Stellmack, Willihnganz, Wightman, & Lutfi, 1997; see Gomes, Molholm, Christodoulou, Ritter, & Cowan, 2000, for a review). Although such testing has often focused on tone detection in laboratory situations, the presumption is that this difficulty in selective attention (or in the use of listening strategies) would translate to greater difficulty attending to a single talker in a multitalker setting as well.

Second, young children require greater SNRs in order to detect and discriminate speech sounds or words than do adults (Fallon et al., 2000; see Mills, 1976, for a review). For example, Fallon and colleagues demonstrated that 5-year-olds required SNRs that were 5 dB more favorable than adults in order to identify words (although see Elliott et al., 1979, for results suggesting that age-related differences may be primarily the result of poorer performance in quiet settings).

Third, children are more sensitive to informational masking than are adults. Informational masking (Durlach, 2006) refers to a failure to segregate sound sources or attend selectively as a result of uncertainty regarding the signal. It can be thought of as any form of difficulty in separating a signal from noise that goes beyond energetic masking (see Kidd, Mason, Richards, Gallun, & Durlach, 2008; Mattys, Brooks, & Cooke, 2009, for further discussions of this distinction). Both preschoolers and school-aged children show more informational masking than do adults (see, e.g., Leibold & Neff, 2007; Lutfi, Kistler, Oh, Wightman, & Callahan, 2003; Oh, Wightman, & Lutfi, 2001; Polka, Rvachew, & Molnar, 2008; Wightman, Callahan, Lutfi, Kistler, & Oh, 2003).

Finally, toddlers and young children have less experience with their native language, and thus cannot rely on prior linguistic knowledge to help them compensate for noise as well as adult listeners can. In this sense, the situation they face is more comparable to that faced by second-language learners than by native adult speakers of a language, and several studies have suggested that second-language learners have particular difficulties comprehending or recognizing speech in the presence of noise and other distracters (Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997; Takata & Nábelek, 1990). Children 13 years and under seem to likewise have poorer ability to use knowledge of lexical and syntactic rules to recognize speech in noise (Elliott, 1979), showing less of an advantage for highly predictable words in noise compared to low-predictable words. Given this range of limitations, we might expect toddlers' speech recognition to be especially handicapped by the presence of noise, including the noise caused by other talkers (Werner & VandenBos, 1993).

Recently, a number of studies have begun exploring young infants' ability to understand speech in the presence of background speech. These studies have suggested that infants require a much higher SNR than adults in order to recognize well-known words (Barker & Newman, 2004; Hollich, Newman, & Jusczyk, 2005; Newman, 2005, 2009; Newman & Jusczyk, 1996). For example, Newman (2005) presented infants of three different ages (5, 9, and 13 months) with the sound of a person repeating a name over and over in the presence of multitalker babble. Infants will typically listen longer to their own name than to other children's names in quiet environments, and Newman (2005) explored the SNR levels at which infants would continue to do so in a multitalker environment. She found that 5- and 9-month-old infants required a minimum of a 10 dB SNR in order to recognize their name in the presence of background speech; 13-month-olds performed slightly better, succeeding with a 5 dB SNR. In contrast, adult listeners' thresholds for speech intelligibility in white noise have been estimated as being in the range of −8 dB SNR (Hawkins & Stevens, 1950), and thresholds for listening in multitalker babble range from −6 to −12 dB SNR (Bronkhorst & Plomp, 1992). Similarly, Trehub, Bull, and Schneider (1981) tested both infants and adults on their ability to detect speech in broadband noise and reported that infants up to 24 months of age show masked thresholds 9–16 dB higher than adults. Thus, infants' ability to recognize and detect speech in the presence of distracters appears to be far worse than that of adults, yet it is unclear what type of developmental change is occurring. Given this fact, examining performance in young children, who are in the midst of rapid vocabulary acquisition, would provide critical insight into the real-world problems faced by children learning language, and this is the primary goal of the present study.

A secondary question revolves around what might drive such developmental change. Obviously, increases in auditory discrimination ability are one potential source of improvement. But Trehub et al. (1981) found little difference in speech-detection-in-noise performance between 6 and 24 months of age. Similarly, Olsho (1985) reports that infants have essentially adult-like frequency resolution (filter widths) by 6 months of age. Such findings, combined with the fact that Newman (2005) found an improvement between 9 and 13 months of age for infants' recognition of their name in background speech is enough to suggest that hearing at the level of the auditory periphery (or pure sensory ability) is unlikely to be the only factor.

Werner and Bargones (1992) suggest that many of the differences in auditory ability between infants and adults are the result of "nonsensory processes" (see also Werner & Bargones, 1991; but see Fallon et al., 2000, for an opposite finding with 5-year-olds). Allen and Wightman (1994) suggest that poor attentional control or the use of nonoptimal strategies (Allen & Wightman, 1995) may help to explain the large individual differences found among 3-year-olds and 4-year-olds in their studies, and this would presumably be the case for younger children as well.

Newman (2005) suggested that the change in streaming performance in her study might be related to improvements in infants' general lexical ability. The improvement in stream segregation occurred at an age at which infants typically begin saying their first words; while this could be coincidental, it could also suggest that as infants' understanding of how words are used to communicate improves, this may result in changes in the ability to separate streams of speech as well. If this is the case, we might expect to see quite significant improvements by the time a child reached 2 years of age (and had a substantially larger vocabulary). We also might expect to find that children who had more advanced vocabularies would do better at understanding speech in difficult listening situations than do children of the same age whose vocabularies are less developed, even when the target words themselves are known to both groups.

Several studies have suggested that children's vocabulary size is not related to their ability to distinguish between different words in quiet settings. For example, Swingley and Aslin (2000) found that 18-month-olds generally did not confuse items that differed by a single phoneme ("tog" and "dog"), and that there was no relationship between this discrimination ability and the children's vocabulary size. Likewise, Bailey and Plunkett (2002) reported that while children (aged 18 and 24 months) discriminated words from mispronunciations, this was unaffected by age or vocabulary size, and Nazzi and colleagues (Havy & Nazzi, 2009; Nazzi, 2005; Nazzi, Floccia, Moquet, & Butler, 2009; Nazzi & New, 2007) found no relationship between vocabulary size and infants' (aged 16, 20, or 30 months) ability to use fine-grained phonetic details in word-learning. In contrast, Werker, Fennell, Corcoran, and Stager (2002) did find such a relationship in a word-learning task, but only among younger infants with smaller vocabulary sizes.

Thus, many studies have failed to find relationships between speech perception or word-learning skills and vocabulary size, particularly with children aged 18 months and up, and the combination of results might suggest that our children (aged 24 months) would likewise not demonstrate any such relationship. However, these prior results all involved listening in a quiet environment; listening in the presence of other speech represents a far more difficult task than does word discrimination in quiet, and language skill could potentially play a larger role in this situation. Perhaps the reason why prior studies with older infants have failed to find relationships between discrimination and vocabulary is that the tasks in those studies were not sufficiently challenging for the infants (as compared to being because the infants themselves no longer differed).

To begin to explore these issues, the present study evaluates the SNRs required for 24-month-old children to understand speech in multitalker babble. We use a variant of the preferential-looking procedure, in which children see two visual objects at each point in time, and hear a voice telling them which object to look at. This voice is combined with a background of multitalker babble, at different SNRs; we expect excellent performance when the SNR is sufficiently high, but decreasing performance as the SNR level becomes more challenging.

Prior work with infants has used a very different task, and this makes direct comparisons across studies difficult. But if toddlers can understand speech at a more difficult SNR than do younger infants, it might suggest that children's developing linguistic knowledge is an important factor in this ability. To further explore this, we also collected vocabulary information from the individual participants, in order to evaluate whether those children who had larger vocabulary sizes were more likely to succeed with the more difficult SNRs.

EXPERIMENT

In the present study, 24-month-olds were presented with pairs of images representing well-known words. They heard a voice telling them to "Find the ___!" in the presence of multitalker babble at one of four different SNRs: 10 dB SNR (in which the target voice was 10 dB more intense than the distracter babble), 5 dB SNR, 0 dB SNR, or −5 dB SNR (in which the babble was 5 dB more intense than the target voice). We predicted that children would look longer at the appropriate (or named) image as long as the SNR was not so great as to preclude their comprehension.

The easier two levels included in this study are the SNR levels used in Newman's (2005) study of infants; the 10 dB level was one at which even the youngest infants in Newman's study were able to succeed, whereas the 5 dB SNR was too difficult for all but the oldest infants in that study (13 months). As the current participants are 24 months of age, we might expect them to succeed at both SNR levels, if the words are sufficiently well-known. We therefore also included two harder SNR levels, 0 dB SNR and −5 dB SNR.

The current task is more difficult in many ways than the one facing the infants in Newman (2005). In the earlier study, the target voice repeated a name, in isolation, 15 times against the backdrop of multitalker babble. The present study involved words presented in the context of fluent speech, rather than in isolation. Moreover, while the words were repeated ("Look at the doggy!" "Can you find the doggy?"), this was done only five times per trial, rather than 15: if children can make out words in multitalker babble occasionally, but not consistently, the increased repetition level in the earlier study could aid performance. Finally, the prior task was merely one of preference: it tested whether infants listened longer to a familiar name than a novel one. In contrast, the present study required not only that the children notice the familiarity of a word, but also that they match the word to its appropriate referent. Thus, while the SNRs used in the present study might lead us to expect successful performance, these levels when combined with a more difficult task might still result in children failing to succeed. On the other hand, one potential advantage to the current task is that the images appeared on introductory trials prior to naming trials; the images could thus potentially prime the to-be-detected word, making it easier to detect in the presence of distracters.

In addition to examining the effect across SNRs, we also examine whether individual differences in parent-reported vocabulary size related to children's performance on the task. While we are not selecting children on the basis of vocabulary differences, finding such correlations might provide insights into the causes of performance variability that would be useful for future research.

METHODS

Participants

A total of 96 toddlers (56 males, 40 females) participated in this study; 24 heard each of the four SNR levels. An additional 54 children were tested, but their data were excluded for the following reasons: not being a native learner of English ($n = 9$), equipment problems ($n = 13$), experimenter error ($n = 14$), fussiness ($n = 14$), difficulty viewing the eyes for coding purposes ($n = 1$), and parental interference ($n = 3$). For the 10 dB SNR condition, the children averaged 2 years, 11 days; for the 5 dB SNR they averaged 2 years, 10 days; for the 0 dB SNR condition, they averaged 2 years, 12 days; and for the −5 dB SNR condition they averaged 2 years, 12 days (overall range: 23 months 4 days to 25 months 15 days). Eight of the parents declined to complete the vocabulary assessment; thus, correlations with vocabulary are based on 88 participants. One child (in the −5 dB condition) did not attend to one of the three animal pairs; data from the cat–dog pair is thus based on 95 participants rather than 96.

Stimuli

In order to ensure that the results would be generalizable, we used three different stimulus sets, each consisting of a pair of video images combined with audio target stimuli and background babble. The three pairs were "kitty" and "doggy," "birdie" and "horsie," and "cow" and "pig." Thus, four target words were bisyllabic, and two were monosyllabic. Animals were presented in the same pairs throughout the study (thus, kitty and doggy were always presented together); these pairings were based on similarity of vocabulary items. The reported results collapse across all six target items, but we also examined the word pairs individually in order to examine generality.

Video images

Visual stimuli consisted of pairs of videotapes, each approximately 10 min long. In order to make the task more interesting for the children, different types of images were used for the three animal pairs. The cat and dog were live videos of relatively calm animals: a typical black cat, and a typical yellow lab. These stimuli had previously been used in Newman (2006). The horse and bird were plastic animals; the (brown) horse was made to "move" back and forth across the screen by means of clear fishnet wire, while the bird (a red cardinal) was battery-operated and moved its beak and tail feathers on its own. The cow and pig were both soft puppets of similar size. All animal pairs were approximately the same size on the screen, and were taped against similar backgrounds.

Target audio recordings

Three different female speakers were recorded producing the target stimuli for this study; one speaker produced the cow/pig recordings, a second produced the kitty/doggy recordings, and a third produced the horsie/birdie stimuli. All were native speakers of American English, and recorded the target sentences in an infant-directed speaking style.

The stimuli consisted of five sentences, each containing a target word ("Can you find the horsie?" "Where is that cow?") along with occasional interjections ("Oh boy!"). All three speakers also recorded neutral stimuli that did not indicate a particular target animal ("Oh look, do you see that?"). On each trial, the child heard the stimulus for that trial played in its entirety; overall stimulus durations were 10 sec for the dog/cat pair, 9.7 sec for the cow/pig pair, and 11.5 sec for the bird/horse pair. A trial ended immediately after the end of the audio file (the end of the fifth sentence). Target words within the sentences averaged 510 msec for cow, 443 msec for pig, 664 msec for horsie, 730 msec for birdie, 597 msec for kitty, and 744 msec for doggy. All recordings were made in a noise-reducing soundbooth, recorded over a Shure SM81 microphone (Shure Incorporated, Niles, IL) at a 44.1 kHz sampling rate and 16 bits precision. Each speaker produced several tokens of each sentence, and these were isolated, edited, and their amplitudes were adjusted to be at the same level (root-mean-square [RMS] amplitude of 25 dB less than the computer's maximum level).

Multitalker babble recording

The background multitalker babble was taken from Newman (2005). It consists of speech from nine different women reading passages aloud from a variety of books. To ensure that all nine talkers were speaking at most moments, long pauses in the individual recordings (as when the speaker turned the page, or stopped to clear her throat) were excised. This does not eliminate amplitude variability, as there were still stop consonant closures, breaks for air, short pauses, etc., as well as the natural variation between amplitude levels of different phonemes. This removal made the individual passages sound a bit more like a memorized monolog than like reading from a book, but did not make them sound unnatural. The nine passages were then adjusted to be of the same overall RMS amplitude, and were blended together at equal ratios to create the competing multitalker babble. None of the three target voices was included in this babble mixture. This combined distracter passage was then edited to be the same length as the different target passages, adjusted to be the appropriate intensity level (10 or 5 dB less intense than the target items, the same intensity as the target items, or 5 dB more intense than the target items), and blended with the target recordings.

Procedure

Each toddler heard only one of the four SNR levels. Toddlers were tested with a variant of the preferential-looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). They were seated on their caregiver's lap in a three-sided booth enclosure. The front panel of the booth contained two 20 in. video monitors, separated by 18 in., and a center light. This light flashed at the beginning of each trial to direct the toddlers' attention to the center of the booth (equidistant from the two videos). When the child was attending to the light, the experimenter pressed a button on a computer-controlled response box, signaling the computer to turn off the light and begin the next trial. Each monitor was connected to a computer-controlled switch box. This box received input from the videocassette recorders playing the stimulus tapes, and a black signal from another camera; the computer controlled when the monitors received the input from the videocassette recorder versus the black signal on the basis of the experimenter's button presses. The auditory stimuli were played over a pair of matched NHT Super One speakers NHT Showroom/Offices, Benicia, CA), one located on each side of the enclosure.

Children were tested with the three pairs of animals in three separate sections of the study; the order of these sections was counterbalanced across participants, such that equal numbers of participants heard all six possible orders. Each section of the study began with two 5-sec introductory trials in which only one of the two videos appeared; since toddlers chose to look at the monitor showing a picture (rather than a black screen), these trials ensured that children had an opportunity to see each object before the test trials; which monitor came on first, and which picture showed on which monitor were both counterbalanced across participants.

This was followed by six test trials in which both images were shown simultaneously; two of those trials consisted of neutral stimuli, and two trials labeled each of the two animals. These six trials were blocked (two blocks each consisting of one of each of the three trial types) and the order of trials within each block was randomized. Video images were synchronized to appear immediately after the completion of the first example of the target word (roughly 1.5 sec after the start of the audio presentation). Thus, from the point at which the video came on, the child could have known which was the appropriate video to attend to.

During this session, the caregiver wore headphones playing masking music, preventing him or her from hearing which item was being named. A second video camera recorded the entire session from behind the parent's head. This camera recorded what appeared on the television screens, allowing us to ensure that the videos played correctly, without static or tracking problems. If such equipment difficulties occurred on any trial for a participant, that participant's data were discarded.

In addition to participating in the experimental session, parents were asked to complete the Language Development Survey (LDS; Rescorla, 1989) for their children. This is a screening checklist for estimating productive vocabulary; it consists of 310 words, and parents were asked to indicate which words their child produced. We chose this vocabulary form because prior work in our lab suggested that its short length would result in a larger proportion of completed questionnaires than would the longer MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994). Moreover, the LDS has been shown to have strong test–retest reliability (Rescorla, 1989; Rescorla & Alley, 2001). Although parents cannot reliably report children's receptive vocabularies after approximately 18 months of age (Fenson et al., 1994), parental reports of production are strongly correlated with receptive language skills. The LDS in particular correlates strongly with the Reynell Receptive Language Age Score (Rescorla & Alley, 2001), so it can be taken as a rough estimate of a child's language ability more generally.

Coding

The sessions were video-taped, and two experimenters (blind to condition) individually coded each child's looking behaviors using Habit 2000

(Cohen, Atkinson, & Chaput, 2000). The averages of these two codings were used as the final data; from this, the children's total duration of looking at each video was calculated. Correlations between the two coders for each individual participant ranged from .869 to .998 (with an average correlation of .975). Only five participants out of 96 had correlations below .9. Average difference scores ranged from .13 sec to .56 sec, with an average of .28 sec. Despite this extremely high reliability, there were occasional trials on which coders disagreed. On any trial in which the two coders disagreed by 2 sec or more, a third experimenter coded the same trial, and the two closest results were selected as the final scores for averaging. (Correlations above are based on the original codings, not these final scores.) There were a total of 17 trials (out of 1,722) that required a third coder (less than 1%). Thus, the final data was based on coding that was highly consistent between raters.

There are a number of different measures that have been taken from preferential-looking studies. Looking time can be calculated based on the single longest look, or based on total looking. It can be measured in seconds, or in the proportion of time spent looking at the appropriate versus inappropriate picture. (The fundamental difference between these has to do with how time spent looking at neither image is incorporated; a measure based on raw seconds will be reduced if the child spends half the trial looking at his or her feet, whereas a measure based on the proportion of time looking at the appropriate versus inappropriate object will not be reduced.) Finally, looking time to the correct object can be compared to a hypothetical baseline (50%), to the time spent looking to the same object on the neutral trials, or the time spent looking at the wrong object on the same trials. Each of these possibilities has valid arguments in its favor, and each has found use in the literature at some point. We chose to focus on the proportion of looking to each object on named trials ("Look at the kitty!") compared to neutral (unnamed) trials ("Look at that!"), but also calculated the average total looking to the matching versus mismatching video; results were comparable across measures. We presume that if children can understand the speech, despite the background babble, they will look longer to the video when it was named than in the baseline condition. We used two-tailed tests for all measures, despite having a directional prediction.

RESULTS AND DISCUSSION

We first report overall results, in which we collapsed looking time measures both across the three target word sets, and across the four SNR levels. In general, toddlers spent a larger proportion of time watching the named video than they spent watching that video in the baseline condition, $t(95) = 10.85$, $p < .0001$.[1] The difference was on the order of 11% overall.

However, this effect differed depending on the SNR, $F(3,92) = 9.50$, $p < .0001$. We therefore explored each SNR separately. Not surprisingly, we found that toddlers' performance in the easier two conditions, 10 and 5 dB SNR, was significantly greater than chance. Children watched the appropriate video a larger proportion of time when named than unnamed, in both the 10 dB SNR (mean difference = 16%), $t(23) = 6.88$, $p < .0001$, and 5 dB SNR (mean difference = 15.2%), $t(23) = 7.78$, $p < .0001$, conditions. More surprisingly, they also showed this pattern in the 0 dB condition (mean difference = 8.6%), $t(23) = 4.76$, $p < .0001$, and even in the −5 dB SNR (mean difference = 3.6%), $t(23) = 2.59$, $p = .016$. These data can be seen in Figure 1. Thus, the interaction between performance and SNR appears to be in the *size* of the listening preference, not in its presence or direction. These findings indicate that 2-year-olds were still able to understand the speech (as demonstrated by longer looking at the appropriate object) when the speech was 5 dB less intense than the background babble. Despite this success, however, the difference in looking was quite small in this condition, suggesting that it may be at the limit of children's ability to recognize speech in noise (or, at least, at the limit of what we are able to demonstrate with this testing paradigm).

Looking at the number of children who showed this pattern (of longer looking to the object when named) we found that 21 children (out of 24) showed this pattern at the 10 dB SNR, 23 showed this pattern at the 5 dB SNR, 19 showed it at 0 dB SNR, and 17 showed it at −5 dB SNR. Thus, as the noise level becomes more difficult, fewer toddlers are able to demonstrate understanding of the target speech signal.

It appears that 2-year-olds are able to recognize speech in the presence of background speech babble, at levels as difficult as −5 dB SNR. Could toddlers also succeed at a −10 dB SNR? Given the very small differences found at −5 dB SNR (3.6% and 0.45 sec), we felt it was unlikely that toddlers would succeed at a greater level of noise in the present task. It would seem safe to conclude that while children can make out some speech at a −5 dB SNR, they are unlikely to succeed at a still harder noise level.

---

[1]Although it might seem that a more appropriate approach would be a 2×4 analysis of variance, with the factors of SNR (4) × phase (named versus neutral), this would violate one of the assumptions of parametric statistical tests, that of independence of measures. The reason is that the looking proportions in the neutral condition are not independent from one another; if a child looks at the cat 55% of the time when they hear "Look at that!", then (by definition) they look at the dog 45% of the time, and the average across the two is necessarily 50%. Doing the analysis on the difference scores avoids using either nonindependent baseline measures (per object) or a constant 50% baseline (averaged across objects).

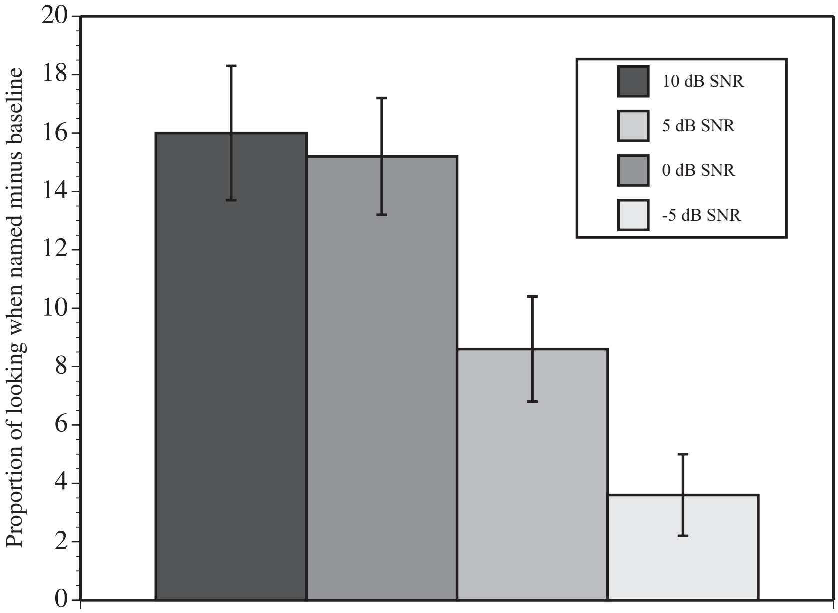

Figure 1 Increase in looking time for the video when named versus when unnamed (baseline condition) at four different signal-to-noise ratios. (Since the baseline condition averages across pairs of animals, baseline itself is 50%.) The error bars represent standard error.

Individual videos

As noted earlier, we used three different word pairs at each SNR in order to ensure that our results were not caused by any single word being easier or harder to hear in noise. However, having found significant overall effects at each SNR, we decided to explore whether this result did differ among the different words. If we find significant differences across the words, this might serve a cautionary role for future research, suggesting a need to test across a range of items. If, on the other hand, the results are comparable across words/videos/speakers, it would suggest that these types of differences across laboratories and studies are less likely to cause apparent differences in the pattern of results.

Looking across the SNRs, we found significant effects (based on the increase in proportion of looking time when named) for each of the six animals, suggesting that children knew and recognized all six items—cat: 8.8%, $t(94) = 3.99$, $p = .0001$; dog: 11.6%, $t(94) = 4.38$, $p < .0001$; bird: 16.9%, $t(95) = 6.78$, $p < .0001$; horse: 13.5%, $t(95) = 5.66$, $p < .0001$; pig: 7.3%, $t(95) = 2.77$, $p = .007$; cow: 6.2%, $t(95) = 2.44$,

$p = .017$. Thus, for all the animals, children spent a longer proportion of time looking to the image when it was named than in the baseline condition. However, when this was examined within each SNR, the effects differed for each of the animals (see Table 1). At 10 dB, the effect was significant for all six animals; at 5 dB, it was significant for five of the six; at 0 dB it was significant for three words; and at −5 dB it was significant only for the word bird. Thus, had we selected only a single pair of animals, we could have found that children's best performance was anything from 5 dB to −5 dB. This highlights the importance of using multiple word pairs in studies with children.

It appears that bird/horse led to the best performance, followed by cat/dog and then pig/cow. This is unlikely to be the result of word-frequency differences, because "kitty" and "dog" are the two words generally learned earliest by children, not "horse" and "bird" (based on the MacArthur-Bates Communicative Development Inventory norms; Dale & Fenson, 1996). Instead, this variability in performance is more likely the result either of differences in how easy the particular phonemes are to hear in noise (e.g., the longer vowel in "horsie" might be easier to hear than the shorter vowel in "kitty"), or the result of differences in how easy these talkers (or these particular productions) are to hear in noise (e.g., acoustic similarity between the target voice and the background speech is likely to have varied across talkers).

To explore this issue, we tested a group of eight adult listeners on a roughly analogous task: Adults were presented with individual sentences and had to select the appropriate referent for each sentence (i.e., the cat or the dog) by pressing a button on a response box. They were presented both with sentences from the recordings that the toddlers heard in the study, and with the same sentences recorded by each of three new talkers. In order to avoid ceiling performance, the SNR was adjusted to be more difficult: −15 dB SNR. Even at this extremely difficult SNR, adult performance was quite high: All listeners performed above 88% correct on each of the animal pairs. Despite that, we found significant differences across animal pairs. Adults showed the same pattern of performance as did the toddlers on the original recordings, demonstrating the strongest performance with the horse–bird pair (94% accuracy), weakest performance on the cow–pig pair (78% accuracy), and intermediate performance on the dog–cat pair (85% accuracy). Two of the three pairs differed significantly: cat–dog versus horse–bird, $t(7) = 3.17$, $p = .024$; cow–pig versus horse–bird, $t(7) = 4.03$, $p = .005$; and cow–pig versus cat–dog, $t(7) = 1.64$, $p > .10$.[2] This replica-

[2]These comparisons are based on rationalized arcsine units; this transformation avoids potential problems caused by using a parametric test on proportions, which are not truly Gaussian.

TABLE 1
Preferences for Individual Videos in Named Minus Baseline Conditions, Across the Four Different SNR Levels. Cells That Are Shaded Represent a Significant Difference, Using Nondirectional t Tests

SNR	*Cat*	*Dog*	*Bird*	*Horse*	*Pig*	*Cow*
10 dB	13.3% $t(23) = 2.59$ $p = .05$	14% $t(23) = 2.14$ $p = .05$	26.5% $t(23) = 4.70$ $p < .0001$	14.4% $t(23) = 2.71$ $p = .05$	14.4% $t(23) = 2.85$ $p = .01$	13.5% $t(23) = 2.89$ $p = .01$
5 dB	9.4% t(23) = 2.09 $p = .05$	26.7% t(23) = 5.52 $p < .0001$	17.5% t(23) = 3.97 $p = .001$	19.4% t(23) = 5.74 $p < .0001$	14.3% t(23) = 3.00 $p = .01$	5.3% t(23) = 0.98 $p > .05$
0 dB	8.0% $t(23) = 2.23$ $p = .05$	6.6% $t(23) = 1.69$ $p = .11$	9.4% $t(23) = 2.14$ $p = .05$	15.8% $t(23) = 3.48$ $p = .002$	1.8% $t(23) = 0.31$ $p > .05$	4.5% $t(23) = 1.10$ $p > .05$
−5 dB	4.6% $t(22) = 1.02$ $p > .05$	−1.2% $t(22) = -0.32$ $p > .05$	13% $t(23) = 2.80$ $p = .01$	5.6% $t(23) = 1.08$ $p > .05$	−0.5% $t(23) = -0.10$ $p > .05$	0.8% $t(23) = 0.14$ $p > .05$

Note. SNR = signal-to-noise ratio.

tion of the basic pattern suggests that the toddlers' poorer performance on the cow–pig pair is not a result of reduced semantic knowledge or poorer visual identification of these images, since presumably these words are all well within the vocabulary knowledge and visual identification skills of our adult participants. If the differences were the result of the discriminability of the words, we might expect to see a similar pattern of performance by the adult listeners when the same sentences were recorded by different voices. However, the data from the three new talkers showed very different results, with significantly lower accuracy for the horse–bird pair (90%), compared to that for cow–pig (96%), $t(7) = 3.29$, $p = .013$, and cat–dog (94%), $t(7) = 2.78$, $p = .039$. This suggests that the acoustic properties of the particular talkers' voices or recordings, rather than simply the discriminability of the word pairs, may have played a role in the toddlers' performance. This may be tied to the choice of masker in the present study; multitalker babble primarily results in energetic (rather than informational) masking, and studies with adults suggest that energetic maskers cause listeners to rely more heavily on acoustic properties of the target speech rather than lexical ones (Mattys et al., 2009).

Our purpose in using different talkers and different words was to assess the generality of our measures; it appears from these findings that children's ability to perceive speech in noise may depend critically on the particular acoustic properties of the talker's productions. Using a range of words and a range of voices may be an important methodological consideration in such studies. In contrast, it does not seem likely that the effects were driven by the particular type of object (e.g., real animal or toy). Prior research with children has not shown such obvious voice or phonetic effects. For example, Newman (2005) tested infants aged 4–13 months on their ability to recognize their name in the presence of background babble. For each age group and condition, a total of five different talkers were presented to the infants, but in none of the four experiments was there any effect of the particular voice or any interaction involving the voice. There was also no difference between the performance of children with monosyllabic names versus performance of those with multisyllabic names. Yet here, with 24-month-olds, we see a much larger effect of the particular talker. It may be that the present task is simply more sensitive to such stimulus differences.

Vocabulary

One possibility is that those children who have more advanced language skills, or larger lexicons, might do better at interpreting speech in the presence of babble. Or, alternatively, those children who are better at listening in babble may have more opportunities to learn language, and thus have larger lexicons as a result. In either case, we might expect to find a relationship between individual children's performance in this task and the size of their lexicons. Across the 96 participants, we saw a wide range of vocabulary scores, with parents reporting that their children said as few as 25 words to as many as 308 words. (The LDS has a total of 310 words, so our participants span nearly the entire range of the assessment.) We examined whether this variability in expressive lexicon size might relate to children's performance in the present study. Given that performance was substantially lower in the −5 dB condition than in the easier SNRs, we conducted these correlations within each SNR condition separately. Performance in the two easier conditions did not seem to relate to children's vocabulary: at a 10 dB SNR, the correlation between reported vocabulary size and the proportion increase in looking to the named object was only $r = .14$ ($p = .26$, one-tailed). Likewise, the correlation at the 5 dB SNR was $r = .25$ ($p = .14$). While both correlations are in the appropriate direction, neither approaches significance. However, the 0 dB SNR did show a strong relationship ($r = .50$, $p = .011$). Yet, at the very hardest SNR (−5 dB), the correlation was again nonsignificant ($r = 0.33$, $p = .063$).

One concern is that the use of "increase in looking" may not be the best choice; if, for example, what matters is whether there *is* an increase in looking, and not the size of this increase, using the proportion measures might be adding additional noise. To address this, we examined how many named trials (out of 12) each individual looked longer to the named item than they did on the baseline measure. Here, too, we found a significant correlation only for the 0 dB condition (10 dB, $r = .14$, $p = .26$; 5 dB, $r = .13$, $p = .29$; 0 dB, $r = .44$, $p = .024$; −5 dB, $r = .16$, $p = .23$).

Given the lack of a relationship at the other SNRs, it is not clear what to make of the significant result at 0 dB; one possibility is that it is simply a Type I error. However, another possibility is that performance in the two easier conditions may have approached ceiling levels, where variability in vocabulary size would have less opportunity to influence the results. Although performance at the most difficult level was clearly not quite at floor levels, there was substantially less variability among the children's looking times at this noise level than there was at the 0 dB SNR (standard deviations of the increase in looking time are 6.85 at −5 dB, and 8.83 at 0 dB, with ranges of 28 versus 42). Yet this cannot be the complete story, since performance was most variable in the easiest condition (10 dB SNR; $SD = 11.4$, range = 42). Thus, it is not clear what may have caused this pattern of results; it may be that vocabulary can influence listening ability in speech noise, but only when the SNRs are sufficient to make the task challenging without being excessively difficult. It may be worth exploring such relationships between vocabulary size and listening ability in noise in the future.

GENERAL DISCUSSION

The studies shown here indicate that (across word pairs), children were able to successfully perceive words in the presence of background babble at SNRs as low as −5 dB. Even when the babble was louder than the target speech, toddlers were able to identify individual words in the target voice, and looked longer to the video that matched what the speaker was saying. This supports prior work with infants, demonstrating that young children are able to separate the target speech stream from the background speech, and to recognize individual words with which they are already familiar even when speech is masked.

The level of performance shown by these toddlers is superior to that shown with younger infants; in studies with infants ranging from 4 to 13 months, no study has demonstrated successful listening in distracter speech for auditory stimuli at lower than a 5 dB SNR. (Hollich et al., 2005, reported successful performance down to 0 dB with a synchronized audiovisual display, but did not find such effects when the supporting visual information was absent.) Thus, the present study suggests a significant improvement between 1 and 2 years of age in children's ability to identify speech in a multitalker background. The size of this improvement is hard to quantify, given the differences across tasks, but would seem to be in the neighborhood of 10 dB SNR.

Performance varied for the different word/voice pairs in the study. Prior work that has investigated the SNRs at which infants are able to identify speech have often been based on only two words, and/or on only a single target voice. Yet, we see here very different results for different pairs of words spoken by different talkers; had we looked only at the words horse and bird, we would think that toddlers perform well at SNR of −5 or below. Had we used only pig and cow, we might estimate children's limits as being near +5 dB SNR. Thus, the particular tokens and voices used appear to have a significant effect on child performance, and future researchers should keep this result in mind when designing studies with this population.

The present work also found that performance related to vocabulary size, at least at some SNRs. Newman (2005) suggested that one possible explanation for why 12-month-olds showed improved recognition of speech in multitalker environments compared to 5- and 9-month-olds was that this is an age when children typically say their first words. This enhanced understanding of the purposes of communication could drive children toward better selective attention in noisy environments. If this were the case, we might expect not only that toddlers would show better performance than young infants, but also that toddlers with larger expressive vocabularies would show better performance than children with smaller vocabularies. While we did find a significant correlation between looking performance and vocabulary size, this relationship was found only at 0 dB SNR, and accounted for only 25% of the variability in children's performance at that noise level.

One possibility is that vocabulary size may only have measurable effects on word recognition when the task is relatively difficult for the child. Such an interpretation would seem to correspond well with the prior literature on the role of vocabulary in phonetic discrimination, where relationships with vocabulary size were only found for those children for whom discrimination was more difficult (those with smaller vocabularies, or of a younger age). Here, we find vocabulary effects only when the task becomes more challenging; there was no correlation between discrimination and vocabulary size at the easier SNRs. Yet, we also do not find this effect at the most difficult noise level, where one might most expect to find it.

It is possible that such effects would have been stronger if we had targeted children at risk for language delay. The children in our study appear to be a representative sample, based on prior norming studies (average vocabulary score on the LDS was 188 words, compared to 177 reported by Klee et al., 1998), but only four children were at the low end of the spectrum (< 50 words; 5% in our study, compared to 8% in Klee et al., 1998). Thus, while we do not appear to have included only children with advanced vocabularies, we also do not have a large number of participants at the lower end of the vocabulary spectrum; perhaps children at particular risk for language delay would show more difficulty perceiving speech in multitalker babble.

It is not clear what might cause such a relationship between vocabulary size and listening performance. One possibility is that poorer vocabulary size is simply an indication of less developed linguistic knowledge. Studies with adults have repeatedly shown that second-language learners (who likewise have less fully developed linguistic knowledge) show poorer performance recognizing speech in noise than do native speakers (Mayo et al., 1997; Takata & Nábelek, 1990), and one might presume that toddlers with less developed linguistic knowledge would likewise be relatively impaired when listening in the presence of distracters. Differences among children that might not be readily observable in easy listening conditions may become more noticeable in tasks that are more demanding. This argument implies that the difficulty in listening in noise is the result of the same underlying cause as the vocabulary differences. However, it is also possible that differential ability at listening in noise leads to the differences in vocabulary size:

Children who are better able to understand speech in noise or babble may have more opportunities to learn language, as they can potentially acquire useful information about their native language even in situations that are not perfectly quiet. The current study cannot speak to the directionality of any effect, but this is an important question for future research.

The present study used only one form of distracter, that of multitalker babble. The results would likely be quite different with other forms of masking. On the one hand, multitalker babble overlaps spectrally with the target speech; this may result in a greater degree of masking than would be found, for example, with a high-frequency sound that provides only informational masking (e.g., bird song; see Polka et al., 2008, for studies with infants using such sounds). On the other hand, multitalker babble also poses far less threat of informational masking than does a distracter consisting of a single voice, even though both contain spectral overlap. In multitalker babble, the individual voices merge, and each individual voice is less distinctive; the combination is likely to be less confusable with the target voice. Studies suggest that young infants (unlike adults) perform more poorly with a single voice as a distracter than with multitalker babble (Newman, 2009), and future work should examine this pattern in toddlers. Moreover, prior work has suggested that informational maskers may encourage listeners to rely on prior lexical knowledge to a greater extent than do energetic maskers (Mattys et al., 2009); if this holds true for toddlers, we might expect increased effects of vocabulary size when the masker consists of a single voice than when it consists of multitalker babble.

CONCLUSION

The ability to recognize speech in multitalker backgrounds is a critical skill for learning language. Children often find themselves in such settings, and little research has examined the extent to which they can understand speech in such difficult environments. Recent work with infants has suggested that they require a minimum of 5 dB SNR in order to recognize speech in multitalker babble. The current set of experiments explores toddlers' performance, and finds that 24-month-olds can identify target words in SNRs as low as −5 dB. There were some hints that children with larger expressive vocabularies are more adept at listening in noise than are children with smaller vocabularies, particularly at moderately difficult SNRs, but such effects are relatively modest, suggesting that vocabulary size, per se, is not the driving factor in children's listening improvement.

These findings suggest that 24-month-olds are better able to selectively attend to an interesting voice in the context of competing distracter voices than are younger infants. Future work will be needed to explore how this change occurs, and what forces might be driving it.

ACKNOWLEDGMENTS

The author thanks Lloyd Frei, Keith Miller, Don Mathis, Ian Palmer, James R. Sawusch, and Jim Garmon for the design and implementation of the test booth and software used in this experiment. I especially thank Amanda McAdoo, Sarah Michael, and Amanda Pasquarella for overseeing substantial parts of the stimulus creation and/or coding of the data reported here, and thank the following students for assistance either in creating the stimuli, scheduling or testing participants, or coding looking time performance: Liz Abramson, Saher Ali, Robyn Amendola, Megan Askew, Shannon Auxier, Emily Callanan, Hannah Couch, Kristin Crawford, Casey DiBiccari, Shoshana Dickler, Justine Dombroski, Lauren Eisenberg, Leah Eiserike, Akiko Elders, Andrea Farina, Melissa Ganginis, Richard Garcia, Sally Grapin, Sarah Haszko, Chelsea Healy, Megan Janssen, Grace Ji, Hannah Kim, Jodi Klempner, Arielle Laurie, Erica Levinson, Rachel Lieberman, Michelle McEntee, Heather McIntosh, Eileen McLaughlin, Giovanna Morini, Abimbola Odukoya, Ashley Orkin, Sabrina Panza, Jillian Parry, Lauren Polovoy, Rachel Ports, Jessica Pressel, Megan Reilly, Shannon Rice, Maria Rodriguez, Alex Roemer, Judith Segal, Kate Shapiro, Emily Singer, Audry Singh, Stephanie Souder, Mara Steinberg, Justine Taweel, Allison Temple, Sarah Thompson, Dena Tran, Sofia Vallila, Robin Weber, Rachel Weitzner, Patricia Yudd, and Michelle Zobel. This work was supported by NSF grant BCS0642294 to the University of Maryland.

REFERENCES

Allen, P., & Wightman, F. L. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech & Hearing Research*, *37*, 205–215.

Allen, P., & Wightman, F. L. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech & Hearing Research*, *38*, 503–511.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*, 1265–1282.

Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; infants do not. *Psychological Science*, *5*, 170–174.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45–B53.

Bronkhorst, A. M., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America*, *92*, 3132–3139.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition* (Version 2.2.5c). Austin, TX: University of Texas.
Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*, 414–427.
Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods & Instrumentation*, *28*, 125–127.
Durlach, N. I. (2006). Auditory masking: Need for improved conceptual structure (L). *Journal of the Acoustical Society of America*, *120*, 1787–1790.
Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America*, *66*, 651–653.
Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., & Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *Journal of the Acoustical Society of America*, *66*, 12–21.
Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America*, *108*, 3023–3029.
Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59* (5, Serial No. 242).
Frank, T., & Golden, M. V. (1999). Acoustical analysis of infant/toddler rooms in daycare centers. *Journal of the Acoustical Society of America*, *106*, 2172.
Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, *119*, 2445–2454.
Golden, M. V., & Frank, T. (2000). Further acoustical analysis of infant/toddler rooms in daycare centers. *Journal of the Acoustical Society of America*, *107*, 2833.
Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23–45.
Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience*, *5*, 108–120.
Havy, M., & Nazzi, T. (2009). Vocalic information in word learning at 16 months of age. *Infancy*, *14*, 439–456.
Hawkins, J. E., Jr., & Stevens, S. S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America*, *22*, 6–13.
Hétu, R., Truchon-Gagnon, C., & Bilodeau, S. A. (1990). Problems of noise in school settings: A review of literature and the results of an exploratory study. *Journal of Speech-Language Pathology & Audiology*, *14*, 31–39.
Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*, 598–613.
Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, 1570–1582.
Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (Vol. 29, pp. 143–190). New York: Springer Science.
Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A. & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech, Language, and Hearing Research*, *41*, 627–641.
Leibold, L. J., & Neff, D. L. (2007). Effects of masker-spectral variability and masker fringes in children and adults. *Journal of the Acoustical Society of America*, *121*, 3666–3676.

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

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

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*, 686–693.

Mills, J. H. (1976). Noise and children: A review of literature. *Journal of the Acoustical Society of America*, *58*, 767–777.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*, 13–30.

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*, 522–537.

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*, 271–279.

Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, *41*, 352–362.

Newman, R. S. (2006). Perceptual restoration in toddlers. *Perception & Psychophysics*, *68*, 625–642.

Newman, R. S. (2009). Infant's listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics*, *71*, 822–836.

Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics*, *58*, 1145–1156.

Oh, E. L., Wightman, F. L., & Lutfi, R. A. (2001). Children's detection of pure-tone signals with random multitone maskers. *Journal of the Acoustical Society of America*, *109*, 2888–2895.

Olsho, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior & Development*, *8*, 371–384.

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America*, *103*, 577–587.

Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology*, *40*, 221–244.

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 143–151). Tokyo: Ohmsha Press.

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

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, *54*, 587–599.

Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in children. *Journal of Speech, Language, and Hearing Research*, *44*, 434–445.

Smyth, V. (1979). Speech reception in the presence of classroom noise. *Language, Speech, and Hearing Services in the Schools*, *10*, 221–230.

Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., & Lutfi, R. A. (1997). Spectral weights in level discrimination by preschool children: Analytic listening conditions. *Journal of the Acoustical Society of America*, *101*, 2811–2921.

Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research*, *47*, 245–256.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

Takata, Y., & Nábelek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America*, *88*, 663–666.

Trehub, S. E., Bull, D., & Schneider, B. A. (1981). Infants' detection of speech in noise. *Journal of Speech and Hearing Research*, *24*, 202–206.

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

Werner, L. A., & Bargones, J. Y. (1991). Sources of auditory masking in infants: Distraction effects. *Perception & Psychophysics*, *50*, 405–412.

Werner, L. A., & Bargones, J. Y. (1992). Psychoacoustic development of human infants. *Advances in Infancy Research*, *7*, 103–145.

Werner, L. A., & VandenBos, G. R. (1993). Developmental psychoacoustics: What infants and children hear. *Hospital and Community Psychiatry*, *44*, 624–626.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., & Oh, E. (2003). Children's detection of pure-tone signals: Informational masking with contralateral maskers. *Journal of the Acoustical Society of America*, *113*, 3297–3305.