Developmental Psychology
2005, Vol. 41, No. 2, 352–362

0012-1649/05/$12.00 DOI: 10.1037/0012-1649.41.2.352

The Cocktail Party Effect in Infants Revisited: Listening to One's Name in Noise

Rochelle S. Newman
University of Maryland

This study examined infants' abilities to separate speech from different talkers and to recognize a familiar word (the infant's own name) in the context of noise. In 4 experiments, infants heard repetitions of either their names or unfamiliar names in the presence of background babble. Five-month-old infants listened longer to their names when the target voice was 10 dB, but not 5 dB, more intense than the background. Nine-month-olds likewise failed to identify their names at a 5-dB signal-to-noise ratio, but 13-month-olds succeeded. Thus, by 5 months, infants possess some capacity to selectively attend to an interesting voice in the context of competing distractor voices. However, this ability is quite limited and develops further when infants near 1 year of age.

One of the first tasks facing an infant is learning his or her native language. This is a difficult enough task in a quiet learning environment; yet infants often are exposed to speech in noisy environments. For example, a caregiver may be talking to an infant while other siblings are playing in the next room. In order to learn from speech in these settings, the infant must first separate that speech from background noise such as that provided by TV shows and siblings. How do infants acquire their native language in such settings?

Although there are few empirical data regarding the frequency with which infants find themselves in these multitalker situations, Barker and Newman (2004) surveyed parents of infants and found that two thirds reported that when they spoke to their infants, other members of the household were "frequently" talking simultaneously.[1] Van de Weijer (1998) recorded all of the language input to which a single child was exposed over the course of 3 weeks. Although he did not report the percentage of time his subject heard speech in silence versus in noise, he did comment that there were multiple people speaking simultaneously during most of the time that the infant was outside of the house (i.e., either in day care or during shopping trips). These findings suggest that at least some infants must find themselves in noisy environments quite often and thus must find a way to compensate for these sources of noise.

One critical aspect of learning language in these settings is the need to separate one stream of speech (e.g., the caregiver's voice) from that of others, a process known as *streaming*. Most of the research on stream segregation has been conducted on adult listeners. This research demonstrates that adults can separate different sound streams on the basis of a variety of acoustic cues, including location in space (Broadbent, 1954; Cherry, 1953; Hirsh, 1950; Pollack & Pickett, 1958; Poulton, 1953; Spieth, Curtis, & Webster, 1954), frequency range (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978), sex of the talkers and their voice pitch for speech (Broadbent, 1952; Brokx & Nooteboom, 1982), onset and offset times (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978), and differences in amplitude modulation (Bregman, Abramson, Doehring, & Darwin, 1985).

Despite this bevy of research on stream segregation in adults, very little is known about infants' abilities in this realm. There are many reasons to believe that infants might have difficulty disentangling concurrent speech signals. Separating streams of speech is a task that even adults find difficult, and adults' ability to hear speech in quiet situations does not always match their ability to do so under real-world conditions (Soli & Nilsson, 1997). Moreover, the ability to separate different streams of speech appears to be one of the first skills lost with aging. Even older listeners who have normal hearing on pure-tone tests tend to have difficulty with

Additional materials for this article are available on the Web at http://dx.doi.org/10.1037/0012.1649.41.2.352.supp

The research reported here was supported by Research Grant HD37822–01 from the National Institute of Child Health and Human Development and Research Grant BCS 0196498 from the National Science Foundation to the University of Maryland. A portion of these results was reported at the 2003 meeting of the Acoustical Society of America in Nashville, Tennessee, and at the 2004 International Conference on Infant Studies in Chicago, Illinois.

I thank Lloyd Frei, Keith Miller, Don Mathis, and Jim Garmon for the design and implementation of the test booth used in this experiment. A large number of talkers were recorded for this study, either as one of the nine distractor voices, or as one of the target voices. These include Jennifer Alexander, Nina Azhdam, Christine Beagle, Jen Bucklin, Sheryl Clouse, Rachael Dolezal, Jen Gerbyshak, Maria Hernandez, Isma Hussain, Lisa Loder, Keren Malaky, Jamie Ratner, Toni Rodriguez, Lauren Simpson, Susan Timm, Andrea Tuttle, Tammy Weppelman, Brooke Werner, and Jenni Zobler. I also thank Jenn Alexander, Christine Beagle, Emilie Clingerman, Hayley Derris, Annie Ferruggiaro, Mike Gordon, Sarah Haszko, Krista Jo Heinlen, Maria Hernandez, Lacey Kahner, Jessica Lisogorski, Lisa Loder, Amy Robbins, Toni Rodriguez, Lauren Simpson, Cheryl Tarbous, Suji Uhm, Donnia Zack-Williams, and Jenni Zobler for assistance in scheduling and testing infants. I also thank Emmanuel Dupoux for encouraging me to test 13-month-old infants, which led to Experiment 4. Finally, I owe a great intellectual debt to Peter Jusczyk for many helpful discussions on infant stream segregation abilities.

Correspondence concerning this article should be addressed to Rochelle S. Newman, Department of Hearing and Speech Sciences and Program in Neurosciences and Cognitive Sciences, 0100 Lefrak Hall, University of Maryland, College Park, MD 20742. E-mail: rnewman@hesp.umd.edu

---

[1] Parents were given the choices "frequently," "sometimes," and "almost never"; this was thus the highest rating possible.

multitalker situations, likely as a result of slight deficits in temporal resolution (Bergman, 1971). Separating streams must be an especially arduous task for infants because it requires both a sensitive auditory system and an ability to selectively attend to a given signal, both of which are still developing in the infant. With regard to the sensitivity of the auditory system, infants have been shown to have poorer auditory thresholds for both pure tones (Bull, Schneider, & Trehub, 1981; Nozza & Wilson, 1984; Sinnott, Pisoni, & Aslin, 1983) and speech (Trehub, Bull, & Schneider, 1981) than do adults. This means that they need for both tones and speech to be louder before they show evidence of having detected them. They also require greater stimulus intensity to discriminate speech sounds both in quiet (Nozza, Rossman, & Bond, 1991) and in noise (Nozza, Rossman, Bond, & Miller, 1990). Less work has been done on infants' ability to attend selectively, but what studies do exist suggest that infants do not appear to listen selectively to particular frequency bands (Bargones & Werner, 1994). These studies all suggest that infants might be at a particular disadvantage when faced with a multitalker environment.

Moreover, location in space is an important cue to stream segregation used by adult listeners, and infants' ability to localize sound is quite poor relative to adults' (Ashmead, Clifton, & Perris, 1987). This suggests that infants may not be able to make use of spatial cues to the same extent as adult listeners. Finally, infants have less experience with their native language and thus cannot rely on prior linguistic knowledge to help them compensate for noise in the way that adult listeners can. In this sense, they are more comparable to second-language learners than to native adult speakers of a language, and several studies have suggested that second-language learners have particular difficulties comprehending speech in noise (Mayo, Florentine, & Buus, 1997; Takata & Nábelek, 1990). Given this range of limitations, we might expect infants' speech recognition to be especially handicapped by the presence of background voices.

Newman and Jusczyk (1996) first demonstrated that infants are capable of separating speech from different talkers, at least in some situations. They suggested that one indication that infants could attend to a particular stream of speech would be if the infants showed some subsequent recognition for information that occurred only in that stream. This recognition would show that the infants processed and remembered at least some of the information in that speech stream. Taking this proposal as a starting point, Newman and Jusczyk (1996) presented infants 7.5 months of age with the speech of two talkers speaking simultaneously. The target voice (a female speaker speaking in an infant-directed speaking style) repeated two words while a male distractor spoke fluently in the background. The signal-to-noise ratio of these two voices was varied across experiments in 5-dB steps. After this familiarization phase, infants heard test sentences produced by the target voice in isolation. These test sentences either contained the familiarized words (the ones the target voice had just been repeating) or did not contain these words. Infants listened longer to sentences containing the familiar words as long as the target voice had been more intense than the distractor during the familiarization phase (either by 5 or 10 dB). They did not listen longer to the familiar words when the two voices had been equally intense (i.e., with a signal-to-noise ratio of 0 dB). These results suggest that infants have some capacity to extract information from speech even in the face of a competing voice. However, their success is limited to situations in which the target speech is more intense than that of the background noises.

There are a number of reasons why a more complete understanding of infants' ability to comprehend speech in noise is relevant to studies on language development. First, as noted above, there is reason to believe that at least some infants are placed in such settings relatively frequently (Barker & Newman, 2004). Being able to separate different voices and attend selectively to a particular voice would greatly enhance the opportunities these infants have to learn language. Second, understanding speech in a noisy environment is one of many ways infants must compensate for variability in the speech signal. A great deal of research has been devoted to exploring these infant abilities (for adjusting to differences in talkers, see, e.g., Jusczyk, Pisoni, & Mullennix, 1992; Kuhl, 1979, 1983; for adjusting to differences in speaking rates, see Eimas & Miller, 1980; Miller & Eimas, 1983), and infants' ability to hear speech despite background noise is yet another example of the complexity of the processing task facing infants. Finally, stream segregation ability is of particular interest because it combines the domains of infant discrimination, perceptual categorization, and attention. In order to listen to one voice in the presence of noise, infants must not only be able to separate different acoustic cues from one another but they must also be able to group together those acoustic properties that "belong together" (those that originate from a single source) and distinguish them from those that do not. Once they've grouped together the appropriate cues into separate streams, they must then choose to attend to the target signal rather than the distractor signal. Thus, stream segregation involves a variety of cognitive processes, and an inability to succeed at any of these would result in poor performance at hearing speech in noise.

Despite work examining the influence of the signal-to-noise ratio on infant performance, many other aspects of infants' stream segregation ability remain unexplored. For example, only two voices occurred in the Newman and Jusczyk (1996) study; can infants likewise focus on one talker's speech when there are many voices speaking simultaneously? There has as yet been no work investigating infants' ability to separate voices in a situation when multiple voices are speaking in the background. Likewise, Newman and Jusczyk purposefully chose two maximally dissimilar voices (a female speaking in the high pitch typical of infant-directed speech and a relatively low-pitched male talker); would infants be able to succeed in this task if the target and distractor streams were more similar acoustically? Some evidence suggests that they cannot; Barker and Newman (2000, 2004) reported that 7.5-month-old infants failed to recognize familiarized words at a 10-dB signal-to-noise ratio when both talkers were female unless the target talker was particularly well-known to the infant. However, the degree of similarity between the two voices was not varied systematically in that study.

Also not yet known is the extent to which infants' streaming abilities might develop over the course of their 1st year of life. This question is difficult to examine using the Newman and Jusczyk (1996) methodology, because that task relies on infants' ability to segment fluent speech into individual words, a skill that does not develop until infants reach approximately 7.5 months of age (Jusczyk & Aslin, 1995). Thus, testing younger infants requires the development of a different stream segregation task.

In the current research, I used another method of testing infant stream segregation: presenting infants with their own names in the context of noise. Infants heard either their own names or another infant's name while other voices spoke in the background. Listening longer to their own names can be taken as an indication that the infants successfully separated that stream of speech from the background noise. Because infants as young as 4.5 months will show recognition of their own names in a quiet setting (Mandel, Jusczyk, & Pisoni, 1995), this paradigm can be tested with younger infants than can the task used by Newman and Jusczyk (1996).

This task also has the advantage that the words being tested are particularly well-known to the infants and have likely been heard uttered by multiple speakers in many different contexts. This avoids one potential problem that has recently been found with the familiarization-and-test paradigm used by Newman and Jusczyk (1996), that of poor generalization across contexts. Singh, Bortfeld, Rathbun, and Morgan (2001) reported that infants familiarized with a target word (in quiet) at one speaking rate often failed to recognize that same word when it was spoken more slowly or more quickly. Similarly, infants familiarized with a word spoken in a happy tone of voice failed to later recognize that word spoken in a sad tone of voice, and vice versa. These results suggest that when infants are familiarized with a target word in a limited context, they may be particularly sensitive to changes in context and thus may not recognize that word in other situations. This is a particular concern in a streaming study, as a change in context from noise to silence (as is the case in the stream segregation studies described above) might well be as great a change in context as a change in speaking rate or prosodic tone. If so, a failure to find significant results in streaming tasks may be an indication of infants' poor generalization abilities for newly learned words rather than a lack of segregation ability. (It is worth noting that when a significant effect in a streaming task is found, poor generalization is clearly not an issue; but when comparing across situations in which significant results are or are not found, it becomes difficult to determine whether the source of the difficulty is poor generalization or poor stream segregation.) Using the infant's own name as the test stimulus, an item that infants have likely heard in a wide variety of contexts (in quiet and in noise, in a happy tone and in a warning tone, and by a wide variety of talkers), avoids this potential difficulty. This is not to say that generalization was not required in the present task; infants entering the laboratory were unlikely to have ever heard the particular talkers used in this study, for example. But because they had presumably heard their own names spoken in a wide variety of ways, they should have been more likely to be able to recognize their names across contextual changes (see, e.g., Gómez, 2002, for work exploring how variability can help infants focus on the important, unvarying properties of signals; Jusczyk et al., 1992).

Thus, in the present study, infants heard a woman's voice repeating a name on each trial. On some trials, the name the woman repeated was that of the infant being tested; on other trials, it was not. At the same time that the target voice was speaking, nine different female voices spoke fluently in the background. All 10 voices came from the same location in auditory space. I measured whether infants would listen longer to those trials in which the target voice repeated their own names than to trials in which the voice repeated other names.

Clearly, detecting one's name in the presence of noise is a less complicated task than following a fluent speech stream in a noisy environment. First, only isolated words needed to be separated from the background and identified. These words were thus clearly segmented for the listener, making the task of identifying them potentially easier (Aslin, 1993). Second, the use of single words means that there was no semantic or syntactic information that the listener needed to detect. Third, the same individual word was here repeated numerous times; thus, an infant who failed to identify his or her name the first time it was spoken could still have done so the second, third, or fourth time it was spoken; in the real world, speech is rarely so redundant. However, this task still required that the listener group together the frequencies that originated from one sound source and separate them from frequencies originating from a different source. As such, the present task might be best thought of as a prerequisite ability for fluent speech streaming: If listeners cannot detect their own names in the presence of noise, it is hard to imagine how they could identify any other form of speech information in that same setting. Thus, although successful performance on this task might not answer the question of whether infants can recognize and learn from a fluent sentence spoken in noise, failure to perform well at recognizing their names in noise would clearly imply that infants would have difficulty interpreting fluent speech in noise.

This task was modeled after one by Mandel and colleagues (1995). In that study, four different names were presented to each infant: the infant's own name, a name matched for stress pattern, and two names with a different stress pattern. The same procedure was used here for comparison purposes. However, it would not be surprising if infants could pick out the general stress pattern of their names even at noise levels in which they could not pick out the particular phonetic pattern. If so, infants might show a preference for their own names over the two nonmatched foils while not showing a preference for their names over the stress-matched foil. Although this would demonstrate some ability to perceive acoustic information in the context of noise, such an ability would not be sufficient to help a child learn the words of his or her native language. Thus, although I compared infants' listening to their own names with their listening to both the foils with matching and mismatching stress patterns, the critical comparison was between an infant's own name and a name matched for stress pattern.

Given the potential similarity between a female target voice and the (female) background voices, it was important to ensure that any effects found were not limited to one particular target voice. To that end, five different female talkers served as target voices in this study. Five infants heard each of the five target voices, for a total of 25 infant participants. To the extent that there is no effect of talker voice, one can be assured that the results are likely to generalize across a range of different talkers.

Experiment 1

As a starting point for this investigation, I tested whether infants 5 months of age would listen longer to their own names than to other names in the context of multitalker babble. The target voice was set to be 10 dB more intense than the combination of background voices. This signal-to-noise ratio is comparable to the easiest signal-to-noise ratio used in the Newman and Jusczyk (1996) study as well as to the level used in the Barker and Newman (2004) study.

Method

Participants. Twenty-five full-term infants (14 boys and 11 girls) participated in this experiment. Participants in this and the following experiments were recruited by means of letters sent to parents who listed birth announcements in the local newspapers. The average age of the infants in this study was 5.2 months (range = 18 weeks 6 days to 25 weeks 3 days). Data from an additional 19 infants were not included for the following reasons: crying/fussiness (*n* = 8); ear infection on the test date (*n* = 2); experimenter error (*n* = 1); stimulus or equipment difficulties (*n* = 4); or failure to orient to the lights or to listen for an average of 3 s per item (*n* = 4). Data from 1 additional infant were excluded because the parent took her headphones off during the experiment. Because this study focused on the infant's own name, and not on particular words in the language, infants were not excluded for being in bilingual homes as long as the target name was one easily pronounced by an English speaker (that is, one using English phonology). One child heard Spanish from a nanny, 3 heard other languages approximately 20%–30% of the time (French or Bengali), and 3 parents reported that their infants "sometimes" or "rarely" heard other languages (French, Spanish, and Grebo). Infants were excluded if the name tested was not the one most commonly heard or if any of the foil names chosen for testing were ones an infant was particularly familiar with (such as names of family members or pets); although we attempted to ascertain such information prior to testing, some parents changed their responses to these questions between the time of initial contact and the final visit. Five infants were excluded for these reasons. Infants were also excluded if they were outside the age range being tested (usually because multiple reschedulings resulted in their being too old). The final participants were primarily European American (71%), with 17% African or African American, 4% Asian, and 8% of mixed ethnicity. Most of the mothers had a bachelor's degree or equivalent (79%), with 38% having a master's degree as their final degree and 17% having a PhD, MD, or JD. The number of siblings in the home ranged from 0 to 2 per participant (average = 0.6). Most of the mothers worked outside of the home, with only 17% reporting that they were currently full-time mothers.

Stimuli. Stimuli consisted of both a target speech stream and a distractor speech stream. All recordings were made in a sound-attenuated room, using a Shure SM51 microphone. The recordings were amplified, low-pass filtered at 44.1 kHz, digitized via a 16-bit analog-to-digital converter, and stored on computer disk. For the distractor speech stream, nine women were recorded reading passages aloud from a variety of books. These passages were adjusted to be of the same overall root-mean-square (RMS) amplitude and then were blended together at equal ratios, resulting in multitalker babble. Five different women were recorded for the target passages; they were instructed to record the names in a lively, animated voice, as if calling a small child. Each target passage consisted of 15 repetitions of a child's name (or nickname). In order to prevent the speaker from producing the target name in a more engaging manner than the foil names, the speakers were always given a variety of names to record at any given time, and were never aware of which names would be target names (as opposed to foil names) in the actual test sessions. Many of the target names also served as foil names for other children. All four names heard by any given child were recorded by the same speaker. Pauses between names were adjusted such that the four sound files (name, stress-matched name, and nonmatched foils) were of the same overall duration. No single recording was used more than twice in any given condition or more than three times in the experiment. The complete lists of names used in the four experiments in this study may be found on the Web at http://dx.doi.org/10.1037/0012.1649.41.2.352.supp.

In order to adjust the signal-to-noise ratio between the stimuli, measurements were needed of the average intensity of both the isolated names and the fluent multivoice distractor passage. Because the children's names were separated by periods of silence, the average intensity level (RMS amplitude, measured on the digital signal) of the entire recording was lower than the intensity of the fluent speech distractors (i.e., the periods of silence served to make the intensity level overall seem lower than the actual level while the individual was talking). In order to adjust for this, an edited version of each name list was created. This version had the pauses between names spliced out. A waveform program on the computer then calculated the average intensity level of these edited versions. The amplitude levels of the original name recordings were then adjusted so that their edited versions had the same average intensity level, and the average intensity level of the multitalker distractor stimulus was adjusted to be 10 dB less than that of this edited version of the name list. This resulted in four test stimuli of the same RMS amplitude, each of which was combined with a distractor stimulus 10 dB less intense than the target voice.

A silence period 500 ms in length was appended to the beginning of each name list; the distractor passage was then adjusted to be at least 500 ms longer than the duration of the name list, with the amplitude tapering after the offset of the final name. In this manner, the distractor passage began prior to the onset of the names and remained at full amplitude until the final name was recorded. The same distractor passage was used for all four name stimuli for any given participant. Then the multitalker distractor passage and the word list were combined into a single sound file. As infants' names varied in length (from one to four syllables), the duration of the stimuli was not constant across children (although all four stimuli were identical in length for any given child); lengths of the four stimuli in this and the next three experiments varied from 22.7 s to 33.8 s.

Two practice stimuli were also created to familiarize infants with the task and setting. These consisted of musical passages, each 14.9 s in length.

Apparatus. The experiment took place in a three-sided test booth constructed out of pegboard panels (4 ft [1.2 m] × 6 ft [1.8 m]). There was a light in the center of the front panel and a hole for the lens of a video camera. The video camera was used to provide a permanent record of each session. An experimenter located behind the front wall of the booth watched the session via a monitor connected to the video camera. The two side walls of the booth each had a red light and a loudspeaker located in the center of the panel. A tan curtain was suspended from the ceiling and prevented the infant from seeing over the top of the booth. A Macintosh computer located behind the front wall of the booth controlled the presentation of the stimuli and recorded the observer's coding of the infant's responses. The experimenter pressed buttons on a response box to signal the computer to start and stop the flashing center and side lights.

Procedure. The infant sat on a caregiver's lap in the center of the test booth. There was an initial practice phase to familiarize the infant with the task; during this phase, the infant heard one of two musical passages on alternating trials until he or she accumulated at least 25 s of listening time to each passage. Listening time was assessed by the amount of time the infant spent looking at the "source" of the sound (the flashing light).

The test phase began immediately after the listening criterion for the practice phase was reached. During this test phase, the infant heard three repetitions of each of four different names. The 12 trials were blocked in groups of four so that each name occurred once in a given block, although the order of the four names within each block was randomized. (Although the two names that mismatched in stress were expected to entail similar listening times, two such items were used so that the test trials as a group had an equivalent number of trials with each stress pattern; this prevented infants from acquiring a preference for a particular stress pattern over the course of the experiment.)

Both familiarization and test trials began with the light in the center of the front panel blinking. Once the infant had oriented in that direction, the light was turned off and one of the two red lights began to flash. Once the infant had oriented toward that light, the stimulus for that trial began to play from the loudspeaker on the same side. The stimulus continued to play until its completion or until the infant had looked away for 2 consecutive seconds, whichever came first. Any time the infant spent looking away (whether it was 2 s or less) was not included when measuring total listening time. The red light continued to flash for the duration of the entire trial. Information about the direction and duration of head turns and the total trial duration was stored in a data file on the computer.

The experimenter behind the center panel pressed a button on the response box whenever the infant looked at or away from the flashing light. Both the experimenter and the caregiver listened to masking music over

Peltor aviation headphones so that they could not influence the infant's behavior or the coding of that behavior.

Results and Discussion

Mean listening times to the four different names were calculated for each infant across the three blocks of trials. Because the two nonmatching names were not expected to differ, listening times to these names were combined for all statistical analyses, resulting in three name types: own name, stress-matched foil, and nonmatch foil.

An analysis of variance (ANOVA) with two factors (name and talker) found no effect of the particular voice used, $F(4, 20) = 1.04$, $p > .05$, and no interaction between talker and name ($F < 1$) but a significant effect of name, $F(2, 40) = 5.94$, $p < .01$. Follow-up directional t tests were used to determine whether infants listened longer to their names than to each of the other foils. These showed that infants listened longer to their own names than to either the stress-matched foil, $t(24) = 3.41$, $p < .005$, or the nonmatch foils, $t(24) = 2.32$ $p < .05$. This pattern can be seen in Figure 1. Overall, infants averaged 14.9 s of listening to their own names and 11.4 s of listening to the stress-matched foil; 18 of 25 infants showed this pattern.

Of the 25 infants, 3 had single-syllable first names. As such, their nonmatch names differed in the actual number of syllables, not just the stress pattern. However, excluding these 3 participants had no effect on the pattern of results. Similarly, excluding all infants whose parents reported that they heard other languages also had no effect on the results.

These findings suggest that infants at this age are capable of separating speech from different talkers and of recognizing individual words in the context of noise. This does not appear to be specific to any particular talker. However, the intensity ratio between the target and distractor voices was quite high. For this reason I decided to examine whether infants were also capable of separating streams of speech at a lower signal-to-noise ratio.

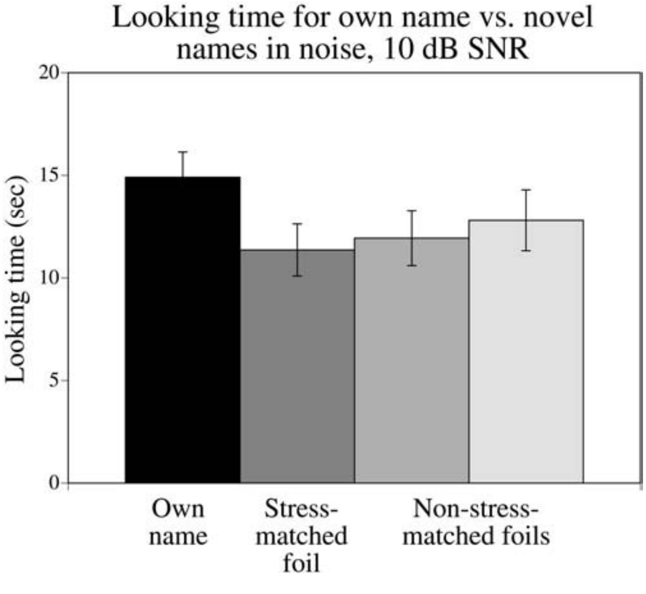

Figure 1. Mean listening times and standard errors to the infant's own name and to the foil names in Experiment 1 (10-dB signal-to-noise ratio [SNR]).

Experiment 2

This experiment was identical to Experiment 1 in all respects except that the target names were adjusted to be 5 dB more intense than the multitalker babble, rather than 10 dB more intense.

Method

Participants. Twenty-five full-term infants (17 boys and 8 girls) participated in this experiment. The average age of the infants was 5.2 months (range = 19 weeks 3 days to 25 weeks 1 day). The data from an additional 13 infants were not included for the following reasons: crying/fussiness ($n = 8$); ear infection on the test date ($n = 1$); experimenter error ($n = 1$); stimulus or equipment difficulties ($n = 1$); parental interference ($n = 1$); or failure to orient to the lights or to listen for an average of 3 s per item ($n = 1$). As before, infants were also excluded if the name tested was not the one most commonly heard, if it was mispronounced, or if any of the foil names chosen for testing were ones with which the infant was particularly familiar; 5 infants were excluded for these reasons. Four of the participants heard languages other than English some of the time (Krio, $n = 1$; Spanish, $n = 3$). Most of the mothers had a bachelor's degree or equivalent (88%), with 44% having a master's degree as their final degree and 8% having a PhD, MD, or JD; 32% of mothers reported that they were currently working as full-time mothers. The number of siblings in the home ranged from 0 to 2 per participant (average = 0.8); 3 of the infants were from twin births. Participants were primarily European American (76%), with 12% African American, 4% Hispanic, and 8% of mixed ethnicity.

Stimuli. Stimuli were identical to those in Experiment 1 with the exception that they were often new names (and thus new recordings). As before, five different talkers participated as target voices; four of these were identical to the talkers in Experiment 1, but one of the five talkers was replaced as a result of a change in laboratory personnel. When the distractor passages were combined with the isolated names, their relative intensity levels were adjusted to be 5 dB less intense than the target names, rather than 10 dB less intense.

Design, apparatus, and procedure. These were identical to those in Experiment 1.

Results and Discussion

The data were analyzed in the same manner as in Experiment 1. The mean listening time to each of the four test passages was calculated for each infant. An ANOVA with two factors (name and talker) found no effect of the particular voice used ($F < 1$) and no interaction between talker and name ($F < 1$). However, unlike in Experiment 1, there was no significant effect of the name, $F(2, 40) = 1.46$, $p > .10$. In order to explore this finding further, follow-up directional t tests were performed as planned comparisons; these showed that infants listened for nearly identical amounts of time to their own names and to stress-matched foils, $t(24) = 0.06$, $p > .05$. They did listen longer to their names than to the nonmatch foils, $t(24) = 1.77$, $p < .05$, which suggests that they may have been able to detect the general stress patterns of their names to some extent. Clearly, however, infants did not demonstrate any ability to detect the phonology of their own names at this more difficult noise level, as shown in Figure 2. Overall, infants averaged 12.8 s of listening to their own names and 12.7 s of listening to the stress-matched foil; only 10 of 25 infants listened longer to their own names.

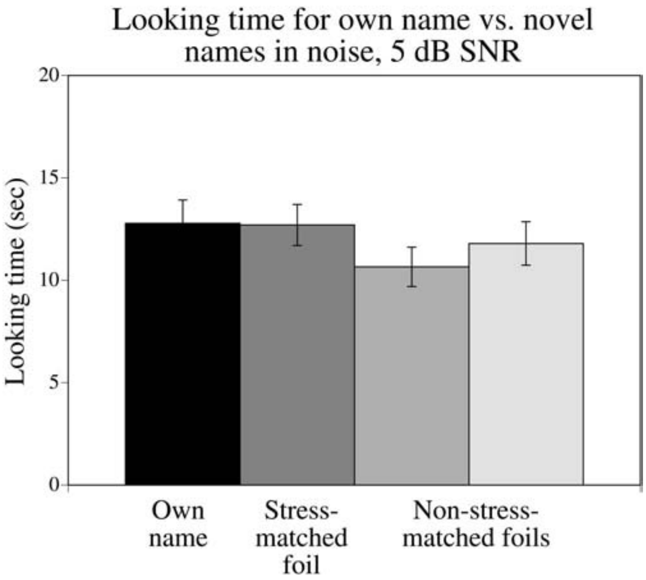

Figure 2. Mean listening times and standard errors to the infant's own name and to the foil names in Experiment 2 (5-dB signal-to-noise ratio [SNR]).

As before, excluding infants with single-syllable names or infants who were bilingual had no effect on the results with the exception that the preference infants showed for their own names over the nonmatching foils disappeared when bilingual infants were excluded, $t(20) = 1.65$, $p < .10$.

These results suggest that infants at 5 months of age are quite limited in the extent to which they are capable of attending to their own names in multitalker babble. Despite the fact that an infant's own name is particularly well-known to him or her, the infant cannot detect it when the intensity of the combination of background noises approaches within 5 dB of the intensity of the name itself. Infants did not fail to listen in the noisy condition (their average listening time of 12.0 s was only slightly below the average of 12.8 s from Experiment 1); they simply did not recognize familiar words in the midst of the noise.

Perhaps these stimuli are simply extremely difficult to hear in the presence of noise even for adult listeners. If so, it might not be surprising that infants failed in the present task. To test this issue, I asked 14 adult listeners to identify the names used in this study; 4 were new members of the laboratory, and 10 were members of the university community who had participated in other speech perception studies in the laboratory. These listeners heard each of the target and distractor names presented in this study, at the same 5-dB signal-to-noise ratio, and were asked to write down the name that they heard. Each name stimulus list was played once, but listeners were allowed to go on to the next trial as soon as they were ready (i.e., they did not need to hear all 15 repetitions of the name within each stimulus list); most listeners identified the name and went on to the next trial after only a few repetitions. When scoring listeners' responses, obvious misspellings were counted as correct, but all other mistakes (regardless of how slight) were counted as errors. Across all participants, performance on both the complete set of names and on just the target names (ignoring distractors) was above 90% accurate. In addition, most of the mistakes involved only a single phoneme, particularly on the less common names. The most common errors were *Bren* for *Brynn* ($n = 4$); *Brian* for *Ryan* ($n = 12$); *Kip* for *Gip* ($n = 4$); *Graham* for *Grant* ($n = 6$), *Carinne* or *Careen* for *Kareem* ($n = 3$), and *Cody* for *Bode* ($n = 7$). These slight errors might have been less likely to have prevented a listener from recognizing his or her own name if it had been present. Regardless of the seriousness of these errors, a finding of over 90% accuracy on an open response test suggests that for adult listeners, the presence of noise did not make these names difficult to hear. Moreover, finding such good performance for adult listeners makes the poor performance of infants in this task all the more striking. (Sample stimuli are available on the Web at http://dx.doi.org/10.1037/0012.1649.41.2.352.supp for readers to make their own judgments regarding their ease of audibility.)

Given these results with adult listeners, why did the infants in this study not appear to recognize their own names at this level of noise? Moreover, Newman and Jusczyk (1996) demonstrated that infants 7.5 months of age could detect a recently familiarized word in noise at this same signal-to-noise ratio. There are several possible reasons for this discrepancy. One difference is the type of noise used: Newman and Jusczyk had a single talker as the background voice, compared with the multitalker babble used here; if infants are able to make use of amplitude dips in the signal, one might expect them to perform better with a single voice as a distractor (where amplitude is likely to vary to a greater degree). In addition, the single talker Newman and Jusczyk used was a male, whereas the distractor talkers used here were all female; when a single female voice was used as a distractor in the Newman and Jusczyk procedure, infants unfamiliar with the talker failed to identify the familiarized word (Barker & Newman, 2000, 2004).

Another potential difference is with the ages of the infants, however. Perhaps infant stream segregation abilities continue to improve throughout the 1st year of life. In Experiment 3, I tested this possibility by presenting older infants with the same 5-dB signal-to-noise ratio stimuli used in Experiment 2.

Experiment 3

This experiment was identical to Experiment 2 in all respects except for the age of the infants.

Method

Participants. Twenty-five infants (13 boys and 12 girls) participated in this experiment. The average age of the infants was 9.2 months (range = 36 weeks 2 days to 43 weeks 1 day). The data from an additional 7 infants were not included for the following reasons: crying/fussiness ($n = 5$); experimenter error ($n = 1$); or failure to orient to the lights or to listen for an average of 3 s per item ($n = 1$). Five of these infants heard other languages part of the time (French, $n = 2$; Burmese, $n = 1$; Greek, $n = 1$, and Spanish, $n = 1$). Most of the mothers had a bachelor's degree or equivalent (84%), with 36% having a master's degree as their final degree and 12% having a PhD, MD, or JD; 44% of mothers reported they were currently working as full-time mothers. The number of siblings in the home ranged from 0 to 5 per participant (average = 0.8); 4 of the infants were from twin births. Participants were primarily European American (84%), with 4% Asian, 4% Hispanic, and 8% of mixed ethnicity.

Stimuli. Stimuli were identical to those in Experiment 2 with the exception that they were often new names (and thus new recordings). As before, five different talkers participated as target voices; one talker was changed from those in Experiment 2.

Design, apparatus, and procedure. These were identical to those in Experiment 2.

Results and Discussion

The data were analyzed in the same manner as in Experiment 2. The mean listening time to each of the four test passages was calculated for each infant. An ANOVA with two factors (name and talker) found no effect of the particular voice used ($F < 1$) and no interaction between talker and name, $F(8, 40) = 1.19$, $p > .05$. As in Experiment 2, there was also no effect of name, $F(2, 40) = 1.99$, $p > .10$. Follow-up *t* tests were again conducted but showed that infants listened nearly identical amounts of time to their own names and to stress-matched foils, $t(24) = 1.30$, $p > .05$, although they did listen longer to their own names than to the non-stress-matched foils, $t(24) = 1.77$, $p < .05$. As in Experiment 2, there was some indication that infants might be able to detect the stress pattern of their own name but no indication that they could pick out its phonology. Moreover, this effect disappeared once infants with single-syllable names were excluded, $t(20) = 1.36$, $p > .05$; it also disappeared when bilingual infants were excluded, $t(19) = 0.86$, $p > .05$. Overall, infants averaged 11.2 s of listening to their own names and 9.5 s of listening to the stress-matched foils; despite this apparent difference, only 14 of 25 infants listened longer to their own names. These data are shown in Figure 3.

These results are quite comparable to those of Experiment 2. In fact, a two-way ANOVA with the factors of name and child age showed no Age × Name interaction ($F < 1$). There was an overall effect of age, $F(1, 48) = 4.82$, $p < .05$, as the younger infants in general listened longer than did the older infants (12.2 s vs. 9.9 s). With the greater number of participants, there was also an overall effect of name, $F(3, 144) = 3.10$, $p < .05$; however, follow-up *t* tests showed that this was a result of infants listening longer to items that matched the typical stress pattern of their names (names vs. non-stress-matched foils), $t(49) = 2.48$, $p < .05$; they listened no longer to their own names than to the other names matching in stress pattern, $t(49) = 0.99$, $p > .05$. Thus, there is no evidence that children of this age are better able to recognize the phonetic pattern of their own names in noise than are younger infants. In Experiment 4, I tested whether this pattern would be any different with yet older infants.

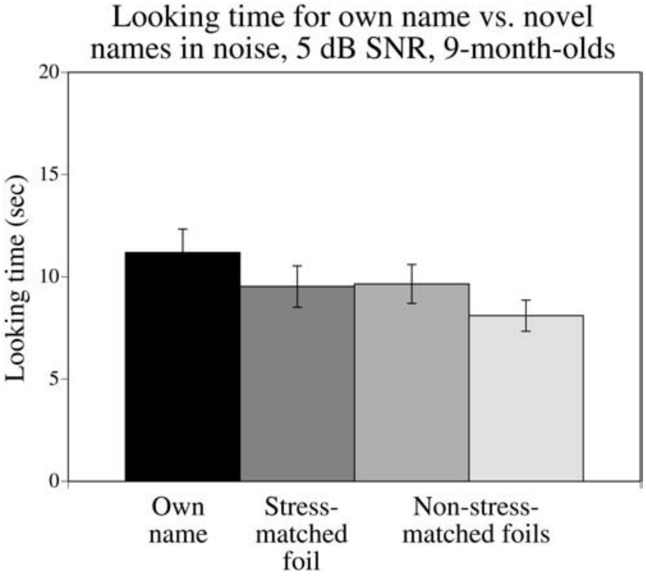

Figure 3. Mean listening times and standard errors to the infant's own name and to the foil names in Experiment 3 (5-dB signal-to-noise ratio [SNR], 9-month-old infants).

Experiment 4

This experiment was identical to Experiments 2 and 3 in all respects except for the age of the infants.

Method

Participants. Twenty-five infants (11 boys and 14 girls) participated in this experiment. The average age of the infants was 13.3 months (range = 55 weeks 5 days to 60 weeks 2 days). The data from an additional 6 infants were not included for the following reasons: crying/fussiness ($n = 4$); ear infection on the test date ($n = 1$); and stimulus or equipment difficulties ($n = 1$). One infant had a sibling's name as a foil and was also excluded. Of the 25 infants, 7 heard other languages part of the time either at home or at day care. Most of the mothers had a bachelor's degree or equivalent (92%), with 24% having a master's degree as their final degree and 28% having a PhD, MD, or JD; 40% of mothers reported they were currently full-time mothers. The number of siblings in the home ranged from 0 to 2 per participant (average = 0.7); 4 of the infants were from twin births. Participants were primarily European American (63%), with 4% African or African American, 4% Hispanic, 4% Asian, and 25% of mixed ethnicity. One infant had also participated at 4 months of age.

Stimuli. Stimuli were identical to those in Experiment 2 with the exception that they were often new names (and thus new recordings). As before, five different talkers participated as target voices; three talkers were changed from those in Experiment 2.

Design, apparatus, and procedure. These were identical to those in Experiment 3.

Results and Discussion

The data were analyzed in the same manner as in Experiment 2. The mean listening time to each of the four test passages was calculated for each infant. An ANOVA with two factors (name and talker) found no effect of the particular voice used, $F(4, 20) = 1.82$, $p > .05$, and no interaction between talker and name ($F < 1$). There was a significant effect of name, however, $F(2, 40) = 5.24$, $p < .01$. Follow-up directional *t* tests showed that infants listened longer to their own names than to stress-matched foils, $t(24) = 1.90$, $p < .05$, or to nonmatched foils, $t(24) = 3.96$, $p < .0005$. Overall, infants averaged 13.6 s of listening to their own names and 10.6 s of listening to the stress-matched foil; 17 of 25 infants listened longer to their own names. Excluding the 8 children who were either being raised bilingual or had monosyllabic names did not alter these results. These data are shown in Figure 4. Although the effect was not consistent across all infants, these results suggest that by the time they reach 13 months of age, infants are better able to detect their own names in the presence of noise, although this remains a difficult task.

Figure 5 shows the differences in mean listening times to the target name versus the stress-matched foil from the individual participants in all four experiments. This graph makes it clear that with the change in signal-to-noise ratio there was a change in the infants' differential listening times. With the lower ratio, there were more infants with negative differentials, which suggests that

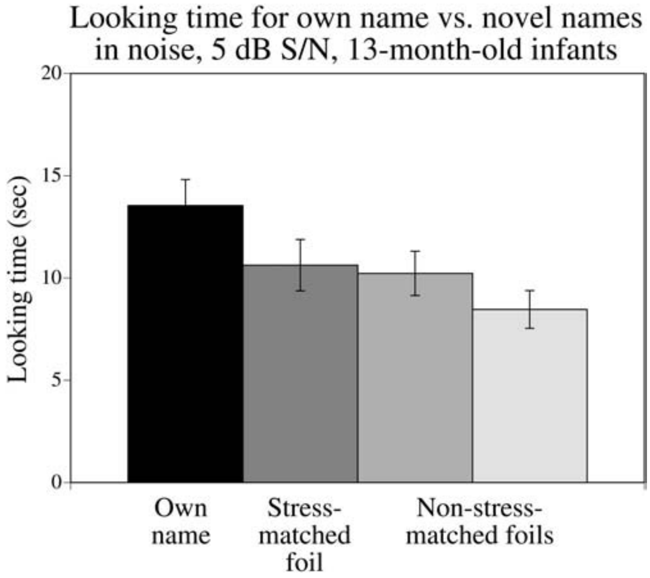

Figure 4. Mean listening times and standard errors to the infant's own name and to the foil names in Experiment 4 (5-dB signal-to-noise ratio [SNR], 13-month-old infants).

more and more infants were failing to recognize their names. Furthermore, the infants' variability increased with the lower ratio, as shown by the larger spread of data points. These data also show that as infants reached 13 months of age, the number of infants with positive differentials increased, but not to the level found with the youngest infants at the easier signal-to-noise ratio.

General Discussion

The present study demonstrates that by the time infants reach 5 months of age, they can separate speech produced by different talkers. Furthermore, even when speech is masked, infants are able to recognize the representations for words with which they are already familiar.

Yet this ability is clearly quite limited for young infants. Infants under 1 year of age failed to recognize their own names unless the combination of distractor voices was 10 dB less intense than the target voice. One cannot be certain whether the infants were unable to selectively attend to a voice at more difficult signal-to-noise ratios or were unwilling to expend the energy necessary to do so. Yet the fact that infants in the 5-dB condition spent nearly as much time listening to the stimuli as did those in the 10-dB condition suggests that the infants were not simply ignoring the sounds and were not distressed by the level of noise. Infants continued to listen to the stimuli; they simply did not recognize that their own names were what was being produced. This failure to recognize a familiar word suggests that infants would likewise have difficulty comprehending speech in this level of noise outside the laboratory. Thus, this finding places strict limitations on the kinds of situations in which young infants might be capable of learning about their native language.

These results also begin to shed light on the developmental time course over which this ability to hear speech in noise improves. For the harder signal-to-noise ratio, infants did not show a significant ability to detect their own names until after they reached 1 year of age. Once they reached 13 months, however, infants were able to recognize their names at the same signal-to-noise ratio at which they had failed 4 months earlier.

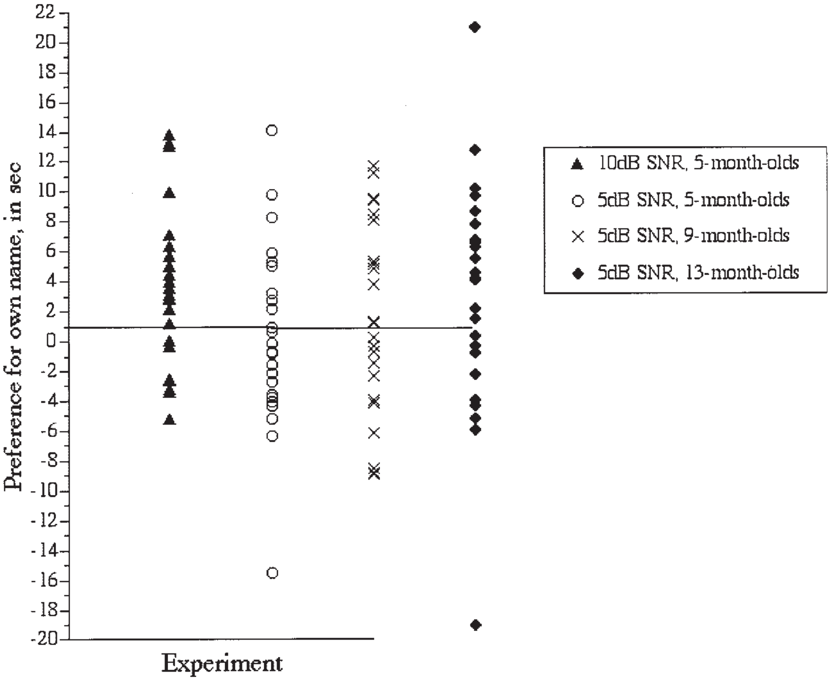

Figure 5. Mean difference in listening times between the infant's own name and the stress-matched foil for each individual participant in all four experiments. SNR = signal-to-noise ratio.

What is less clear is the reason for this improvement at 13 months. Obviously, increases in auditory discrimination ability, or in stream segregation ability, are possible explanations for this improvement. As infants become older, their perceptual abilities may increase. However, a number of studies have found either comparable auditory skills in infants across a range of ages or only relatively slight improvements with age (Bull et al., 1981; Nozza & Wilson, 1984; Trehub et al., 1981). For example, Nozza and Wilson (1984) found only a 2-dB nonsignificant difference in pure-tone thresholds for 6- versus 12-month-old infants, and Trehub et al. (1981) found little difference in infant performance from 6 to 24 months of age in a speech-detection-in-noise task. Given these results, perhaps a more likely cause of the change in streaming performance is related to improvements in infants' general lexical ability. The fact that this improvement in performance in noise co-occurs with the point at which infants typically begin talking themselves suggests the possibility that these improvements may be tied to improvements in the understanding of how words are used to communicate. Unfortunately, information on whether the particular infants tested had begun to say their first words was not collected at the time of the study, so it is impossible to determine whether the 15 infants who succeeded at 13 months (those who listened longer to their own names than to the stress-matched foils) were different in this respect from the 10 who did not succeed. Still, the fact remains that improvements in lexical comprehension, or in understanding the properties of words that are important (including such basic knowledge as the fact that background noise should be excluded from the lexical representation), could be the reason for older infants' improved performance.

It is not clear to what extent the current results might generalize across infants. Infants raised in quiet home settings may differ from those raised in noisier environs. Moreover, although the effects in the current study did not change when only monolingual infants were included, I did not explicitly compare monolingual versus bilingual performance or the performances of infants from different cultures. If the improvement at 13 months of age is related to increases in lexical ability, it may depend on the particular word-learning environment in which infants are raised.

In none of the four experiments was there any effect of the particular talker. To the extent that greater similarity between talker and distractor voices makes the stream segregation task more difficult, it might have been expected that one of the talkers would, by chance, be more difficult to separate from the background than the others. Yet this did not appear to be the case. Whether this is an indication that talker similarity has little effect on infant performance or is an artifact of the particular talkers used is unclear. The number of background talkers was sufficient that they merged into a background babble; talker similarity effects may be more likely when individual voices can be discerned in the background noise.

Comparing this study to previous work, in particular, to that by Newman and Jusczyk (1996), is complicated by the number of factors that differ between these studies. First, Newman and Jusczyk used a procedure in which infants were familiarized with words as part of the study itself. In contrast, the present study used words with which infants were presumably already familiar. It might be expected that infants would be able to identify words they were already familiar with at lesser signal-to-noise ratios than words they had only just learned. However, these two tasks differ not only in the recency of that familiarization but also in the amount of familiarization and the type of familiarization (the extent to which generalization across different instances was encouraged). In terms of the latter aspect, familiarizing infants with a word as part of a task raises the possibility that infants may not generalize this familiarization to other tokens of the word or to the same word in other contexts. In fact, a number of recent studies have suggested that infants do not generalize across different tokens after such familiarization (Houston & Jusczyk, 2000; Singh et al., 2001). Presumably, a word that has been familiarized via real-world (rather than laboratory) exposure has been presented to the infant by multiple voices and in multiple contexts, increasing the likelihood that infants would exclude these token-specific properties from their representation of the word. This might also result in an expectation that infants' performance in noise on the present task would be better than infants' performance on the Newman and Jusczyk version. However, it remains unclear exactly *how* familiar infants are with their own names and how this familiarity increases over time. Some families may tend to use a child's name more often than other families; similarly, although the names used in the current study were always the ones reported as used most frequently by the parents, they were the sole names for some infants while being one of a large set of nicknames for other infants. Although using an already familiar name avoids the problem of specificity in representation, it raises the problem of a lack of control of the degree of familiarity.

There are a number of other differences between the current study and the Newman and Jusczyk (1996) study that make comparisons complicated. First, the current study did not use infants of the same age (7.5 months) as the infants in the prior study. This does not appear to pose a problem, because infants both younger (5 months) and older (9 months) performed comparably in the current task. It seems unlikely that 7.5-month-old infants would be different in this regard. However, the Newman and Jusczyk study also used only a single, male background voice, whereas the current study used 9 female voices blended together. Future work will need to explore each of these possible contributions to the task systematically, investigating the roles of voice similarity between target and background voices and of the number of background voices.

It is therefore unclear exactly why infants in the present study could not identify words at a 5-dB signal-to-noise ratio until 13 months of age, whereas those in the Newman and Jusczyk (1996) study could do so at 7.5 months of age. Regardless of the reasons for this difference, however, the results from both studies suggest that infants will likely have difficulty attending to speech in noise as the level of the signal approaches that of the background noise.

Despite the limitations on infant stream segregation found here, there are some advantages infants might have outside the laboratory to help them comprehend speech in noisy environments. One such advantage is that infants may already know the voice speaking to them. Although we found no effect of talker in the present study, all of the voices were ones that were unknown to the infant prior to the test session. In the real world, infants often are spoken to by a relatively small number of talkers with whom they have had extensive experience. There is evidence to suggest that this familiarity with a target voice makes it easier for the infant to separate that voice from the background signal (Barker & Newman, 2000, 2004).

Another advantage infants may have outside of the laboratory is that of visual information. Recent findings from Hollich, Newman, and Jusczyk (in press) suggest that infants can make use of visual information on a talker's face to help them separate that talker's speech from the speech of different talkers. In that study, infants presented with a video of a person talking were able to recognize words at signal-to-noise levels 10 dB worse than they could with auditory information alone. Infants also succeeded with unfamiliar visual cues (such as an oscilloscope pattern) as long as these cues correlated with the auditory signal; when presented with video signals that did not match what they were hearing, or when presented with static faces, infants failed to segregate the speech streams. These findings suggest that infants can use dynamic visual information to aid segregation of speech. The combination of familiar voices and the presence of visual cues could potentially allow infants to successfully recognize, and learn from, speech in noisier environments than those tested here.

There are many issues in infant stream segregation yet to be examined. Although the differences between the present findings and those of Newman and Jusczyk (1996) suggest that infants might find a single distracting voice less difficult than multitalker babble, this needs to be explored in more detail. A single voice is likely to have a much more varying amplitude level than does a combination of multiple voices, and adult listeners have the ability to listen within these amplitude dips, thus improving their resistance to masking. Yet studies have suggested that infants have poor selective attention (Bargones & Werner, 1994); this might be taken to suggest that they likewise do not have the same ability to focus their attention at points where the distractor signal is less intense. Future research will be needed to explore this issue directly.

Another issue is that of spatial location. In the experiments reported here, and in those in Newman and Jusczyk (1996), both the distractor voice and the target voice were presented from the same speaker, which was located on either the infant's right or left side. Thus there was no spatial location difference between the two voices. Outside of the laboratory, different voices usually come from different locations in space, and adult listeners perform better at streaming tasks when the voices differ in spatial location (Broadbent, 1954; Cherry, 1953; Hartmann & Johnson, 1991; Hirsh, 1950; Pollack & Pickett, 1958; Poulton, 1953; Spieth et al., 1954). However, the ability to detect and use auditory localization information has a relatively slow developmental progression. Ashmead et al. (1987) estimated that 7-month-old infants could only detect differences in localization down to approximately 19°, compared with values of 1° to 2° for adult listeners. Moreover, there is reason to believe that these minimal auditory angles may actually overestimate infants' ability to use spatial location information for stream segregation. Litovsky (1997) has shown that although children's minimal audible angles for single sounds have reached levels comparable to those of adult listeners by 5 years of age, children of this age still perform more poorly than adults on localization tasks when faced with multiple sound sources. This suggests that basic auditory abilities (such as the ability to distinguish locations in auditory space) may not provide a true measure of children's ability to use this information in more sophisticated auditory tasks. Thus, many real-world situations may be, for the infant, akin to our laboratory situation—although talkers may actually be located in different locations in space, young infants may not be capable of picking up on these differences auditorily. Still, future research will be needed to explore infants' ability to use spatial location differences to aid their stream segregation.

Another issue is when stream segregation abilities might develop further. The present results suggest that infants' abilities remain at a relatively constant level for much of their 1st year of life and improve only at around 13 months. Yet this level does not seem to match adult performance; adults could identify even unfamiliar names at this level quite easily, which suggests that there might be further development after 13 months of age. When such development takes place and whether this improvement is the result of auditory maturation, cognitive development, or greater experience with language are issues for future research.

It also is unclear exactly how often infants find themselves in multitalker environments. Although parents report that this is a common phenomenon (Barker & Newman, 2004), there have not been good measures of the proportion of time that infants actually spend in such settings. Presumably, infants in multichild households and infants who spend time in day-care centers might experience such noisy environments more often than would infants in single-family, one-child homes. However, without good measures of the signal-to-noise ratios that infants typically face, it remains unclear exactly how important of a skill stream segregation is to the developing infant. Future work will need to explore this issue in more depth.

One final issue is whether individual infants' ability to recognize speech in noise might predict their later language development. Although the current research shows that the majority of infants were able to recognize their names in noise as long as the distractor speech was 10 dB less intense, not all infants showed this pattern. Some recent research suggests that variation among infants in laboratory tasks may have implications for later language development. For example, infants' performance on such tasks as speech segmentation and their ability to recognize phonological regularities can predict their vocabulary development at age 2 and their syntactic and semantic abilities at ages 3 to 5 (Bernstein Ratner, Newman, Dow, Jusczyk, & Jusczyk, 2004). The ability to discriminate speech in noise might also be an important underpinning for later language development; poorer ability to segregate streams of speech could potentially lead to slowed language acquisition, at least for those infants who frequently find themselves in noisy environments. I hope to examine whether variability in infant performance in this task is related to long-term linguistic development.

In conclusion, the present study represents an initial step in the exploration of infants' ability to recognize spoken words in the midst of multitalker babble. The findings suggest that by 5 months of age, infants possess at least some capacity to selectively attend to an interesting voice in the context of competing distractor voices. However, this ability appears quite limited and does not appear to develop further for many months. Future work will be needed to explore the limits of these abilities in greater depth.

References

Ashmead, D. H., Clifton, R. K., & Perris, E. E. (1987). Precision of auditory localization in human infants. *Developmental Psychology, 23,* 641–647.

Aslin, R. N. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 305–315). Dordrecht, The Netherlands: Kluwer Academic.

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

Barker, B. A., & Newman, R. S. (2000). The cocktail party effect in infants: Following one's mother's voice. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *BUCLD 24: Proceedings of the 24th annual Boston University Conference on Language Development* (pp. 92–103). Somerville, MA: Cascadilla Press.

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

Bergman, M. (1971). Hearing and aging: Implications of recent research findings. *Audiology, 10,* 164–171.

Bernstein Ratner, N., Newman, R., Dow, K., Jusczyk, A. M., & Jusczyk, P. W. (2004, November). *Infant speech segmentation ability predicts later language development.* Paper presented at the Boston University Conference on Child Language Development, Boston, MA.

Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics, 37,* 483–493.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19–31.

Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology, 44,* 51–55.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology, 47,* 191–196.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–36.

Bull, D., Schneider, B. A., & Trehub, S. E. (1981). The masking of octave-band noise by broad-spectrum noise: A comparison of infant and adult thresholds. *Perception & Psychophysics, 30,* 101–106.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Dannenbring, G. L., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics, 24,* 369–376.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209,* 1140–1141.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–184.

Hirsh, I. J. (1950). The relation between localization and intelligibility. *Journal of the Acoustical Society of America, 22,* 196–200.

Hollich, G. J., Newman, R. S., Jusczyk, P. W. (in press). Infants' use of synchronized visual information to separate streams of speech. *Child Development.*

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570–1582.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 28,* 1–23.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of talker variability on speech perception by 2-month-old infants. *Cognition, 43,* 253–291.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66,* 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech during early infancy. *Infant Behavior & Development, 6,* 263–285.

Litovsky, R. Y. (1997). Developmental changes in the precedence effect: Estimates of minimum audible angle. *Journal of the Acoustical Society of America, 102,* 1739–1745.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6,* 314–317.

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.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135–165.

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

Nozza, R. J., Rossman, R. N. F., & Bond, L. C. (1991). Infant–adult differences in unmasked thresholds for the discrimination of consonant–vowel syllable pairs. *Audiology, 30,* 102–112.

Nozza, R. J., Rossman, R. N. F., Bond, L. C., & Miller, S. L. (1990). Infant speech–sound discrimination in noise. *Journal of the Acoustical Society of America, 87,* 339–350.

Nozza, R. J., & Wilson, W. R. (1984). Masked and unmasked pure-tone thresholds of infants and adults: Development of auditory frequency selectivity and sensitivity. *Journal of Speech and Hearing Research, 27,* 613–622.

Pollack, I., & Pickett, J. M. (1958). Stereophonic listening and speech intelligibility against voice babble. *Journal of the Acoustical Society of America, 30,* 131–133.

Poulton, E. C. (1953). Two-channel listening. *Journal of Experimental Psychology, 46,* 91–96.

Singh, L., Bortfeld, H., Rathbun, K., & Morgan, J. (2001, November). *Effects of speech rate and sentence position on infant word recognition.* Paper presented at the biennial meeting of the Society for Research in Child Development. Minneapolis, MN.

Sinnott, J. M., Pisoni, D. B., & Aslin, R. N. (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior & Development, 6,* 3–17.

Soli, S. D., & Nilsson, M. J. (1997, June). *Predicting speech intelligibility in noise: The role of factors other than pure-tone sensitivity.* Paper presented at the meeting of the Acoustical Society of America, State College, PA.

Spieth, W., Curtis, J. F., & Webster, J. C. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America, 26,* 391–396.

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.

Van de Weijer, J. (1998). *Language input for word discovery.* Wageningen, the Netherlands: Ponsen & Looijen, BV.

Received January 13, 2004
Revision received June 3, 2004
Accepted July 26, 2004 ■