Infants' listening in multitalker environments: Effect of the number of background talkers

ROCHELLE S. NEWMAN
University of Maryland, College Park, Maryland

Infants are often spoken to in the presence of background sounds, including speech from other talkers. In the present study, we compared 5- and 8.5-month-olds' abilities to recognize their own names in the context of three different types of background speech: that of a single talker, multitalker babble, and that of a single talker played backward. Infants recognized their names at a 10-dB signal-to-noise ratio in the multiple-voice condition but not in the single-voice (nonreversed) condition, a pattern opposite to that of typical adult performance. Infants similarly failed to recognize their names when the background talker's voice was reversed—that is, unintelligible, but with speech-like acoustic properties. These data suggest that infants may have difficulty segregating the components of different speech streams when those streams are acoustically too similar. Alternatively, infants' attention may be drawn to the time-varying acoustic properties associated with a single talker's speech, causing difficulties when a single talker is the competing sound.

Infants often find themselves being spoken to in the context of background sound, including speech from one or more other talkers. For example, van de Weijer (1998) recorded all of the language input to which a single child was exposed over the course of 3 weeks and reported that there were multiple people speaking simultaneously during most of the time that the infant was outside of the house (e.g., in daycare or during shopping trips). Golden and Frank (2000) 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 SNRs for different teachers averaged only 5–6 dB. These findings suggest that young children frequently are spoken to in multitalker environments (see also Manlove, Frank, & Vernon-Feagans, 2001).

In order to learn language in these multitalker settings, infants must be able to separate one stream of speech, such as that of their caregiver's voice, from others. Although a great deal of research has been focused on how adult listeners separate speech from multiple talkers (Broadbent, 1952; Brokx & Nooteboom, 1982; Cherry, 1953; Hirsh, 1950; Pollack & Pickett, 1958; Poulton, 1953; Spieth, Curtis, & Webster, 1954), there has been much less research on infants' ability to do so. Furthermore, the aspects of the signal that might make the task of separating a talker's voice from the background easier for infants have not been well studied.

Understanding the factors that affect infant performance in a multitalker environment provides information about both the acoustic cues on which infants rely in their day-to-day listening and infants' processing abilities. In order to understand what one voice says despite the presence of background speech, infants must perform a number of tasks. Understanding the limitations of these processes forms an underpinning to many theories on infant language learning.

In order to separate streams of speech, infants must first cope with *energetic masking*, the masking of one sound by another at the auditory periphery. Because the frequency ranges of multiple streams of speech overlap, the competing signals mask one another. For very young infants, poor spectral resolution could make this form of masking even more difficult, although several studies suggest that spectral resolution is essentially adult-like by the time an infant is 6 months of age (see, e.g., Abdala & Folsom, 1995; Olsho, 1985; Schneider, Morrongiello, & Trehub, 1990; Spetner & Olsho, 1990; cf. Werner & Bargones, 1992, for a review).

In addition to coping with energetic masking and resolving different frequency components, infants must segregate the two sources of sound, which involves analyzing a complex sound into its components and grouping acoustic properties that "belong together" (those that originate from a single source) to distinguish them from those that do not.

Separating streams of speech requires a sensitive auditory system; infants have been shown to have poorer auditory thresholds and, more critically, poorer detection of speech and other sounds in the presence of noise. Studies suggest that 5- to 12-month-olds are more susceptible to masking of both tones and octave-band noises than are adults (Olsho, 1985; Schneider, Trehub, Morrongiello, & Thorpe, 1989). Infants show higher tonal thresholds

R. S. Newman, rnewman@hesp.umd.edu

in noise of approximately 8 dB, with performance being slightly more similar to adults for broadband noise targets (Bargones, Werner, & Marean, 1995; Nozza & Wilson, 1984; Werner & Boike, 2001). Infants also require greater stimulus intensity than do adults to discriminate among different speech sounds both in quiet and in noise (Nozza, Miller, Rossman, & Bond, 1991; Nozza, Rossman, & Bond, 1991; Nozza, Rossman, Bond, & Miller, 1990). Infants' unmasked thresholds for distinguishing speech sounds are roughly 25–30 dB higher than those of adults (Nozza, Rossman, & Bond, 1991), and, although adults can discriminate speech sounds in band-passed noise at SNRs of approximately −9 dB, infants require a minimum SNR of between −2 and −3 dB (Nozza, Miller et al., 1991; Nozza et al., 1990). Moreover, these results were found for the relatively simple task of discriminating two different CV syllables (such as /ba/ vs. /ga/). The authors suggest that the age-related differences would have a much greater effect in a more complex task, such as recognizing meaningful speech in noise. Although these speech-sound studies all used band-passed noise as the masker, such deficits would also likely result in poorer ability to resolve the different components in a multiple-voice setting.

After infants have grouped the appropriate cues into separate streams, they must choose to attend to the target signal as compared with the competing signal ("selective attention"). Thus, separating speech streams depends on the ability to attend selectively, another skill that is still developing in infants (Bargones & Werner, 1994). Failure to segregate sound sources or to attend selectively results in a type of masking that has been termed *informational masking* (IM; Durlach, 2006). When two voices can be separated and the listener can selectively attend to one (such as when there is an obvious perceptual distinction that listeners can focus on), there is less potential for uncertainty and thus less IM (Kidd, Mason, Deliwala, Woods, & Colburn, 1994). Infants may be more susceptible to IM than adults are. Several studies suggest that preschoolers and school-age children show more IM than adults do (see, e.g., Lutfi, Kistler, Oh, Wightman, & Callahan, 2003; Oh, Wightman, & Lutfi, 2001; Wightman, Callahan, Lutfi, Kistler, & Oh, 2003), and recent studies with infants suggest that they, too, are very susceptible to IM (Leibold & Werner, 2006). Even information in a nonoverlapping frequency range can interfere with infants' speech perception. For example, using a habituation procedure, Polka, Rvachew, and Molnar (2008) presented infants with a discrimination task involving two speech sounds, either in quiet or in the presence of background noise. The background noise was a high-frequency sound that did not overlap spectrally with the speech sounds, and thus there was no possibility of energetic masking. Despite that lack of overlap, infants showed poorer discrimination performance in the presence of this background noise; the authors suggest, similarly to Leibold and Werner, that IM may play a substantial role in infant perception. Such a susceptibility to IM could help to explain why infants appear to show greater effects of masking than would be expected on the basis of their fre-

quency resolution abilities (see Werner & Bargones, 1992, for a discussion of this "excess" masking).

Finally, adult listeners typically use their linguistic knowledge to help them extract information in a multitalker environment. Infants have limited experience with their native language. In this regard, they may be more comparable to second-language learners than to native adult speakers of a language, and several studies suggest that second-language learners have particular difficulties comprehending speech in noise or in the presence of other speech (Cooke, Garcia Lecumberri, & Barker, 2008; Garcia Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997; Takata & Nábelek, 1990).

Thus, separating a single voice from background speech is likely to depend on a variety of cognitive and perceptual processes, and an inability to succeed at any of these would result in poor performance in hearing speech in multitalker environments. In the present article, we refer to the general problem, which includes aspects of the segregation of sound sources, selective attention, and spectral resolution, combined, as *separating streams of speech*.

Given the range of limitations described above, we might expect infants' speech recognition to be especially handicapped by the presence of background voices. Indeed, although the first study to examine this issue found that infants were capable of separating streams of speech (Newman & Jusczyk, 1996), doing so was very difficult for them. Infants required substantially higher SNRs than would adult listeners, even with two talkers whose voices were very distinct acoustically. Such high SNRs may not always be available in the real world. As a result, follow-up work has begun exploring the types of cues that infants might use to help them separate different streams of speech.

Infants have been shown to make use of audiovisual synchrony in stream segregation (Hollich, Newman, & Jusczyk, 2005). In this study, 7.5-month-olds were familiarized with two voices presented simultaneously at a 0-dB SNR and were then tested on individual words that had occurred in the target speech stream. When the infants had been shown a static picture along with the audio signal during the familiarization phase, they later failed to recognize the words that the talker had said. Likewise, if shown a dynamic movie not synchronized with the speech, they failed to learn the words. However, if they were shown a synchronous signal (either of the talker's face while the talker was speaking or of an oscilloscope pattern matching the audio signal), infants later showed recognition of the target word, suggesting that audiovisual synchrony is one cue that infants use to separate streams of speech.

Infants also perform better if the target voice is more familiar to them. Barker and Newman (2004) found that infants were better able to understand speech in the presence of background talkers if the target voice was that of their own mother, as opposed to that of someone else's mother.

One cue that has yet to be investigated is the number of background voices—in particular, competition from a single voice versus multiple voices. This factor is interesting for a variety of reasons. First, a single competing voice is likely to be more similar acoustically to a single target

voice than is a multiple-voice combination. More specifically, a single voice contains patterns of amplitude and frequency modulation that are similar to those of other single voices but quite different from those of multitalker babble (Brungart & Simpson, 2002). This acoustic similarity could cause confusion as to which components of the input "belong" to one stream as opposed to the other; increased similarity between a masker and a target should increase IM (Kidd, Mason, & Arbogast, 2002).

Second, a single voice may attract attention, either because of its semantic content or because of the need to process that content. Adult listeners likely would recognize words spoken by a single competing voice and shift their attention away from the target to that stream of speech. This is less likely to be an issue for young infants than it is for adults, as a result of infants' limited lexical knowledge. However, listeners (including infants) may automatically process all speech signals that are present (Eich, 1984; MacKay, 1973; Treisman, 1960; Wood & Cowan, 1995; but see Wood, Stadler, & Cowan, 1997), as when fluent readers find it difficult to avoid reading words in the Stroop (1935) task. Such processing of the competing voice could take resources away from processing of the target voice. In a combination of many voices, each individual voice is less distinctive.

Finally, a single voice may attract attention on the basis of its acoustic, nonsemantic properties in a way that multiple voices might not (Brungart & Simpson, 2002; Brungart, Simpson, Darwin, Arbogast, & Kidd, 2005). It is possible that listeners could have an automatic response to attend to time-varying signals that have the properties of a single voice. Infants, in particular, have already been shown to attend to a single talker over white noise (Colombo & Bundy, 1981) and over sine wave analogues (Vouloumanos & Werker, 2004, 2007), and they might similarly attend to a single voice over a combination of voices. This could cause them to shift attention to the wrong stream in a task involving a single competing voice.

Thus, a single-voice competing sound differs from a multiple-voice competitor in many ways: It can be confused with the single target voice, its semantic content might attract attention or processing, and its time-varying acoustic properties could attract attention. These factors might cause infants to perform better in the presence of a multiple-voice background than in the presence of a single-voice background.

Surprisingly, though, adults perform better when a single background voice serves as a competing signal than when a multiple-voice background does so (Bronkhorst & Plomp, 1992; Drullman & Bronkhorst, 2004; Pollack & Pickett, 1958; Simpson & Cooke, 2005). This is in part because a single voice also varies in its moment-to-moment amplitude to a far greater extent than does multitalker babble (Bronkhorst & Plomp, 1992; Drullman & Bronkhorst, 2000, 2004). With a single talker as a competitor, there are periods of relative silence (e.g., when that talker pauses or during stop consonant closure) and times of more intense speech (e.g., during vowels). During the low-amplitude periods in the background, the target speech may be more audible. But as more talkers are added to the mix, these variations in amplitude tend to flatten out. Amplitude variations, such as those in the single-voice condition, are helpful for adult listeners, who perform better with a background noise that has amplitude modulations than with background noise that does not (e.g., Carhart, Tillman, & Greetis, 1969; Carhart, Tillman, & Johnson, 1966; Dirks & Bower, 1969; Festen & Plomp, 1990; Gustafsson & Arlinger, 1994; Miller & Licklider, 1950; Takahashi & Bacon, 1992; Wilson & Carhart, 1969). If infants similarly can capitalize on these gaps in masking, they might likewise be able to recognize the target voice's words during the amplitude dips and thus perform better with a varying-level masker, particularly in situations in which the target voice repeats words over and over, providing multiple opportunities to "catch" the target words.

Thus, there are a variety of hypothetical reasons to predict that a single-voice competing sound would differ from a multiple-voice background in terms of its effect on an infant listener. There is one final reason to investigate the effect of the number of voices, and that relates to a discrepancy in the literature. Newman and Jusczyk (1996) reported that 7.5-month-olds could separate speech from different talkers when the target signal was at least 5 dB more intense than the background speech. In contrast, a more recent study (Newman, 2005) reported that infants could not succeed at that level of noise until after their first birthday, although both 5- and 9-month-olds succeeded at a slightly easier 10-dB SNR. There were three major differences between these two studies. One was methodological: Newman presented infants with their own names as the target words (an item that was already well known to the infants), whereas Newman and Jusczyk first familiarized the infants with target words and then tested for those. This difference should have made the task easier in Newman and thus cannot explain why infants performed more poorly in that study. The other two differences have to do with the competing speech: Although both studies used a female talker as the target voice, Newman and Jusczyk used a single male voice as the background, whereas Newman used a combination of female voices. Thus, the number of the voices, the gender of the voices, or both, might explain the poorer performance in the later study. It is quite likely that gender was a factor in the discrepancy between studies, since other studies with infants and adults have reported effects of background gender (Barker & Newman, 2004; Brungart, 2001; see also Brokx & Nooteboom, 1982). Yet it is possible that the number of voices also played a role in this discrepancy. Examining whether infants perform better with a single voice as the competing sound could help explain the conflict between these two studies.

The present study investigates this issue of the effect of the number of background voices on infants' speech perception in a multitalker background. The methodology was identical to that of Newman (2005). Infants heard a woman's voice repeating a name on each trial. On some trials, the name that 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, either one female voice or nine different female voices spoke fluently in the

background. We measured whether infants listened longer on trials in which the target voice repeated their own name than on trials in which the voice repeated other names. Longer listening to their own names would indicate that the infants had been able to separate the target voice from the background and recognize the words that the target voice was repeating.

EXPERIMENT 1

Infants in the present study heard a target voice repeating their own name or a foil name in the context of competing speech sounds. These competing sounds consisted of either a single voice or a nine-voice combination (multiple-voice condition). The SNR was set at 10 dB, the level at which infants of a similar age in the Newman (2005) study succeeded at recognizing their names in the context of multi-talker babble. The multiple-voice condition is a direct replication of that study, and we would expect infants to listen longer to their names in this situation. Of interest is whether infants would also listen longer to their names when there was only a single background voice at the same SNR. If not, it would suggest that a single background voice is more difficult for infants than are multiple voices.

Method

Subjects. Of the 72 infant participants, 46 were male and 26 were female. Although this represents a gender imbalance, there is no reason to expect gender differences in audition at this age.[1] However, we tested effects of gender in order to ensure that there were no differences. Participants were recruited by means of letters sent to parents who listed birth announcements in the local newspapers. The infants had an average age of 4.8 months on the basis of due date (range, 17 weeks, 0 days, to 24 weeks, 3 days). Data from an additional 19 infants were excluded for the following reasons: crying/fussiness ($n = 8$), prematurity ($n = 5$), falling asleep ($n = 1$), or failure to orient to the lights or to listen for an average of 3 sec per item ($n = 5$). Because the present study focused on the infants' own names and not on particular words in the language, infants were not excluded for being in bilingual homes. 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 with which an infant was particularly familiar (such as family members or pets; $n = 3$); although we attempted to ascertain such information before testing, some parents changed their responses to these questions between the time of initial contact and the final visit.

Half of the infants were assigned to the single-voice condition (16 female, 20 male) and half (10 female, 26 male) were assigned to the multiple-voice condition, as described below. Infants in the single- and multiple-voice conditions did not differ in age [$t(70) = 1.21$, $p > .05$]. Although proportions of gender differed across conditions, this did not result in any interactions in the later analyses. [For the single-voice condition, gender × name and gender × talker × name, both $Fs < 1$; for talker × gender, $F(3,28) = 2.29$, $p > .10$; for the multiple-voice condition, gender × name and gender × talker × name, both $Fs < 1$; for talker × gender, $F(3,28) = 1.22$, $p > .10$]. We therefore ignored gender in the final analyses.

Stimuli. Stimuli consisted of both a target speech stream and a competing (or background) speech stream. All recordings were made with a Shure SM51 microphone in a sound-attenuated room. They were amplified, low-pass filtered at 44.1 kHz, digitized via a 16-bit analog-to-digital converter, and stored on computer disk.

The target speech stream consisted of a talker repeating a name over and over. Each infant heard trials consisting of four different names:

his/her own name, a name matched for stress pattern, and two names with different stress patterns. Infants might be able to pick out the stress pattern of their names at SNRs at which they could not pick out the particular phonetic pattern; if so, they should prefer listening to their names rather than to the nonmatching foils. The more critical question, however, is whether the child can also make out the phonology of his/her own name; to examine this, we compared listening time to the child's own name with listening time to the name matched for stress pattern. Because both names occurred in the presence of the same competing speech, listening to their own names longer than to the foil names can be taken as an indication that infants were able to identify the name despite the background speech.

The background speech stream varied across groups of infants, consisting of either a single female talker or a combination of nine female talkers. To control for acoustic variability among the background voices, each of the nine voices in the multiple-voice condition appeared in the single-voice condition one ninth of the time. In this way, the acoustic properties of the talkers presented to infants in the single-voice condition were matched to (or included in) those presented in the multiple-voice condition.

To create these competing sounds, we recorded nine women reading passages aloud from a variety of books. Long pauses in the recordings (as when the speaker turned the page or stopped to clear her throat) were excised to ensure that the talkers were speaking at most moments. This does not eliminate amplitude variability, because there were still stop consonant closures, breaks for air, short pauses, and so forth, as well as the natural variations between amplitude levels of different phonemes. The removal of long breaks made the individual passages sound a bit more like a memorized monologue 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 root mean square (RMS) amplitude. These passages served as the competing sounds for the single-voice condition; they were then blended together at equal ratios to create the competing sound for the multiple-voice condition.

In order to ensure that our effects generalized across talkers, four different talkers served as target voices. The same target voices were heard in the single- and multiple-voice conditions. This resulted in the need for 36 infants for the single-voice condition (4 target voices × 9 background voices = 36) and an additional 36 infants for the multiple-voice condition, for a total of 72 infants.

The women recorded in the target passages 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 talker from producing the target name in a more engaging manner than the foil names, the talkers 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 name recordings also served as foil names for other children (see the Appendix). All four names heard by any given child were recorded by the same talker. Pauses between names were adjusted so that the four sound files (name, stress-matched name, and nonmatched foils) were of the same overall duration. No single recording occurred more than three times in any condition. The complete list of names used in the present study is given in the Appendix.

In order to adjust the SNR between the stimuli, we needed measurements of the average intensities of both the isolated names and the fluent competing passages. Because the children's names and the foil names were separated by periods of silence, the average intensity level (RMS amplitude, measured on the digital signal) of the recording as a whole was lower than the intensity of the fluent speech background; the periods of silence served to make the intensity level overall seem lower than the actual level while the individual was talking. To adjust for this, we created an edited version of each name list with 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 four original name recordings were then adjusted so that their edited versions had the same average intensity level, and the average intensity level of the competing stimulus was adjusted to be 10 dB less than that of these edited versions of the name lists. This resulted in four test stimuli of the same RMS amplitude, each of which were combined with a competing stimulus 10 dB less intense than the target voice.

A 500-msec silent period was prepended to each name list; the competing passage was then adjusted to be at least 500 msec longer than the duration of the name list. In this manner, the competing passage began prior to the onset of the names and remained at full amplitude until the final name was recorded. The same competing passage was used for all four name stimuli for any given participant. Finally, the competing passage and the name list were combined into a single sound file. Because infants' names varied in length (from one to four syllables), the duration of stimuli was not constant across children (varying from 21.4 to 35.4 sec), although all four stimuli were identical in length for each child. Although some infants occasionally did listen to the end of a trial, no infant did so regularly; thus, the overall performance of different infants was never as long as any of these trial lengths, and this variation across infants would not affect the final results.

As noted in the introduction, amplitude variability tends to be far greater for a single voice speaking than for nine voices speaking; this was certainly the case here. The difference between the minimum and maximum power levels across the competing sound sequence was only 14.9 dB for the nine-voice mixture. For the single voices, this difference between minimum and maximum values ranged from 45 to 73 dB for different talkers, with an average of 57 dB. Clearly, then, the single-voice backgrounds differed from the mixture, not only in being simply one voice but also in the degree to which amplitude levels varied throughout the recording.

Apparatus. The experiment took place in a three-sided test booth constructed from pegboard panels (4 ft × 6 ft). In the center of the front panel was a light for attracting the child's attention and a hole for the lens of a video camera used to record each session. The experimenter sat behind the front wall of the booth and watched the session via a monitor connected to the camera. Each side panel had a red light and a loudspeaker (NHT Super One, with a frequency response from 57 Hz to 25 kHz, ±3 dB), which was located behind the center of the panel. A tan curtain suspended from the ceiling prevented the infant from seeing over the top of the booth. A Macintosh G3 computer located behind the front panel was used to control 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. Each infant sat on a caregiver's lap in the center of the test booth. Each trial began with the blinking of the light in the center of the front panel. As soon as the infant oriented in that direction, this light was turned off and one of the two side lights began to flash. When the infant 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. Listening time was assessed as the amount of time the infant spent looking at the "source" of the sound (the flashing light). Any time the infant spent looking away (whether 2 sec or less) was excluded when measuring total listening time. The red light continued to flash for the duration of the trial. Data about the duration and direction of head turns and total trial durations were stored on the computer.

To familiarize the infants with the task, there was an initial practice phase, during which the infants heard one of two musical passages (14.9 sec long) on alternating trials until they accumulated at least 25 sec of listening time to each passage.

The test phase began immediately after the listening criteria for the practice phase was reached. During the test phase, infants heard three repetitions of each of four different name passages (the passages containing the child's name, a stress-matched foil, and two

nonmatched foils). These 12 trials were grouped into four blocks, such that each name occurred once in a given block, with the order of the four names randomized within the block. Although the two names that mismatch in stress were expected to entail identical listening times, we used two such items 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.
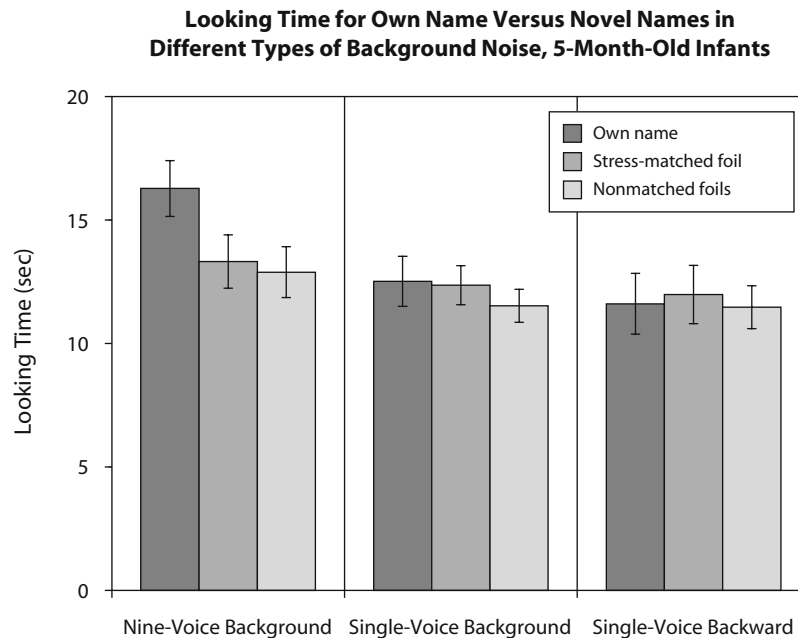
The experimenter behind the center panel pressed a button on the response box whenever the infant looked at or away from the flashing light. Both the experimenter and the caregiver listened to masking music over Peltor aviation headphones so that they could not differentially influence the infant's behavior 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. (Two children completed only two blocks.) Because the two nonmatching names were not expected to differ, listening times to these names were averaged for all statistical analyses, resulting in three name types: own name, stress-matched foil, and nonmatched foils.

We examined the data from the multiple-voice condition first. Because this condition was a direct replication of Newman (2005), we expected that infants would listen significantly longer to their own names than to the foil names and that there would be no effect of the particular talker. Indeed, a 4 (talker) × 3 (name) ANOVA found no effect of the particular voice used [$F(3,32) = 1.55, p > .05$] and no interaction between talker and name [$F(6,64) < 1$], but a significant effect of name [$F(2,64) = 5.36, p < .001$]. We used follow-up directional t tests to determine whether infants listened longer to their own names than to each of the other foils. These showed that infants listened longer to their own names than to both the stress-matched foil [$t(35) = 2.49, p < .05$] and the nonmatched foils [$t(35) = 3.02, p < .005$]. This pattern can be seen in the left panel of Figure 1. Overall, infants averaged 16.3 sec of listening to their own names, but only 13.3 sec listening to the stress-matched foil. Nevertheless, only 23 of 36 infants showed this pattern. This variable performance is fairly typical of research with infants, but does suggest that the task was indeed difficult for infants.

We then examined the data from the single-voice condition. There was a significant effect of target voice [$F(3,32) = 3.30, p = .033$]; infants appeared to listen longer to one of the target voices than to the others. However, there was no interaction between this factor and the name [$F(6,64) < 1$], suggesting that it is not likely to be an effect of speech stream separation but more an overall preference for that target's voice. Interestingly, there was also no effect of the name itself [$F(2,64) < 1$]; infants listened 12.5 sec on average to their own names and 12.4 sec to the stress-matched foil (see Figure 1, center panel). Although their average listening was slightly lower to the nonmatched foils (11.5 sec), none of the three names differed significantly from one another [own name vs. stress-matched foil $t(35) = 0.13$; own name vs. nonmatched foil, $t(35) = 0.89$; stress-matched foil vs. nonmatched foil, $t(35) = 0.94$; all ps $> .05$].

**Looking Time for Own Name Versus Novel Names in
Different Types of Background Noise, 5-Month-Old Infants**

Figure 1. Mean listening times and standard errors to the child's own name and
to the foil names in the three conditions (multiple voice, single voice, and single voice
backward) for the 5-month-olds.

Thus, the present experiment suggested that, although infants *are* able to separate the speech of a target voice repeating a word from multitalker babble and recognize their own names in this type of competing sound, they do not appear to do so at the same SNR when there is a single voice speaking in the background. Why might this be the case? One possibility is that the infants were able to separate the streams of speech, but that they did not have a preference for listening to the list of names over a single competing voice, whereas they did prefer listening to the name list over the multiple voice babble. Another possibility is that the single background voice was attracting their attention away from the target voice despite their preferences. However, if infants actually preferred listening to the competing passage, one might expect that the infants would have listened longer overall in the single-voice condition than in the multiple-voice condition, at least when the child's own name was not being presented. This was not the case, however; to the extent that there was any difference at all, infants tended to listen longer to the foil names in the multiple-voice condition than in the single-voice condition (12.9 vs. 11.5 sec). Rather, it appears that the single-voice background made it more difficult for infants to make out what the target voice was saying, perhaps as a result of attentional capture, but not that it actually caused infants to listen longer.

Before going on to examine why this might have occurred, it was important to address one concern regarding the age of the infants tested. The results clearly suggest that there is an age-related change in the effect of the number of background voices: 5-month-olds appear to have performed better when the competing sound consisted of multiple talkers, whereas adults perform better when the competing sound consists of a single talker. When might such a change take place? As noted above, Newman and Jusczyk (1996) had found better performance in their study by using a single-voice competing sound than was found in the present experiment. Their study involved 7.5-month-olds. One concern, then, was that the effect of the number of background voices actually might change between 5 and 8 months of age. That is, the discrepancy between study results might have occurred because of the age of the infants tested. To test this, we replicated Experiment 1 with older infants, closer in age to (but slightly older than) those tested in Newman and Jusczyk.

EXPERIMENT 2

This experiment was identical to Experiment 1, except in the age of the infants being tested.

Method

Subjects. Of the 54 infants who participated in this experiment, 27 (14 male, 13 female) participated in the multiple-voice (control) condition and 27 (17 male, 10 female) participated in the single-voice condition. The infants had an average age of 8.5 months on the basis of due date (range, 7.4–9.7 months). The data from an additional 15 infants were not included for the following reasons: crying/fussiness ($n = 9$), ear infection on test date ($n = 1$), wrong age at test date (usually the result of multiple reschedulings, $n = 3$), or experimenter error ($n = 2$).

Stimuli. The stimuli were identical to those in Experiment 1, with the exception that they were often new names and thus were new target voice recordings. Three speakers were recorded for the target speech in the present experiment; of these, one had also been a target speaker in Experiment 1.[2] The same nine single-voice background talker recordings were used, as well as the same nine-voice combination recording.

Design, Apparatus, and Procedure. These were identical to those in the previous experiment.

Results and Discussion

Mean listening times to the four different names were calculated for each infant, as in Experiment 1. We first examined the data from the multiple-voice condition, and, as in Experiment 1, we found no effect of talker and no interaction with talker (both $Fs < 1$). Surprisingly, however, the effect of name missed significance [$F(2,48) = 2.60, p = .085$]. It is not clear why this would be the case; perhaps the older infants were simply more variable in their performance than were the 5-month-olds. Despite the marginal nature of this result, we still performed the follow-up directional t tests, given our strong predictions for this multiple-voice condition (on the basis of both Experiment 1 and our prior research, of which this condition is a direct replication). As expected, we found that infants listened longer to their own names than to both the stress-matched foils [$t(27) = 1.75, p < .05$] and the nonmatched foils [$t(26) = 2.14, p < .05$]. This pattern can be seen in Figure 2. Overall, infants averaged 11.3 sec of listening to their own names, but only 9.5 sec listening to the stress-matched foils and 9.5 sec listening to the nonmatched foils; however, only 15 of 27 infants showed this pattern, suggesting that this was a difficult task.

We then examined the data from the single-voice condition. There was no effect of target voice [$F(2,24) = 1.71, p > .05$] nor were there any interactions between this and the name ($F < 1$). As in Experiment 1, there was no effect of the name itself [$F(2,48) = 1.79, p > .05$], although there was what may appear to be a trend in that direction, as can be seen in the right-hand panel of Figure 2. Infants listened 11.3 sec on average to their own names,

9.8 sec to the stress-matched names, and 9.8 sec to the foil names.[3] This trend was not significant; however, it may be that some infants at this age were able to succeed in the single-voice condition, just as they were in the multiple-voice condition. Perhaps still older infants would show a significant effect at this age. Critically, though, these 8.5-month-olds were not performing *better* in this condition than in the multiple-voice condition.

In general, the findings from the 8.5-month-olds are somewhat similar to those for the younger infants, with the exception that the overall ANOVA just missed significance in the nine-voice condition, perhaps as a result of greater variability. There was an apparent trend toward a preference for their own names in the single-voice condition, but this did not reach significance. There is no evidence to suggest that listening in a single-voice background was easier for these infants than listening in a multitalker voice background. This, in turn, suggests that the difference between the results from Experiment 1 (in which infants failed to recognize speech with a single-talker competing sound) and that of Newman and Jusczyk (1996) was unlikely to be simply the result of the difference in age across studies.

The results from Experiments 1 and 2 suggest that listening in the context of a competing sound is at least as difficult for infants when that competition consists of a single talker than when it consists of multiple talkers. This is somewhat surprising, given the pattern of results in earlier infant literature. For example, Newman (2005) found poorer performance for infants when a nine-voice background masker was used than Newman and Jusczyk (1996) found when a single-voice masker was used. One possible explanation that Newman gave for this difference in performance was that she used multitalker babble. That explanation hinges on the idea that a single

Looking Time for Own Name Versus Novel Names in Different Types of Background Noise, 8.5-Month-Olds
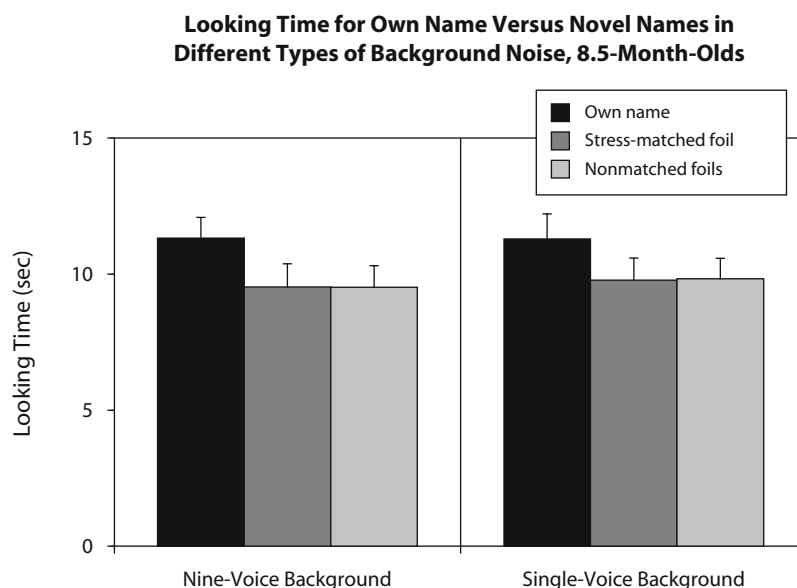
Figure 2. Mean listening times and standard errors to the child's own name and to the foil names in the two conditions (multiple and single voice) for the 8.5-month-olds.

talker is easier to segregate, but the present results suggest that a single talker is actually harder to segregate. It appears more likely that the improved performance in the Newman and Jusczyk study was a result not of their use of a single voice as the competition but of their use of a male voice, rather than the female voices used here. This is consistent with the suggestion by Barker and Newman (2004) that a change in gender between target and background voices may be an especially potent cue to infants' separation of speech streams, just as it is for adults' stream segregation (for effects of gender, see Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001; Cullington & Zeng, 2008; for effects of the associated feature of fundamental frequency range, see Assmann & Summerfield, 1990; Brokx & Nooteboom, 1982; Chalikia & Bregman, 1989; Scheffers, 1982, 1983; Zwicker, 1984).

It appears that infants find it slightly more difficult to separate a single talker from a single competing voice than from multiple competing voices. There are several possibilities why this might occur, but each revolves around the notion that a single voice attracts infant attention in some manner. It seems unlikely that the semantic content of the single voice attracts attention, because infants 5 months of age have limited semantic knowledge. Rather, attention could be attracted by a spectral property of the signal, a temporal property of the signal, or simply the fact that it sounds like a voice talking and, thus, like something that could be interesting. As a first step toward narrowing these possibilities, Experiment 3 investigated infant performance in a condition where the competing sound consisted of a single voice played in reverse. This resulted in an acoustic signal that had very similar spectral composition to that of the forward signal and the same time-varying nature as that of the forward signal, but that did not sound like normal speech (Gallun, Mason, & Kidd, 2007; Hygge, Rönnberg, Larsby, & Arlinger, 1992). If infants were attending to the single background voice because it sounded potentially interesting, this effect should be reduced when the competing sound did not seem to be real speech, and we would expect them to listen longer to their own names in Experiment 3.

EXPERIMENT 3

This experiment was identical to the single-voice condition in Experiment 1, except that the background speech was reversed before being blended with the target voice. Because the only apparent difference between the two ages from the prior experiments had been that the results from the older infants were more variable, we returned to testing the younger infants.

Method

Subjects. Twenty-seven infants (14 male, 13 female) participated in this experiment. The infants had an average age of 5.1 months on the basis of due date (range, 17 weeks, 0 days, to 24 weeks, 4 days). The data from an additional 8 infants were not included for the following reasons: crying/fussiness ($n = 2$), failure to attend ($n = 2$), ear infection on test date ($n = 1$), and wrong age or excessive prematurity ($n = 3$).

Stimuli. The stimuli were identical to those in Experiment 1, with the exception that often they were new names and thus new target voice recordings. Three of the four target speakers from Experiment 1 were recorded for the target speech in the present experiment. The same nine background talker recordings were used but were reversed prior to being blended with the target voice.

Design, Apparatus, and Procedure. These were identical to those in the previous experiments.

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 target voice) found an effect of the particular target voice used, as in the prior experiment [$F(2,24) = 3.42$, $p < .05$], but no effect of name [$F(2,48) < 1$] and no interaction [$F(2,48) < 1$]. Infants spent nearly identical amounts of time listening to the three names in this experiment (11.6 sec to their own names, 12.0 sec to the stress-matched foils, and 11.5 sec to the nonmatched foils), as is shown in the far right panel of Figure 1. Thus, even when the speech from the background talker was reversed and no longer sounded like normal speech, it still appeared to prevent infants from recognizing what the target voice was saying. Had the infants in Experiment 1 recognized individual words in the single-voice background speech stream, causing them to shift their attention away from the target stream, we would have expected infants to show a different pattern in the present experiment, in which the background stream sounded less speech-like and no individual words could be identified. However, we did not find any such difference. Indeed, a comparison of the single-voice (Experiment 1) and reversed single-voice (Experiment 3) results, using a 4 (talker) × 2 (direction: forward vs. backward) × 3 (name) ANOVA showed no effect of the name [$F(2,112) < 1$] and no name × direction interaction ($F < 1$).

These results imply that infants' difficulty with the single-voice condition in Experiment 1 was not likely to have been the result of having recognized words in the background speech stream or because they were choosing to attend to signals that sounded like fluent speech. Rather, it would appear that some aspect of the acoustic properties of the signal that were maintained when speech was reversed made it more difficult for infants to separate the speech streams and to selectively attend to the target voice.

It is possible, of course, that we found the same effects but for different reasons. Perhaps in Experiment 1, infants were distracted by recognizing words in the competing voice, whereas here they were distracted by the fact that the background sounded unnatural. It is impossible to explore fully all the potential reasons for the lack of an effect in the present study or all the factors that distinguish a single voice speaking from multitalker babble. However, it appears most likely that a single-voice competing sound contains some acoustic or temporal property that attracts infant attention away from the target voice or precludes recognizing the target speech.

GENERAL DISCUSSION

The present study replicated previous findings suggesting that by the time infants reach 5 months of age, they can separate speech produced by different talkers. When faced with a single target voice that is at least 10 dB more intense than a background of multitalker babble, infants listened longer to their own names than to other names. This suggests that infants were able to separate the target speech stream from the background speech and to recognize individual words in that stream. Furthermore, the present results suggest that, even when speech is masked, infants can recognize the representations for words with which they are familiar.

However, the type of competing sound is clearly a critical factor in infants' abilities. Infants at both 5 and 8.5 months of age appeared to be able to recognize their own names at a 10-dB SNR when the competing speech contained multiple voices but not when it was a single voice. This was so even when the background speech was played backward, which suggests that infants were not merely recognizing individual words in the single-voice background speech stream and shifting their attention toward that stream in response. Rather, the properties that make a single voice more distracting to infants appear to be related to its time-varying acoustic properties, rather than anything related to meaning or content.

Surprisingly, the pattern shown by the infant listeners in the present study is exactly the opposite of the pattern typically shown by adults (Bronkhorst & Plomp, 1992; Pollack & Pickett, 1958). Adults consistently perform better in the single-voice masker condition than in the multiple-voice masker condition, given equal SNRs. Infants show the opposite pattern. The age at which the adult pattern of performance emerges during development and what mechanisms support this change are not known; however, the present results suggest that this change occurs after 9 months. Future work should explore more directly the question of when this change occurs.

There are a variety of possible explanations for the differences between the single- and multiple-voice conditions tested here. As noted, single voices tend to have much more varying amplitude levels than does a combination of voices. The present results suggest that infants are unable to take advantage of those dips, perhaps because they lack an ability to focus attention at points when the competing signal is less intense. In fact, if infants are unable to focus their attention selectively, their efforts might be hindered by the time-varying nature of the signal: They might experience greater masking from the more intense portions of the competing signal without the counteracting benefit of being able to take advantage of the less intense portions of the signal. This could be explored in the future by comparing infant performance with background babble that had greater or lesser degrees of amplitude variation; according to this account, a flatter amplitude contour would result in better separation of speech streams.

A second possibility is that the properties of the single-voice background signal attracted infant attention away from the target voice. Perhaps the single voice not only masked the target speech stream but caused infants to attend to the wrong signal. Previous work has suggested that even very young infants attend preferentially to speech over nonspeech signals (Colombo & Bundy, 1981; Glenn, Cunningham, & Joyce, 1981; Vouloumanos & Werker, 2004, 2007). Perhaps infant attention is drawn not toward speech in particular but toward anything that has the general acoustic patterns associated with human speech signals. A single voice may have more of these properties than does background babble. Although playing a speech signal backward alters some of these acoustic properties, the general prosodic and spectral patterns of forward speech are likely to be quite similar to those of backward speech. This would result in infants' attention being drawn toward the competing signal both in Experiment 3 and in the single-voice conditions of Experiments 1 and 2.

Despite these acoustic similarities, there are subtle acoustic differences between reversed and forward speech to which adults and infants could have been sensitive. Indeed, newborns have been shown to be able to discriminate different languages when speech samples are played forward, but not when the samples are played backward (Ramus, Hauser, Miller, Morris, & Mehler, 2000). This implies that infants must be sensitive to some cues that differ between forward and backward speech. Reversing the direction of speech has different effects on different segments (having less influence on steady-state portions than on transitions, for example), and these differential patterns provide cues that infants may be able to pick up on. Yet the present results suggest that these cues do not influence the extent to which a signal can be ignored. Instead, some aspect of the acoustic signal that is the same between reversed and forward speech, but differs between one voice and multiple voices, appears to make speech distracting to infants.

Brungart et al. (2005) argued that normal speech and time-reversed speech "share a common set of acoustic features that interfere in some way with central speech processing" (p. 293). One likely candidate for such an acoustic feature is the spectral modulation in the signal. The authors tested adults' ability to identify words in a target stream while a second speech or nonspeech stream occurred simultaneously. They measured the amount of across-ear interference from the competing stream. Whenever the competing signal contained "speech-like [spectral] modulations," it generated across-ear interference. On the basis of the present results, it seems likely that a similar factor influences infant selective attention as well: When a competing signal has the time-varying changes of a single talker's voice, it prevents infants from attending to the target speaker.

Infants in the present study showed similar patterns with forward versus reversed speech. Studies with adult listeners comparing the effectiveness of forward and backward speech maskers have shown a mixed pattern of results: Some show improved performance when a competing speech stream is played in reverse, whereas others do not (e.g., Dirks & Bower, 1969; Hygge et al., 1992; Summers & Molis, 2004), depending on the particulars of the task and the other cues available for segregation. One expla-

nation for such differences, when found, has focused on the notion that intelligibility of the distractor signal causes additional IM. When listening conditions are easier, listeners can sometimes take advantage of semantic differences between forward and reversed speech to aid in their segregation (Brungart & Simpson, 2002). Some have argued that, although forward speech has both informational and energetic masking components, backward speech contains only energetic components, because it is not intelligible as speech (Summers & Molis, 2004). Because 5-month-olds are unlikely to be able to interpret much of the speech signal, they would be less likely to show a difference between reversed and forward speech than are adults.

In summary, it appears that infants, like adults, experience the IM effects caused by a time-varying, speech-like signal. This results in poorer performance for more speech-like maskers. For adults, such effects are complicated by two other factors: Adults can focus their attention to amplitude dips in the masker, improving performance in a single-voice condition relative to a multiple-voice condition, and they can use semantic differences in the distractor to help them separate different streams, improving performance in a reversed-speech condition relative to a forward-speech condition. Infants do not appear to demonstrate the latter effect and perhaps are also unable to listen in the dips.

The present findings help to explain the apparent conflict between two previous studies. Newman (2005) found poorer detection performance for infants using a nine-voice background masker than Newman and Jusczyk (1996) found using a single-voice masker. One possible explanation Newman has given for this difference in performance is the fact that she used multitalker babble. However, the present results clearly suggest the opposite: A single talker is harder to segregate.

Instead, the discrepancy is likely to have been caused by the other major difference between the two studies: the gender of the background talker(s). Both the present study and Newman (2005) used female foreground and background talkers. In contrast, Newman and Jusczyk (1996) used a male talker as their competing sound. Given the present findings, it appears likely that the improved performance in the Newman and Jusczyk study resulted not from their use of a *single* voice as the distractor but from their use of a *male* voice. This corroborates the suggestion by Barker and Newman (2004) that a difference in gender between target and background voices may be an especially potent cue to infant separation of speech streams, as it is for adults. This, too, leads to potential directions for future work. There are many ways in which a male talker and a female talker typically differ—for example, in $f0$ range and formant range. These factors lie on a continuum, and it would be possible to create voices that were more versus less distinct on the basis of these factors, treating them as continuously varying, rather than dichotomous with respect to gender. By manipulating the similarity between the background voice and the target voice in a stepwise manner, we could explore which aspects of the signal are most critical.

In conclusion, the results of the present study suggest that infants find it more difficult to understand speech in competing sounds when the background consists of a single talker rather than multitalker babble, a pattern opposite that typically shown by adults. This difference does not appear to be the result of infants detecting individual words in the background speech signal and switching their attention as a result; rather, it appears that the acoustic properties of a single-voice background either attract infant attention or serve as a better masker of a target speech stream. Regardless of the reason, the results suggest that infants could have difficulty understanding speech in the noise levels typical of many infant settings.

REFERENCES

Abdala, C., & Folsom, R. C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *Journal of the Acoustical Society of America*, **96**, 921-930.

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America*, **88**, 680-697.

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

Bargones, J. Y., Werner, L. A., & Marean, G. C. (1995). Infant psychometric functions for detection: Mechanisms of immature sensitivity. *Journal of the Acoustical Society of America*, **98**, 99-111.

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

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

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

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.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, **109**, 1101-1109.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustical Society of America*, **112**, 2985-2995.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G., Jr. (2005). Across-ear interference from parametrically de-

graded synthetic speech signals in a dichotic cocktail-party listening task. *Journal of the Acoustical Society of America*, **117**, 292-304.

BRUNGART, D. S., SIMPSON, B. D., ERICSON, M. A., & SCOTT, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, **110**, 2527-2538.

CARHART, R., TILLMAN, T. W., & GREETIS, E. S. (1969). Perceptual masking in multiple sound backgrounds. *Journal of the Acoustical Society of America*, **45**, 694-703.

CARHART, R., TILLMAN, T. W., & JOHNSON, K. R. (1966). Binaural masking of speech by periodically modulated noise. *Journal of the Acoustical Society of America*, **39**, 1037-1050.

CHALIKIA, M. H., & BREGMAN, A. S. (1989). The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation. *Perception & Psychophysics*, **46**, 487-496.

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.

COLOMBO, J. A., & BUNDY, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior & Development*, **4**, 219-223.

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.

CULLINGTON, H. E., & ZENG, F.-G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *Journal of the Acoustical Society of America*, **123**, 450-461.

DIRKS, D. D., & BOWER, D. R. (1969). Masking effects of speech competing messages. *Journal of Speech & Hearing Research*, **12**, 229-245.

DRULLMAN, R., & BRONKHORST, A. W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America*, **107**, 2224-2235.

DRULLMAN, R., & BRONKHORST, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *Journal of the Acoustical Society of America*, **116**, 3090-3098.

DURLACH, N. (2006). Auditory masking: Need for improved conceptual structure. *Journal of the Acoustical Society of America*, **120**, 1787-1790.

EICH, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, **12**, 105-111.

FESTEN, J. M., & PLOMP, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, **88**, 1725-1736.

GALLUN, F. J., MASON, C. R., & KIDD, G., JR. (2007). The ability to listen with independent ears. *Journal of the Acoustical Society of America*, **122**, 2814-2825.

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.

GLENN, S. M., CUNNINGHAM, C. C., & JOYCE, P. F. (1981). A study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Child Development*, **52**, 1303-1307.

GOLDEN, M. V., & FRANK, T. (2000). Further acoustic analysis of infant/toddler rooms in daycare centers. *Journal of the Acoustical Society of America*, **107**, 2833.

GUSTAFSSON, H. Å., & ARLINGER, S. D. (1994). Masking of speech by amplitude-modulated noise. *Journal of the Acoustical Society of America*, **95**, 518-529.

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

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.

HYGGE, S., RÖNNBERG, J., LARSBY, B., & ARLINGER, S. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *Journal of Speech & Hearing Research*, **35**, 208-215.

KIDD, G., JR., MASON, C. R., & ARBOGAST, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America*, **111**, 1367-1376.

KIDD, G., JR., MASON, C. R., DELIWALA, P. S., WOODS, W. S., & COLBURN, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America*, **95**, 3475-3480.

LEIBOLD, L. J., & WERNER, L. A. (2006). Effect of masker-frequency variability on the detection performance of infants and adults. *Journal of the Acoustical Society of America*, **119**, 3960-3970.

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.

MACKAY, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology*, **25**, 22-40.

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.

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

MILLER, G. A., & LICKLIDER, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, **22**, 167-173.

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., & JUSCZYK, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics*, **58**, 1145-1156.

NOZZA, R. J., MILLER, S. L., ROSSMAN, R. N. F., & BOND, L. C. (1991). Reliability and validity of infant speech-sound discrimination-in-noise thresholds. *Journal of Speech & Hearing Research*, **34**, 643-650.

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 & Hearing Research*, **27**, 613-622.

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.

POLKA, L., RVACHEW, S., & MOLNAR, M. (2008). Perceiving speech in the presence of other sounds poses a cognitive challenge for young infants. *Infancy*, **13**, 421-439.

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.

RAMUS, F., HAUSER, M. D., MILLER, C., MORRIS, D., & MEHLER, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, **28**, 349-351.

RIBEIRO, F. M., & CARVALLO, R. M. (2008). Tone-evoked ABR in full-term and preterm neonates with normal hearing. *International Journal of Audiology*, **47**, 21-29.

SCHEFFERS, M. T. M. (1982). The role of pitch in the perceptual separation of simultaneous vowels II. *IPO Annual Progress Report*, **17**, 41-45.

SCHEFFERS, M. T. M. (1983). *Sifting vowels: Auditory pitch analysis and sound segregation*. Unpublished thesis, University of Groningen.

SCHNEIDER, B. A., MORRONGIELLO, B. A., & TREHUB, S. E. (1990). Size of critical band in infants, children, and adults. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 642-652.

SCHNEIDER, B. A., TREHUB, S. E., MORRONGIELLO, B. A., & THORPE, L. A. (1989). Developmental changes in masked thresholds. *Journal of the Acoustical Society of America*, **86**, 1733-1742.

SIMPSON, S. A., & COOKE, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N*. *Journal of the Acoustical Society of America*, **118**, 2775-2778.

SPETNER, N. B., & OLSHO, L. W. (1990). Auditory frequency resolution in human infancy. *Child Development*, **61**, 632-652.

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.

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

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

TAKAHASHI, G. A., & BACON, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech & Hearing Research*, **35**, 1410-1421.

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.

TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242-248.

VAN DE WEIJER, J. (1998). *Language input for word discovery*. Wageningen, The Netherlands: Ponsen & Looijen.

VOULOUMANOS, A., & WERKER, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, **7**, 270-276.

VOULOUMANOS, A., & WERKER, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, **10**, 159-164.

WERNER, L. A., & BARGONES, J. Y. (1992). Psychoacoustic development of human infants. *Advances in Infancy Research*, **7**, 103-145.

WERNER, L. A., & BOIKE, K. (2001). Infants' sensitivity to broadband noise. *Journal of the Acoustical Society of America*, **109**, 2103-2111.

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.

WILSON, R. H., & CARHART, R. (1969). Influence of pulsed masking on the threshold for spondees. *Journal of the Acoustical Society of America*, **46**, 998-1010.

WOOD, N. [L.], & COWAN, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 255-260.

WOOD, N. L., STADLER, M. A., & COWAN, N. (1997). Is there implicit memory without attention? A reexamination of task demands in Eich's (1984) procedure. *Memory & Cognition*, **25**, 772-779.

ZWICKER, U. T. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Communication*, **3**, 265-277.

NOTES

1. One place where gender differences have been found among infants is in auditory brainstem response latencies. However, this has been argued to be the result of faster neural travel times as a result of a shorter cochlea in girls; there is no expectation of developmental hearing differences among male and female infants (Ribeiro & Carvallo, 2008).

2. One of the voices we originally had planned to use in both this experiment and Experiment 1 was that of a student who graduated before the present experiment was completed, making it difficult to finish the study using her voice. The data presented above include the remaining three voices, only one of which overlaps with those used in Experiment 1. However, we also performed the analysis a second way. This voice had also been one of the five used in Newman (2005); the multiple-voice condition of the present experiment is an exact replication of that earlier study. There were 5 participants in that earlier study who heard this particular talker's voice. We managed to test 13 new participants with this voice (9 in the single-voice condition, 4 in the multiple-voice condition) and then included those 5 infants from the prior study in the present multiple-voice condition analysis. The pattern of results including this additional talker is identical to those reported above, including the marginal main effect. The concern with reporting these results is due to the fact that the data for 5 of the participants had been reported previously; however, the conditions were equivalent, and all infants from that prior study were used.

3. The statistical null results suggest no effect of the child's name, but these averages might imply such a trend. However, the addition of the fourth talker described earlier reduces this pattern: Listening time for the 36 infants averaged 10.7 sec for their own name, 9.4 sec for the stress-matched foil, and 10.1 sec for the foil names. This suggests that what might appear to be a trend in the listening time data is not simply a null result for lack of power.

APPENDIX
Names Used

Target	Foils		Speaker	
	Stress-Matched	Nonmatched	Target	Background
Single Voice, 5-Month-Olds				
Juan David	Beatriz	Ann Rosa, Guillermo	M.H.	A.T.
Peter	Nathan	Keon, Roseanne	J.A.	A.T.
Adam	Ilse	Lynette, Nicole	L.S.	A.T.
Kyle	Eden	Shereen, Marvelle	A.R.	A.T.
Ella	Michael	Norit, Tamar	M.H.	B.W.
Keon	Jeanine	Mehdi, Peter	J.A.	B.W.
Ilse	Morgan	Danielle, Lynette	L.S.	B.W.
Michael	Joey	Kareem, Enriq	A.R.	B.W.
Hadassah	Mackenna	Annabelle, Tabitha	M.H.	I.H.
AJ	David	Adelle, Babette	J.A.	I.H.
Nicholas	Joshua	Melissa, Samantha	L.S.	I.H.
Casey	Deena	Roseanne, Elaine	A.R.	I.H.
Isabel	Juan David	Guillermo, Ann Rosa	M.H.	J.B.
Owen	David	Babette, Jeanine	J.A.	J.B.
Elizabeth	Victoria	Alexander, Jeremiah	L.S.	J.B.
Louisa	Delaney	Jillian, Cynthia	A.R.	J.B.
Beatrice	Monica	Deonte, Sophia	M.H.	J.G.
Kevin	Mia	Roseanne, Jeanine	J.A.	J.G.
Morgan	Adam	Nicole, Nadine	L.S.	J.G.
Vincent	Riley	Denise, Maureen	A.R.	J.G.

APPENDIX (Continued)

	Foils		Speaker	
Target	Stress-Matched	Nonmatched	Target	Background
Single Voice, 5-Month-Olds (Continued)				
Jacob	Chesney	Eileen, Simone	M.H.	R.D.
Ty	Em	David, AJ	J.A.	R.D.
Samantha	Melissa	Killian, Christopher	L.S.	R.D.
Zephra	Kayleigh	Shereen, Marvelle	A.R.	R.D.
Toosa	Jordan	Norit, Tamar	M.H.	S.C.
Cole	Kate	Shakti, Jarred	J.A.	S.C.
Greer	Max	Phoebe, Alex	L.S.	S.C.
Kian	Russell	Malone, Tashae	A.R.	S.C.
Annabelle	Tabitha	Mackenna, Hadassah	M.H.	S.T.
Lillian	Gabriel	Amanda, Mackenzie	J.A.	S.T.
Melissa	Samantha	Christopher, Killian	L.S.	S.T.
William	Eden	Shereen, Marvelle	A.R.	S.T.
Mackenna	Hadassah	Tabitha, Annabelle	M.H.	T.W.
Peter	Mehdi	Jeanine, Keon	J.A.	T.W.
Killian	Christopher	Samantha, Melissa	L.S.	T.W.
Josh	Will	Joey, Michael	A.R.	T.W.
Nine Voices, 5-Month-Olds				
Devin	Matthew	Norit, Tamar	M.H.	9 voices
Jack	Reese	Matthew, Sarah	M.H.	9 voices
Brookie	Ella	Tamar, Norit	M.H.	9 voices
Reese	Jack	Devon, Emma	M.H.	9 voices
Boobuh	Luka	Simone, Eileen	M.H.	9 voices
Devon	Emma	Norit, Tamar	M.H.	9 voices
Adebola	Valentino	Veronica, Elizabeth	M.H.	9 voices
Matteo	Ann Rosa	Isabel, Juan David	M.H.	9 voices
Luka	Shaney	Simone, Eileen	M.H.	9 voices
Brady	Isaac	Earline, Babette	J.A.	9 voices
Isaac	Brady	Earline, Babette	J.A.	9 voices
Gabriel	Anthony	Ricardo, Mackenzie	J.A.	9 voices
Robert	Gabby	Purnell, Khalil	J.A.	9 voices
Oliver	Monica	Mackenzie, Felicia	J.A.	9 voices
Kieran	Maya	Khalil, Purnell	J.A.	9 voices
Luke	Paul	Nathan, Mehdi	J.A.	9 voices
Peter	Nathan	Jeanine, Roseanne	J.A.	9 voices
Kenzie	Aggie	Earline, Babette	J.A.	9 voices
Logan	Maddie	Nadine, Nicole	L.S.	9 voices
Melody	Julian	Daniela, Alonzo	L.S.	9 voices
Joshua	Nicholas	Maryann, Gabrielle	L.S.	9 voices
Nicholas	Joshua	Gabrielle, Maryann	L.S.	9 voices
Julian	Melody	Daniela, Alonzo	L.S.	9 voices
Joshua	Christopher	Melissa, Samantha	L.S.	9 voices
Max	Blake	Nora, Ryan	L.S.	9 voices
Drew	Blake	Nora, Ryan	L.S.	9 voices
Alex	Phoebe	Leanne, Annette	L.S.	9 voices
Riley	Aidan	Denise, Maureen	A.R.	9 voices
Beck	Miles	Cody, Lena	A.R.	9 voices
Lena	Thomas	Michelle, Doreen	A.R.	9 voices
Delaney	Melissa	Cynthia, Jillian	A.R.	9 voices
Taryn	Riley	Denise, Maureen	A.R.	9 voices
Maxim	Anna	Elaine, Roseanne	A.R.	9 voices
Eden	Kayleigh	Shereen, Marvelle	A.R.	9 voices
Ezra	Thomas	Doreen, Michelle	A.R.	9 voices
Anna	Cody	Doreen, Michelle	A.R.	9 voices
Single Voice, 8.5-Month-Olds				
Anjali	Broderick	Kameko, Zipora	H.C.	A.T.
Allison	Zachary	Sabrina, Sophia	L.S.	A.T.
Kutti	Prima	Suzanne, Leanne	S.R.	A.T.
Mason	Cora	Renee, Denise	H.C.	B.W
Alex	Phoebe	Leanne, Annette	L.S.	B.W.

APPENDIX (Continued)

	Foils		Speaker	
Target	Stress-Matched	Nonmatched	Target	Background
Single Voice, 8.5-Month-Olds (Continued)				
Dalton	Simon	Jamal, Roseanne	S.R.	B.W.
Caroline	Jillian	Fiona, Allanah	H.C.	I.H.
Maddie	Andrew	Danielle, Nicole	L.S.	I.H.
Elliot	Caroline	Diana, Miranda	S.R.	I.H.
Fiona	Allanah	Caroline, Jillian	H.C.	J.B.
Jackson	Patrick	Elaine, Marie	L.S.	J.B.
Simon	Colin	Yvette, Jamal	S.R.	J.B.
Olympia	Penelope	Alexander, Jeremiah	H.C.	J.G.
Andrew	Maddie	Danielle, Nicole	L.S.	J.G.
Connor	Braden	Leanne, Suzanne	S.R.	J.G.
Grant	Miles	Tayo, Cora	H.C.	R.D.
Aiden	Hunter	Philippe, Renee	L.S.	R.D.
T.J.	Dalton	J.R., Renee	S.R.	R.D.
Kahlan	Jesse	Denise, Renee	H.C.	S.C.
Max	Blake	Jamie, Patrick	L.S.	S.C.
Clara	Kirsten	Marie, Yvette	S.R.	S.C.
Jesse	Duncan	Denise, Renee	H.C.	S.T.
Blake	Sam	Patrick, Jamie	L.S.	S.T.
Jackie	Sydney	Elaine, Michelle	S.R.	S.T.
Ibrahim	Bathsheba	Phineas, Eferhard	H.C.	T.W.
Alex	Phoebe	Annette, Leanne	L.S.	T.W.
Dante	Simon	Jamal, Roseanne	S.R.	T.W.
Nine Voices, 8.5-Month-Olds				
Jade	Claire	Mason, Cora	H.C.	9 voices
Zachariah	Marianna	Olympia, Penelope	H.C.	9 voices
Claire	Greg	Cora, Tayo	H.C.	9 voices
Mia	Lily	Shereen, Suzanne	H.C.	9 voices
Micah	Grayson	Renee, Denise	H.C.	9 voices
Keira	Mason	Renee, Denise	H.C.	9 voices
Bug	Greg	Mason, Kira	H.C.	9 voices
Kira	Mason	Denise, Renee	H.C.	9 voices
Micah	Grayson	Renee, Denise	H.C.	9 voices
Nicholas	Christopher	Maryanne, Gabrielle	L.S.	9 voices
Aiden	Hunter	Philippe, Renee	L.S.	9 voices
Abby	Gavin	Renee, Philippe	L.S.	9 voices
Patrick	Jamie	Marade, Elaine	L.S.	9 voices
Henry	Spencer	Nicole, Jolene	L.S.	9 voices
Ryan	Nora	Marie, Elaine	L.S.	9 voices
Isabella	Alexandria	Angelica, Penelope	L.S.	9 voices
James	Blake	Patrick, Ryan	L.S.	9 voices
Alex	Phoebe	Annette, Leanne	L.S.	9 voices
Bridget	Sena	Suzanne, Leanne	S.R.	9 voices
Mason	Larkin	Michelle, Elaine	S.R.	9 voices
Larkin	Mason	Elaine, Michelle	S.R.	9 voices
Sena	Bridget	Leanne, Suzanne	S.R.	9 voices
Rosie	Katie	Michelle, Elaine	S.R.	9 voices
Zach	Fred	Simon, Dalton	S.R.	9 voices
Piglet	Dalton	Rene, J.R.	S.R.	9 voices
Fe-fe	Jo-jo	Rene, J.R.	S.R.	9 voices
Chase	Zach	Dalton, Simon	S.R.	9 voices
Reversed Single Voice, 5-Month-Olds				
Sarah	Devon	Norit, Tamar	M.H.	A.T. reversed
Ethan	Cassie	Babette, Adelle	J.A.	A.T. reversed
Sweetheart	Puddin'	Elaine, Marade	L.S.	A.T. reversed
Annabelle	Tabitha	Meira, Hadassah	M.H.	B.W. reversed
Josh	Paul	Ethan, Kevin	J.A.	B.W. reversed
Max	James	Nora, Ryan	L.S.	B.W. reversed
Will	Grace	Maya, Jacob	M.H.	I.H. reversed
Katerina	Elliana	Ezekial, Penelope	J.A.	I.H. reversed

APPENDIX (Continued)

	Foils		Speaker	
Target	Stress-Matched	Nonmatched	Target	Background
Reversed Single Voice, 5-Month-Olds (Continued)				
Max	Sam	Jamie, Patrick	L.S.	I.H. reversed
Ian	Darcy	Eileen, Simone	M.H.	J.B. reversed
Matthew	David	Denise, Jeanine	J.A.	J.B. reversed
Max	Sam	Ryan, Nora	L.S.	J.B. reversed
Hannah	Chesney	Simone, Eileen	M.H.	J.G. reversed
Gabby	Robert	Purnell, Khalil	J.A.	J.G. reversed
Sam	Max	Nora, Ryan	L.S.	J.G. reversed
Lydia	Cameron	La'Asia, Ayonte	M.H.	R.D. reversed
Verlyn	Milan	Charlie, Phillip	J.A.	R.D. reversed
Aidan	Hunter	René, Phillippe	L.S.	R.D. reversed
Will	Grace	Jacob, Maya	M.H.	S.C. reversed
Nicholas	Jonathan	Allana, Melissa	J.A.	S.C. reversed
Phoebe	Alex	Leanne, Akhil	L.S.	S.C. reversed
Matthew	Sarah	Tamar, Norit	M.H.	S.T. reversed
Frankie	Jarred	Earline, Babette	J.A.	S.T. reversed
Gabrielle	Maryann	Chrisopher, Nicholas	L.S.	S.T. reversed
Sarah	Matthew	Tamar, Norit	M.H.	T.W. reversed
Charlie	Phillip	Verlyn, Milan	J.A.	T.W. reversed
Alex	Phoebe	Annette, Leanne	L.S.	T.W. reversed

Note—Target names in bold also occurred as a foil for another child in the same condition.