

The cocktail party effect in infants

ROCHELLE S. NEWMAN and PETER W. JUSCZYK
State University of New York, Buffalo, New York

Most speech research with infants occurs in quiet laboratory rooms with no outside distractions. However, in the real world, speech directed to infants often occurs in the presence of other competing acoustic signals. To learn language, infants need to attend to their caregiver's speech even under less than ideal listening conditions. We examined 7.5-month-old infants' abilities to selectively attend to a female talker's voice when a male voice was talking simultaneously. In three experiments, infants heard a target voice repeating isolated words while a distractor voice spoke fluently at one of three different intensities. Subsequently, infants heard passages produced by the target voice containing either the familiar words or novel words. Infants listened longer to the familiar words when the target voice was 10 dB or 5 dB more intense than the distractor, but not when the two voices were equally intense. In a fourth experiment, the assignment of words and passages to the familiarization and testing phases was reversed so that the passages and distractors were presented simultaneously during familiarization, and the infants were tested on the familiar and unfamiliar isolated words. During familiarization, the passages were 10 dB more intense than the distractors. The results suggest that this may be at the limits of what infants at this age can do in separating two different streams of speech. In conclusion, infants have some capacity to extract information from speech even in the face of a competing acoustic voice.

As adults, we are constantly faced with the problem of listening to one speaker amidst various other noises. Often, we have to deal with competition from other speakers as well. This is classically known as the *cocktail party effect*—an allusion to our ability to follow a conversation even in the midst of a party environment, when many people are speaking simultaneously in a relatively small area of space. This type of selective attention is related to a more general phenomenon known as *streaming*, which refers to our ability to group together sounds originating from one source and separate them from sounds originating from other sources.

The cocktail party effect and streaming have been the subject of extensive research since the 1950s. Listeners are known to separate different sound streams on the basis of a variety of acoustic cues, such as 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 & Nootboom, 1982; Treisman, 1960), onset and offset times (Bregman &

Pinker, 1978; Dannenbring & Bregman, 1978), and differences in amplitude modulation (Bregman, Abramson, Doehring, & Darwin, 1985).

For all the work on streaming with adults, there is a surprising dearth of research with infants. Like adults, infants often are engaged in communication in noisy situations. Their caregiver may speak to them at the same time that a car drives down the street outside or the phone rings in the next room. Infants also have to deal with competition from other speech, such as that from the television down the hall and from their siblings in the next room. Yet there has been almost no research on infants demonstrating the extent to which they can succeed at this difficult task. Of course, infants are known to display a range of other abilities that are useful for compensating for variability in the speech signal. For example, like adults, their discrimination of many speech contrasts is categorical (e.g., Eimas, 1974, 1975; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). In addition, infants 6 months of age (Kuhl, 1979, 1983) and younger (Jusczyk, Pisoni, & Mullennix, 1992) show some ability to adjust to differences in speaking voices. Similarly, there is evidence that 2- to 3-month-olds compensate for changes in speaking rate (Eimas & Miller, 1980; Miller & Eimas, 1983). Given the evidence of these speech perception capacities, it is interesting to determine the extent to which infants are also able to attend to the speech of one talker when there is competition from the speech of other talkers.

Infants have a higher auditory threshold than adults, for both pure tones (Nozza & Wilson, 1984; Ruben, 1992; Sinnott, Pisoni, & Aslin, 1983) and speech (Nozza, Rossman, & Bond, 1991; Nozza, Wagner, & Crandell, 1988), often demonstrating a 15- to 25-dB difference in threshold. Some research has suggested that 6- to 8-month-old

The research reported here was supported by Research Grant HD-15795 from NICHD to the second author and by an NSF Predoctoral Fellowship to the first author. Portions of this research were reported in a poster presented at the 129th meeting of the Acoustical Society of America in Washington, D.C., May 1995. The authors wish to thank Ann Marie Jusczyk, Nancy Redanz, Aileen Worden, and Amy Gambon-Dennis for their help in running and scheduling subjects, and Denise Mandel and Steven Elsis for their assistance in stimulus creation. P.W.J. is currently at Johns Hopkins University. Correspondence should be addressed to R. Newman, Department of Psychology, Park Hall, SUNY at Buffalo, Buffalo, NY 14260 (e-mail: rochelle%art@ubvms.cc.buffalo.edu).

infants also need a higher signal-to-noise (S/N) ratio than do adults in order to detect speech (Nozza et al., 1988; Trehub, Bull, & Schneider, 1981) and make phonetic distinctions (Nozza, Miller, Rossman, & Bond, 1991; Nozza, Rossman, Bond, & Miller, 1990). Nozza et al. (1990) tested 7- to 11-month-old infants on a /ba-/ga/ distinction and found that although they could perform above chance with S/N ratios as low as 0 dB, they needed ratios as high as 8 dB in order to reach 85% correct. Nozza et al. note that this discrimination task is fairly simple, and they suggest that good performance on fluent speech recognition may require even larger S/N ratios. There is some reason to believe that the nature of the noise masker could influence infants' abilities to discriminate speech sounds. For example, Nozza et al. (1990; Nozza et al., 1988) and Trehub et al. (1981) used random or broadband noise as a masker. Some research with adults (Young, Parker, & Carhart, 1975) has suggested that speech is a more efficient masker than noise (although other research has suggested that fluctuating signals, such as speech, are less efficient maskers than steady noise; see Festen & Plomp, 1990).

Another indication that segregating competing speech signals may be an especially difficult task comes from research with elderly adults (Bergman, 1971). Even elderly listeners who have normal hearing tend to have difficulty with the cocktail party effect. That is, speech segregation can be disrupted even by hearing difficulties too small to be detected by pure-tone hearing tests. Speech segregation seems to be very easily disrupted by hearing loss or reductions in the S/N ratio. Consequently, infants, whose detection of pure tones is worse than that of adults, might be especially handicapped by the presence of additional background voices.

All of this suggests that infants could have difficulty attending to one voice amidst other competing voices, unless the S/N ratio is very substantial. If so, this would place strong limits on the kinds of settings in which infants could be expected to acquire useful information about their native language. Put simply, in order to learn a language, infants must be able to hear it, and this requires the ability to separate it from background noise. Although there are certainly many instances when no other voices compete with that of the caregiver, there are also many instances when there is competition. If infants cannot successfully separate different voices, then none of these occasions could help them learn language. If infants *can* separate out different voices, and choose to attend to the one directed to them, then this would greatly increase their opportunities to listen to their native language and thus hasten the language-learning process. In this respect, information about infants' capacities to attend to speech in noisy settings is important for determining the impact of speech input on the course of language acquisition.

There have been a few studies that examined infant streaming abilities with simple musical stimuli, rather than speech. McAdams, Bertoincini, and Bobrow (1990) tested 3- to 4-day-old infants on a musical streaming task. Infants were habituated to a 4-tone sequence of ascending

frequency repeated multiple times. The 1st and 3rd tones were played on one instrument, and the 2nd and 4th were played on another. Adults tend to separate these sequences by timbre, hearing two streams, each of 2 alternating notes (e.g., 1 3 1 3 1 3 and 2 4 2 4 2 4 rather than 1 2 3 4 1 2 3 4 1 2 3 4). Because of this streaming, they no longer are aware of the ascending nature of the overall sequence. If infants segregate the sequences by timbre, one would expect that they would similarly be unaware of the ascending nature and not notice a switch to a descending pattern (4 3 2 1) after the habituation point. Unfortunately, the results were somewhat equivocal. The newborns did display some release of habituation in the correct direction, but it was not significant. However, the change was as large as that of another series in which there was a significant release of habituation. This result hints that newborns might not separate streams as strongly as adults do, but it is certainly not conclusive.

Demany (1982) performed a similar study using 7- to 15-week-old infants, although his streams differed in frequency range rather than timbre. He found that infants of this age did segregate streams much in the same way as adults do, suggesting that streaming abilities, if not present at birth, develop by 3 months of age.

Both studies just described used rather simple musical stimuli. More importantly, although the different streams overlapped in time, the actual notes did not. There was never more than one signal being presented at any given instant, so there was no possibility of masking. (This situation is known as *sequential streaming*, as opposed to the simultaneous segregation typical of the cocktail party effect; see Bregman, 1990, pp. 30–31, for a discussion of this point.) Therefore, it is not clear how these results with musical stimuli would transfer to streaming speech. In the typical cocktail party effect, information from two different speech sources arrives at the ear at exactly the same instant in time. Furthermore, each speech signal consists of information at a variety of different frequencies, which must all be grouped together into a coherent whole.¹ In the melodic sequences used by McAdams et al. and by Demany, there was only one tone being played at any given time. The segregation and attentional requirements, then, would seem to be far simpler for the musical stimuli than for speech segregation.

Given the importance of the ability to segregate speech to adults, it seems somewhat surprising that so little research on infants' abilities has been done. We decided to address this lacuna in the literature by examining infants' ability to separate two streams of speech. The first step was to devise an appropriate measure of infants' abilities to segregate speech. One indication that infants attend to one stream of speech as opposed to another would be if they showed some subsequent recognition for information that occurred in that particular stream. This would be an indication that the infants processed and remembered at least some of the information in that stream of speech. A procedure used in a recent study by Juszczyk and Aslin (1995) appears promising in this respect. Their objective was to examine infants' capacity to segment

words from fluent speech. Using the headturn preference paradigm (Kemler Nelson et al., 1995), they presented infants with a particular word that was repeated several times in isolation and then tested the infants' preferences (assessed by listening time) for sentences containing the word versus sentences without the word. They assumed that if the infants listened longer to the sentences containing the word, they must have noticed some similarity between the word in the sentences and the word in isolation. Their results suggested that infants do recognize words in the different contexts at 7.5 months of age, although they do not show any such recognition at 6 months of age.

In the present study, we used the same methodology as that used by Jusczyk and Aslin (1995), but we added a second, distracting voice presented simultaneously with the target voice repeating the isolated words. If infants can still notice the similarity between the words in isolation and those in fluent speech, they must be able to separate the speech of this target speaker from that of the distracting voice. In some ways, this experiment is an extreme test. We are not only asking infants to separate the two streams of speech but also to selectively attend to one of the two streams, detect the words, and remember them for later recognition. Because of this memory requirement, this task may be harder than more typical tests of streaming, such as those discussed earlier. That is, if infants are unable to perform this task well, it may be that the combined requirements of segregation, storage and retrieval are too difficult rather than that segregation by itself is too hard a task.

In addition to the memory requirement, another factor could potentially make our task more difficult for the infants. Because we did not know whether sounds emanating from two different locations would have an impact on infants' behavior in the headturn preference paradigm, we elected to use only a single sound location to present both the target and the distractor voices. Thus, there was no spatial separation between the voices. In the real world, when infants hear more than one person speaking simultaneously, these voices generally originate from different locations in space. Studies with adults indicate that spatial cues do help in separating different streams of speech (Broadbent, 1954; Cherry, 1953; Hirsh, 1950; Pollack & Pickett, 1958; Poulton, 1953; Spieth et al., 1954). While less is known about infants' abilities, even very young infants will turn toward a source of a sound (Clifton, Morrongiello, Kulig, & Dowd, 1981; Muir & Clifton, 1985; Muir & Field, 1979), suggesting some ability to localize. Hence, it is possible that information about the spatial location of two competing voices is helpful to them in separating speech streams in their normal environments.

To maximize the distinctiveness of the target and distractor materials for the infants, a relatively high-pitched female voice was used to record the target materials in a child-directed speech register, whereas a male voice was used to record the distractor passages in an adult-directed speech register. The difference in voice pitch (primarily a difference in F_0) between the talkers should be a strong cue for voice separation. It has already been demonstrated

to be an important cue for adult listeners (Assmann & Summerfield, 1990; Broadbent, 1952; Brox & Nootboom, 1982; Chalikia & Bregman, 1989; Darwin, 1981; Summerfield & Assmann, 1991; Treisman, 1960). Similarly, the difference between the child-directed speech of the target voice (with its greater changes in amplitude and frequency) and the adult-directed, more monotonic, distractor voice should facilitate stream segregation, as well as the infants' listening preferences. Moreover, the fact that one stream consisted of fluent speech and the other of a list of words should also have enhanced their separability, since there are clear differences in onsets and offsets between the streams (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978).

EXPERIMENT 1

As a starting point for our investigation, we decided to examine whether infants would be able to attend to a target voice that was 10 dB more intense than a competing background voice. We also selected our stimulus materials from those used by Jusczyk and Aslin (1995). Their study indicated that 7.5-month-olds who were familiarized with isolated words were able to recognize these words in fluent speech passages. Hence, these results provide a baseline measure of what infants at this age are capable of when listening to only one voice at a time. If 7.5-month-olds are able to separate the two streams of speech, then they too should listen longer to fluent speech passages that contain the words they heard during the familiarization period. The decision to test 7.5-month-olds was also motivated by previous investigations that have indicated that sensitivity to many aspects of native language sound structure develops between 6 and 9 months of age. For example, although 6-month-olds have relatively little difficulty discriminating speech contrasts in any language, sensitivity to certain non-native speech contrasts begins to decline toward the end of this period (see Best, 1995, and Werker & Desjardins, 1995, for recent reviews). During the same period, infants appear to develop sensitivity to characteristics of native language sound organization, such as typical word stress patterns (Jusczyk, Cutler, & Redanz, 1993) and phonotactic structure (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994). Hence, infants at this age might be particularly attentive to sound patterns in speech that is directed toward them.

Method

Subjects. Twenty-four American infants (14 males, 10 females) from monolingual homes participated in this experiment. The infants had an average age of 33 weeks 2 days (range = 30 weeks 1 day to 34 weeks 6 days), or approximately 7.5 months. The data from an additional 7 infants were not included for the following reasons: 3 for crying, and 4 for failing to listen for an average of 3 sec to each passage (long enough to have heard at least one of the six sentences in the passage) or failure to complete the experiment.

Stimuli. Our target items were identical to those used in Jusczyk and Aslin (1995). A female native speaker of American English recorded 15 repetitions in a row of each of the four target words (*cup*,

Table 1
The Four Six-Sentence Passages Used in These Experiments

Cup
The cup was bright and shiny. A clown drank from the red cup. The other one picked up the big cup. His cup was filled with milk. Meg put her cup back on the table. Some milk from your cup spilled on the rug.
Dog
The dog ran around the yard. The mailman called to the big dog. He patted his dog on the head. The happy red dog was very friendly. Her dog barked only at squirrels. The neighborhood kids played with your dog.
Feet
The feet were all different sizes. This girl has very big feet. Even the toes on her feet are large. The shoes gave the man red feet. His feet get sore from standing all day. The doctor wants your feet to be clean.
Bike
His bike had big black wheels. The girl rode her big bike. Her bike could go very fast. The bell on the bike was really loud. The boy had a new red bike. Your bike always stays in the garage.

dog, feet, bike) and recorded four different six-sentence test passages (see Table 1). The talker was asked to speak in a lively, animated voice, as if speaking to a small child. The target words appeared in a variety of sentential locations in the six sentences, so as to prevent their being spoken with similar intonation contours in the different sentences. Furthermore, the target words were generally not the most emphasized word in any given sentence (see Jusczyk & Aslin, 1995, for more details on this point).

The recordings were made in a sound-attenuated room, using a Shure microphone. They were amplified, low-pass filtered at 4.8 kHz, digitized via a 12-bit analog-to-digital converter at a 10-kHz sampling rate, and stored on a VAX Station model 3176 computer. The average durations were 19.72 and 26.53 sec, for the target-passages and the target-word lists, respectively. For additional details of the target items, the reader is referred to Jusczyk and Aslin (1995).

A male native speaker of American English recorded the distractor passages. These passages consisted of text from the Methods section of Fernald's (1985) paper on child language. These were recorded in the same manner as the target passages, except that the male speaker was not instructed to speak in a lively, animated manner.

To adjust the S/N ratio in the familiarization phase, we needed measurements of the average intensity of both the isolated target words and the fluent distractor passage. Since the target words were separated by periods of silence, the average intensity level (r.m.s. 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). The distractor sentences had far fewer pauses, making the intensity level of the target seem artificially low in comparison. To adjust for this, we created an edited version of the word list. In this version, the pauses between words had been spliced out. A waveform program on the computer then calculated the average intensity level of this

edited version. The average intensity level of the distractor stimulus was adjusted to be 10 dB less than that of this edited version of the word list. This distractor passage and the original (unedited) word list were then combined into one file. The (combined) stimuli were presented to the infant at an average intensity level of 74 dB(C) (± 2 dB).

Although the distractor passage did not have many pauses in it, it did have a few that were generally located between sentences. We were concerned that a repetition of the target word might be presented during one of these pauses, making it effectively unmasked. To avoid this possibility, we carefully edited the distractor passage by moving sentences around and shortening pauses.

Design. Half of the infants heard the words *cup* and *dog* during the familiarization phase, and the other half heard the words *feet* and *bike*. During the test phase, all of the infants heard four blocks of the same four test passages. Each block contained a different random ordering of the test passages containing the words *cup*, *dog*, *feet*, and *bike*.

Apparatus. A PDP-11/73 controlled the presentation of the stimuli and recorded the observer's coding of the infant's responses. A 12-bit D/A converter was used to recreate the audio signal. The resulting audio signal was then filtered at 4.8 kHz, amplified, and played from two loudspeakers mounted on the side walls of the testing booth.

The experiment took place in a three-sided test booth constructed out of pegboard panels (4 × 6 ft). An experimenter, located behind the front wall of the booth, looked through the existing holes in the pegboard to observe the infant. The remainder of the pegboard panels were backed with white cardboard so that the infant could not see movements behind the panels. There was a green light in the center of the front panel and a 5-cm hole for the lens of a video camera. The video camera was used to provide a permanent record of each session. Each of the two side boards had a red light and a loudspeaker located in the center of the panel. A white curtain was suspended from the ceiling and prevented the infant from seeing over the top of the booth. A computer terminal and six-button response box were located behind the front wall of the booth. The experimenter pressed buttons to signal the computer to start and stop the flashing center and side lights.

Procedure. Given the potential difficulty posed by the absence of spatial location differences for the target and distractor voices, we tried to ensure that there were other kinds of cues that might facilitate the task of attending to the target voice. We thought that the infants would prefer listening to the higher pitched, animated female voice of the target. However, to further strengthen this preference, we presented the infants with a 2.5-min videotape prior to the experiment. This tape was of the target speaker (accompanied by two puppets) telling an interesting story; we hoped this positive experience would further prime the infants to listen more closely to the target voice in the actual experiment.² The infants were held on the lap of the caregiver, approximately 3 ft from a video monitor. The woman on the videotape was recorded telling a story to the infants, with the assistance of two hand-held puppets. The room had dim lighting and few other distractions.

Following the videotape presentation, the infant and caregiver were shown to the room with the test booth. The infants were again held on the lap of the caregiver, who was seated in a chair in the center of the booth. During the familiarization phase, the infants heard two of the target words (and distractor passages) on alternating trials until they accumulated at least 30 sec of listening time to each word. Listening time was assessed by the amount of time the infant spent looking at the "source" of the sound (the flashing light). The loudspeaker used for each word was randomly varied from trial to trial. A trial was terminated when the infant looked away for more than 2 sec or when the end of the word list was reached. Information about the direction and duration of headturns and the total trial duration were stored in a data file on the computer.

The test phase began immediately after the familiarization criterion was reached. The test trials were blocked in groups of four so that each test passage occurred once in a given block, although the order of the four passages within each block was randomized. Each infant was tested on four blocks, for a total of 16 test trials.

Both familiarization and test trials began with the blinking of the green light in the center of the front panel. 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 (regardless of whether it was a familiarization stimulus or a test stimulus) 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 sec or less) was not included when measuring the total listening time. The red light continued to flash for the duration of the entire trial.

The experimenter behind the center panel pressed a button on the response box whenever the infant looked at or away from the flashing light. The experimenter was not told ahead of time which items were used in the familiarization phase, and both the experimenter and the caregiver listened to masking music over headphones (Sony MDR-V600). Reliability checks between the experimenter and observers of the video tapes of each session are high, with correlations ranging from .92 to .96 (see Kemler Nelson et al., 1995); however, it is important to note that these estimates of reliability refer to the procedure in general, not to these specific experiments.

Results and Discussion

Mean listening times to the four different test passages were calculated for each infant across the four blocks of trials; these times were then averaged for the test passages containing the familiar words and those containing the unfamiliar words. Twenty-one infants listened longer to the passages containing familiar words. The average listening times for these passages across all infants was 7.71 sec, and the average for passages containing unfam-

iliar words was 6.21 sec (see Figure 1). This difference was significant by a paired *t* test [$t(23) = 5.27, p < .0001$].

These results suggest that infants are capable of separating different streams of speech and are capable of listening selectively. However, in the present experiment, the intensity ratio between the two streams of speech was relatively high. For this reason, we decided to explore whether infants are also capable of separating competing streams of speech at lower S/N ratios.

EXPERIMENT 2

Experiment 2 was identical to Experiment 1 in all respects except for the intensity ratio between the two streams of speech. In Experiment 1, we had used a 10-dB S/N ratio; in Experiment 2, the target voice was only 5 dB more intense than the distractor voice. If infants can still recognize words presented for familiarization with this S/N ratio, it would suggest that infants' abilities to segregate speech sources are fairly robust.

Method

Subjects. Twenty-four American infants (17 males, 7 females) from monolingual homes participated in this experiment. The average age of these infants was 31 weeks 2 days (range = 30 weeks 3 days to 35 weeks 1 day), or approximately 7.5 months. Thirteen additional infants were tested but were not included for the following reasons: 5 for crying, 1 for restlessness, 3 for failing to look for an average of 3 sec to each passage, 2 for equipment error, and 2 for experimenter error.

Stimuli. The original recordings used to prepare the stimuli in Experiment 1 were used to prepare the stimuli for Experiment 2. However, when the distractor passages were combined with the repetitions of the isolated words into a single stimulus file, their relative intensity levels were adjusted differently. In Experiment 1, the average intensity level of the male (distractor) voice was adjusted to be 10 dB less than that of the isolated words spoken by the female (target) voice; however, in Experiment 2, it was adjusted to be only 5 dB less intense.

Design, Apparatus, and Procedure. The design, apparatus, and procedure 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. The average listening times were then found for the test passages containing familiar words and for the test passages containing unfamiliar words. Eighteen infants listened longer to the passages containing familiar words. The average listening times for these passages across all infants was 8.01 sec, and the average for passages containing unfamiliar words was 6.90 sec (see Figure 2). Again, this difference was significant by a paired *t* test [$t(23) = 3.092, p = .0051$].

Despite the lower S/N ratio, the infants still seemed able to separate the two streams of speech. Yet the question still remains as to the minimal S/N ratio that infants require to display evidence of having segregated the target voice from the competing voice. With this in mind, we decided to lower the S/N ratio again.

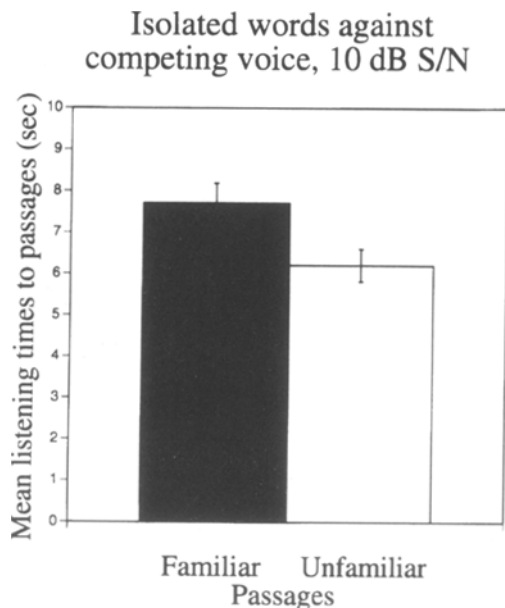


Figure 1. Mean listening times and standard errors to the familiar and unfamiliar passages in Experiment 1.

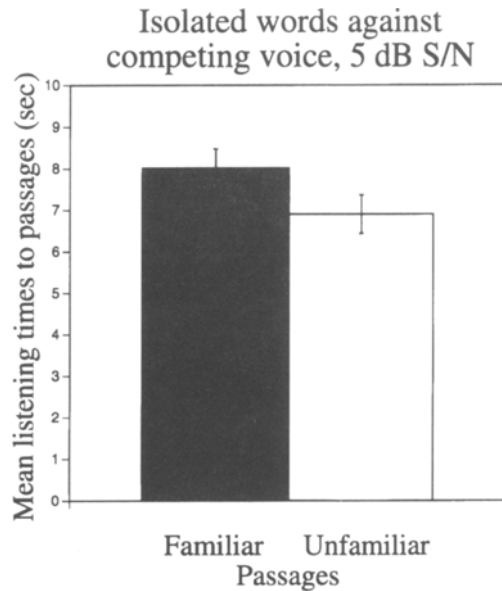


Figure 2. Mean listening times and standard errors to the familiar and unfamiliar passages in Experiment 2.

EXPERIMENT 3

Experiment 3 was identical to Experiments 1 and 2 except for the intensity ratio between the two streams of speech. In Experiment 1, we used a 10-dB S/N ratio; in Experiment 2, we used a 5-dB S/N ratio. In Experiment 3, we lowered the ratio to 0 dB. That is, in this experiment, the target voice and the distractor voice were at exactly the same intensity level.

Method

Subjects. Twenty-four American infants (16 males, 8 females) from monolingual English-speaking homes participated in this experiment. The average age of these infants was 33 weeks 1 day (range = 31 weeks 5 days to 35 weeks 4 days), or approximately 7.5 months. Eleven additional infants were tested but were not included for the following reasons: 3 for crying, 7 for failing to look for an average of 3 sec to each passage, and 1 for equipment failure.

Stimuli. The same recordings used in Experiments 1 and 2 were used to prepare the combined stimulus files for Experiment 3. However, rather than make the distractor voice less intense than the target voice, the male voice was adjusted to be of the same average intensity as the female voice.

Design, Apparatus, and Procedure. The design, apparatus, and procedure were identical to those in Experiments 1 and 2.

Results and Discussion

The data were analyzed in the same manner as in Experiments 1 and 2. Mean listening time to each of the four test passages was calculated for each infant. The average listening times were then found for the test passages containing familiar words and for the test passages containing unfamiliar words. Only 10 of the 24 infants listened longer to the passages containing familiar words. The average listening times for these passages across all infants was 6.74 sec, and the average for passages containing unfa-

miliar words was 6.65 sec (see Figure 3). This difference was not significant by a paired *t* test [$t(23) = 0.23, p > .05$].

In contrast to the results of Experiments 1 and 2, there was no evidence that the infants recognized the similarity between the words in the test passages and the words that had been presented in isolation during the familiarization phase. Figure 4 shows the differences in mean listening times from the individual subjects in Experiments 1–3 (each subject's data are represented by a separate point).³ This graph makes it clear that with the change in S/N ratio, there was a change in the infants' differential listening times. With lower ratios, there were more infants with negative differentials, suggesting that more and more infants were failing to notice the similarity between the masked words and the later words in the passages. Furthermore, the infants' variability increased with the lower ratios, as shown by the larger spread of data points.

To obtain an index of how difficult it was to detect the words in the target voice in the presence of the distractor at a 0-dB S/N ratio in our test setting, we tested a group of 6 adults on a version of the same task. The adults were tested individually. Each subject was seated in the test room in the chair normally occupied by the caregiver during the infant testing sessions. Subjects were instructed that they would be listening to materials spoken simultaneously by a male and a female voice. They were told to listen carefully to the materials but were *not* explicitly instructed to attend to the female voice. Subjects then heard two familiarization trials: each consisted of a repeated isolated target word in combination with the simultaneous distractor passage. (In contrast to the infant testing situation, the combined target words and distractor passages were played only once.)

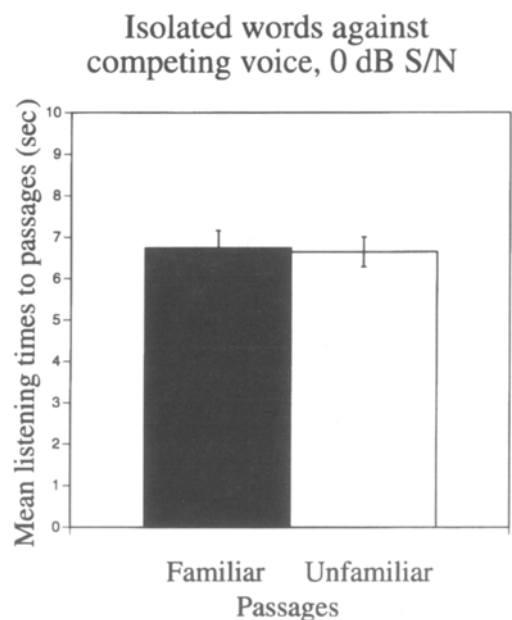


Figure 3. Mean listening times and standard errors to the familiar and unfamiliar passages in Experiment 3.

Following these two familiarization trials, the subjects were told that they would hear four test passages by the female talker and that two of these included words that she had spoken during the familiarization trials. Four test trials (two with the familiar words and two with the unfamiliar words) occurred. After each trial, subjects were asked whether the passage contained a familiar word or not. Overall performance for these 6 adults was 67%, significantly above chance level responding (i.e., 50%), yet far from perfect. In fact, only 1 subject produced four correct responses.⁴ Hence, although adults extracted and remembered sufficient information about the target words to perform at better than chance levels during the test phase, the task appears to have been relatively difficult for them.

Although it is tempting to suggest that the low S/N ratio made it impossible for the infants to selectively attend to one of the two voices (i.e., that we had reached some sensory threshold), alternative explanations cannot be completely ruled out. For example, one possibility is that there is a cognitive explanation for the infants' performance—one based on the infants' cognitive limitations or strategies. The task is quite demanding, and the infants may simply have been unwilling to devote the cognitive resources necessary to perform well. Alternatively, infants may normally choose to listen to whatever stream of speech is loudest in their environment; in Experiments 1 and 2, this would have been the target voice. Here, there would be no reason for the infants to pick the female voice consistently over the male voice. We find this explanation less likely, because there were other biasing factors that should have led the infants to attend to the target voice (familiarity with the voice from the video tape, the greater pitch variation of child-directed speech, etc.). Nevertheless, we cannot tell from these results whether the infants were incapable of recognizing words presented simultaneously with an equal-intensity distractor (a sensory failure) or simply did not choose to do so (a cognitive one).

Whether infants cannot selectively attend to a voice in this situation or simply find the task too demanding, one might assume that this would hold true outside of the laboratory as well. That is, when a caregiver speaks to an infant in a noisy situation, the infant will likely find paying attention to the caregiver easy if his or her voice is substantially more intense than that of the distracting noise (corresponding to Experiments 1 and 2), but the infant may find the task too difficult if the distraction is equally intense as the caregiver's voice. On the other hand, this real-world situation would provide visual, as well as auditory, cues. Seeing the caregiver's mouth moving may encourage the infant to pay attention even when the task is demanding. There is also evidence that 4- to 5-month-olds do have some capacity to match speech sounds to the correct visual displays of faces articulating these sounds (Kuhl & Meltzoff, 1982, 1984; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Since there were no such visual

cues in our laboratory experiment, it might be the case that our results underestimate infant abilities. Regardless, infants should have little trouble attending to a caregiver's voice in noise when the distractor voice is substantially quieter.

However, another issue to consider has to do with the nature of the competing stimuli in the present experiment. Specifically, the target voice in this situation was saying isolated words against the background of the distracting voice. Yet, as Woodward and Aslin (1990) have noted, caregiver's speech to infants rarely contains isolated words. Hence, claiming that infants are able to learn much about their native language in the presence of competition from other voices is not really justified. A more convincing demonstration would be to show that they are capable of attending to one fluent stream in the environment of another competing stream.

How well would infants be able to separate two fluent passages of speech? Although, at first glance, this may not seem to be any different, in reality it is a much more difficult task. The infants did not have to segment the words in the familiarization list produced by our target voice: the words were already clearly separated by pauses. In fluent speech, this is not the case. Segmenting words from fluent speech is far from trivial, since there are often no clearly marked acoustic discontinuities (Cole & Jakimik, 1980; Klatt, 1980; Reddy, 1976). Masking (from the distractor voice) may make segmentation especially difficult, since it may make it harder for infants to find the subtle allophonic and prosodic clues that could aid them in this task. For these reasons, we might predict that, even though infants can later recognize words that were initially presented in a list fashion in the presence of a distractor, they may not be able to do the same with words presented in fluent speech against a fluent speech distractor. To examine this more fully, we decided to investigate whether infants who were initially exposed to fluent passages in the context of a distractor would later be able to recognize those words in isolation.

EXPERIMENT 4

In this experiment, we reversed the order of the target stimuli. That is, in Experiments 1–3, we presented the isolated words for familiarization (along with the masking stimulus) and then presented the passages during the test phase. In Experiment 4, we presented the passages during familiarization (along with the masking stimulus) and then presented the isolated words during the test phase. This manipulation corresponds to the final experiment by Jusczyk and Aslin (1995), although they did not use a distracting voice in their experiment. Even without the distractor, this task is much harder. As they note, learning to recognize the individual words is likely to be easier when presented first in isolation (with word boundaries clearly marked) than when presented first in fluent speech (requiring that infants segment the words themselves). Jus-

czyk and Aslin found that infants could still recognize the words presented in the familiarization phase in this situation. However, the addition of a distractor voice might exacerbate the difficulty of this task by making it even more difficult to segment the individual words correctly. Thus, if infants can still perform this task well, it would be an even more powerful demonstration of infants' ability to separate different streams of speech.

Method

Subjects. Twenty-four American infants (11 males, 13 females) from monolingual English-speaking homes participated in this experiment. The average age of these infants was 32 weeks 6 days (range = 30 weeks 5 days to 35 weeks 4 days), or approximately 7.5 months. Six additional infants were tested but were not included for the following reasons: 2 for crying, 2 for failing to look for an average of 3 sec to each type of word list, 1 for falling asleep during the experiment, and 1 for experimenter error.

Stimuli. The same recordings used in Experiments 1–3 were used here. However, rather than combining the target voice saying the list of words with the distractor voice reading the passage, we combined the target voice's passages with the distractor voice. We adjusted the average intensity of the distractor passage to be 10 dB less than that of the target passage. Also, to aid the infants in the start of the task, we had the target voice start 1 sec before the distractor voice (i.e., we added 1 sec of silence to the start of the distractor passage). This time lag for the distractor passage would presumably encourage the infants to listen to the target voice and might also aid in segregation (Dannenbring & Bregman, 1978). We adjusted the order of the sentences within the female voice's passage so that the target word did not occur before this 1-sec lead-in ended. That is, although the target voice had 1 sec in which it was not masked by the distractor voice, the target word did not occur during this period. As in Experiment 1, we made sure that the target word did not occur during any of the pauses in the distractor passage. That is, it never occurred without masking.

Design. Half of the infants heard the passages containing *cup* and *dog* during the familiarization phase, and the other half heard the passages containing *feet* and *bike*. During the test phase, all of the infants heard four blocks of the same four lists of words. Each block contained a different random ordering of the lists containing the words *cup*, *dog*, *feet*, and *bike*.

Apparatus. The apparatus was identical to that in Experiments 1–3.

Procedure. The major change in procedure from that of Experiments 1–3 was that the passages containing the target words were presented during familiarization, and the lists of words were presented during the test phase. During the familiarization phase, the infants heard two target passages (each containing six sentences) on alternate trials. Following Jusczyk and Aslin (1995), this phase continued until they accumulated at least 45 sec of listening time to each of the passages containing the target words. This requirement ensured that the infants were exposed to the target word in the passage at least 12 times during the familiarization phase. During the test phase, the infants heard repetitions of one of the target words on each trial. Over the course of four trials (or one block), they heard all four lists of words. In order to reduce the length of the session (and compensate for the increased time in the familiarization phase), each infant was tested on three blocks, rather than four, for a total of 12 test trials (the same procedure followed by Jusczyk & Aslin, 1995).

Results and Discussion

Mean listening time to the four word lists was calculated for each infant across the three blocks of trials. The mean listening times for the two familiar word lists were then averaged together, as were the mean listening times

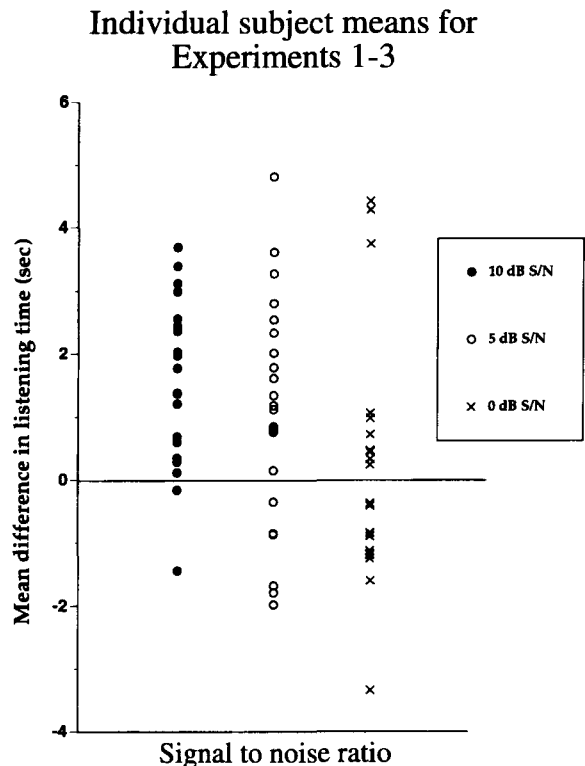


Figure 4. Mean difference in listening times for each individual subject in Experiments 1–3.

for the two unfamiliar word lists. Only 13 infants listened longer to the lists containing the familiar words. However, the average listening time for these lists across all infants was 11.34 sec, and the average for lists containing unfamiliar words was 9.85 sec (see Figure 5).⁵ This difference was significant by a paired *t* test [$t(23) = 2.28, p = .0322$], even though only half the infants showed the effect.

Only half of the infants had longer listening times for the familiar stimuli. We were therefore somewhat surprised that the difference in listening times was statistically significant. Closer scrutiny of the data indicated that the performance of individual subjects tended to be bimodally distributed. As is evident in Figure 6, there is a cluster of infants hovering around 0 (no difference in listening time) and another group of subjects who are listening longer to the familiar stimuli (positive differences in listening time). This figure suggests that this task may be just on the cusp of what the infants can (or will) do. That is, infants *can* segregate passages of fluent speech into separate streams and selectively listen to one passage, but they appear to find this task quite difficult and therefore do not consistently perform well. Or, perhaps, half of the infants can perform the task quite easily and the other half simply find it too difficult. Unfortunately, we do not have any other measures on these infants (such as hearing test results) that might correlate with their different results and thus explain why the task seemed to be easier for some than for others. Whatever the cause of the bimodal distribution, the fact that the overall analysis was

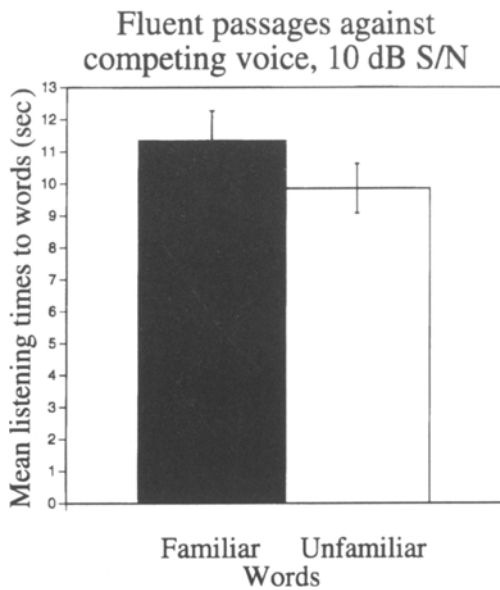


Figure 5. Mean listening times and standard errors to the familiar and unfamiliar words in Experiment 4.

significant suggests that even when speech is masked, infants have some capacity for recognizing words in isolation that they previously have heard only in sentential contexts.⁶

As in the previous experiment, we also obtained information regarding adults' abilities to extract the target items in the face of a competing distractor passage. Six new adult subjects were tested individually. All subjects received two familiarization trials, each of which consisted of a target passage paired with a simultaneous distractor passage. As in the previous experiment with adults a 0-dB S/N ratio between the target and distractor passages was used.⁷ These familiarization trials were followed by four test trials (each consisting of repetitions of isolated familiar or unfamiliar words). On each test trial, the subject had to indicate whether the isolated word had appeared in one of the familiarization passages spoken by the female voice. Again, the overall performance level for these 6 adults was significantly above chance, but far from ceiling level. In fact, their level of performance was identical to that found in the previous experiment (i.e., 67% correct). Three of the subjects were correct on all four test trials, whereas 1 subject was incorrect on all four trials. Hence, even though adults heard the stimuli at an S/N ratio that was 10 dB lower than that for the infants, they were still able to extract and remember sufficient information about the target words to perform at better than chance levels during the test phase.

Why do infants appear to have more difficulty with this version of the selective attention task than do adults? One possible explanation has to do with differences between infants and adults in thresholds for perceiving speech in noise. Infants' thresholds are often measured at much higher levels than adults' thresholds. As noted in the introduction, infants' thresholds for speech detection in quiet have been measured at 15–25 dB higher than adults'

thresholds (e.g., Nozza, Rossman, & Bond, 1991; Nozza et al., 1988). For speech presented in noise, infants' thresholds have been estimated to be 6–16 dB higher than adults' thresholds (Nozza et al., 1988; Trehub et al., 1981). This difference in infant and adult thresholds fits well with the fact that adults were still performing at above chance levels with S/N ratios 10 dB lower than for infants in the present task. In this light, our difference between adults and infants was actually predicted by the literature.

It is worth noting that task demands will influence the way in which S/N ratios will affect both infants' and adults' abilities to process speech in noisy environments (although adult performance should still be better than infant performance at any given S/N ratio). For example, previous research with adults has suggested that the threshold for detectability of speech in noise is at an S/N ratio of approximately -17 dB (Hawkins & Stevens, 1950). However, this estimate of adult thresholds is based on correct performance in *detection* tasks, as are the estimates in previous studies with infants (Nozza et al., 1988; Trehub et al., 1981). By comparison, *distinguishing* between speech sounds (which is required in our task) normally requires higher intensity levels than does simple detection. Hawkins and Stevens found the threshold for intelligibility of speech in noise to be at approximately a -8-dB S/N ratio, or about 9 dB higher than the threshold for detection. Presumably, the threshold for speech intelligibility in noise for infants would likewise be larger than their detection thresholds. Additionally, in order to

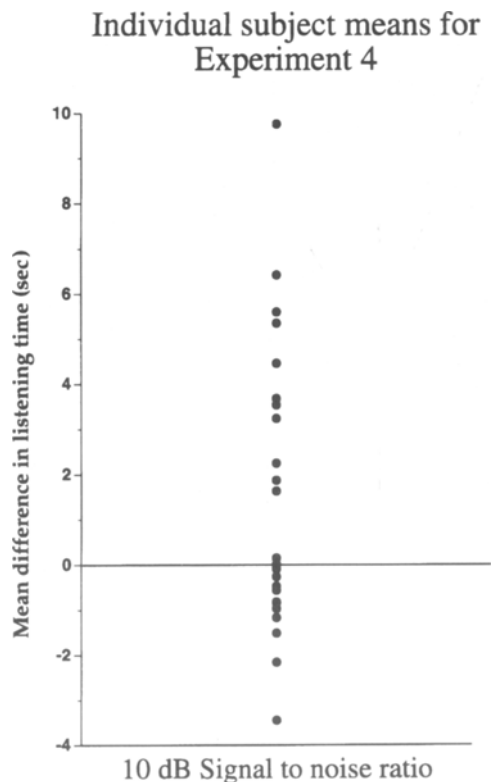


Figure 6. Mean difference in listening times for each individual subject in Experiment 4.

perform consistently above chance, infants would presumably require greater S/N ratios than they do to perform at chance (threshold performance). Thus, we might expect that tasks such as ours (which do require performance consistently above chance and also involve distinguishing between speech sounds rather than simple detection) would require a greater S/N ratio than would the studies with infants reported earlier (Nozza et al., 1988; Trehub et al., 1981). In fact, this is the case: Nozza et al. (1990) found that infants needed an 8-dB S/N ratio in order to reach an 85% hit rate on a /ba/-/ga/ distinction in the presence of noise. This S/N ratio is close to the 10-dB ratio we used here. It is not surprising, then, that the 10-dB S/N ratio seems to be just barely at a point at which infants can selectively attend to one fluent speech passage against a background of a competing distractor.

One might wonder whether slightly older infants (e.g., 10-month-olds) would have performed better on this task. Although we have not examined this, prior literature suggests that there may be little difference. Nozza and Wilson (1984) examined pure-tone detection thresholds in quiet for both 6- and 12-month-old infants. They found only a 2-dB nonsignificant difference and stated that this "is not evidence that auditory function changes during the second half of the first year of life" (Nozza & Wilson, 1984, p. 619). Similarly, Trehub et al. (1981) tested infants at 6, 12, 18, and 24 months of age in a speech-detection-in-noise task and found little difference between the groups. If tone and speech detection thresholds in noise do not change significantly during this time period, speech identification in noise is also unlikely to change appreciably. This is because the latter abilities are presumably limited by speech detection thresholds in noise.

GENERAL DISCUSSION

The present study demonstrated that 7.5-month-olds can separate streams of speech produced by different talkers. Furthermore, even when speech is masked, infants are able to store representations of words they are exposed to and later recognize these same items in different contexts.

It is certainly beneficial to the infants to be able to extract information from their caregiver's speech even when there are other noises in the background. As anyone who has ever tried to record vocal interactions between infants and their mothers can attest to, infants frequently have to deal with less than ideal listening situations. They may have to "tune out" all sorts of competing noises around the home to gain any information from their caregiver's speech. If they were unable to do this, they would have to draw on a much smaller set of utterances in order to discover the structure and organization of their native language.

Most laboratory testing has taken place in perfectly quiet situations that are unlikely to be an accurate representation of what infants typically experience. One might argue that our testing situation has gone too far in the opposite extreme. Our testing situation was quite difficult and might therefore be considered to test the limits of infant streaming, rather than examining situations that in-

fants are likely to experience at home. For example, both our distractor voice and our target voice came from the same location in space, and, thus, the infant could not depend on localization cues to segregate the different voices. Furthermore, we did not provide any visual cues, and infants have been shown to be sensitive to lip movements (Kuhl & Meltzoff, 1982, 1984; MacKain et al., 1983). Perhaps, when listening to speech around the home, infants can use cues from the speaker's face to help them focus on only one voice.

That some infants could perform well with the fluent passages masked is quite remarkable, especially given that it is only at this age that infants first begin to show an ability to detect words in fluent speech (Jusczyk & Aslin, 1995). This result demonstrates the robustness of infants' ability to segment, extract, and store words from fluent speech. In turn, these findings should provide some reassurance to those investigating the impact of linguistic input on the course of language development. Even when investigators focus only on child-directed input, there might be questions as to how effective such input is in the face of competition from other acoustic signals. The pattern of findings reported here suggests that it may not be necessary to restrict what counts as input to only those utterances that are produced "in the clear" (without competition from other noises or other voices).

There are many issues yet to be examined. Although we did not provide infants with any localization cues to separate the signal from the noise, the cues we did provide were quite distinctive. For instance, our voices were very different in fundamental frequency: one was a low-pitched male voice and the other a high-pitched female voice (the F_0 for the male tended to range from 80 to 95 Hz; the female's range tended to be from about 180 to 450 Hz). This difference made the voices very easy to distinguish. In follow-up studies, we are examining the importance of this cue by comparing infants' performance when voices are judged to be very dissimilar with their performance when the voices are judged to be quite similar.

Another issue worth exploring has to do with the number of background voices. In a typical cocktail party, we do not have only one target voice and only one distractor voice to contend with. Rather, there are many different voices in the background, all speaking simultaneously. Adults performance is typically U-shaped, with poorer discrimination of the target voice as more voices are added to the background up to a certain point and then better discrimination with additional voices beyond that point (Carhart, Johnson, & Goodman, 1975). Do infants follow the same pattern?

Another issue that remains to be addressed concerns the extent to which spatial cues might help infants segregate different voices. In the experiments reported here, both the distractor voice and the target voice were presented from the same speaker, which was located on the infants' right or left side. What if we were to separate the two voices in space? Would infants be able to segregate different voices at a lower S/N ratio in this condition than they would in the ones presented here? The headturn

procedure we used severely limits the possible locations for the distractor voice. We would not want it to be at the midpoint, for instance, since it would then require less effort for the infant to look at the distractor than at the target. Moreover, locating the distracting voice to the same side might make it difficult to judge whether the infant is orienting to the target or the distractor. We are currently exploring other possibilities that will allow us to spatially separate the voices. In any event, by following these directions for future research, we should be able to determine more accurately when and how much infants benefit from hearing caregivers' speech.

Finally, to what degree might these abilities to segregate different voices be evident in even younger infants? The age group tested in the present study, 7.5-month-olds, is the youngest to display evidence of word detection abilities. Using a version of the same test paradigm, Jusczyk and Aslin (1995) found no evidence for word detection abilities in 6-month-olds. Although it is possible that the use of a different task might reveal evidence of the ability to separate voices in younger infants, it is also possible that this ability develops in conjunction with the beginnings of lexical development. A growing attentiveness to the presence of words in fluent speech may provide additional motivation for following a particular voice in a noisy background. One way to explore whether there is some linguistic motivation that provides the impetus for streaming voices apart would be to undertake comparable studies using nonspeech materials, such as instrumental music.

In conclusion, the present study takes an initial step forward in exploring infants' capacities for processing and encoding speech information in the presence of competing auditory signals. Our findings suggest that, by 7.5 months, infants possess at least some limited capacity to selectively attend to an interesting voice in the context of a competing distractor voice. That is, infants can discriminate between voices and can extract information from the speech signal even in the face of other competing acoustic signals. Moreover, it is significant that the task used in the present study—recognizing familiar sound patterns of words—is of the type that seems likely to be associated with building up a lexicon for the native language.

REFERENCES

ANDERSON, D. R., ALWITT, L. F., LORCH, E. P., & LEVIN, S. R. (1979). Watching children watch television. In G. Hale & M. Lewis (Eds.), *Attention and cognitive development* (pp. 331-361). New York: Plenum Press.

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**, 680-696.

BERGMAN, M. (1971). Hearing and aging. *Audiology*, **10**, 164-171.

BEST, C. T. (1995). Learning to perceive the sound patterns of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 217-304). Norwood, NJ: Ablex.

BERGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

BERGMAN, A. S., ABRAMSON, J., DOEHRING, P., & DARWIN, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics*, **37**, 483-493.

BERGMAN, 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.

CARHART, R., JOHNSON, C., & GOODMAN, J. (1975). Perceptual masking of spondees by combinations of talkers [Abstract]. *Journal of the Acoustical Society of America*, **58**, S35.

CHALIKIA, M. H., & BERGMAN, 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.

CLIFTON, M. G., MORRONGIELLO, B., KULIG, J., & DOWD, K. (1981). Newborns' orientation towards sound: Possible implications for cortical development. *Child Development*, **52**, 833-838.

COLE, R. A., & JAKIMIK, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NJ: Erlbaum.

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

DARWIN, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, **33A**, 185-207.

DEMANY, L. (1982). Auditory stream segregation in infancy. *Infant Behavioral Development*, **5**, 261-276.

EIMAS, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, **16**, 513-521.

EIMAS, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, **18**, 341-347.

EIMAS, P. D., & MILLER, J. L. (1980, September 5). Contextual effects in infant speech perception. *Science*, **209**, 1140-1141.

EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P., & VIGORITO, J. (1971, January 22). Speech perception in infants. *Science*, **171**, 304-306.

FERGUSON, G. A., & TAKANE, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw Hill.

FERNALD, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, **8**, 181-195.

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.

HAWKINS, J. E., JR., & STEVENS, S. S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America*, **22**, 6-13.

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

JUSCZYK, P. W., & ASLIN, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, **29**, 1-23.

JUSCZYK, P. W., CUTLER, A., & REDANZ, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675-687.

JUSCZYK, P. W., FRIEDERICI, A. D., WESSELS, J., SVENKERUD, V. Y., & JUSCZYK, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language*, **32**, 402-420.

JUSCZYK, P. W., LUCE, P. A., & CHARLES-LUCE, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language*, **33**, 630-645.

JUSCZYK, P. W., PISONI, D. B., & MULLENNIX, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, **43**, 253-291.

KEMLER NELSON, D. G., JUSCZYK, P. W., MANDEL, D. R., MYERS, J., TURK, A., & GERKEN, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior & Development*, **18**, 111-116.

- KLATT, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.
- KUHL, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, **66**, 1168-1679.
- KUHL, P. K. (1983). Perception of auditory equivalence classes for speech during early infancy. *Infant Behavior & Development*, **6**, 263-285.
- KUHL, P. K., & MELTZOFF, A. N. (1982, December 10). The bimodal perception of speech in infancy. *Science*, **218**, 1138-1141.
- KUHL, P. K., & MELTZOFF, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, **7**, 361-381.
- LINDMAN, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman.
- MACKAIN, K., STUDDERT-KENNEDY, M., SPIEKER, S., & STERN, D. (1983, March 18). Intermodal perception is a left hemisphere function. *Science*, **219**, 1347-1349.
- MCADAMS, S., BERTONCINI, J., & BOBROW, J. (1990, November). *Organization and discrimination of repeating sound sequences by newborn infants*. Paper presented at the 120th Meeting of the Acoustical Society of America, San Diego, CA.
- MILLER, J. L., & EIMAS, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, **13**, 135-165.
- MUIR, D., & CLIFTON, R. K. (1985). Infants' orientation to the location of sound sources. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the 1st year of postnatal life: A methodological overview* (pp. 171-193). Norwood, NJ: Ablex.
- MUIR, D., & FIELD, J. (1979). Newborn infants orient to sounds. *Child Development*, **50**, 431-436.
- 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., WAGNER, E. F., & CRANDELL, M. A. (1988). Binaural release from masking for a speech sound in infants, preschool children and adults. *Journal of Speech & Hearing Research*, **31**, 212-218.
- 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.
- 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.
- REDDY, R. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE*, **64**, 501-531.
- REMEZ, R. E., RUBIN, P. E., BERNS, S. M., PARDO, J. S., & LANG, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, **101**, 129-156.
- RUBEN, R. J. (1992). The ontogeny of human hearing. *Acta Otolaryngologica*, **112**, 192-196.
- 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.
- 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.
- SUMMERFIELD, Q., & ASSMANN, P. F. (1991). Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony. *Journal of the Acoustical Society of America*, **89**, 1364-1377.
- TREHUB, S. E., BULL, D., & SCHNEIDER, B. A. (1981). Infants' detection of speech in noise. *Journal of Speech & Hearing Research*, **24**, 202-206.
- TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242-248.
- WERKER, J. F., & DESJARDINS, R. N. (1995). Listening to speech in the 1st year of life: Experiential influences on phoneme perception. *Current Directions in Psychological Science*, **4**, 76-80.
- WOODWARD, J. Z., & ASLIN, R. N. (1990, April). *Segmentation cues in maternal speech to infants*. Paper presented at the 7th Biennial Meeting of the International Conference on Infant Studies, Montreal.
- YOUNG, L. L., JR., PARKER, C., & CARHART, R. (1975). Effectiveness of speech and noise maskers on numbers embedded in continuous discourse [Abstract]. *Journal of the Acoustical Society of America*, **58**, S35.

NOTES

1. It is worth mentioning that not all researchers agree with this notion of grouping information together. Remez, Rubin, Berns, Pardo, and Lang (1994) have argued that the kinds of Gestalt principles that are commonly cited as a means of this grouping actually will not work in speech. Speech is a very complex signal, and its acoustic constituents often would not be grouped together according to these rules. While this debate is intriguing, it is not directly relevant to the research we present here. That is, we are trying to determine whether infants have the ability to segregate voices at all; we are not yet ready to join the debate as to what processing mechanisms they use to do so.

2. Anderson, Alwitt, Lorch, and Levin (1979) have shown that both women and puppets encourage attention from children watching television.

3. We thank Daphne Maurer for first calling our attention to this method of plotting the data.

4. Several of the subjects reported at the end of the experiment that they actually listened more attentively to the male voice because there was more semantic content in the passage that he was reading.

5. These are similar to the listening times found by Jusczyk and Aslin (1995) for isolated words following familiarization with passages containing those words. They found times of 10.43 sec for the familiar words and 8.32 sec for the unfamiliar words.

6. Although the data are bimodal, *t* tests are relatively insensitive to violations of normality as long as the number of subjects is large enough. Ferguson and Takane (1989) considered 25 subjects to be "large enough"; our 24 is presumably sufficient in this regard. Also, see Lindman (1974), for a discussion of the use of parametric statistical tests when normality is violated.

7. We chose to test adults in this study at a 0-dB S/N ratio because this is the same level we had used with the adult subjects in the study described in the discussion of Experiment 3. While this may make comparison with the infant data more difficult, we felt that the better hearing abilities of adults would make a 10-dB condition too easy for them.

(Manuscript received June 9, 1995;
revision accepted for publication February 14, 1996.)