9- and 12-month-olds Fail to Perceive Infant-Directed Speech in an Ecologically Valid Multi-Talker Background

Dana Elizabeth Bernier

A Thesis submitted to the
Faculty of Graduate Studies of The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology
University of Manitoba
WINNIPEG, MANITOBA, CANADA

Abstract

Little is known about infants' ability to deal with commonly encountered situations in which speech from one individual occurs simultaneously with that of others. Previous research has shown that while age of the infant and intensity of the background matter, so does the number of background speakers. The read-aloud multi-talker speech used in previous studies is perceptually different from conversational speech typically encountered by infants. To test generalizability, this study used a background of ecologically valid multi-talker speech. Using the head-turn preference procedure, infants were presented with passages of background noise with and without target infant-directed speech at a 10 dB SNR. Results show that while 9-month-olds prefer passages containing target speech with a white noise background, both 9- and 12-month-olds failed to show a preference with a multi-talker background. This inability to segregate speech streams under ecologically valid conditions demonstrates the adversity infants face to learn their community's language.

Acknowledgments

I would like to thank Dr. Melanie Soderstrom for her guidance and mentorship. Her feedback and support throughout the research process and during the preparation of this manuscript were invaluable. Additionally, I would also like to thank the other members of my supervisory committee, Dr. Melanie Glenwright, and Dr. Kevin Russell for their helpful comments. Finally, but certainly not least, I would also like to thank the people in the Baby Language Lab, in particular Melissa Reimchen, Anne Marie Heard, Michelle Donato, and Lindsay Bacala, for their support, and their help along the way.

This research was funded by a Post-Graduate Scholarship from the Natural Science and Engineering Research Council of Canada, a Manitoba Graduate Scholarship with a University of Manitoba Graduate Fellowship Top-Up, and a University of Manitoba Faculty of Arts Thesis Write-Up Award.

Table of Contents

List of Tables

List of Figures

9- and 12-month-olds Fail to Perceive Infant-Directed Speech in an Ecologically Valid Multi-Talker Background

In order for infants to learn the language(s) they are exposed to, speech needs to be acoustically discriminable from all other sounds in the environment. In the best case scenario (home with a single caregiver) 24-53% of a toddler's day is taken up by human vocalizations, with 31-62% of these vocalizations being overlapped with other sounds, including other speech (Bernier & Soderstrom, 2013). Given that such a high proportion of the speech infants are exposed to is overlapped with other sounds, it is important to understand whether these occurrences of overlapping speech are something infants can use as language input, or whether they are perceived as nothing more than a jumbled mess of random noise without rules wherein the target speech blends into the background speech. In order to examine this question, infants' ability to perceive infant-directed, speech under naturalistic overlapping speech conditions was assessed in a laboratory setting.

In situations of overlapping speech infants must perceptually group the sounds from one source (the target) separately from another (the masker or background). The first hurdle that must be overcome is that of energetic masking, in which acoustic properties of the target are rendered undetectable due to them being "swamped" by the competing background sounds. Once this type of interference has been accounted for and target stimuli are detectable, there remains the need to analyze the complex sound mixture into its constituent parts, and group the acoustic properties that belong together (those originating from each source) into separate

streams. Some of these processes will be signal-driven (bottom-up) and already present in infancy (Smith & Trainor, 2011), while others will be knowledge-driven (top-down) and subject to learning. Considering that infants have little in the way of a learned language, these knowledge-driven (top-down) processes will be limited if not wholly absent, likely resulting in situations of overlapping speech that adults find manageable and infants find impossible. These top-down, knowledge-driven processes are particularly evident in cases of second language learners listening in multi-talker environments. Second Language learners have limited experience with the language in question, just like infants, and have more difficulty perceiving target speech in these multi-talker situations than do native speakers (Cooke, Lecumberri, & Barker, 2008; Van Engen, 2010), thus demonstrating the power of language experience in stream segregation.

Not only must infants contend with an immature knowledge of the language system, they are further challenged by having an auditory system that is not yet fully developed. Studies have found that infant auditory perception is not as keen as that of adults, needing louder stimuli in order to detect its presence, as well as to discriminate speech sounds. Infants have been found to need a higher threshold (minimum intensity) for the detection of masked and unmasked pure tones compared to adults (Nozza & Wilson, 1984; Sinnott, Pisoni, & Aslin, 1983; Werner & Boike, 2001). The same holds true for broadband noise bursts (Werner & Boike, 2001). Accordingly, infants' ability to detect speech, a broadband noise, is inferior to that of adults in both quiet (Nozza, Wagner, & Crandell, 1988) and noise (Trehub, Bull, & Schneider, 1981).

Detecting the presence of speech is of course only the beginning. In order to make sense of speech, we must be able to discriminate between the various speech sounds (e.g. /ba/ vs. /ta/ vs. /da/), otherwise we wouldn't know if someone was saying ball, tall, or doll; words with very different meanings though phonetically similar. It is unsurprising, given infants' need for a higher auditory threshold to detect speech and their minimal linguistic experience, that their ability to discriminate speech-sounds in both quiet and noise is also not as keen as adults (Nozza, Miller, Rossman, & Bond, 1991; Nozza, Rossman, & Bond, 1991; Nozza, Rossman, Bond, & Miller, 1990). In total, it would seem that infants face a great many challenges in perceiving the speech information they are exposed to in order to acquire their first language.

Although much has been learnt about adults' ability to segregate their native speech in multi-talker environments (Broadbent, 1952; Cherry, 1953; Hirsh, 1950; Miller, 1947; Pollack & Pickett, 1958; Poulton, 1953), very few studies have examined infants' abilities to perceive continuous speech under noisy conditions, with the bulk of the research to date stemming from the work of Newman and colleagues (Barker & Newman, 2004; Newman, 2005; 2009; Newman & Jusczyk, 1996; Newman & Morini, 2010). The big question of what constitutes acceptable background noise – as type and intensity – for language acquisition thus remains incomplete.

One way to examine infants' perceptual capabilities in noise is to see if they are able to recognize the properties of their own name. Not only is one's own name a familiar word, it is also highly salient and therefore likely to induce the proverbial cocktail party effect; a phenomenon first identified 60 years ago where highly salient

or familiar words draw our attention under noisy conditions (Cherry, 1953; Cherry & Taylor, 1954). Irrespective of noise level, the ability to recognize familiar words, such as one's own name, relies on the ability to perceive the word's stress pattern (e.g. reCORD vs. REcord or MAry vs. maRIE) as well as the more subtle phonetic details that distinguish individual speech sounds, or phonemes (e.g. bunny vs. funny or Robby vs. Bobby). With regard to infants' ability to recognize their own name, this is something already in place by 4.5 months (Mandel, Jusczyk, & Pisoni, 1995). Without any background interference, young infants are able to discriminate their own name (e.g. MElody) from foils with different stress patterns (e.g. daNIELa), as well as from the more difficult stress-matched foils (e.g. JUlia). Infants' ability to discriminate their own name from the non-stress-matched foils (e.g. MElody vs. daNIELa) indicates that they are able to recognize the stress pattern of their name and possibly the phonetic details. Infants' ability to discriminate their own name from the stress-matched foil (e.g. MElody vs. JUlia), on the other hand, indicates that they are able to recognize the subtle phonetic details of their names.

Newman (2005; 2009) examined infants' ability to recognize their own names under conditions of multi-talker babble consisting of nine overlapping female voices reading aloud from a book in a child-directed speaking style. Newman found that, 5-month-old infants were capable of discriminating their names from both non-stress-matched and stress-matched foils when the target speech (signal) was 10 dB higher than the background speech (noise); a signal-to-noise ratio (SNR) of 10 dB. However, with a more difficult SNR of 5 dB, they were only able to discriminate their names from the non-stress-matched foils. Their inability to discriminate two names with the

same stress pattern indicates that phonetic details are either too difficult for the infants to make out, or simply not heard at all. Either way, an SNR of 5 dB appears to present too much noise for 5-month-olds to perceive their own names clearly.

To examine when infants' perceptual capabilities improve at the lower 5 dB SNR level, Newman (2005) examined both 9- and 13-month-olds, and found that not until 13 months were infants able to discriminate their names from both non-stress-match and stress-match foils. Thus between 9 and 13 months infants develop the perceptual sensitivity to be able to discriminate phonetic details at noise levels that younger infants were not.

Although 5- to 9-month-olds are able to discriminate their names from both non-stress-matched and stress-matched foils in a multi-talker background at a 10 db SNR, when they are presented with a single-talker background at this SNR they are unable to discriminate their names from either foil (Newman, 2005; 2009). Thus single speakers present an even greater challenge to infants than multiple speakers.

For adults, the reverse is true; they have an easier time with single speaker backgrounds (Pollack & Pickett, 1958). The release from masking seen in these situations has been found to be due to listeners taking advantage of silent periods (dips) that occur in the background (masking) speech (Miller, 1947) called "listening in the dips". To accomplish this, listeners must focus their attention to the dips in the background speech and use their lexical knowledge to fill in the missing information. With multiple speakers, the silent periods begin to disappear as the fluctuations found with any one speaker blends with those of other speakers producing a

cacophony of overlapping sounds eliminating even fleeting information that could be used to infer the content of target speech.

Infants, lacking lexical knowledge and having poor selective attention, therefore appear to be doubly disadvantage with single speaker backgrounds. . The physical properties of the background speech, as opposed to its contextual nature, appear to be the source of their difficulties in these situations since reversed single-talker speech poses an equally challenging problem to 5-month-olds as forward speech (Newman, 2009). Birdsong and cricket chirps, equally melodic but outside the frequency range of human speech, also interfered with 6- to 8-month-olds' ability to discriminate speech sounds (Polka, Rvachew, & Molnar, 2008), demonstrating further that energetic masking is not what is driving the effect. One possibility is that the rhythmic pulsating natures of these sounds – their fluctuation amplitudes – is what is diverting infants' attention away from a task they are otherwise able to perform. Multiple speakers, with their cacophony of overlapping sounds, mesh together providing infants with a more uniform, and thus less distracting, sound. This is exemplified by infants' failure to discriminate their name from foils in amplitude modulated white noise, despite successful discrimination in constant amplitude white noise (Newman & Morini, 2010). Therefor with infants, the problem of segregating speech in single-talker backgrounds is equally an issue of selective attention, but one in which the "dips" are distracting them away from the task.

So far we have only considered infants' ability to discriminate speech sounds and recognize their own name in isolation. Language is more complex than this, requiring that infants perform a number of linguistic tasks simultaneously, including

mapping meaning onto words. However, to do this, infants must already have some knowledge of what a word is. In its simplest form, a word is a group of sounds (phonemes) that a language community agrees can be uttered in isolation, having semantic or pragmatic content. To complicate matters, infants rarely hear words in isolation. They hear them in continuous speech surrounded by many other sounds (words). This means that the infant must chunk sounds together into words using only the information present in the speech stream. The infant's task is to first learn which sounds should be grouped together (words) and which should be separated; in other words they must learn to segment the speech stream.

Jusczyk and Aslin (1995) tested infants on their ability to do just that. After first familiarizing infants with words in isolation, they tested whether infants were able to recognize those words, versus other non-familiarized words, in sentences. They found that 7½-month-olds listened significantly longer to sentences containing familiarized words than to those containing non-familiarized words. Their ability to recognize the familiarized words was not due to a general loose matching of the stress pattern or the particular vowel used (e.g. the /u/ in cup) but rather to detailed representations of the words since they did not confound similar sounding words (e.g. "tup" for "cup"). Recognizing sound patterns of words was something Jusczyk and Aslin found that 6-month-olds could not do. However, more recent research suggests that 6-month-olds are able to use top-down processing, taking advantage of highly salient familiar words such as their own name or the moniker used for their mother (Mommy or Mama) to recognize the word immediately following it (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005).

Jusczyk & Aslin's (1995) results were obtained without any background interference. Just like previously discussed studies, when 7½-month-olds were given this task with a single-speaker background at a 10 dB SNR, they failed to recognize familiarized words in sentences. However, when the target speaker was their own mother, instead of an unfamiliar female, infants showed a preference for familiarized words (Barker & Newman, 2004). Familiarity therefore helps infants segregate target and background speech.

Adults are also better at identifying novel words in noise produced by a familiar speaker compared to a novel speaker (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). With adults, speaker familiarity induces a perceptual illusion such that they will estimate background noise to be less intense for familiarized words produced by the same (familiar) speaker, compared to when they are produced by a novel speaker (Goldinger, Kleider, & Shelley, 1999; Jacoby, Allan, Collins, & Larwill, 1988). In other words, familiarity with a speaker induces the perception of a (falsely) lower SNR. Infants, just like adults, appear to be using these top-down processes when segregating familiar voices from background voices.

Unlike previously discussed findings, when the background was a single *male* speaker, instead of the *female* speaker of previously discussed studies (i.e. the opposite gender of the target speaker), 7½-month-olds fared quite well at both a 10 dB, and a 5 dB SNR, showing a preference for familiarized words. Only when both background, and target speech were of the same intensity did 7½-month-olds fail (Newman & Jusczyk, 1996). Not only did 7½-month-olds succeed with a male background speaker at an intensity they failed with a female background speaker (10

dB SNR), they also succeeded with *a single male* speaker at an intensity they failed with *multiple female* speakers (5 dB SNR). What this suggests is that as the SNR gets more difficult, there is a point at which speakers of the same gender mesh together, but which opposite gender speakers continue to stand out even when taking number of background speakers into account. It appears then that the gender of the background and target speakers plays a pivotal role in segregating speech streams; speakers of the same gender presenting a more difficult task than speakers of different genders. A finding that is also evident with adult listeners (Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001).

What this tells us is that amplitude modulations of the background speech can't be the only factor involved with infants' failure to segregate 2 unfamiliar female speakers. In order to segregate target and background speech, infants must be able to perceptually group the two streams into distinct units. At 7½-month infants are able to generalize, from familiarization to test, across different speakers of the same gender (all males or all females), but not across speakers of different genders (male to female or female to male) (Houston & Jusczyk, 2000). Although all speakers differ from one another, voices used in this study of the same gender were judged by adults to be more alike than voices of different genders. Together, this implies that infants rely on physical differences in the speech streams to perceptually separate target and background speech into distinct units; a reliance that gives them an edge over number of speakers.

In each of these cases, the infants were not only required to segregate two different streams of speech, they were also required to segment the target speech

stream into chunks (words). In order to locate words in fluent speech, infants must break up the speech stream in some way. English-learning infants have been shown to use a variety of cues, such as transitional probabilities (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996), phonotactics (Mattys, Jusczyk, Luce, & Morgan, 1999), coarticulation (Johnson & Jusczyk, 2001; Jusczyk, Hohne, & Bauman, 1999; Mattys & Jusczyk, 2001), and prosody (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Houston, Santelmann, & Jusczyk, 2004; Jusczyk, Houston, & Newsome, 1999; Polka & Sundara, 2012; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003), to accomplish this task.

The first cue, transitional probabilities, are the statistical probabilities that two syllables occur contiguously in the input, relying on the fact that sounds that belong together as words are more likely to be heard together than sound combinations that only occur a word boundaries. For example, in [pɹɪ·ti·bej·bi], the syllable sequence [pɹɪ·ti] has a much higher probability of co-occurrence than [ti·bej], therefore the sequence is understood as *pretty baby*. The second cue, phonotactics, is the permissible combinations of syllables, consonant clusters, and vowel sequences found in a language. In English, as in other languages, the clusters of consonants and vowels found within a word differ from those found between words, with each language having different permissible combinations. Clusters that are found primarily within words, such as the [ŋ·k] cluster in *blanket*, indicate perceptual cohesion, whereas clusters that are found primarily between words, such as the [ŋ·t] cluster of *bring two*, indicate perceptual separation, and therefore a word boundary.

The final two cues are based on the sound characteristics present in the speech signal proper. Coarticulation is the influence a speech sound has on its neighbors. Since speech sounds are not made independently of one another, the articulation pattern of a sound will be influenced by the preceding sound, and will in turn influence, and be influenced by, the one that follows. This leads to variations in the precise articulation of individual phonemes depending on its neighbors; effects that are stronger within words than across words. For sequences such as [sɛl·fɪʃ], the precise articulation of the [l·f] sequence results in either *selfish* or *sell fish*. The final cue, prosody, refers to the intonation, rhythm, and stress patterns of speech. These acoustic features, such as word stress discussed earlier, are what give speech its characteristic melodic quality.

Since infants' ability to segregate target and background speech has been shown to rely on physical characteristics inherent to the speech stream (e.g. speakers' gender), other physical characteristics such as prosody should equally influence infants' ability to segregate target and background speech. Similar prosodic characteristics present in both target and background speech would therefore promote perceptual grouping as a single unit, whereas different prosodic characteristics would promote perceptual separation and thus stream segregation.

Research done to date has relied on recordings of passages being read aloud. As most people can attest, there is a difference between the prepared speech of someone reading aloud and the spontaneous speech produced during conversations. This has been shown to be due to differences in intonation and rhythm across the two

speaking styles that remains discernable when the speech is filtered (Levin, Schaffer, & Snow, 1982).

What makes these two speaking styles different is the 'on-line' nature of spontaneous speech that results in many disfuencies (fillers, hesitations, repetitions, false starts, repairs) as well as incomplete sentences and long pauses, which are absent from the structured organization of written text and therefore absent when it is read aloud. Such differences result in distinct combination of pitch contour, segmental durations, and spectral features between the two speaking styles (Batliner, Kompe, Kießling, Nöth, & Niemann, 1995; Laan, 1997), with the position of stress (Howell & Kadi-Hanifi, 1991), and the location of prosodic and tone boundaries (Blaauw, 1994; Howell & Kadi-Hanifi, 1991) differing across the two speaking styles as well. And although both speaking styles contain pauses, the location, frequency, and duration of these pauses also differs across the two speaking styles with longer, and more frequent pauses occurring during spontaneous speech (Howell & Kadi-Hanifi, 1991; Levin, Schaffer, & Snow, 1982).

In addition to these intonational differences, rhythmic organizational differences have also been noted across speaking styles (Guaïtella, 1999). Whereas the structured order of written text results in read speech being produced with a regular (metrical) rhythm, the broken-up disfluent nature of spontaneous speech favors a different type of rhythmic organization. Instead of being organized around the regularity of prosodic and syntactic units the way read speech is, spontaneous speech is instead organized around the production of events; in other words around pragmatic content.

Thus the prosodic realization of naturally occurring spontaneous speech will be quite different from the read speech used in previous studies. This is especially true when you consider overlapping speech. The speech of each individual in a natural multi-talker situation is likely to be dynamic in pitch, intensity and rhythm changes, with many pauses which will affect the prosody of the overlapped speech sample by allowing individual speakers to stand out momentarily, while leaving the amplitude unscathed relative to read aloud multi-talker speech. It is with this in mind, that the present study has been designed.

Using the Head-Turn Preference procedure, 9- and 12-month old infants were tested on their ability to detect target infant-directed speech in a background of multi-talker adult-directed conversational speech. Given the dynamic nature of overlapping conversational speech, it was expected that infants would have difficulty with this background, but not with white noise. Their difficulty was expected to be due to informational masking resulting from similar prosodic contours found in both target and background speech.

General Method

Participant Recruitment

All participants were full-term (at least 37 weeks gestation), Canadian English-learning infants with no more than 10% exposure to other languages or other dialects of English in the home. Recruitment was done through Manitoba Health via a blind mailing, through community events, or through the lab's existing database of volunteer participants.

Stimuli

Stimuli were recorded to a computer with Bliss© 9 (Mertus, 2011) by a native female speaker of Canadian English using an audio-technica AT2020 USB cardioid condenser microphone in an infant directed speech register with a sampling rate of 22050 Hz at a 16 bit resolution. Stimuli were the same as those used in Jusczyk and Aslin (1995; see Appendix A). Multiple tokens of each sentence and word were generated and later extracted from the recordings using Praat (Boersma & Weenink, 2011), ensuring that there were no sound artifacts such as peak clipping or page rustling.

Fifteen tokens of each word were selected ensuring a variety of pronunciations (see Table 1 for a summary of the acoustic properties). The *cup-dog* tokens do not differ from the *feet-bike* tokens in duration or fundamental frequency, *p*s = ns. The *cup-dog* tokens are however more intense than the *feet-bike* tokens, $t(58) = 3.85$, $p < .001$. Although this difference is large (Cohen's d of 1.01), the intensities of the tokens all fall within an acceptable range and were not judged to be different by lab staff.

One token of each sentence was selected (see Table 1 for a summary of the acoustic properties). The cup-dog sentences do not differ from the feet-bike sentences in either duration or fundamental frequency, *p*s = ns. The *cup-dog* sentences are however more intense than the *feet-bike* sentences, $t(22) = 2.48$, $p = .021$. As with the word tokens, despite the large difference (Cohen's d of 1.06), the intensities of the sentences all fall within an acceptable range and were not judged to be different by lab staff.

Table 1: Acoustic Properties of Stimuli used in Experiments 1 and 2

		Duration (ms)		Frequency (Hz)		Intensity (dB)	
		Mean	*Range*	*Mean*	*Range*	*Mean*	*Range*
tokens	**CUP**	517	430-647	206.4	161.4-245.9	74.2	71.2-76.5
	DOG	583	515-757	169.8	135.3-223.3	73.8	70.3-76.8
	FEET	640	525-743	197.3	147.2-224.4	70.4	63.9-76.5
	BIKE	523	469-582	188.6	153.0-230.1	72.2	65.8-76.1
sentences	**CUP**	2581	2125-2998	194.3	187.6-201.5	70.6	64.5-75.1
	DOG	2512	2046-2778	187.9	176.5-197.0	67.8	65.5-70.8
	FEET	2630	2180-2997	186.9	169.4-212.8	64.8	63.3-65.4
	BIKE	2420	2204-2595	194.4	188.7-198.7	68.0	65.6-71.3

Apparatus

The testing room has three 19 inch LCD flatscreen monitors imbedded into the wall. The infant sat on a caregiver's lap on a chair in the middle of the room approximately 3 feet away from the screens; one screen in front of the infant (and caregiver), one to the left, and another to the right. To avoid unintentionally influencing the infant, the caregiver listened to masking music through circumaural headphones. The experimenter, situated in another room, observed the infant's looking behavior via a closed circuit video feed from a camera beneath the monitor in front of the infant hidden from view.

Procedure

The following experiments use the head turn preference procedure (HPP). In all cases, each trial starts with a flashing yellow circle on the front screen to obtain the infant's attention to center. Once the infant is looking to the front screen, the experimenter indicates this via mouse button press. A flashing yellow circle then appears on one of the side screens; left versus right presentation randomized across trials. Once the infant looks to that screen, the experimenter indicates this by holding

the mouse button for that side, at which point the flashing yellow circle is replaced with a colorful checkerboard pattern and audio files begins playing from the speaker associated with that side. Whenever the infant looks away from the screen, the experimenter releases the mouse button, which stops the sound from playing, and holds the mouse button down again to continue playing the sound once the infant looks back to the screen. Trials end when either the stimulus passage has played for 20 seconds or the infant has looked away for more than 2 consecutive seconds. To avoid cutting audio files part way through while the infant is attending to the screen, files continue to play to the end of the sound file once the 20 second limit is reached. Trials in which an infant looks for less than 1 second total during familiarization (Experiments 1 and 2) or warm-up (Experiments 3 and 4), and 2 seconds total during testing are immediately repeated. A minimum 1 second pause occurs before the next trial can begin. Stimuli are randomized such that passages are presented in random order within a block. For Experiments 1 and 2, sentences within passages are randomly ordered on each trial. For Experiments 3 and 4, sentences are presented in the order listed in Appendix A. The dependent measure is the time the infant spends looking at the side monitor while sound is playing, excluding any time the infant looks away for less than 2 consecutive seconds.

Caregivers were also asked to fill out a questionnaire asking them for basic demographic data and caregiver arrangements.

Experiment 1

In order to replicate previous findings as the basis for a control condition, 9-month-olds were tested on their ability to locate familiarized words in sentences

under two conditions: no background, and white noise. Since the no background condition is a replication of Jusczyk and Aslin's (1995) findings, 9-month-olds were expected to succeed at this task, showing a preference for passages containing familiarized words. Based on previous findings (Colombo, Frick, Ryther, Coldren, & Mitchell, 1995), 9-month-olds were also expected to succeed when target speech was presented in a background of white noise 10 dB quieter than the target speech.

Method

Participants. A total of 40 healthy full-term infants participated in this experiment; 25 in the no background condition (13 M, 12 F) and 15 in the white noise condition (11 M, 4 F). Infants were between 260 and 290 days old, or 9 months 5 days on average. An additional 11 infants were discarded due to signs of their unwillingness to participate such as fussiness/crying/sleepiness (n = 7), parent request (n = 2), external noise (n = 1), and technical difficulties (n = 1).

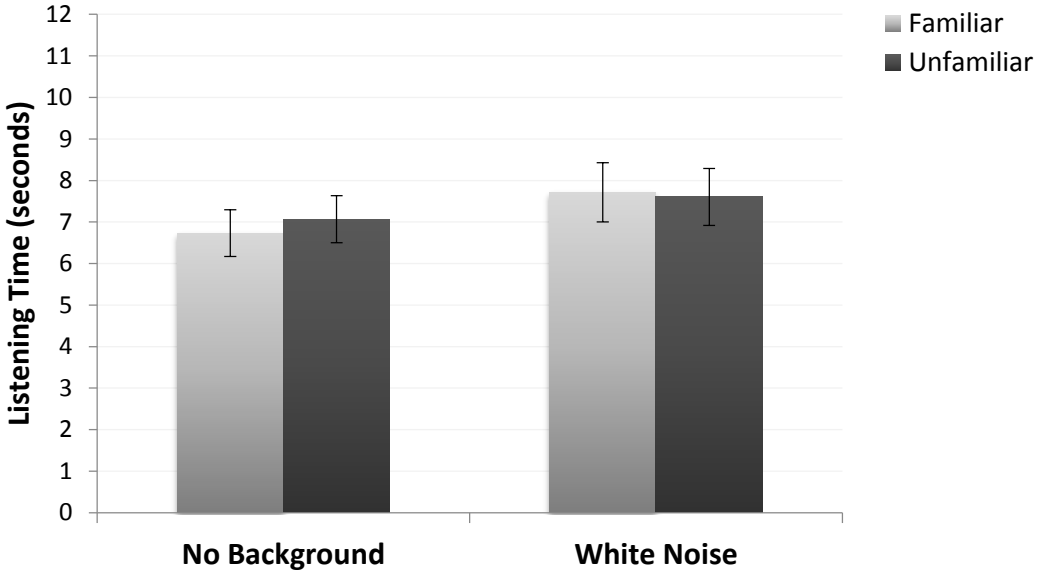
Stimuli. To create the white noise background stimuli, single channel white noise sound files were generated with Praat (Boersma & Weenink, 2011) using a random Gauss formula with a mean of 0, and a standard deviation of 0.25 over a range of +1 to -1 (formula: randomGauss (0,0.25)) with a sampling frequency of 44100 Hz. The mean intensity of each file was then scaled to be 10 dB less than the mean of their respective target sentence. The sentences were then overlapped with their respective white noise files with Bliss© 9 (Mertus, 2011), thereby generating a single audio file of overlapping speech with a 10 decibel (dB) signal-to-noise ratio (SNR) for each sentence.

Design. A modified HPP was used for this experiment (Jusczyk & Aslin, 1995). During the familiarization phase, infants were exposed to two of the four target words. Each trial consists of repetitions of the 15 tokens separated by 500 ms pauses. Half of the infants in each condition were familiarized to *cup* and *dog*, the other half to *feet* and *bike* making the familiar passages of one group the unfamiliar passages of the other thus counterbalancing the two sub-conditions. No background noise was used during the familiarization phase. The test phase began once a minimum of 30 seconds of listening time was reach for each familiarization word. The testing phase consisted of three blocks of four passages (one passage for each of the target words) each containing 6 sentences separated by 500 ms pauses. Half the infants were tested with no background, half with a background of white noise at a 10 dB SNR.

Results & Discussion

Infants listened to familiar passages for 6.73 seconds (*SD* = 2.82), and to unfamiliar passages for 7.07 seconds (*SD* = 2.83) when there was no background; they listened to familiar passages for 7.72 seconds (*SD* = 2.76), and unfamiliar passages for 7.60 seconds (*SD* = 2.66) when the background was white noise (Figure 1). A 2 (familiarity: familiar, unfamiliar) x 2 (background: no background, white noise) mixed model ANOVA was conducted with familiarity as the within subjects factor. Unlike what was found in previous studies, infants did not show a preference for familiar passages, $F(1,38) = 0.076$, $p = .785$. There was also no effect of background, $F(1,38) = 0.875$, $p = .355$, nor was there a familiarity x background interaction, $F(1,38) = 0.309$, $p = .582$. Since no difference was found across the two background conditions, further analyses were performed collapsing across them.

Figure 1: 9-month-olds' listening times to familiar and unfamiliar passages with and without the presence of white noise at a 10 dB SNR



In order to inspect these results, data was then analyzed block by block (Figure 2) with a 2 (familiarity) x 3 (block) repeated measures ANOVA. No familiarity x block interaction was found, $F(2,78) = 2.009$, $p = .141$, indicating that infants' preference did not change across blocks. Planned analyses for each block revealed no effect of familiarity for Block 1, $t(39) = -1.554$, $p = .128$ (familiar: $M = 8.06$, $SD = 3.70$; unfamiliar: $M = 9.10$, $SD = 3.63$), Block 2, $t(39) = -.078$, $p = .938$ (familiar: $M = 7.04$, $SD = 3.43$; unfamiliar: $M = 7.09$, $SD = 3.82$), or Block 3, $t(39) = 1.018$, $p = .315$ (familiar: $M = 6.20$, $SD = 3.63$; unfamiliar: $M = 5.62$, $SD = 3.01$), demonstrating that from the beginning infants showed no preference for the familiar passages.

One possibility for these null results might be that half the infants are showing a novelty preference, while the other half are showing a familiarity preference, in which case the preference of one group would nullify the preference of the other. This

did not however, prove to be the case, since the distribution of infants' difference scores (Figure 3) is unimodal and centered on 0.

Figure 2: 9-month-olds' listening times to familiar and unfamiliar passages block by block collapsed across conditions

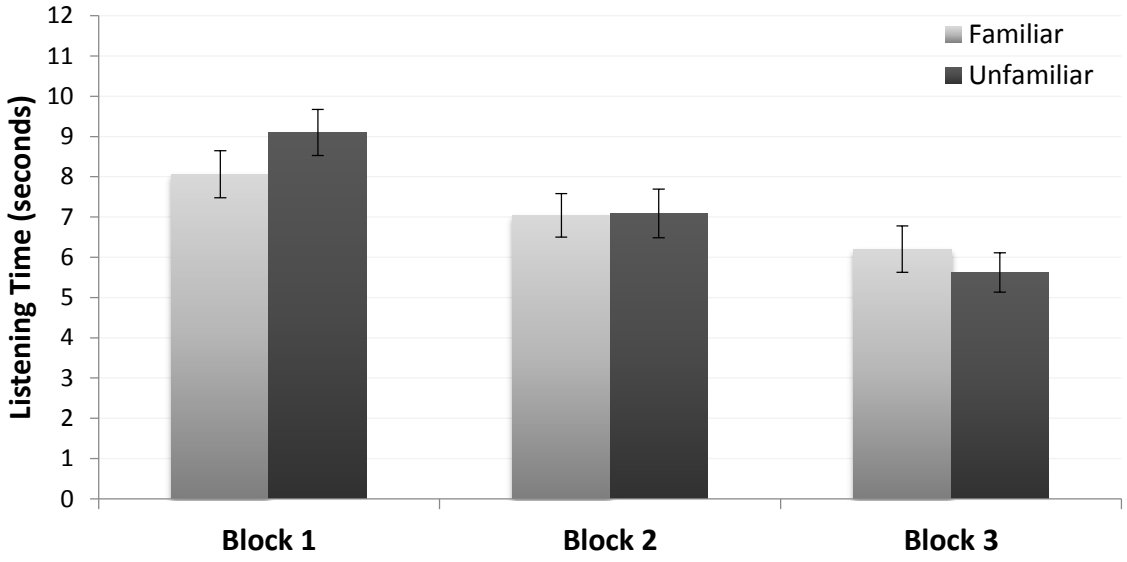
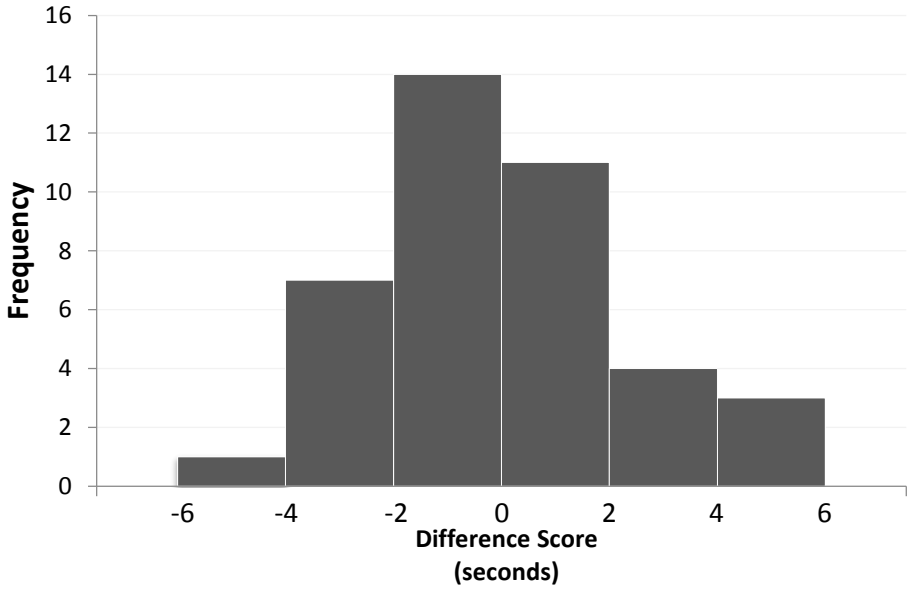


Figure 3: Histogram of 9-month-olds' difference scores for which a novelty preference is negative and a familiarity preference is positive



Given these null results, it was decided to test somewhat older infants, to see if they would segment the target speech.

Experiment 2

Experiment 2 is a replication of the previous experiment with 12-month-olds.

Method

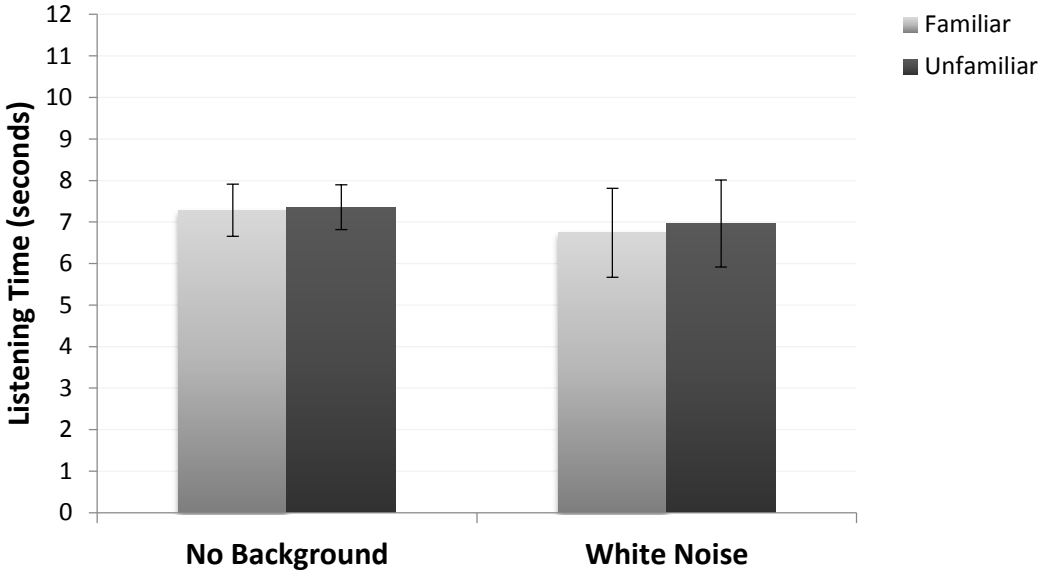
Participants. A total of 18 healthy full-term infants participated in this experiment; 12 in the no background condition (9 M, 3 F) and 6 in the white noise condition (2 M, 4 F). Infants were between 345 and 381 days old, or 11 months 26 days on average. An additional 12 infants were discarded due to, signs of their unwillingness to participate such as fussiness/crying/sleepiness of the infant (n = 9), parent request (n = 2), and external noise (n = 1).

Stimuli & Design. Same as Experiment 1.

Results & Discussion

Infants listened to familiar passages for 7.28 seconds (SD = 2.18), and unfamiliar passages for 7.36 seconds (SD = 1.86) when there was no background noise; they listened to familiar passages for 6.74 seconds (SD = 2.62), and unfamiliar passages for 6.96 (SD = 2.56) when the background was white noise (Figure 4). A 2 (familiarity) x 2 (background) mixed model ANOVA was conducted with familiarity as the within subjects factor. Just as with the 9-month-olds, there was no effect of familiarity, $F(1,16) = 0.077$, $p = .785$, background, $F(1,16) = 0.233$, $p = .636$, or familiarity x background interaction, $F(1,16) = 0.019$, $p = .892$. As with the 9-month-olds, further analyses were performed collapsing across conditions.

Figure 4: 12-month-olds' listening times to familiar and unfamiliar passages with and without the presence of white noise at a 10 dB SNR



As with Experiment 1, data was further analyzed block by block (Figure 5). A 2 (familiarity) x 3 (block) repeated measures ANOVA was conducted. Again, no familiarity x block interaction, $F(2,34) = 0.702$, $p = .503$, was found. Planned analyses for each block revealed no effect of familiarity for either Block 1, $t(17) = 0.688$, $p = .501$ (familiar: $M = 9.14$, $SD = 3.61$; unfamiliar: $M = 8.21$, $SD = 4.45$), Block 2, $t(17) = -1.318$, $p = .205$ (familiar: $M = 5.80$, $SD = 2.72$; unfamiliar: $M = 6.74$, $SD = 2.63$), or Block 3, $t(17) = -0.361$, $p = .722$ (familiar: $M = 6.36$, $SD = 3.82$; unfamiliar: $M = 6.73$, $SD = 3.82$). And, just like the 9-month-olds, the distribution of infants' difference scores (Figure 6) was unimodal and centered on 0.

These results, coupled with those of the Experiment 1 are quite unexpected since Jusczyk and Aslin's (1995) findings have been well replicated. Nevertheless, random looking behavior like that found here, cannot always be equated with an inability to discriminate stimuli (Houston-Price & Nakai, 2004). Random looking

behavior can also occur when familiarization times are sufficiently long to push infants into the "sweet spot" between showing a preference for familiar stimuli, and

Figure 5: 12-month-olds' listening times to familiar and unfamiliar passages block by block collapsed across conditions

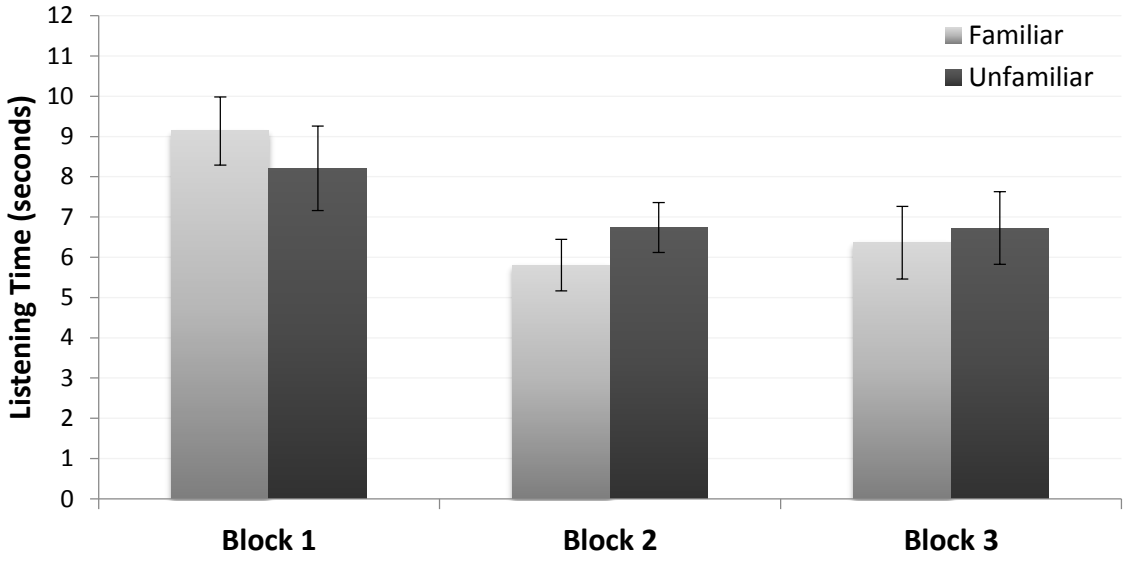
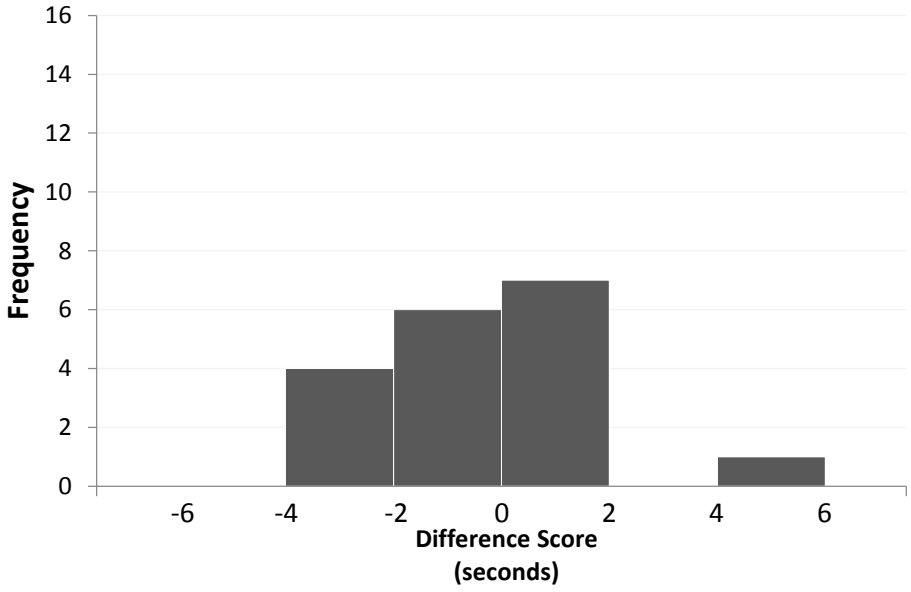


Figure 6: Histogram of 12-month-olds' difference scores for which a novelty preference is negative and a familiarity preference is positive



showing a preference for novel stimuli; something that occurs with maturation as a task becomes easier. However, previous studies with both 9- and 10½-month-olds (Houston & Jusczyk, 2000; Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000) employing the same procedure with identical familiarization criteria as Jusczyk and Aslin's, have found that infants show a preference for familiarized target words in fluent speech. It is therefore unlikely to be the case that the infants in the current study are in the "sweet spot", and more likely that the null results found here are due to this task being too difficult. Probable sources of difficulty lie with the testing room and the stimuli themselves.

With regard to the testing room itself, low-level mechanical noise present during testing may have interfered with infants' ability to perform the task. However, as will be shown in Experiment 3, infants this age are able to "ignore" white noise, a similar types of sound, in favor of attending to target speech. This doesn't rule out the possibility that starts and stops of extraneous noises during the presentation of passages could affect infants' attention. Nevertheless no extraneous distractions were evident by the experimenter for infants who completed the testing. Infants exposed to known extraneous noises were discarded, either because the experiment heard the noise or the infant displayed signs of distraction to some irrelevant corner of the testing room.

The stimuli themselves represent the most likely source of infants' difficulties. Some potential issues include, the target words not being salient to the infant, or the stimuli not being sufficiently engaging to entice the infants to perform the task. Another possible reason revolves around the position of the target word within the

first few sentences. Due to the randomized order of sentences, there is a 33% chance on each trial that infants were presented with the target word in an utterance-medial position. This position has been shown to be more difficult for infants to segment when presented with familiarization times only slightly shorter than the one used here (Seidl & Johnson, 2006). A fixed order presentation thus may not only have made the passages more consistent across trials (and more like a story), presenting them with the target word in an utterance-initial position in the first sentence, followed by an utterance-final position in the second, as Jusczyk and Aslin (1995) did, may have helped promote successful segmentation and ensure that the target word appeared during the first 2 seconds of each trial.

Given the inability to replicate Jusczyk and Aslin's (1995) finding with both 9- and 12-month olds, a different tactic was needed to test the question of interest; that of infants' ability to perceive target speech under natural multi-talker conditions.

Experiment 3

Infants have been shown to prefer single-talker speech to white noise (Colombo & Bundy, 1981), complex non-speech analogues (Vouloumanos & Werker, 2004), as well as Macaque calls, and other human vocalizations (Shultz & Vouloumanos, 2010). It is therefore likely that infants will also prefer single-talker to multi-talker speech. Infants' ability to discriminate their own name from other names in a multi-talker background (Newman, 2005; 2009) lends strength to this argument by demonstrating that infants are able to selectively ignore this kind of background and attend solely to their name.

Unlike Newman (2005; 2009), this study uses sentences produced in an infant-directed (ID) speaking style as the target speech, something less salient than one's own name, contrasting passages of noise with and without target speech. Since infants show a bias for single-talker speech (Colombo & Bundy, 1981; Newman, 2005; 2009; Shultz & Vouloumanos, 2010; Vouloumanos & Werker, 2004), if they are able to perceive target speech in background noise at a 10 dB SNR, they should show a preference for passages that contain target speech over passages that don't. Background noise used in this study are white noise and natural multi-talker adult-directed speech.

Method

Participants. A total of 32 healthy full-term infants participated in this experiment; 16 in the white noise condition (7 M, 9 F) and 16 in the multi-talker speech condition (9 M, 7 F). Infants were between 260 and 295 days old, or 9 months 4 days on average. An additional 8 infants were discarded due to signs of their unwillingness to participate such as fussiness/lack of interest of the infant (n = 7), pre-term birth (n = 1).

Stimuli. Sentences were those used in Experiments 1 and 2 with background noise consisting of either white noise or multi-talker speech (described below). To allow the inter-stimulus interval (ISI) to consist of background noise sentences were concatenated in a fixed order (as presented in Appendix) with the intervening ISI. The ISI was set to 1 second to ensure that passages were a minimum of 20 seconds in length with no sentence repetition. An additional 0.5 seconds of background noise was concatenated to the beginning and end of each passage to ensure that target

speech was surrounded by background noise. Passages consisted of either background noise alone (Noise Only passages) or background noise with target speech (Target and Noise passages). For the Target & Noise passages, sentences were overlapped with their respective background prior to concatenation. Table 2 presents a summary of the acoustic properties of the Noise Only passages. With the exception of target speech, the Noise Only passages were identical to the Target & Noise passages.

White noise. White noise clips for each sentence were those used in Experiments 1 and 2. The ISIs were generated in the same manner described in Experiment 1. The intensity of each ISI was adjusted to avoid any sudden changes in as the passage shifted from the preceding sentence to the ISI, and from the ISI to the following sentence, with the intensity of the ISI linearly rising or falling as appropriate. The mean intensity of each ISI was equal to the mean of the background sample of the preceding and following sentences. The white noise concatenated to the

Table 2: Acoustic Properties of Noise Only Stimuli used in Experiments 3 and 4

		Duration (ms)	Frequency (Hz)	Intensity (dB)
white noise passages				
	CUP	2179	- - -	61.69
	DOG	2107	- - -	58.03
	FEET	2178	- - -	5476
	BIKE	2052	- - -	58.38
multi-talker passages				
	CUP	2179	270.79	61.69
	DOG	2107	268.73	58.03
	FEET	2178	262.77	54.76
	BIKE	2052	271.91	58.38

beginning and end of each passage was adjusted to be equal to the intensity of the white noise of the first or last sentence as appropriate.

Multi-talker speech. To create the multi-talker stimuli, multiple adults and young children were recorded in a semi-naturalistic setting. To obtain this recording a 2 hour playdate with 4 mothers with their 6 young children (4 under 2 years, 2 between 3 and 5 years), 3 adult female volunteers, and the experimenter was recorded in the lab's waiting area. Additional toys were brought in and added to the lab's existing array of toys to ensure the children had enough to keep them entertained. Clips of naturalistic overlapping speech were cut from this recording for use as background speech.

A LENA (Language ENvironment Analysis) digital language processor (dlp) was used to record the playdate (www.lenafoundation.org). A dlp is a lightweight (2oz) recording device with advanced technology to record up to 16 hours of high quality audio. The dlp is placed inside a pocket of a specially designed vest worn by the child, thus providing a child's perspective of the audio environment. LENA software uploads and processes the recording, classifying segments of it into categories such as clear meaningful speech, overlapping speech, and non-verbal noise. Only segments classified by LENA as being overlapping speech (i.e. speech that is close to the child but is masked by other noise) were considered.

Segment selection was made based on number and kind of speakers (multiple adults, no infants) and amount of non-speech noise (none or minimal) occurring concomitantly for at least 4 seconds. Six segments that met this criteria were selected and extracted from the recording. Each segment served as the background for 1 of the

sentences in each passage (cup, dog, feet, bike). Thus 4 clips were generated from each segment; one for each target word.

For each of the 6 segments, the target words (cup, dog, feet, bike) of the respective sentences, occurred in the same location within the segment. The location of the target word was chosen to ensure that it did not occur during a dip in the background speech. The length of each clip was equal to their respective target sentence, resulting in each clip beginning and ending at different locations within the segment. The mean intensity of each clip was then scaled to be 10 dB less than the mean of their respective target sentence.

One additional segment of shorter duration was selected, with a single clip extracted for use as the ISI. This single clip was used in all instances. The intensity of each ISI was adjusted to be the mean of the background sample of the preceding and following sentences. No sudden changes in intensity were detectable at either ISI boundary. To create the beginning and end background noise, the first half of the ISI clip was used at the beginning of each passage (with the intensity equal to the background of the first sentence), and the second half was used at the end of each passage (with the intensity equal to the background of the last sentence).

Procedure. Due to the addition of 0.5 seconds of background noise at the beginning of the passages, the minimum trial duration during the test phase was increased by this amount to 2.5 seconds.

Design. A standard HPP was used with a 2-trial warm-up phase to instrumental music (in lieu of the familiarization phase used in Experiments 1 and 2), and a test phase with three blocks of four passages. Infants were tested on passages

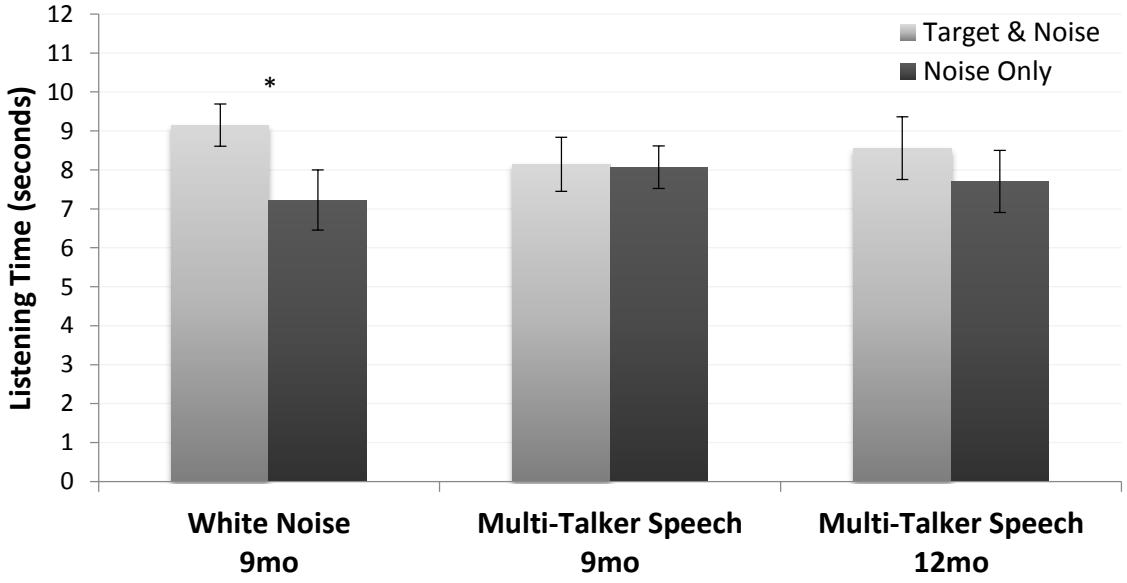
of background noise alone (noise only) and passages of background noise with overlapped target speech (target & noise) where half the infants were tested with a background of white noise, the other half with a background of natural multi-talker speech. Half the infants in each condition were exposed to *cup* and *dog* passages, the other half to *feet* and *bike* passages. Across both Experiment 3 and 4, no differences were found between infants who were exposed to the cup-dog passages and those exposed to the feet-bike passages, with no main effect of group, $F(1,46) = 0.006$, $p = .938$, and no group x passage type interaction, $F(1,46) = 2.227$, $p = .142$. Further analyses were therefore collapsed across group.

Results & Discussion

Infants listened to the target & noise passages for 9.15 seconds ($SD = 2.16$), and to the noise only passages for 7.23 seconds ($SD = 3.07$) when the background was white noise; they listened to the target & noise passages for 8.14 seconds ($SD = 2.76$) and to the noise only passages for 8.07 seconds ($SD = 2.19$) when the background was natural multi-talker speech (Figure 7). A 2 (passage type: target & noise, noise only) x 2 (background: white noise, multi-talker speech) mixed model ANOVA was conducted with passage type as the within subjects factor. Analyses revealed a marginal effect of passage type, $F(1,30) = 3.585$, $p = .068$, with no effect of background, $F(1,30) = 0.012$, $p = .912$. Crucially, there was a marginal background x passage type interaction, $F(1,30) = 3.075$, $p = .09$. Planned analyses for each background condition revealed a significant effect for passage type when the background was white noise, $t(15) = 2.577$, $p = .021$, but not when the background was multi-talker speech, $t(15) = 0.099$, $p = .922$.

One thing that makes grouping target and background into separate streams difficult is target-background similarity. The more physical characteristics that are similar across target and background, the more difficult they will be to perceptually separate (Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001). Previous infant studies where target and background were perceptually *similar* (e.g. unfamiliar target speaker with *same* gender single-talker background) have found that infants have difficulties segregating target and background (Barker & Newman, 2004; Newman, 2009). However, when the target and background were perceptually *different* (e.g. unfamiliar target speaker with *opposite* gender single-talker background, or with read-aloud multi-talker background), infants readily segregate target and background speech (Newman, 2009; Newman & Jusczyk, 1996). The natural multi-talker speech used here, in contrast to the read-aloud multi-talker speech used in

Figure 7: 9- and 12-month-olds' listening times to passages of background noise with and without the addition of target speech at a 10 dB SNR



previous studies, appears to represent a background that is perceptually *similar* to the target speech, thus impeding stream segregation. Consistent with this idea, the white noise background presents infants with target-background *dissimilarity*, thereby promoting successful segregation.

Due to the false starts, repairs, and pauses present in conversational speech, the ebbs and flows of one speaker are not always filled in by other speakers in a multi-talker situation. So, although there may be five individuals speaking, there are many pauses which can occur simultaneously across different individuals, momentarily reducing the number of speakers, and thus producing a highly variable prosodic contour. This is in stark contrast to read-aloud multi-talker speech, which maintains a consistent overlap of all speakers, and thus a more uniform prosodic contour. This would suggest that natural multi-talker speech is more like single-talker speech than read-aloud multi-talker speech, and thus the source of infants' difficulty in this condition.

However, background is not the only thing that distinguishes this experiment from Newman's (2005; 2009). The use of a highly salient and familiar word (the child's own name) may itself be a reason for infants' success with a read-aloud multi-talker speech background. Considering that saliency helps adults detect target speech in background speech (Cherry, 1953; Cherry & Taylor, 1954) it is also possible that the infants are able to perceive the presence of the target speech, but that it is insufficiently salient to garner their attention above that of the background alone. Alternately, it is possible that in Newman's previous studies infants were unable to perceive the unfamiliar names, but were able to perceive their own name since it is

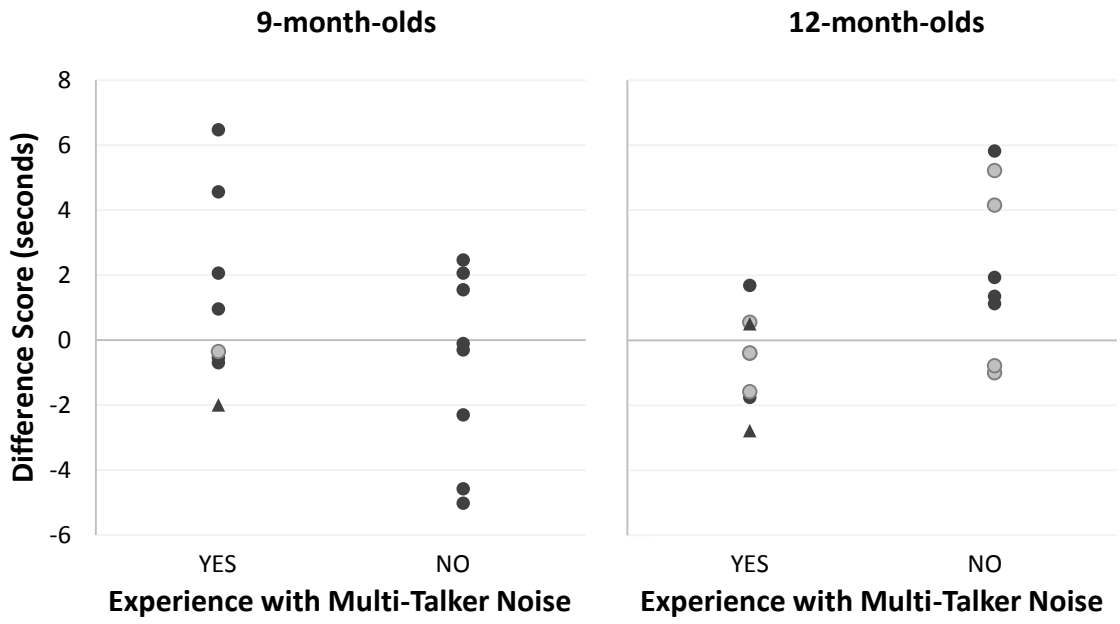
familiar to them. If this is the case, the sentences used here would be equally unperceivable by infants in both read-aloud and natural multi-talker speech. In either interpretation, only salient speech is being attended to.

In order to resolve these possibilities, infants' ability to detect their own name in this background will need to be tested. Successful segregation would indicate that the target stimuli were not sufficiently salient to gain the infants' attention, whereas a failure would indicate that target-background similarity is the source of infants' difficulties with natural multi-talker speech.

Although the 9-month-olds as a group did not show a preference for either passage type with the multi-talker speech background, it remains possible that individual differences in experience with noisy, multi-talker environments may be masked by the overall null result. Experience with these kinds of environments may provide infants with more opportunities in which to learn to separate the relevant target speech directed at them from the irrelevant background speech noise. Eight of the infants spent a large amount of time with other pre-school-age children and were therefore frequently exposed to these types of environments – 1 in child care more than 2 days per week, 7 at home with at least 1 older sibling 5 years or less (Target & Noise: M = 9.25, SD = 3.54; Noise Only: M = 7.94, SD = 2.09). The other eight infants were home alone with a parent during the week – all 8 had no siblings, 1 of which also had child care 1 day per week (Target & Noise: M = 7.47, SD = 1.21; Noise Only: M = 8.25, SD = 2.40). A 2 (passage type) x 2 (noise experience: yes, no) mixed model ANOVA was conducted with passage type as the between subjects factor.

It was predicted that infants with frequent exposure to noisy environments would succeed at segregating target and background speech (and thus show a preference for passages with target speech), whereas those with little or infrequent experience would not show a preference. This however did not turn out to be the case (Figure 8), as there was no experience x passage type interaction, $F(1,14) = 2.060$, $p = .173$. Although difference scores (mean looking time for target & noise passages minus mean looking time for noise only passages) indicate a potential trend in the predicted direction, 9-month-olds with regular exposure to noisy environments did not perform any differently than those who did not have this experience. No other effects were significant, $ps > .05$.

Figure 8: Individual differences with the multi-talker speech background based on experience with noisy of environments



Note: [1]a positive difference score denotes a preference for target & noise passages, a negative difference score denotes a preference for noise only passages
[2]triangles denote infants whose noise experience stems from formal child care, light colored circles denote those with school-age siblings, dark circles denote all other infants

Overall, 9-month-olds were unable to perceive target speech under natural multi-talker background conditions, regardless of noise experience, despite succeeding with a white noise background. To see if older infants would succeed under these conditions, 12-month-olds were tested with the multi-talker background.

Experiment 4

Experiment 4 is a replication of the natural multi-talker condition of the previous experiment with 12-month-olds.

Method

Participants. A total of 16 healthy full-term infants participated in this experiment (10 M, 6 F). Infants were between 354 and 383 days old, or 12 months 5 days on average. One additional infant was discarded due their lack of interest in participating in the study (n = 1).

Stimuli, Procedure & Design. Same as Experiment 3.

Results & Discussion

Infants listened to the target & noise passages for 8.56 seconds (SD = 3.22) and to the noise only passages for 7.70 seconds (SD = 3.18; Figure 7). Overall the 12-month-olds did not show a preference for either passage type, $t(15) = 1.379$, $p = .188$, however the difference was in the predicted direction. It is possible that with a larger sample size a significant difference will emerge. Irrespective of this outcome, it is likely that even older infants will succeed at separating target speech from this background.

Despite showing no preference overall, just as with the 9-month-olds, experience with noisy environments is expected to affect infants' responses. Eight of

the infants spent a large amount of time with other pre-school-age children and thus have more experience with noisy environments – 2 in child care, 6 with at least 1 older sibling 5 years or less (Target & Noise: $M = 7.10$, $SD = 2.10$; Noise Only: $M = 7.62$, $SD = 2.40$). The other eight infants were home with a parent during the week – 3 had no siblings, 4 had only school-age siblings, and 1 had school-age siblings and a newborn 2-week-old sibling (Target & Noise: $M = 10.01$, $SD = 3.59$; Noise Only: $M = 7.78$, $SD = 3.99$). A 2 (passage type) x 2 (noise experience) mixed model ANOVA was conducted with passage type as the within subjects factor. As with the younger infants, it was predicted that infants with more noise experience would segregate target and background speech showing a preference for passages with target speech, whereas those without extensive experience would not. Analyses indicate a significant experience x passage type interaction, $F(1,14) = 6.821$, $p = .021$ (Figure 8). Planned post-hoc analyses reveal that, counter to what was predicted, infants who had little experience with noisy environments showed a preference for passages with target speech, $t(7) = 2.429$, $p = .045$, whereas those who spent a lot of time with other pre-school children did not, $t(7) = -1.008$, $p = .347$. No other effects were significant, $ps > .05$.

Although it is possible that only those without extensive experience with noisy environments are capable of segregating target and background speech, this is unlikely to be to be the case. Another, more probable, alternative is that 12-month olds in both groups can segregate the two speech streams, but only those *without* noise experience show a preference. In this case, 12-month-olds in frequent proximity to other pre-school aged children (who are also spoken to in an ID speech

register) show little interest in attending to generic ID speech; in other words, they have learnt that not all ID speech is for them. Those with little or no experience with other pre-school aged children on the other hand, would not have had this opportunity, since the ID speech in their environment would only be directed at them.

This is particularly evident when you consider those who have only school-aged siblings (Figure 8, light circles). One possibility is that their sibling is at school and therefore has little impact on the amount of overlapping speech noise in the infants' environment. However, by the time children reach 5 years, maternal speech to them has already shifted towards a more adult-directed (AD)-like register (Amano, Nakatani, & Kondo, 2006; Liu, Tsao, & Kuhl, 2009; Warren-Leubecker & Bohannon, 1984), such that by the time they enter school full-time, one would expect full AD-like speech to be directed at them. Infants with both school and pre-school aged siblings (light circles in the Yes group) will therefore have the opportunity to habituate to the presence of irrelevant ID speech in their environment due to their younger sibling. Those with only school aged siblings (light circles in the No group) on the other hand would have no such experience, since they are the only pre-school-aged child being spoken to.

Given the data at hand, it appears then, that experience with overlapping speech is not what is driving these effects, but experience with ID speech directed at other children in the environment. Thus, 12-month-olds with extensive experience listening to ID speech being directed at other children have grown accustomed to (habituated to) its presence in their environment, needing some indication that the speech is being directed at them before preferentially attending to it.

Finally, individual differences with noise experience were examined across ages with a 2 (passage type) x 2 (experience) x 2 (age: 9-month, 12-month) mixed model ANOVA with passage type as the within subjects factor. The analysis revealed a significant 3-way interaction, $F(1,28) = 7.257$, $p = .012$, indicating that the pattern of results for the 12-month-olds differed significantly from that of the 9-month-olds, and therefore represent a developmental trend. No other effects were significant across age groups, all ps > .05.

Overall, only 12-month-olds without noise experience showed a preference for passages with target speech under natural multi-talker conditions. Those with noise experience showed no such preference, possibly due to being habituated to irrelevant ID speech in their environment. This pattern differed from that of the 9-month-olds who showed no preference regardless of noise experience.

General Discussion

Despite the failure to replicate Jusczyk and Aslin's (1995) findings, a change in procedure allowed the investigation of the central question of whether infants would be able to detect target speech among a background of naturally derived multi-talker babble. With this new procedure, both 9- and 12-month-olds failed to show a preference for passages containing target speech in a natural multi-talker background. Since 9-month-olds succeeded in a white noise background at the same SNR, energetic masking could be eliminated in favor of informational masking, in particular target-masker similarity, as the source of infants' difficulty. There were also no individual differences found for the 9-month-olds, while for the 12-month-

olds with little or no experience with speech noise showed a preference for passages with target speech, while those with experience did not.

It is however important to explore these finding further by increasing the number of participants in each condition, thereby increasing power and reducing the chance of error. Despite this, the primary results of interest – infants' ability to perceive speech in noise – do appear to be solid and likely to hold up with a larger sample size. In addition to increasing sample size, a third condition examining infants' ability to perceive their own name in a natural multi-talker background is needed to determine whether these results are due to target-background similarity or to the target speech not being sufficiently salient to induce a cocktail party effect.

Although the procedure appears to works well based on the white noise condition, it is still possible (though unlikely) that it could fail under conditions of background speech noise. In particular, one cannot rule out the possibility that the infants' apparent difficulty with the natural multi-talker speech was due to a lack of preference between target and background speech, and not to an inability to segregate them. Although there is research indicating that infants prefer single-talker speech to other human vocalizations (Shultz & Vouloumanos, 2010), infants' preference regarding single- versus multi-talker speech is indirect (Newman, 2005; 2009). Further investigation into infants' ability to perceive their own name in a natural multi-talker background will help resolve this matter.

Setting aside methodological issues, both 9- and 12- month olds appear to be unable to segregate target speech from natural multi-talker background speech at a 10 dB SNR. If this, and similar findings with single-talker backgrounds, is due to

similarity in prosodic characteristics between target and background, then it is likely that background speech in a different language may help infants group them into separate streams as it does for adults (Van Engen & Bradlow, 2007). French (a syllable-timed language) and Japanese (a mora-timed language) are prosodically different from English (a stress-timed language) which could provide English-learning infants with the cues they need to separate English target speech from foreign single-talker background speech. Dutch, a stress-timed language, prosodically similar to English, could give English-learning infants difficulties if they are relying on prosody to separate target and background speech. If, however they also use more basic phonetic information, then the foreign Dutch sounds and its foreign phonetic pattern could provide infants with enough cues to separate out English target speech. Alternatively, if infants' difficulties lie not with prosody, or phonotactics but with other more language-general characteristics of speech, such as spectral content, then the language of background speech would make no difference.

On a similar note, it is also possible that child speech, with its distinct physical properties, will be treated differently than female adult speech. In other words, differences in pitch and formant frequencies between child speech and female adult speech should provide infants with enough cues to successfully segment target speech. This would also mean that experience with this child background speech, which occurs with siblings and daycare, would not be the same as experience with female adult background speech. It remains possible that infants presented with multi-talker child speech as a background would show a distinction between those

with and without sibling/daycare experience, if not wholly succeeding, at segmenting target speech from this background.

Regardless of the precise mechanism for infants' lack of stream segregation, being able to distinguish phonetic details is a crucial element for learning the sound patterns of words and their associated meanings; without which ball, tall, and doll become indistinguishable sound patterns referring simultaneously to three very different things. Natural multi-talker speech, just like same gender single-talker speech, present infants with just this kind of situation. The overlapped target and background speech, appears to mesh together resulting in two indistinguishable streams, thereby rendering phonetic information of the target speech indistinguishable, and language learning difficult to impossible.

Although the overlapping speech samples used in this study represent only a small fraction of the possible scenarios encountered by infants in their daily lives, their apparent lack of ability to segregate target infant-directed speech under ecologically valid multi-talker background conditions demonstrates the limitations of infants' language learning abilities. Given that these types of scenarios are frequently encountered by infants and toddlers (Bernier & Soderstrom, 2013), this means that infants are learning the language of their communities with a more limited input than one would expect, making language acquisition all the more impressive.

References

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *Journal of the Acoustical Society of America, 119*, 1636-1647.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

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

Batliner, A., Kompe, R., Kießling, A., Nöth, E., & Niemann, H. (1995). Can you tell apart spontaneous and read speech if you just look at prosody? In A. J. Rubio Ayuso, & J. M. López Soler, *Speech Recognition and Coding: New Advances and Trends* (pp. 101-104). Berlin: Springer.

Bernier, D. E., & Soderstrom, M. (2013). Clarity of language input to toddlers across childcare settings. *LENA international conference.* Denver, CO.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14*, 359-375.

Boersma, P., & Weenink, D. (2011, March). *Praat: doing phonetics by computer (Version 5.2.x) [computer program].* Retrieved March 2011, from http://www.praat.org/

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298-304.

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

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., 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.

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.

Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 26*, 554-559.

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

Colombo, J., Frick, J. E., Ryther, J. S., Coldren, J. T., & Mitchell, D. W. (1995). Infants' detection of analogs of "Motherses" in noise. *Merrill-Palmer Quarterly, 41*, 104-113.

Cooke, M., Lecumberri, M., & 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.

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition, 27*, 328-338.

Guaïtella, I. (1999). Rhythm in speech: What rhythmic organizations reveal about cognitive processes in spontaneous speech production versus reading aloud. *Journal of Pragmatics, 31*, 509-523.

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

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

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language segmentation by 9-month-olds. *Psychonomic Bulletin & Review, 7*, 504-509.

Houston, D. M., Santelmann, L. M., & Jusczyk, P. W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and cognitive processes, 19*, 97-136.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity in infants preference procedures. *Infant and Child Development, 13*, 341-348.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication, 10*, 163-169.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: noise judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 240-247.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548-567.

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., Hohne, E. A., & Bauman, A. (1999). Intants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61*, 1465-1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

Laan, G. P. (1997). The contributions of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication, 22*, 43-65.

Levin, H., Schaffer, C., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language andspeech, 25*, 43-54.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language, 36*, 909-922.

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

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring continuous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27*, 644-655.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactics and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

Mertus, J. (2011, March). *Welcome to Bliss: The Brown Lab Interactive Speech System (version 9.2.1) [computer program].* Retrieved March 2011, from http://mertus.org/Bliss/index.html

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*, 105-129.

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. (2009). Infants' listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics, 71*, 822-836.

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

Newman, R. S., & Morini, G. (2010). Infants' ability to recognize speech in the presence of amplitude-modulated background noise. *Journal of the Acoustical Society of America, 127*, 1905-1905.

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

Nozza, R. J., Miller, S. L., Rossman, R. 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. 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. 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.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 2-46.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy, 17*, 198-232.

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

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.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisisted: edge alignment facilitates target extraction. *Developmental Science, 9*, 565-573.

Shultz, S., & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development, 6*, 241-257.

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.

Smith, N. A., & Trainor, L. J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distractors. *Infancy, 16*, 655-668.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249-267.

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

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication, 52*, 943-953.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America, 121*, 519-526.

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

Warren-Leubecker, A., & Bohannon, J. N. (1984). Intonation pattern in child-directed speech: mother-father differences. *Child Development, 55*, 1379-1385.

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

Appendix

Target Stimuli

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