

BRIEF RESEARCH REPORT

# Was that my name? Infants' listening in conversational multi-talker backgrounds

Dana E. BERNIER<sup>1\*</sup> and Melanie SODERSTROM<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Waterloo, 200 University Avenue West (PAS 3020), Waterloo, Ontario N2L 3G1 and <sup>2</sup>Department of Psychology, University of Manitoba, P404 Duff Roblin Bldg, 190 Dysart Rd, Winnipeg, Manitoba R3T 2N2

\*Corresponding author. E-mail: [dbernier@uwaterloo.ca](mailto:dbernier@uwaterloo.ca)

(Received 10 September 2017; revised 7 February 2018; accepted 31 May 2018;  
first published online 17 July 2018)

## Abstract

This study tested infants' ability to segregate target speech from a background of ecologically valid multi-talker speech at a 10 dB SNR. Using the Headturn Preference Procedure, 72 English-learning 5-, 9-, and 12-month-old monolinguals were tested on their ability to detect and perceive their own name. At all three ages infants were able to detect the presence of the target speech, but only at 9 months did they show sensitivity to the phonetic details that distinguished their own name from other names. These results extend previous findings on infants' speech perception in noise to more naturalistic forms of background speech.

**Keywords:** infant speech perception; speech in noise; infant-directed speech

Over half the speech heard by infants in real-world environments is overlapping with other sounds (Bernier & Soderstrom, 2014). In these conditions, infants must perceptually group the sounds from one source (the target) separately from the other (the masker or background). Infants are faced with two separable problems – *ENERGETIC MASKING*, in which the background interferes directly (i.e., acoustically) with the target (for example, a broad spectrum white noise that directly interferes with a speech signal), and *INFORMATIONAL MASKING*, in which the presence of the competing signal provides perceptual and/or attentional interference (e.g., Leek, Brown, & Dorman, 1991; for example, an interesting background conversation that prevents the listener from attending to an important target signal). Multiple processes for disentangling the signal from the background are therefore involved. Some of these processes are signal-driven (bottom-up) and may already be present in early infancy (Smith & Trainor, 2011), while others are knowledge-driven (top-down) and subject to learning.

Infants are further challenged by having an auditory system that is not yet fully developed. They need a higher intensity for the detection of masked and unmasked pure tones and broadband noise bursts compared to adults (Nozza & Wilson, 1984; Sinnott, Pisoni, & Aslin, 1983; Werner & Boike, 2001). Accordingly, infants' ability to detect speech is inferior to that of adults in both quiet (Nozza, Wagner, &

Crandell, 1988) and noise (Trehub, Bull, & Schneider, 1981), as is their ability to discriminate speech sounds (Nozza, Miller, Rossman, & Bond, 1991; Nozza, Rossman, & Bond, 1991; Nozza, Rossman, Bond, & Miller, 1990).

Although much has been learned 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' ability to perceive speech under noisy conditions. With older infants and toddlers, studies have shown that background speech impacts their ability to recognize the intended referent of a familiar label, with recognition decreasing as the intensity of the target speech drops from 5 dB louder than background speech to 5 dB quieter (Newman, 2011). And additionally, their ability to learn novel words is also impacted by background speech, even with different-sex background speech at a signal-to-noise ratio (SNR) as high as 5 dB (McMillan & Saffran, 2016). With younger infants, Newman and Jusczyk (1996) first showed that 7.5-month-olds were able to perceive target speech against a speech background with an SNR of as little as 5 dB in a speech segmentation task when the speakers were different sexes. However, infants failed to discriminate with an SNR as high as 10 dB when the speech involved same-sex speakers who were unfamiliar to the child (Barker & Newman, 2004).

Other studies have examined infants' recognition of their own name against a speech background; something they can do in quiet settings as early as 4.5 months (Mandel, Jusczyk, & Pisoni, 1995). For adults, one's own name is well known to play a unique role in inducing the so-called cocktail party effect, where salient speech pops out against a background (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 phonetic details that distinguish individual phonemes (e.g., *Robby* vs. *Bobby*), in addition to the word's stress pattern (e.g., *reCORD* vs. *REcord*).

Newman (2005, 2009) examined infants' ability to recognize their own name against a speech background consisting of one or many female voices reading aloud from a book. Infants as young as 5 months were able to discriminate their own name (e.g., *JACob*) from both stress-matched (e.g., *KAYden*) and non-stress-matched (e.g., *miCHELLE*) foil names against a multi-voice background at a 10 dB SNR. However, when the background consisted of a single speaker, the infants did not show discrimination. This is the opposite of what is found with adults. To succeed in single-speaker backgrounds, adults focus their attention on the 'dips' in the background speech, using their lexical knowledge to fill in the missing information (Miller, 1947; Pollack & Pickett, 1958). For infants, their poor selective attention and limited lexical knowledge may not allow them to take advantage of these dips when listening in single-talker backgrounds. By contrast, multi-talker speech may minimize the effects of attention capture because no single speaker may be discriminable by the infant from within the babble.

This body of work has provided important information about the limits of infant speech perception in noise under ideal, controlled conditions. However, in these prior studies, background speech consisted of read-aloud speech which is different in important ways from the conversational speech infants experience every day. This is due to differences in intonation and rhythm across the two speaking styles that remains discernable even when the speech is filtered (Levin, Schaffer, & Snow, 1982). A key component of what makes these two speaking styles different is the 'on-line' nature of spontaneous speech that results in disfluencies (fillers, hesitations,

repetitions, false starts, repairs), and incomplete sentences, as well as longer and more frequent pauses (Blaauw, 1994). Such differences result in a distinct combination of pitch contours, segmental durations, and spectral features between the two speaking styles (Batliner, Kompe, Kießling, Nöth, & Niemann, 1995; Laan, 1997).

Therefore, the prosodic realization of naturally occurring spontaneous speech will differ substantially as a form of background noise from the read-aloud speech used in previous studies. The speech of each individual in a conversational multi-talker situation is likely to be dynamic in pitch, intensity, and rhythm, with many pauses which will affect the prosody of the overlapped speech sample by allowing individual speakers to stand out momentarily. On the one hand, this more dynamic background may reduce the effects of energetic masking. On the other, informational masking may become more significant, as more dynamic speech may provide attentional capture that is absent from typical laboratory-generated background speech.

We therefore performed a conceptual replication of the Newman (2005, 2009) studies, but with a more naturalistic background, to examine whether the infants would perform similarly under the prosodic conditions of naturalistic speech. As with the studies described above, target speech took one of three forms: the child's own name, a name of equal length with the same stress pattern, and a name of equal length with a different stress pattern. In addition, we included a background only control, which allowed us to ask whether infants' ability to identify speech against a naturalistic background is limited to their own name.

To address these questions, we first examined whether infants could detect the presence of speech in the form of name repetitions against a naturalistic speech background (infants' listening behavior to the three name conditions compared with the background only condition). We next examined, as per the Newman (2005, 2009) studies, at which age infants begin to recognize their own name as distinct from the other names, and how prosodically and phonetically fine-grained this recognition is (own name versus stress-matched name and own name versus non-matched name). Finally, we examined the extent to which the ability to detect name repetitions can be found for names other than their own (stress-matched and non-matched name conditions versus background only).

## Method

### *Participants*

A total of 72 healthy full-term Canadian English-learning infants from a moderately large city and surrounding area in south-central Canada, with no more than 10% exposure to other languages or dialects of English in the home (via parent report), participated in this experiment. There were twenty-four 5-month-olds (17 M, 7 F; mean 157 days, range 138–175), twenty-four 9-month-olds (15 M, 9 F; mean 273 days, range 254–289), and twenty-four 12-month-olds (16 M, 8 F; mean 361 days, range 344–380) who participated between March 2014 and October 2016. Participant recruitment was done through provincial health records via a blind mailing process, community events, online advertisement, or the lab's existing database of volunteer participants. An additional four 5-month-olds and two 12-month-olds were excluded due to crying ( $n=3$ ), and fussiness or lack of interest ( $n=3$ ). Decisions regarding participant exclusion were made prior to examining the data.

### *Stimuli*

Stimuli consisted of a string of multiple tokens of a name overlapped with conversational multi-talker speech at an SNR of 10 dB. Passages were created such that the name occurred at regular intervals (every 1.5, 1.6, or 1.7 seconds depending on number of syllables in the name), with an additional 0.5 seconds at the beginning and end of each passage. Background speech occurred continuously throughout each passage, with the same selection used across all passages. Final passage length ranged from 20.4 to 21.4 seconds ( $M = 21.1$ ). Sample passages with common names can be found at <<https://osf.io/6qf6y/>> (to preserve participant confidentiality we have not made the full stimuli publicly available).

### *Background multi-talker speech*

To create the multi-talker background stimuli, multiple adults and young children were recorded in a semi-naturalistic setting with the goal of recording overlapped adult speech that a young infant would typically be exposed to. To obtain this recording a 2-hour playdate with four mothers and their six young children (four under 2 years, two between 3 and 5 years), and four adult female lab members was recorded in the lab's waiting area. 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 (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011). 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.

Clip selection was made based on the competing requirements of speech samples that were naturalistic but also sufficiently consistent across the sample. We therefore selected clips in which there was clear multi-talker babble throughout, no detectable child noises, and no or minimal detectable non-speech noise occurring concomitantly for at least 4 seconds. Although some 'pop out' occurred in which an individual speaker's voice could be detected by the careful adult listener, this was restricted to speech perceived to be adult-directed and was at a much lower perceived signal-to-noise ratio than the target speech. Note that, by design, non-speech noise, child vocalization, and child-directed speech may well have contributed to the underlying multi-talker babble but were not detectable by the adult listener.

Six clips that met these criteria were extracted from the full recording, adjusted to be of equal intensity, then concatenated to form a single background noise sample 21.5 seconds in duration. We used PRAAT's (Boersma & Weenink, 2014) TO PITCH and TO INTENSITY functions (with a pitch floor of 75 Hz and ceiling of 600 Hz) to assess the background stimuli's acoustic features. Across the full duration, the background speech had a mean  $f_0$  of 261.2 Hz ( $SD = 84.3$ ), and mean intensity of 62.84 dB ( $SD = 3.63$ ). Intensity ranged from 8.73 to 74.53 dB.

### *Names*

Target stimuli were recorded using 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. Twelve to fourteen tokens were selected depending on the number of syllables in the name (12 for 4-syllable names, 13 for 3-syllable names, 14 for 1- and 2-syllable names). Using PRAAT scripts (Boersma & Weenink, 2014), token intensity was adjusted as necessary so that it fell within 2 SD of the other tokens in the same passage, and was 8–12 dB louder than the background speech it occurred with. The mean token SNR for each passage ranged from 9.63 to 10.51 dB ( $M = 10.07$ ). As with previous studies (Newman, 2005, 2009), mean intensity of the background speech for the full passage was 10 dB lower than the mean intensity of the string of names concatenated without pauses. The final string of names (with pauses) was overlapped with the background speech using Bliss 9's AudioTransmuter (Mertus, 2011). Mean intensity of the final passages ranged from 68.61 to 74.65 dB ( $M = 71.45$ ).

### Design

The Headturn Preference Procedure (HPP) was used with a 2-trial warm-up phase to instrumental music, followed by a test phase with three blocks of four passages. Infants were tested on passages of background noise alone (Background Only) and passages of background noise with overlapped target speech consisting of a multi-token string of one of three names – the child's own name (Own Name), a name with the same stress pattern as the child's (Stress-Matched Name), or a name with a different stress pattern than the child's (Non-Matched Name). Order of presentation of each passage within a block was randomized for each participant. A sample list of the names used is presented in Table 1. There were no differences in overall passage intensity or mean token SNR across the three names used ( $F_s < 1$ ).

### Apparatus

Infants were tested in a small dimly lit room with three 19-inch LCD flatscreen monitors embedded 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. The caregiver listened to masking music through circumaural, aviator-style headphones. The experimenter, situated in another room, observed the infant's looking behavior via a closed-circuit video feed hidden from the infant's view.

**Table 1.** Sample Name Pairings

Own	Stress-Matched	Non-Matched
Zach	Drew	Niko
Jacob	Kayden	Michelle
Matthew	Henry	Simone
Amy	Mason	Danielle
Oliver	Madelyn	Ricardo
Dominik	Christopher	Aleeya
Isabella	Jeremiah	Veronica

### Procedure

Each trial started with a flashing yellow circle on the front screen to obtain the infant's attention to center. Once the infant was looking to the front screen, a flashing yellow circle appeared on one of the side screens; left versus right presentation was randomized across trials. Once the infant looked to that screen, the flashing yellow circle was replaced with a colorful checkerboard pattern and audio files began playing from the speaker associated with that side. Whenever the infant looked away from the screen, the audio stopped playing, resuming if the infant looked back. Trials ended when either the stimulus passage had played through to the end or the infant had looked away for more than 2 consecutive seconds. Trials in which an infant looked for less than 1 second in total during warm-up, and 2 seconds in total during testing, were immediately repeated and the data from the first trial attempt excluded from analysis. A minimum 1-second pause occurred between trials. The dependent measure was the time the infant spent looking at the side monitor while the audio was playing.

### Results

Overall, mean listening times increased with age. However, the magnitude of this increase differed across the four Stimulus Conditions (Figure 1). When no target speech was present (Background Only) there was a minimal increase in listening time across the three age groups (5-months:  $M = 7.79$  s [ $SD = 2.67$ ]; 9-months:  $M = 8.18$  s [ $SD = 3.30$ ]; 12-months:  $M = 8.90$  s [ $SD = 3.39$ ]). But, when target speech was present, the increase in listening time was larger, and differed by condition. When hearing their Own Name, 5-month-olds listened for 9.42 s ( $SD = 3.27$ ), while 9-month-olds listened for 12.18 s ( $SD = 3.88$ ), and 12-month-olds for 12.84 s ( $SD = 4.55$ ). Whereas when they heard a foil name with the same stress pattern (Stress-Matched Name), 5-month-olds listened for 8.68 s ( $SD = 2.08$ ), 9-month-olds for 9.97 s ( $SD = 3.81$ ), and 12-month-olds for 11.07 s ( $SD = 4.38$ ). And when they heard a foil name with a different stress pattern (Non-Matched Name), 5-month-olds listened for 9.37 s ( $SD = 3.01$ ), 9-months for 10.61 s ( $SD = 4.10$ ), and 12-month-olds for 12.35 s ( $SD = 5.08$ ).

A three (Age)  $\times$  four (Stimulus Condition) mixed ANOVA revealed main effects of Condition ( $F(3,207) = 18.83$ ,  $p < .001$ ,  $\eta_p^2 = .214$ ) and Age ( $F(2,69) = 4.410$ ,  $p = .016$ ,

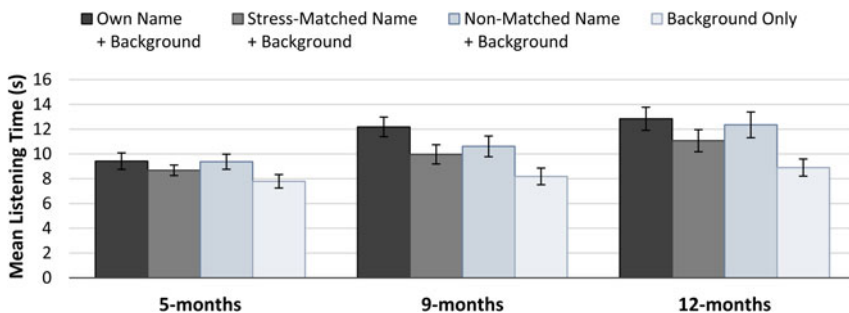


Figure 1. Mean listening times to passages with and without target speech for 5-, 9-, and 12-month-olds. Error bars represent standard error of the mean.

$\eta_p^2 = .113$ ), but no interaction between them ( $F(6,207) = 1.227$ ,  $p = .294$ ,  $\eta_p^2 = .034$ ). Having established that infants showed differential preferences for the stimulus types, we examined our specific research questions with a series of four planned contrasts, applying a Bonferroni correction to give an adjusted  $\alpha$  of .0125 for all statistics reported below.

### *Planned contrast 1: detection of speech against a naturalistic background*

We first examined whether infants were able to detect the presence of the target speech (i.e., the names) against the naturalistic background. The three conditions that contained a name (Own Name, Stress-Matched Name, Non-Matched Name) were contrasted with the Background Only condition, with age as a between-subjects factor. For this contrast, there was a significant effect of Condition ( $F(1,69) = 50.303$ ,  $p < .001$ ,  $\eta_p^2 = .422$ ), but no interaction with Age ( $F(2,69) = 2.553$ ,  $p = .085$ ,  $\eta_p^2 = .069$ ). As an additional confirmatory analysis, we ran separate analyses at each age group. All three were significant (5-months:  $F(1,23) = 10.546$ ,  $p = .004$ ,  $\eta_p^2 = .314$ ; 9-months:  $F(1,23) = 14.809$ ,  $p = .001$ ,  $\eta_p^2 = .392$ ; 12-months:  $F(1,23) = 27.186$ ,  $p < .001$ ,  $\eta_p^2 = .542$ ). Thus, all three age groups were able to detect the presence of the names against a naturalistic speech background.

### *Planned contrasts 2 and 3: discrimination of own name versus non-matched and stress-matched foil names*

Having determined that infants at all three ages could detect the presence of the target speech, we next turned to determining the nature of their perception. As per Newman (2005, 2009), we conducted separate comparisons between the infants' own name and the two foil names. While the non-matched name differed on both stress and phonetic details, the stress-matched name differed only on phonetic details. Based on prior studies, we expect infants to show a preference for their own name compared to both the stress-matched and the non-matched foils if the phonetic details are perceivable. However, if only prosodic information is perceivable, they should only discriminate their own name from the non-matched foil. Finally, if neither of these cues are perceivable, they should not show a preference for their own name compared with either foil.

For the contrast with the Non-Matched Name, there was neither a main effect of Condition ( $F(1,69) = 2.053$ ,  $p = .156$ ,  $\eta_p^2 = .029$ ), nor an interaction with Age ( $F(2,69) = 0.849$ ,  $p = .432$ ,  $\eta_p^2 = .024$ ). Separate analyses for each age group also did not reach significance ( $ps \geq .027$ ,  $\eta_p^2s \leq .196$ ). For the contrast with the Stress-Matched Name, there was a significant effect of Condition ( $F(1,69) = 11.598$ ,  $p = .001$ ,  $\eta_p^2 = .144$ ), but no interaction with Age ( $F(2,69) = 0.888$ ,  $p = .416$ ,  $\eta_p^2 = .025$ ). However, in separate analyses at each age, the 9-month-olds ( $F(1,23) = 13.517$ ,  $p = .001$ ,  $\eta_p^2 = .370$ ), but neither the 5-month-olds ( $p = .304$ ,  $\eta_p^2 = .046$ ) nor the 12-month-olds ( $p = .100$ ,  $\eta_p^2 = .113$ ), reached significance. However, upon post-hoc inspection, two of the 12-month-olds showed a substantially larger difference score than the others ( $\geq 1.92$  SD below the mean vs.  $0.95$  SD below to  $1.65$  SD above the mean for the remaining participants). Excluding these two outliers resulted in a significant effect ( $p = .005$ ,  $\eta_p^2 = .318$ ). Application of this post-hoc rule more broadly did not alter the conclusions of any other analyses. This strongly suggests that the original null result at 12 months was a Type II error. Thus, contrary to the possible

outcomes articulated above, by 9 months infants only showed a significant preference for their own name when the foil name was SIMILAR to their own (i.e., stress-matched), but not when the foil name differed in both phonetic content and stress pattern. We will return to this unexpected finding in the 'Discussion'.

#### *Planned contrast 4: detection of an unfamiliar name against a naturalistic background*

We next asked whether infants' detection of their own name was driving the observed preference in our first contrast. To assess whether infants show a preference for target speech signals aside from their own name, the two conditions that contained an unfamiliar name (Stress-Matched Name, Non-Matched Name) were contrasted with the Background Only condition. There was a significant effect of Condition ( $F(1,69) = 29.272$ ,  $p < .001$ ,  $\eta_p^2 = .298$ ), but no interaction with Age ( $F(2,69) = 1.444$ ,  $p = .243$ ,  $\eta_p^2 = .040$ ). In separate analyses, 9-month-olds ( $F(1,23) = 8.150$ ,  $p = .009$ ,  $\eta_p^2 = .262$ ) and 12-month-olds ( $F(1,23) = 14.618$ ,  $p = .001$ ,  $\eta_p^2 = .389$ ) showed a significant preference for the unfamiliar names, while the 5-month-olds' preference was marginal ( $F(1,23) = 7.334$ ,  $p = .013$ ,  $\eta_p^2 = .242$ ), with the Bonferroni correction ( $\alpha = .0125$ ). This suggests that infants are able to detect words other than their own name and show a preference for this speech over a background babble.

## Discussion

With a signal-to-noise ratio of 10 dB, infants as young as 5 months showed a preference for target speech against a naturalistic multi-talker background. These results also replicate Newman's (2005; 2009) main findings that by 9 months infants are sensitive to the phonetic detail of their own name against a multi-talker background, and extend these findings to a background more characteristic of the multi-talker speech infants are likely to encounter. However, unlike earlier work, our results suggest that 5-month-olds may have difficulty with a naturalistic multi-talker speech more typical of everyday life. Our findings further provide evidence that by 9 months (and possibly as young as 5 months) infants can detect speech other than their own name under these noise conditions (see also Newman and Hussain, 2006, for indirect evidence of this).

The ability to detect the presence of a target speech signal is an important first step towards being able to discriminate one name or word from another in background noise. A preference for target speech (embedded within background speech) was present in infants as young as 5 months and is consistent with prior work (Shultz & Vouloumanos, 2010; Vouloumanos & Werker, 2004) showing that infants as young as 3 months prefer listening to human speech over complex non-speech analogues and macaque calls, as well as other human non-speech sounds (e.g., laughter). Our results extend this preference for human speech to a preference for single-talker speech (heard over a multi-talker background) over multi-talker speech alone.

The next step to discriminating one name or word from another in background noise is the ability to perceive the linguistic detail of the target speech. Unexpectedly, infants only showed a preference for their own name over the more similar stress-matched name (at least by 9 months), but not the non-matched name. This contrasts with Newman's (2005, 2009) findings, in which infants preferred their own name to both types of foil names, and was not predicted by our hypotheses. One



alternative explanation might be that the non-matched foil represented a more appealing target (multi-syllabic with a trochaic stress pattern) than the stress-matched foil. However, re-analysis of the two foil names based on stress-pattern (trochaic vs. iambic or monosyllabic) showed that infants at all three ages did not show a preference for the more appealing trochaic stress pattern, which was the stress-match foil for the majority of infants.

Aside from the difference in the background speech, a key design difference in our study is the inclusion of a no-target control, rather than a second non-matched name. It is possible that this difference altered infants' preferences by drawing particular attention to the presence/absence of target speech, such that they ALSO show a preference for the DIFFERENT prosodic pattern. However, this explanation is post hoc in nature and thus only speculative. Regardless, our findings do suggest that infants are showing some sensitivity to phonetic details under these noise conditions, at least by 9 months of age.

A key question in our study was the age at which this discrimination emerges. Our findings suggest that, although infants are able to DETECT the presence of the target speech at 5 months, they did not show a preference for their own name when they were analyzed separately from the other age groups. Though this developmental pattern must be interpreted with caution (particularly given the lack of interaction with age), it may indicate either a greater difficulty segregating the signal from the noise or a less fine-grained phonetic representation of their own name. While infants recognize their own name by 4.5 months (Mandel *et al.*, 1995), and are able to use their own name to segment other words from the speech stream (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005) by 6 months under quiet conditions, their representations may nonetheless be less robust than those of older infants (Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015), and thus more challenged under noisy conditions.

Despite these complexities, our findings provide clear evidence that infants as young as 5 months can detect speech at a 10 dB SNR against the type of background speech used here. This background speech differed in important ways from prior multi-talker speech in that it was produced under highly naturalistic conditions and recorded live, rather than recorded and mixed in the lab. This means that it includes prosodic and dynamic variations not found in read-aloud speech, and that the speakers contributing to the speech may have been responding to the nature of the acoustic environment itself in their productions. Nonetheless, even our more naturalistic stimuli differed from 'real-world' speech noise. For reasons of experimental control, significant pauses and readily identifiable individual speech streams (particularly child, or child-directed) were excluded from the sample, despite being significant aspects of the speech contexts infants experience in everyday life and part of the larger recording from which our background was sampled. While a systematic study of the impact of perceptually salience pauses and/or the perceptual emergence of individual streams from the background will be important going forward, our study is an important step in understanding infants' perceptual capabilities in real-world noise conditions.

**Acknowledgements.** We would like to thank the participating families and the members of the Baby Language Lab, especially Robin O'Hagan, Jacquelyn Klassen, Alex Holt, and Laurisa Adams. We would also like to thank Manitoba Health for their assistance with participant recruitment. The results and conclusions are those of the authors and no official endorsement by Manitoba Health is intended or

should be inferred. This research was funded by the University of Manitoba, and NSERC Discovery grant 371683–2010 to the second author.

## References

- Barker, B. A., & Newman, R. S.** (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), 45–53.
- 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? *Speech Recognition and Coding*, 321–24.
- Bernier, D. E., & Soderstrom, M.** (2014). *Clarity of language input to toddlers across child care settings*. Poster Presented at Development 2014: A Canadian Conference on Developmental Psychology, Ottawa, ON.
- Blaauw, E.** (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*(4), 359–75.
- Boersma, P., & Weenink, D.** (2014). Praat: doing phonetics by computer [computer program]. Retrieved 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*(4), 298–304.
- Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T.** (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–98.
- Broadbent, D. E.** (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, *44*(1), 51–5.
- 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*(5), 975–79.
- 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*(4), 554–9.
- Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J.** (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*, 83–92.
- Hirsh, I. J.** (1950). The relation between localization and intelligibility. *Journal of the Acoustical Society of America*, *22*(2), 196–200.
- Laan, G. P. M.** (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*, 43–65.
- Leek, M. R., Brown, M. E., & Dorman, M. F.** (1991). Informational masking and auditory attention. *Perception & Psychophysics*, *50*(3), 205–14.
- Levin, H., Schaffer, C. A., & Snow, C.** (1982). The prosodic and paralinguistic features of reading and telling stories. *Language & Speech*, *25*(1), 43–54.
- Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B.** (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–17.
- McMillan, B. T. M., & Saffran, J. R.** (2016). Learning in complex environments: the effects of background speech on early word learning. *Child Development*, *87*(6), 1841–55.
- Mertus, J.** (2011). Welcome to Bliss: The Brown Lab Interactive Speech System [computer program]. Retrieved from <<http://mertus.org/Bliss/index.html>>.
- Miller, G. A.** (1947). The masking of speech. *Psychological Bulletin*, *51*(4), 327–58.
- Newman, R. S.** (2005). The cocktail party effect in infants revisited: listening to one's name in noise. *Developmental Psychology*, *41*(2), 352–62.
- Newman, R. S.** (2009). Infants' listening in multitalker environments: effect of the number of background talkers. *Attention, Perception, & Psychophysics*, *71*(4), 822–36.
- Newman, R. S.** (2011). 2-year-olds' speech understanding in multitalker environments. *Infancy*, *16*(5), 447–70.
- Newman, R. S., & Hussain, I.** (2006). Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. *Infancy*, *10*(1), 61–76.
- Newman, R. S., & Jusczyk, P. W.** (1996). The cocktail party effect in infants. *Perception & Psychophysics*, *58*(8), 1145–56.

- 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, Language, and Hearing Research*, 34(3), 643–50.
- 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(2), 102–12.
- 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(1), 339–50.
- Nozza, R. J., Wagner, E. F., & Crandell, M. A.** (1988). Binaural release from masking for a speech sound in infants, preschoolers, and adults. *Journal of Speech, Language, and Hearing Research*, 31(2), 212–18.
- Nozza, R. J., & Wilson, W. R.** (1984). Masked and unmasked pure-tone thresholds of infants and adults. *Journal of Speech, Language, and Hearing Research*, 27(4), 613–22.
- Pollack, I., & Pickett, J. M.** (1958). Stereophonic listening and speech intelligibility against voice babble. *Journal of the Acoustical Society of America*, 30(2), 131–33.
- Poulton, E. C.** (1953). Two-channel listening. *Journal of Experimental Psychology*, 46(2), 91–6.
- Shultz, S., & Vouloumanos, A.** (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development*, 6(4), 241–57.
- Sinnott, J. M., Pisoni, D. B., & Aslin, R. N.** (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior and Development*, 6(1), 3–17.
- Smith, N. A., & Trainor, L. J.** (2011). Auditory stream segregation improves infants' selective attention to target tones amid distracters. *Infancy*, 16(6), 655–68.
- Trehub, S. E., Bull, D., & Schneider, B. A.** (1981). Infants' detection of speech in noise. *Journal of Speech, Language, and Hearing Research*, 24(2), 202–6.
- Vouloumanos, A., & Werker, J. F.** (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, 7(3), 270–6.
- Werner, L. A., & Boike, K.** (2001). Infants' sensitivity to broadband noise. *Journal of the Acoustical Society of America*, 109(5), 2103–11.