Infants' name recognition in on- and off-channel noise

Rochelle S. Newman and Giovanna Morini
Department of Hearing and Speech Sciences, 0100 Lefrak Hall, University of Maryland, College Park, Maryland 20742
rnewman1@umd.edu, gmorini@umd.edu

Monita Chatterjee
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131
monita.chatterjee@boystown.org

Abstract: Previous work by Polka, Rvachew, and Molnar [Infancy **13**(5), 421-439 (2008)] has reported that infants are poor at focusing their attention on a particular frequency range, and, as a result, are distracted by maskers that are outside of the target frequency range. The current study explores this effect of irrelevant distractors further and finds that 8-month-old infants are significantly less affected by maskers outside the frequency range (off-channel maskers) than by on-channel maskers. Thus while infants may display difficulty ignoring irrelevant distractors, they are able to do so to at least some degree, suggesting some ability to perceive speech from spectrally remote maskers, despite the demonstrated presence of greater informational masking at this age.

PACS numbers: 43.71.Ft [AC]
Date Received: January 10, 2013 **Date Accepted:** March 12, 2013

1. Introduction

Infants frequently hear speech in noisy environments (e.g., Barker and Newman, 2004), which can adversely affect their ability to detect and discriminate speech sounds (Nozza *et al.*, 1990; Nozza *et al.*, 1991; Trehub *et al.*, 1981). Several studies have explored factors that might make it easier or harder for infants to recognize speech in noise. For example, Newman (2005) showed that infants can recognize known words in the presence of multitalker babble, but they failed to do so in the presence of a single distracting talker (Newman, 2009). Barker and Newman (2004) found that infants were better able to understand speech in the presence of background talkers when the target voice was familiar (their own mother) compared to an unfamiliar talker. These studies have focused particularly on situations in which the background masker consisted of one or more other voices—signals that are quite similar acoustically to the target talker.

Another important factor that influences stream segregation, at least for adult listeners, is the spectral overlap of the target and masker signals. Adults show greater interference when a nonspeech masker overlaps the target speech signal in frequency range than when it does not (e.g., Kidd *et al.*, 2005). There are likely several reasons for this. First, sounds that are in the same frequency range have the potential to cause perceptual (or "energetic") masking (masking at the level of the auditory periphery), but sounds outside that range do not. Second, sounds with the same frequency range are likely to result in greater stimulus uncertainty, resulting in more "informational masking" (IM; e.g., Brungart, 2001; Kidd *et al.*, 1994; Mattys *et al.*, 2009) or masking caused by non-perceptual factors, such as difficulties in attention and segregation.

Although infants' spectral resolution is essentially adult-like by 6 months (e.g., Olsho, 1985; Schneider *et al.*, 1990), there are reasons to suspect that they might not show the same patterns as do adults. First, Bargones and Werner (1994) demonstrated that infants have difficulty focusing attention selectively to particular frequency ranges, suggesting that they might show less benefit when a potential masker was off-channel compared to when it was on-channel. Moreover, informational masking may play a greater role in children's listening success. Even preschoolers and school-aged children show more informational masking than adults (e.g., Oh *et al.*, 2001; Wightman *et al.*, 2003), and infants are particularly sensitive to IM (Leibold and Werner, 2006). This suggests that infants would be more easily distracted by extraneous information and would show IM even from off-channel sounds.

Recently, Polka *et al.* (2008) examined 6- to 8-month-old infants' use of frequency range as a cue for selective attention and found that sounds outside the frequency range of a target speech signal could cause interference. Infants were tested in a habituation-dishabituation procedure on their ability to discriminate speech syllables. Some infants were habituated and tested in quiet, others were habituated and tested with the syllables mixed with a high-frequency background noise, and still others were habituated with stimuli in the noise and tested in quiet. The noise consisted of a combination of high-frequency cricket and bird songs and had no overlapping frequencies with the speech syllable. It thus could not cause energetic masking. Yet far fewer children appeared to recognize the speech syllables during the test phase (as indexed by a preference for the novel syllable) when noise had been present during habituation.

Polka *et al.* interpreted their findings as suggesting that "perceiving speech in the presence of irrelevant sounds poses a cognitive challenge for young infants" (p. 435) and that this difficulty cannot be the result of energetic masking alone. This greater susceptibility to informational masking could help to explain why infants appear to show greater effects of masking than would be expected based on their spectral resolution abilities (see Werner and Bargones, 1992).

However, Polka *et al.* did not compare the *degree* of distraction caused by sounds outside the target's frequency range with that caused by sounds within the target's frequency range. If infants can segregate streams by frequency, they might be *better* able to segregate a competing noise from the target stream when the noise is off-channel (or off-frequency). This does not mean that sounds outside the target frequency range will cause *no* masking, but only that they will cause significantly *less* masking than sounds within the frequency range.

The current study explored this issue of frequency range using on- and off-frequency band-passed white noise as the competing stream. Infants were presented with the sound of a name being repeated on each trial; on some trials, the name was their own, whereas on other trials, it was a foil name. Infants typically will listen longer to their name than to a foil name (Mandel *et al.*, 1995; Newman, 2005). We predicted that infants would show better segregation with a frequency-range difference, as shown by longer listening to their own name than a foil name in the presence of an off-channel masker, but no listening difference with an on-channel masker.

2. Methods

2.1 Participants

A total of 40 infants (22 male, 18 female) participated (mean age 8.3 months; range 7.0-9.2). Data from an additional 14 infants were excluded due to prematurity ($n = 2$), being tested with an incorrect name pronunciation ($n = 4$), or crying/fussiness ($n = 8$). Because the present study focused on infants' names rather than English words, exposure to multiple languages was not a reason for exclusion; 12 infants heard a language besides English some of the time.

2.2 Stimuli

Stimuli consisted of a target speech stream and a competing masker. All speech recordings were made using a Shure SM81 microphone, at a sampling rate of 44.1 kHz, in a sound-attenuated room. Target passages consisted of 15 repetitions of the name or nickname reported to be used most often when speaking to the child; foil passages were similar but consisted of another name matched for stress pattern. To ensure that effects generalized across talkers, recordings were made by four different female speakers, who were instructed to produce the names as if calling a small child in a lively voice.[1] To ensure that both target and foil names sounded equally engaging, talkers recorded a variety of names in each recording session, without any knowledge of which names would be used as targets and which as foils during test sessions. The same speaker recorded both of the names heard by any given child, and these were matched for overall duration by adjusting the pauses between name repetitions. In some cases, the target name for one child was used as the foil name for another, further reducing the stimulus differences across conditions for the set of infants as a whole. If a child listens longer to his or her own name than to the foil name, it is taken as an indication that the child was able to identify the name despite the masker.

The masker consisted of white noise band-pass filtered to be approximately one-half octave in bandwidth. The use of a bandwidth measured in octaves results in noise that covers an equal number of critical bands in both frequency ranges; infants and adults appear to have similar critical bandwidths (Schneider and Trehub, 1992) although age-related differences would not pose a problem unless they were also frequency-dependent. For the on-channel condition, the noise band was centered at 1 kHz (range: 828–1172 Hz), masking an important region of the speech signal. For the off-channel condition, the noise was centered at 8 kHz (range: 6.6 to 9.4 kHz). The band-pass filter slopes were sufficiently steep to ensure that the off-channel noise did not overlap spectrally with the speech signal (Fig. 1). These two noise signals were adjusted to have the same RMS amplitude. Although adults typically show higher free-field thresholds at 8 kHz than at 1 kHz, thresholds for both 6- and 12-month-olds are actually quite similar at these two frequencies (Olsho *et al.*, 1988); we therefore did not adjust the relative amplitude of the two noise bands to ensure equivalent degrees of perceived loudness.[2]

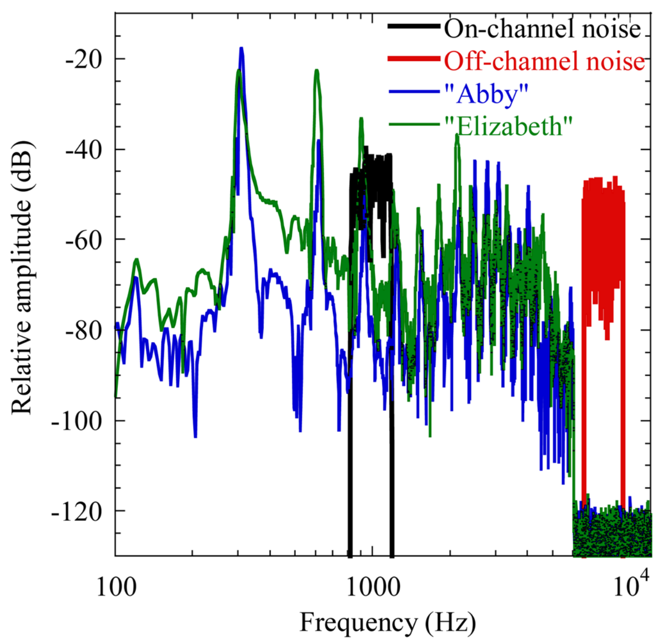

Fig. 1. (Color online) Overall average spectra for the two band-pass noises and for two example target (name) stimuli.

The critical difference between the two noises, then, is that one overlaps the frequencies that are important for speech, while the other should be too remote to cause energetic masking of the speech stream. To ensure that the off-channel noise truly did not overlap with the speech signal, all name recordings were low-pass filtered at 6 kHz, so that the only signal present was at frequencies below the noise spectrum (Fig. 1). Use of 6 kHz as a cut-off allows information for fricatives to remain present in the signal, permitting us to test children with names like "Silas" without the name sounding distorted. That said, the low-pass filtering did reduce the relative intensity of some phonemes relative to others in the same word. However, this was the case for both the on-channel and off-channel noise conditions, and thus the alteration of the higher-frequency speech sounds cannot explain any difference across the two noise conditions.

We then adjusted the noise signals to be 10 dB less intense than the target speech. As periods of silence caused the average intensity of the recordings to be much lower than the speech segments, the r.m.s. amplitude of the utterances was measured after removal of the silent gaps. The intensity levels of the original name recordings (including the pauses) were then adjusted such that the intensity of the speech was matched across the target and foil names and was exactly 10 dB more intense than the maskers.

The masking noise began 500 ms before the first repetition of the child's name and continued until the final repetition was complete. All sound files for a given child were identical in length, but sound files for the different children ranged from 28 to 38 s, primarily as a result of variation in the length of children's names.

This resulted in four sound files for each child: One contained the child's own name combined with a noise that did not overlap in frequency range (the off-channel masker), one consisted of a foil name with the same masker, and the other two contained the child's own name or the foil name combined with the on-channel masker. Two practice stimuli were used to familiarize infants with the task and setting. These consisted of musical passages used in previous name-identification studies (Newman, 2005, 2009), each 14.9 s in length.

2.3 Procedure

The apparatus and procedure were identical to that used in Newman (2009). Participants sat on their caregivers' laps in the center of the booth, facing the front panel. Each trial began with blinking of the center-panel light. Once the child was oriented in that direction, that light stopped flashing, and one of the two side-lights began to blink. Once the child looked at that side light, the auditory stimulus began playing from a speaker located behind the light and continued to play until its completion or until the infant had looked away for a minimum of two consecutive seconds. The experimenter pressed buttons on the response box to indicate the infant's looking behavior; listening time was determined by the amount of time that the child remained looking at the flashing light and excluded any time spent looking away.

Infants completed a practice phase to familiarize them with the task; they heard two musical passages on alternating trials until they accumulated 25 s of listening time to each. This was followed by a 16-trial test phase with four blocks, each containing a random order of one trial of each of the four types. The caregiver and experimenter listened to masking music over Peltor aviation headphones to avoid influencing the child's behavior or the coding.

3. Results and discussion

A 2 (name) $\times$ 2 (noise type) $\times$ 4 (talker) repeated measures ANOVA showed no effect of talker nor any interaction involving talker; we therefore collapsed across talkers and conducted a 2 (name) $\times$ 2 (noise) repeated measures ANOVA. We found a significant overall effect of noise [$F(1,39) = 9.49$, $p < 0.005$] as well as a name $\times$ noise interaction ($F(1,39) = 4.56$, $p = 0.039$]. There was no overall effect of name [$F(1,39) = 1.25$, $p > 0.25$]. Given the significant interaction, we conducted follow-up *t*-tests exploring the effect of name (own name vs foil name) separately for the two noise conditions. This showed a significant effect of name in the off-channel condition [$t(39) = 2.18$, $p = 0.035$], but no effect of name in the on-channel condition [$t(39) = 0.90$, $p > 0.35$]. In the off-channel condition, infants listened an average of 12.1 s to the trials presenting their own name, but only 9.94 s to trials presenting a foil name, suggesting that they were able to distinguish the two names and showed a preference for their own name. When the noise was on-channel, infants listened longer overall but made no distinction between trials including their own name (12.3 s) and those that did not (13.0 s), as shown in Fig. 2.

Thus the current study suggests that the off-channel noise was less disruptive to infant speech recognition than was the on-channel masker. This does not negate the findings by Polka *et al.* (2008) suggesting that infants have difficulty with informational masking; but it does place limits on the extent of this difficulty. Infants do seem to be less affected by maskers that are off-channel than those that are on-channel, suggesting at least some limited ability to attend selectively to a given frequency range.

An important difference between our study and that of Polka *et al.* is the nature of the off-channel noise. While both studies used maskers containing only high-frequency spectral components, our masker was based on white noise, whereas Polka *et al.* used high-frequency cricket and bird songs. The noise-based masker would result in a more constant amplitude over time, further reducing opportunities for informational masking. This is an important difference given recent work suggesting that infants may experience informational masking caused by time-varying signals (Newman, 2009). In addition, the current masker remained present throughout the test trial, whereas that used by Polka *et al.* was gated on and off in synchrony with the speech signal; the latter may have encouraged grouping of the two signals, despite their frequency range differences. Third, our masker was set at a 10 dB SNR based on acoustic r.m.s. measurements, whereas their stimuli were first adjusted to be of equal perceived loudness (by adult listeners) and then presented at a much harder SNR (presentation levels of 61–63 dBA for the speech, 69–72 dBA for the nonspeech, or roughly −8 to −9 dB SNR); this greater SNR could have caused greater IM. Finally, our masker remained constant throughout the experiment, whereas their masker varied from token to token. These differences may have helped infants in our study selectively ignore the masking signal. As noted by Polka *et al.*, exploration of how the timing of a masking signal affects infant performance is an important direction for future research.

Looking across these two studies, we see that by 8 months, infants show some ability to perceive speech despite the presence of spectrally non-overlapping maskers but are affected by these maskers to a greater degree than are adults. However, the

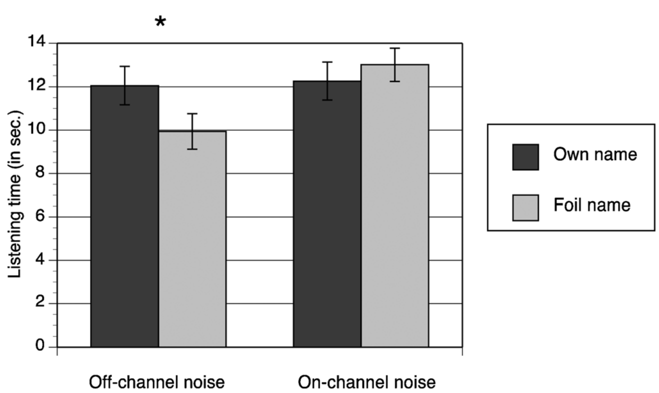

Fig. 2. Mean listening times and standard errors to the child's own name and to the foil name in the presence of a white noise masker band-passed to be in a non-overlapping frequency range (off-channel) or an overlapping frequency range (on-channel).

reason for this difference remains unclear. One possibility is that young infants may be less adept at using frequency information as a cue for either stream segregation or selective attention than are adult listeners. Another possibility is that they show greater effects of informational masking more generally, perhaps as a result of greater stimulus uncertainty. A third possibility is that infants are less adept at inhibiting a competing stimulus. Future work will need to explore these distinctions in more depth, and examine how this ability changes with development.

In summary, we found that infants are better able to recognize speech in the presence of an off-channel noise masker than in the presence of an on-channel masker, suggesting some ability to separate streams of sound by frequency range.

Acknowledgments

We thank Lloyd Frei, Keith Miller, Don Mathis, James Sawusch, and Jim Garmon for design and implementation of our test booth and D. Brunson, J. Dombroski, M. O'Fallon, and A. Pasquarella for recording names; we thank these people plus A. Arnold, C. Bender, T. Bipat, A. Cook, J. Coon, S. Dougherty, L. Evans, A. Gandee, M. Javid, A. Jensen, S. Lee, P. and R. Lieberman, V. Moussavi, M. Nasuta, R. Rhodes, A. Rodriguez, K. Shniderman, V. Son, K. Voelmle, K. Wilson, and C. Wu for assistance scheduling and testing infants. This work was supported by NSF Grant No. 0642294 to UMD. M.C. was supported by NIH Grants No. R01 DC004786 and No. R21 DC011905; G.M. was supported by an NRSA T32 and an NSF IGERT to UMD.

References and links

[1]Although our intent was to have infants evenly split across talkers, the use of children's own name as stimuli required that recordings be made in advance; it was impossible to predict at the time of the recording session which families would keep their appointment and thus how many names we would need to record in each voice. One infant too many heard one of the voices (11 infants) and one too few heard another (9 infants). However, there was no main effect of or interaction with talker, suggesting that the general pattern held across talkers. This in turn suggests that the minor variation from the goal of 10 infants per talker could not affect results.

[2]In contrast, threshold values at these two frequencies are quite different at 3 months (Olsho *et al.*, 1988); this is a primary reason for testing children greater than 6 months of age in this study.

Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively; infants do not," Psychol. Sci. **5**, 170–174.

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

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**(3), 1103–1109.

Kidd, G., Jr., Mason, C. R., Deliwala, P., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.

Leibold, L. J., and Werner, L. A. (**2006**). "Effect of masker-frequency variability on the detection performance of infants and adults," J. Acoust. Soc. Am. **119**, 3960–3970.

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

Mattys, S. L., Brooks, J., and Cooke, M. (**2009**). "Recognizing speech under a processing load: Dissociating energetic from informational factors," Cognit. Psychol. **59**, 203–243.

Newman, R. S. (**2005**). "The cocktail party effect in infants revisited: Listening to one's name in noise," Dev. Psychol. **41**(2), 352–362.

Newman, R. S. (**2009**). "Infants' listening in multitalker environments: Effect of the number of background talkers," Attention, Percept. Psychophys. **71**(4), 822–836.

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

Nozza, R. J., Rossman, R. N. F., Bond, L. C., and Miller, S. L. (**1990**). "Infant speech-sound discrimination in noise," J. Acoust. Soc. Am. **87**, 339–350.

Oh, E. L., Wightman, F. L., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**(6), 2888–2895.
Olsho, L. W. (**1985**). "Infant auditory perception: Tonal masking," Infant Behav. Dev. **8**, 371–384.
Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., and Spetner, N. B. (**1988**). "Pure-tone sensitivity of human infants," J. Acoust. Soc. Am. **84**(4), 1316–1323.
Polka, L., Rvachew, S., and Molnar, M. (**2008**). "Speech perception by 6- to 8-month-olds in the presence of distracting sounds," Infancy **13**(5), 421–439.
Schneider, B. A., Morrongiello, B. A., and Trehub, S. E. (**1990**). "Size of critical band in infants, children, and adults," JEP: Hum. Percept. Perform. **16**, 642–652.
Schneider, B. A., and Trehub, S. E. (**1992**). "Sources of developmental change in auditory sensitivity," in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC).
Trehub, S. E., Bull, D., and Schneider, B. A. (**1981**). "Infants' detection of speech in noise,". J. Speech Hear. Res.**24**, 202–206.
Werner, L. A., and Bargones, J. Y. (**1992**). "Psychoacoustic development of human infants," Adv. Infancy Res. **7**, 103–145.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.