

# A comparison of monolingual and bilingual toddlers' word recognition in noise

International Journal of Bilingualism

1–14

© The Author(s) 2021

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/13670069211028664

[journals.sagepub.com/home/ijb](https://journals.sagepub.com/home/ijb)**Giovanna Morini** 

University of Delaware, USA

**Rochelle S. Newman**

University of Maryland, College Park, USA

## Abstract

**Aims and objectives:** The purpose of this study was to examine whether differences in language exposure (i.e., being raised in a bilingual versus a monolingual environment) influence young children's ability to comprehend words when speech is heard in the presence of background noise.

**Methodology:** Forty-four children (22 monolinguals and 22 bilinguals) between the ages of 29 and 31 months completed a preferential looking task where they saw picture-pairs of familiar objects (e.g., *balloon* and *apple*) on a screen and simultaneously heard sentences instructing them to locate one of the objects (e.g., look at the *apple!*). Speech was heard in quiet and in the presence of competing white noise.

**Data and analyses:** Children's eye-movements were coded off-line to identify the proportion of time they fixated on the correct object on the screen and performance across groups was compared using a 2 × 3 mixed analysis of variance.

**Findings:** Bilingual toddlers performed worse than monolinguals during the task. This group difference in performance was particularly clear when the listening condition contained background noise.

**Originality:** There are clear differences in how infants and adults process speech in noise. To date, developmental work on this topic has mainly been carried out with monolingual infants. This study is one of the first to examine how background noise might influence word identification in young bilingual children who are just starting to acquire their languages.

**Significance:** High noise levels are often reported in daycares and classrooms where bilingual children are present. Therefore, this work has important implications for learning and education practices with young bilinguals.

## Keywords

Bilingual toddlers, listening in noise, word recognition, speech perception

---

### Corresponding author:

Giovanna Morini, Communication Sciences and Disorders Program, University of Delaware, 100 Discovery Boulevard, 5th Floor, Newark, DE 19713, USA.

Email: [gmorini@udel.edu](mailto:gmorini@udel.edu)

## Introduction

An important skill in language acquisition is being able to understand speech in different listening conditions – including noisy settings. From a very early age, children are exposed to high levels of ambient noise found in hospitals and preschools (Busch-Vishniac et al., 2005; Frank & Golden, 1999), as well as in homes, where background speech is often present along with a variety of other noises (including heating, ventilation, and air conditioning noise, traffic noise, and speech addressed to other children). To comprehend speech in noise, the listener must succeed at “stream segregation”, which refers to the process of separating two competing sound streams into the specific components that make up each signal, and grouping together the elements that make up one stream.

There are high task demands associated with processing speech in noise, and even adult listeners (who have had many years of experience using their native language) still find it difficult to process competing acoustic information (e.g., Cooke, 2006; Festen & Plomp, 1990; Miller et al., 1951; Pollack & Pickett, 1958; Simpson & Cooke, 2005). Furthermore, the language background of the listeners (i.e., whether they grew up with one versus two languages) also plays a role in stream segregation, with bilinguals having greater difficulty than monolinguals understanding speech in the presence of background noise (Florentine, 1985a, 1985b; Florentine et al., 1984; Mayo et al., 1997; Meador et al., 2000; Morini & Newman, 2020; Rogers et al., 2006; Shi, 2010). This is the case even for extremely balanced bilingual adults who acquired both their languages before the age of six (Rogers et al., 2006; Mayo et al., 1997; Meador et al., 2000; Morini & Newman, 2020; Tabri et al., 2011). It remains unclear, however, whether differences in cognitive and linguistic abilities associated with bilingualism in adulthood – that might influence performance on stream-segregation tasks – are already present during early childhood. The present work explores this topic.

There is a high incidence of bilingualism worldwide (Grosjean, 2010), and even in primarily monolingual countries such as the United States at least 20% of children grow up in homes where a language other than English is spoken (Shin & Ortman, 2011). Hence, understanding how language-related processes develop in this population (including word recognition across listening environments) is of great importance. Based on theories of bilingual language processing, the previously-reported bilingual “disadvantage” during listening-in-noise tasks is attributed to additional language processing demands that bilinguals (but not monolinguals) are faced with. To successfully identify words in a particular language, bilinguals must inhibit the non-target language (Grosjean, 1997; MacKay & Flege, 2004). This is an extra step, and one that leads to additional demands on the attentional resources that bilingual listeners have available for speech processing. Carrying out this extra step during tasks with a greater processing load (as in the case of comprehension of speech in noise) is particularly problematic. In these situations, bilinguals must: (a) inhibit information from the non-target language; (b) process and segregate the simultaneously occurring complex acoustic signals; and (c) extract the relevant information included in the target speech.

There are clear differences in how infants and adults process speech in noise in general – regardless of language background (Newman, 2005, 2009; Newman & Jusczyk, 1996). For example, infants are better at segregating speech signals when there are multiple voices in the background than when there is a single voice (Newman, 2009), while adults show the opposite pattern, performing better with a single voice (Gustafsson & Arlinger, 1994). A possible explanation is that a single talker’s voice varies in amplitude over time, and adults are able to take advantage of such fluctuations by “listening in the dips” (Festen & Plomp, 1990; Wilson & Carhart, 1969). This is especially the case when the variation occurs in a slow, predictable manner (Gustafsson & Arlinger, 1994). Infants, on the other hand, do not appear to be able to benefit from the dips in the signal

(Newman, 2009). These age-related differences in performance have been associated with selective attention (i.e., focusing on those time periods or portions of the signal that are likely to be most beneficial), a cognitive ability that does not appear to be fully developed in young children (Bargones & Werner, 1994; Garon et al., 2008; Jones et al., 2003). However, this developmental work has mainly been carried out with monolingual infants, and hence the way in which background noise might influence word identification in young bilingual children who are just starting to acquire their languages remains greatly unknown.

To our knowledge, only two studies have explored bilingual children's stream segregation abilities, but these studies were conducted with older children and provided conflicting findings. Krizman et al. (2017) asked monolingual and bilingual 14-year-olds to identify sentences-in-noise, words-in-noise, and tones-in-noise. They found that monolinguals outperformed bilinguals during the sentences-in-noise task, while bilinguals outperformed monolinguals when the degraded auditory target was non-linguistic (i.e., during the tones-in-noise measure). They found no difference between monolinguals and bilinguals during the words-in-noise condition. Krizman et al. (2017) concluded that differences in performance across groups were linked to the quantity of linguistic information available in the auditory signal. More specifically, they concluded that bilingualism might have multiple effects: on the one hand improving stream segregation abilities (when linguistic information is not available), but also reducing accuracy during tasks that involve lexical identification (and hence lexical competition). When the task is highly linguistic (e.g., recognizing sentences), the deficit outweighs the advantage, but when the task is nonlinguistic, the improved stream segregation proves more important.

In another study, Reetzke et al. (2016) administered a comprehensive speech-in-noise battery to school-age children (age 6–10 years), who were raised either as monolinguals or as balanced bilinguals. The battery included recognition of English sentences across different modalities (e.g., audio-only and audiovisual), across different types of background noise (e.g., steady-state pink noise and two-talker babble), and across different signal-to-noise ratios (SNRs). They found no differences in performance across the two groups in any of the measures. This is, however, the only study to have reported no group differences between monolinguals and bilinguals in a task that involves stream segregation.

Despite many years of experience using both languages, bilingual adolescents and adults still appear to be at a disadvantage during word-recognition-in-noise tasks. Furthermore, this bilingual disadvantage has been reported across different types of background noises including: talker babble (Mayo et al., 1997); pink noise (Meador et al., 2000); reverberation and speech-spectrum noise (Rogers et al., 2006); and white noise (Morini & Newman, 2020). In other words, the effect is consistent and does not appear to be linked to a certain type of noise. This might suggest that young bilingual children, with far less experience negotiating their two languages, would similarly struggle more than monolinguals when processing speech in noise. However, bilingual adults and adolescents have large lexicons in both languages, with many words that may be generating active competition. Young bilinguals are in an early stage of learning their languages, with far smaller lexicons in either language, and could be different in this regard. In other words, variations in lexical development could lead to age-related differences in bilingual performance during word comprehension in noise. An important next step is to evaluate listening-in-noise skills in much younger bilingual listeners.

The present study used a version of the preferential looking procedure (Golinkoff et al., 1987) to test monolingual and bilingual toddlers on their ability to identify familiar words in the presence of white noise, as well as in a quiet setting. Additionally, participants' caregivers completed the Language Development Survey (LDS; Rescorla, 1989), which provides a measure of children's productive vocabulary. The goal of this study was to examine whether the language background

(specifically being raised monolingual versus bilingual) plays a role on young children's ability to comprehend familiar words when speech is heard in the presence of background noise. We hypothesize that, like adults and older children, both monolinguals and bilingual toddlers will show greater difficulty (i.e., lower accuracy) understanding words that are heard in the presence of competing background noise, compared to words that are presented in a quiet condition. More importantly, we expect that the two groups will differ in their performance, particularly when there is noise in the background, with bilinguals showing lower accuracy compared to monolinguals – likely as a result of the additional language processing demands associated with having to sort through two (as opposed to one) language systems during a relatively demanding task.

## Methods

### Participants

Forty-four children (22 monolinguals and 22 bilinguals) between the ages of 29 and 31 months (mean ( $M$ ) = 30.1, standard deviation ( $SD$ ) = 0.65) recruited from two testing sites (one in the United States and one in Canada) completed the study. Based on responses from a background questionnaire that parents were asked to complete, participants had no known developmental or physiological diagnoses. Additionally, parents reported that all participants had normal hearing and had not had any recent ear infections. Data from an additional 22 participants were excluded for the following reasons: fussiness/crying ( $n = 17$ ); failure to meet language requirement ( $n = 2$ ); motor and language delays ( $n = 1$ ); equipment failure ( $n = 1$ ); or experimenter error ( $n = 1$ ).

Monolinguals (11 males (m), 11 females (f)) were born in the United States ( $n = 15$ ) or in Canada ( $n = 7$ ), and were being raised in households where English was spoken at least 90% of the time. Based on the LDS, monolingual children had an average of 231 words in their productive vocabulary (range: 13–309 words). Bilinguals (12 m, 10 f) were being exposed to a minimum of 30% and a maximum of 70% of each of two languages since birth (one of the two languages being English). These parameters were based on the definition of bilingualism used in previous child studies (Fennell et al., 2007). Children in this group were also born either in the United States ( $n = 5$ ) or in Canada ( $n = 17$ ). Language background was measured through a Language History Questionnaire. A detailed distribution of the non-English language (i.e., the second language of exposure) is provided in Table 1. Additionally, five of the bilingual participants had also been exposed to a third language, but only for 5% or less of the time. LDS scores for the bilingual participants revealed an average English productive vocabulary of 208 words (range: 65–300 words). Groups were not expected to be matched in their vocabulary, given that scores were only calculated in English (and not the two languages combined in the case of bilinguals). Surprisingly though, there was no significant difference in LDS scores between monolingual and bilingual children ( $t(42) = 1.47, p > 0.05$ , Cohen's  $d = 0.29$ ), suggesting that participants in both groups had similar abilities in terms of their use of English vocabulary.

Additionally, both groups of participants were being raised in middle- to high-socioeconomic-status (SES) homes, as determined by maternal education. Maternal education was used as the main measure of SES because it has been previously correlated with other indices of social class and because according to prior work it is a highly predictive component of SES when examining developmental outcomes (Hurtado et al., 2007; Noble et al., 2005). On average, monolingual children's mothers had completed 17.9 years of education, while bilingual children's mothers had completed 17 years. This difference was not significant ( $t(42) = 1.66, p > 0.05$ , Cohen's  $d = 0.29$ ). Previous work suggests that the linguistic input that children receive varies as a function of SES (Hoff et al., 2002), and that this in turn influences children's language development (Hoff &

**Table 1.** Distribution of the non-English language for bilingual participants.

Language background	Number of participants
Arabic	1
Cantonese	2
Farsi	1
French	1
Greek	2
Gujarati	2
Malayan	1
Mandarin	1
Polish	1
Portuguese	1
Punjabi	2
Spanish	5
Tagalog	1
Urdu	1

Naigles, 2002; Huttenlocher et al., 1991; Naigles & Hoff-Ginsberg, 1998). Recruitment for both groups of participants was conducted in the two sites (i.e., in the United States and in Canada), which allowed for both a larger bilingual sample size and better matching of groups for SES.

### Stimuli

The auditory stimuli consisted of a target speech stream and a competing noise signal. The speech stimuli consisted of short sentences instructing participants to look at an item on the screen, and included a familiar two-syllable word (e.g., *apple*, *flower*, and *cookie*) presented in sentence-final position and repeated three times per trial (e.g., Look at the *apple!* Can you find the *apple?* *Apple!*). Target words were chosen based on English lexical norms for 30-month-olds (Dale & Fenson, 1996) to ensure that both the words and the objects that they represent were familiar to the children. Speech stimuli were recorded by a female native speaker of American English in a sound-attenuated booth using child-directed-speech prosody at a 44.1 kHz sampling rate with a 16-bit analog-to-digital converter. Sentences were edited to have the same root mean square amplitude. Words specific to American or Canadian English were avoided, which meant that all target words included in the study had comparable pronunciations across the two dialects. While subtle accent differences could still have had an impact, we found no difference in performance between Canadian and American children, suggesting this is not a concern.

The competing signal was white noise presented with a steady-state amplitude envelope. White noise was chosen for a number of reasons: (a) it has been previously used during speech-perception tasks with young children; (b) it does not contain language-specific features that would share similarity with speech in a particular language (an important factor since the non-English language of our bilinguals varied across participants); (c) background noise that contains familiar lexical information such as a speaker of the same language in the background (Cooke et al., 2008; Van Engen & Bradlow, 2007) or babble that is produced by intelligible voices (Simpson & Cooke, 2005) leads to greater interference during stream-segregation tasks; and (d) as mentioned earlier in the introduction, the bilingual disadvantage associated with speech-processing in noise has been consistently reported across different types of noise (including white noise). Hence, as a starting point, we

wanted to select a background noise that we anticipated would primarily cause energetic, rather than informational, masking. If group differences in performance were observed with this “easier” type of noise, we could expect harder noise types (e.g., talker babble) to simply magnify the effect.

The noise signal was delivered in combination with the speech at approximately 70–75 dB sound pressure level through speakers, with the target speech being either 5 dB more intense than the background noise (+5 dB SNR) or the same intensity (0 dB SNR). Two noise-levels were selected to avoid the concern that a single noise level might unintentionally lead to either ceiling (e.g., the same level as in quiet) or floor performance such that results could not be interpreted. The background signal always began 500 milliseconds (ms) prior to the target speech and continued until the end of the trial.

The visual stimuli consisted of colored pictures of familiar objects comparable in size and overall shape. Images were presented in pairs on a television (TV) screen. Between trials a short attention-getter (a laughing baby) appeared on the screen to reorient participants’ attention to the center of the screen. All trials were 7500 ms, with the onset of the first repetition of the target word occurring 2000 ms from the appearance of the images. The carrier phrase started at least 1100 ms after trial onset, which meant that when noise was present, it began before the speech.

### Procedure

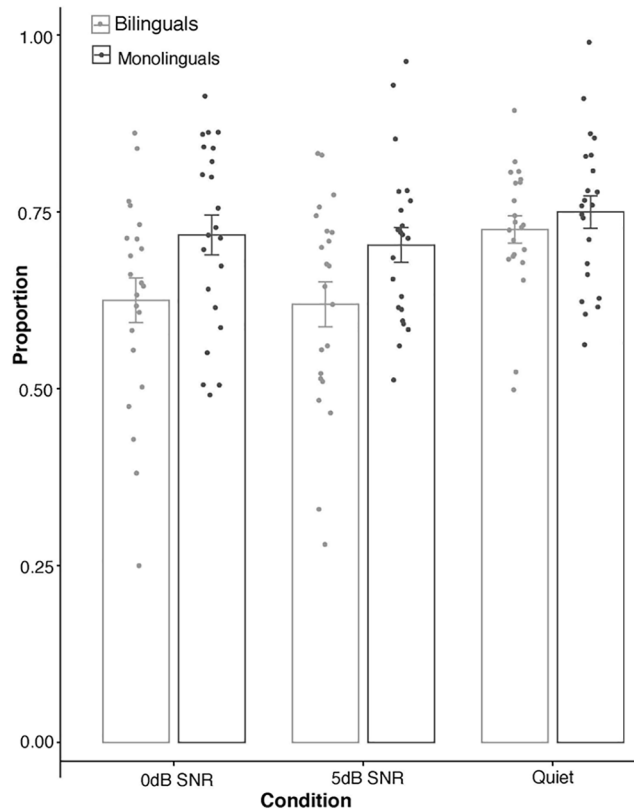
Both testing sites had comparable facilities and equipment, and the procedure was identical across locations. Testing was conducted in a quiet room with participants seated on their caregiver’s lap four feet from the screen. A camera recorded the child’s eyes, while an experimenter controlled the testing paradigm from outside the testing room. Each trial began with the attention-getter, which continued to play until the child was looking at the screen. Participants then saw two pictures presented side-by-side on a white background. Testing began with three familiarization trials (one for each object-pair), in which “generic” sentences were heard (*Look at that! Do you see that! How neat!*). These trials familiarized participants with the images of the objects and with the general procedure of the task.

This was followed by 18 test trials, six in each of three listening conditions (0 dB SNR, 5 dB SNR, and quiet), in random order. The position of the objects (left versus right) and the target noun were counterbalanced across trials. The experimental design (including the number of trials presented during the task) was selected based on previous studies with toddlers that relied on the same preferential looking procedure, and found significant results with the same or fewer trials per condition (e.g., Byers-Heinlein et al., 2017; Fernald & Hurtado, 2006; Morini & Newman, 2019). Caregivers listened to masking music over noise-reducing headphones to prevent them from inadvertently influencing the children’s looking behavior.

Children’s eye movements (left versus right) were coded offline, on a frame-by-frame basis using SuperCoder coding software (Hollich, 2005). A second coder coded 10% of the videos from each group. Reliability correlations across the coders ranged from  $r = 0.99$  to  $r = 1.0$ . All coders were blind to the location of the target object.

### Data analysis

Data on the overall time course of eye movements were linked to the auditory stimuli in order to calculate fixation duration and shifts in gaze between the images on screen. Data were analyzed for accuracy, defined as the amount of time that the participants remained fixated on the appropriate image, as a proportion of the total time spent fixating on either of the two pictures, averaged over a time window of 367 to 2000 ms after the onset of the first repetition of the target word, across all



**Figure 1.** Proportion of looking time to the correct object in monolinguals and bilinguals across the three types of trials.

trials of the same condition. Similar analysis windows have been used in the child literature (e.g., Byers-Heinlein et al., 2017; Fernald & Hurtado, 2006; Fernald et al., 2001) and are based on the notion that gaze shifts before 367 ms occur before the child has had enough time to process the auditory stimulus and thus are not the result of lexical processing (Haith et al., 1993). Accuracy scores were used to compare participants' performance to chance (i.e., 50%), as well as to compare monolingual and bilingual performance during quiet and noise trials. As part of an exploratory analysis, we also examined whether there would be a difference in the amount of time each group required to perform the task, and whether bilinguals might have delayed recognition. Time course data from both groups indicated that this was not the case. Hence, all analyses were conducted using the same time window.

## Results

Figure 1 outlines the accuracy scores during the task across groups and conditions. Examination of the proportions of children's fixation patterns revealed that in general, accuracy was highest in the quiet condition (monolinguals:  $M = 0.74$ ,  $SD = 0.10$ ; and bilinguals:  $M = 0.73$ ,  $SD = 0.08$ ). Accuracy across the two noise levels was similar within the monolingual group (0 dB SNR:  $M = 0.72$ ,  $SD = 0.13$ ; and 5 dB SNR:  $M = 0.71$ ,  $SD = 0.12$ ), as well as in the bilingual participants (0

dB SNR:  $M = 0.63$ ,  $SD = 0.12$ ; and 5 dB SNR:  $M = 0.62$ ,  $SD = 0.13$ ), and neither level was at floor nor ceiling. One-tailed single-sample  $t$ -tests indicated that the two groups performed significantly above chance in the quiet trials (monolinguals:  $t(21) = 11.31$ ,  $p < 0.0001$ , Cohen's  $d = 2.41$ ; bilinguals:  $t(21) = 13.09$ ,  $p < 0.0001$ , Cohen's  $d = 2.79$ ), the 0 dB SNR noise trials (monolinguals:  $t(21) = 7.97$ ,  $p < 0.0001$ , Cohen's  $d = 1.70$ ; bilinguals:  $t(21) = 5.08$ ,  $p < 0.0001$ , Cohen's  $d = 1.08$ ), and the 5 dB SNR noise trials (monolinguals:  $t(21) = 8.54$ ,  $p < 0.0001$ , Cohen's  $d = 1.82$ ; bilinguals:  $t(21) = 4.16$ ,  $p < 0.0001$ , Cohen's  $d = 0.89$ ).

A  $2 \times 3$  mixed analysis of variance with Group as a between-subjects factor (monolingual versus bilingual) and Listening Condition as a within-subjects factor (0 dB SNR, 5 dB SNR, and Quiet) revealed significant main effects of listening condition ( $F(2, 84) = 8.91$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.18$ ) and group ( $F(1, 42) = 5.72$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.12$ ), as well as a significant interaction ( $F(2, 84) = 3.26$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.07$ ). *Post-hoc t*-tests showed that bilinguals performed more poorly than monolinguals in the two noise conditions (0 dB SNR:  $t(42) = 2.47$ ,  $p < 0.05$ , Cohen's  $d = 0.74$ ; 5 dB SNR:  $t(42) = 2.45$ ,  $p < 0.05$ , Cohen's  $d = 0.74$ ), but not in quiet ( $t(42) = 0.38$ ,  $p > 0.05$ , Cohen's  $d = 0.12$ ). While overall performance across conditions was above chance for both groups of children, these analyses suggest that 30-month-olds were more accurate at recognizing words during trials presented in quiet compared to trials that contained noise. More importantly, accuracy scores were lower for bilinguals compared to monolinguals when noise was present in the background. Prior work has suggested that bilingual adults have particular difficulty with speech perception in noise, and the aim of this study was to explore whether this would also be the case for toddlers.

## Discussion

This experiment examined how language background interacts with the ability to understand speech in noise, early in development. Prior work has suggested that bilingual adults have particular difficulty with speech perception in noise, and the aim of this study was to explore whether this would also be the case for toddlers. While general accuracy was high for both groups, bilingual children performed significantly worse than monolinguals of the same age when there was a competing auditory signal present. There were, however, no significant differences between monolinguals' and bilinguals' English LDS scores. Additionally, we found no significant correlations between LDS scores and performance on the word recognition task in the 0 dB SNR (monolinguals:  $r = -0.07$ ,  $p > 0.05$ ; bilinguals:  $r = 0.13$ ,  $p > 0.05$ ), 5 dB SNR (monolinguals:  $r = 0.19$ ,  $p > 0.05$ ; bilinguals:  $r = 0.34$ ,  $p > 0.05$ ), nor in the quiet condition (monolinguals:  $r = -0.10$ ,  $p > 0.05$ ; bilinguals:  $r = 0.36$ ,  $p > 0.05$ ). Furthermore, at the end of the study, caregivers were asked whether they felt their child could understand the specific words that were included in the study. Based on parental report, there was no difference in familiarity with the target words across monolinguals and bilinguals. Together, this suggests that the observed group differences are not the result of dissimilarities in the size of the children's English vocabulary.

This bilingual "disadvantage" is comparable to what has consistently been reported in the adult literature. Most widely supported accounts from adult work attribute this bilingual disadvantage to factors such as reduced word recency (the amount of time since a word was used) and/or frequency (the number of times a word has been perceived), as well as cross-linguistic competition (resulting from joint activation of the two languages). Together, these factors are thought to contribute to weaker lexical representations and pathways (Ecke, 2004; Gollan et al., 2008), which become more evident when task demands are high (e.g., during listening-in-noise tasks). In other words, since bilinguals can only produce one language at a time, the amount of practice using representations in each language will be less compared to monolinguals (for whom lexical entries in their sole



language are used 100% of the time). In addition, based on inhibitory control theories (e.g., Green, 1986, 1998), there are lexical nodes for each language that are simultaneously activated by a single concept and can compete; that is, when a Spanish–English bilingual sees a cat, both the lexical entries in English (*cat*) and Spanish (*gato*) are activated (Shook & Marian, 2013). In order to then retrieve the desired word (in the target language), bilinguals must suppress the lexical nodes in the non-target language. Based on this view, bilinguals may have greater difficulty during lexical retrieval because of the additional need to inhibit the lexical nodes in the non-target language.

One important consideration is whether the pattern of results observed can be generalized to other “types” of background noises, specifically to noises which children might encounter in their everyday life (e.g., talker babble). Prior work suggests that when there is competing noise present, the segments of the speech signal can be covered (or masked) by the competing noise; this leads to a reduction in the linguistic and acoustic cues that are available to the listener, making it harder to understand the message (Helfer & Wilber, 1990). Regardless of the type of noise, this “energetic masking” effect appears to be more and more pronounced as the SNR decreases, and the target signal becomes too weak to be reliably understood amid the competing noise (Miller et al., 1951).

Background noise that contains meaningful information (e.g., speech in a known language), in addition leads to what is known as “informational masking” (review in Kidd et al., 2007). Previous findings suggest that this informational masking results in: (a) greater interference during stream-segregation tasks (Cooke et al., 2008; Simpson & Cooke, 2005; Van Engen & Bradlow, 2007); and (b) processing difficulties at a higher level of language processing compared to noise that only produces energetic masking (e.g., white noise). The present study only used white noise, and our findings suggest that the bilingual disadvantage during speech-processing in noise was present even when the noise would be expected to only result in energetic masking. There is no reason to think that this effect would disappear with competing noise that also resulted in informational masking (i.e., an arguably harder listening situation). If anything, we would expect that the bilingual disadvantage would be even greater with a type of noise that contained meaning.

Another important factor to further discuss is the SNR. Here we attempted to explore this topic by including two different SNRs. In general, we had expected accuracy to be lower when the background noise was the same intensity as the target speech (i.e., the 0 dB SNR), compared to when the target speech was 5 dB more intense than the background noise. Instead, we found that performance across the two SNRs was the same in both monolinguals and bilinguals. One possibility is that the two SNRs were not different enough to result in a measurable difference during our task. Given young children’s limited attention span, it was not feasible to test additional SNRs within the same testing session. Future studies should, however, expand on this topic by evaluating performance with different SNRs.

In addition to limitations associated with the type of noise and the number of trials that were included in our study, there are other elements that should be addressed in follow-up work. First, our study only included children being raised in mid-SES homes, which limits the ability to generalize the findings to larger groups of bilingual children. In fact, Latino dual-language learners are the largest segment of bilingual children in the United States, and they are three times more likely to grow up in poverty than non-Latino children (Haskins et al., 2004). Poverty makes this group more likely to experience not only low-quality education and housing, but considerably more noise in their environment (Evans, 2004; Evans & Kantrowitz, 2002). It is therefore critical to understand how the ability to process speech in noise and language background interact in a more diverse group, and how experience with greater noise in the home might impact subsequent cognitive behavior in a laboratory setting.

Second, our study only examined word recognition, and it is important to consider that the demands associated with the type of linguistic task may play a role in bilingual performance. Most

studies to date, including this one, have focused on whether or not monolingual and bilingual listeners can comprehend familiar words heard in noise – a task that relies primarily on accessing previously-stored knowledge to identify particular words. Word learning, on the other hand, relies initially on a different set of processes that make use of mostly newly-acquired information, where factors such as cross-linguistic competition and recency of lexical use may play less of a role. It is therefore worth expanding this work to other types of language tasks, as they might shed greater light on the underlying cause of these differences (e.g., weaker representations via lack of use, versus inherent competition between languages, versus competition particular to words for which two language variants are known).

Taken together, the fact that bilinguals as young as 30 months show poorer word recognition (compared to monolinguals) when noise is present suggests that while children's productive vocabulary is still relatively small at this age, like adults, bilingual toddlers already have weaker lexical representations. Bilinguals' poorer performance at speech perception in noise at such a young age clearly has implications for school and daycare settings. While this work does not provide information about the exact point in development when bilingualism begins to play a role in word recognition in noise, finding group differences between monolinguals and bilinguals at 30 months does suggest that this pattern emerges very early on. It also suggests that lexical competition is a fundamental part of the bilingual experience from toddlerhood and across the lifespan. Our findings indicate that the development of linguistic and cognitive abilities in bilinguals are not primarily tied to elements such as vocabulary size (which changes over time), but instead might be associated with the early organization of the lexical system.

More specifically, our findings, as well as the prior literature, suggest that bilingualism (from an early age) leads to some challenges for understanding speech in noise. This has important implications for educational settings, which typically are quite noisy – and for learning. Several studies have evaluated the types of noise that are present in classrooms, and the effects that these might have on young learners. This work suggests: (a) that noise in classrooms comprises both internal sounds produced by the children and/or teacher, but also external noises in the environment that are transmitted through the building envelope (i.e., children are exposed to noise from a variety of sources including noise from the ventilation system and from traffic near the school); and (b) noise has a direct relation to children's performance in the classroom (Dockrell & Shield, 2006; Hygge et al., 2002; Shield & Dockrell, 2003). Researchers have suggested that second-language learners might benefit from modifications of classroom acoustics (e.g., Picard & Bradley, 2001). The current work suggests that even early/balanced bilinguals are hampered by noise and would hence also benefit from these modifications. This work can also inform clinical practices, given that for early intervention (i.e., treating speech and language delays in children from birth to 3 years) speech-language-pathologists often conduct treatment in homes and daycares, where noise is present. Based on our work, finding a relatively quiet space to conduct therapy might be particularly important when working with bilingual clients, and discussing the effects of noise with caregivers of bilingual children, along with options on how to reduce internal sound sources in the home (e.g., limiting the amount of time the TV is on in the background) would be beneficial.

To conclude, the present work supports the notion that even in early childhood, the language background of the individual plays a role in the organization and accessing of lexical knowledge. This leads to differences in performance between monolingual and bilingual children (with bilinguals experiencing greater processing costs) in tasks that require the listener to segregate competing sounds and recognize previously acquired words.

## Acknowledgements

This work was part of the first author's PhD dissertation; we thank committee members: Colin Phillips; Nan Bernstein Ratner; Yi Ting Huang; and Jared Novick for helpful advice. We thank Elizabeth Johnson for guidance throughout the data-collection process, for providing access to participants and testing facilities in Canada, and for the coding-scripts utilized for data analysis. We also thank Molly Nasuta for help with coding, and Maura O'Fallon for audio recording. We are grateful to the members of the Language Development Laboratory and the Infant Language and Speech Laboratory for assistance with scheduling and testing.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by a Dissertation Improvement Grant (Grant Number: NSF BCS-1322565), by a Postdoctoral Fellowship (Grant Number: SBE-SMA-1514493) from the National Science Foundation, by a University of Maryland Center for Comparative and Evolutionary Biology of Hearing Training Grant (Grant Number: NIH T32 DC000046-17), and by start-up funds from the University of Delaware.

## ORCID iD

Giovanna Morini  <https://orcid.org/0000-0002-0034-8213>

## References

- Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; infants do not. *Psychological Science*, 5(3), 170–174. <https://doi.org/10.1111/j.1467-9280.1994.tb00655.x>
- Busch-Vishniac, I. J., West, J. E., Barnhill, C., Hunter, T., Orellana, D., & Chivukula, R. (2005). Noise levels in Johns Hopkins Hospital. *Journal of the Acoustical Society of America*, 118(6), 3629–3645. <https://doi.org/10.1121/1.2118327>
- Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences of the United States of America*, 114(34), 9032–9037. <https://doi.org/10.1073/pnas.1703220114>
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, 119(3), 1562–1573. <https://doi.org/10.1121/1.2166600>
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, 123(1), 414–427. <https://doi.org/10.1121/1.2804952>
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1), 125–127. <https://doi.org/10.3758/BF03203646>
- Dockrell, J. E., & Shield, B. M. (2006). Acoustical barriers in classrooms: The impact of noise on performance in the classroom. *British Educational Research Journal*, 32(3), 509–525. <https://doi.org/10.1080/01411920600635494>
- Ecke, P. (2004). Words on the tip of the tongue: A study of lexical retrieval failures in Spanish-English bilinguals. *Southwest Journal of Linguistics*, 23, 33–63.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, 59(2), 77–92. <https://doi.org/10.1037/0003-066X.59.2.77>
- Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health*, 23(1), 303–331. <https://doi.org/10.1146/annurev.publhealth.23.112001.112349>

- Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development, 78*(5), 1510–1525. <https://doi.org/10.1111/j.1467-8624.2007.01080.x>
- Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science, 9*(3), F33–F40. <https://doi.org/10.1111/j.1467-7687.2006.00482.x>
- Fernald, A., Swingle, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*(4), 1003–1015. <https://doi.org/10.1111/1467-8624.00331>
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception SRT for impaired and normal hearing. *Journal of the Acoustical Society of America, 88*(4), 1725–1736. <https://doi.org/10.1121/1.400247>
- Florentine, M. (1985a). Non-native listeners' perception of American-English in noise. *Proceedings of Inter-noise, 85*, 1021–1024.
- Florentine, M. (1985b). Speech perception in noise by fluent, non-native listeners. *Proceedings of the Acoustical Society of Japan*, H-85-16.
- Florentine, M., Buus, S., Scharf, B., & Canévet, G. (1984). Speech perception thresholds in noise for native and non-native listeners. *Journal of the Acoustical Society of America, 75*, 84. <https://doi.org/10.1121/1.2021645>
- Frank, T., & Golden, M. V. (1999). Acoustical analysis of infant/toddler rooms in daycare centers. *Journal of the Acoustical Society of America, 106*(4), 2172. <https://doi.org/10.1121/1.427242>
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. <https://doi.org/10.1037/0033-2909.134.1.31>
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45. <https://doi.org/10.1017/S030500090001271X>
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787–814. <https://doi.org/10.1016/j.jml.2007.07.001>
- Green, D. W. (1986). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language, 27*(2), 210–223. [https://doi.org/10.1016/0093-934X\(86\)90016-7](https://doi.org/10.1016/0093-934X(86)90016-7)
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language & Cognition, 1*(2), 67–81. <https://doi.org/10.1017/S1366728998000133>
- Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. de Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225–253). Erlbaum.
- Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.
- Gustafsson, H. Å., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *Journal of the Acoustical Society of America, 95*(1), 518–529. <https://doi.org/10.1121/1.408346>
- Haith, M., Wentworth, N., & Canfield, R. (1993). The formation of expectations in early infancy. In C. Rovee Collier, & L. P. Lipsitt (Eds.), *Advances in infancy research (Vol 11)*, pp. 251–297). Ablex.
- Haskins, R., Greenberg, M., & Fremstad, S. (2004). Federal policy for immigrant children: Room for common ground? In *The future of children (Vol. 14)*, pp. 1–5). Brookings Institution.
- Helfer, K., & Wilber, L. (1990). Hearing loss, aging, and speech perception in reverberation and in noise. *Journal of Speech & Hearing Research, 33*(1), 149–155. <https://doi.org/10.1044/jshr.3301.149>
- Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting* (2nd ed., pp. 231–252). Erlbaum.
- Hoff, E., & Naigles, L. (2002). How children use input in acquiring a lexicon. *Child Development, 73*(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- Hollich, G. (2005). *Supercoder: A program for coding preferential looking* (Version 1.5) [Computer software]. Purdue University.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language, 34*(2), 227–249. <https://doi.org/10.1017/S0305000906007896>

- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236–248. <https://doi.org/10.1037/0012-1649.27.2.236>
- Hygge, S., Evans, G. W., & Bullinger, M. (2002). A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren. *Psychological Science, 13*(5), 469–474. <https://doi.org/10.1111/1467-9280.00483>
- Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science, 6*(5), 498–504. <https://doi.org/10.1111/1467-7687.00307>
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational masking. In W. Yost (Ed.), *Springer handbook of auditory research, 29: Auditory perception of sound sources* (pp. 143–190). Springer.
- Krizman, J., Bradlow, A. R., Lam, S. S. Y., & Kraus, N. (2017). How bilinguals listen in noise: Linguistic and non-linguistic factors. *Bilingualism: Language and Cognition, 20*(4), 834–843. <https://doi.org/10.1017/S1366728916000444>
- MacKay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics, 25*(3), 373–396. <https://doi.org/10.1017/S0142716404001171>
- Mayo, L., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3), 686–693. <https://doi.org/10.1044/jslhr.4003.686>
- Meador, D., Flege, J. E., & Mackay, I. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language & Cognition, 3*(1), 55–67. <https://doi.org/10.1017/S1366728900000134>
- Miller, G., Heise, G., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the speech materials. *Journal of Experimental Psychology, 41*(5), 329–335. <https://doi.org/10.1037/h0062491>
- Morini, G., & Newman, R. S. (2019). Dónde está la ball? Examining the effect of code switching on bilingual children's word recognition. *Journal of Child Language, 46*(6), 1238–1248. <https://doi.org/10.1017/S0305000919000400>
- Morini, G., & Newman, R. S. (2020). Monolingual and bilingual word recognition and word learning in background noise. *Language and Speech, 63*(2), 381–403. <https://doi.org/10.1177/0023830919846158>
- Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25*(1), 95–120. <https://doi.org/10.1017/S0305000997003358>
- Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology, 41*(2), 352–362. <https://doi.org/10.1037/0012-1649.41.2.352>
- Newman, R. S. (2009). Infants' listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics, 71*(4), 822–836. <https://doi.org/10.3758/APP.71.4.822>
- Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics, 58*(8), 1145–1156. <https://doi.org/10.3758/bf03207548>
- Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*(1), 74–87. <https://doi.org/10.1111/j.1467-7687.2005.00394.x>
- Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms: Revisando la interferencia en el habla dentro del salón de clases. *International Journal of Audiology, 40*(5), 221–244. <https://doi.org/10.3109/00206090109073117>
- Pollack, I., & Pickett, J. M. (1958). Stereophonic listening and speech intelligibility against voice babble. *Journal of the Acoustical Society of America, 30*(2), 131–133. <https://doi.org/10.1121/1.1909505>
- Reetzke, R., Lam, B. P. W., Xie, Z., Sheng, L., & Chandrasekaran, B. (2016). Effect of simultaneous bilingualism on speech intelligibility across different masker types, modalities, and signal-to-noise ratios in school-age children. *PLoS One, 11*(12), e0168048. <https://doi.org/10.1371/journal.pone.0168048>. eCollection 2016
- Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*(4), 587–599. <https://doi.org/10.1044/jshd.5404.587>

- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, 27(3), 465–485. <https://doi.org/10.1017/S014271640606036X>
- Shi, L.-F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech, Language, and Hearing Research*, 53(4), 821–835. [https://doi.org/10.1044/1092-4388\(2010/09-0081\)](https://doi.org/10.1044/1092-4388(2010/09-0081))
- Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics*, 10(2), 97–116. <https://doi.org/10.1260/135101003768965960>
- Shin, H. B., & Ortman, J. M. (2011). *Language projections: 2010 to 2020* [Conference session]. Federal Forecasters Conference, Washington, DC. <https://www.census.gov/content/dam/Census/library/working-papers/2011/demo/2011-Ortman-Shin.pdf>
- Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, 16(2), 304–324. <https://doi.org/10.1017/S1366728912000466>
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America*, 118(5), 2775–2778. <https://doi.org/10.1121/1.2062650>
- Tabri, D., Chacra, K. M. S. A., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders*, 46(4), 411–422. <https://doi.org/10.3109/13682822.2010.519372>
- 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(1), 519–526. <https://doi.org/10.1121/1.2400666>
- Wilson, R. H., & Carhart, R. (1969). Influence of pulsed masking on the threshold for spondees. *Journal of the Acoustical Society of America*, 46(4), 998–1010. <https://doi.org/10.1121/1.1911820>

### Author biographies

Giovanna Morini is an assistant professor in the Department of Communication Sciences and Disorders at the University of Delaware. She studies child language development and bilingualism.

Rochelle S. Newman is a professor and department chair in the Department of Hearing & Speech Sciences at the University of Maryland, College Park, as well as a member of the Maryland Language Science Center. She studies speech perception and word recognition in difficult listening conditions (such as noise), particularly in young children.