Effects of Auditory Masking on Children's Use of Semantic Context

Lauren Staiger

1. Abstract

More often than not, children spend their lives in noisy environments where they must learn and listen to language. Although spoken word recognition is a well-documented process in quiet conditions, it remains unclear how background noise influences the mechanisms involved in child speech perception The current study investigated how children's ability to use informative semantic sentence cues to facilitate lexical access of an upcoming word and suppress similar sounding competitors is impacted by speech-shaped noise and two-talker babble noise. We hypothesized that children would use semantic context in the informative condition to facilitate lexical access of the target and suppress cohort competitors despite the presence of maskers. Additionally, we hypothesized that over time, children would be slower to fixate on the target image and would sustain fixations to the cohort image for longer in the two-talker babble condition. Eye tracking equipment monitored eye gazes across four images during the presentation of a sentence stimulus masked with background noise. In both the two-talker babble and speech-shaped noise maskers, children were able to use semantic context to facilitate access of a target word and suppress cohort competition. Although children demonstrated a trend toward prolonged consideration of the cohort competitor in the two-talker babble condition, there was no significant effect of noise type when considering target fixations, potentially suggesting that at high signal-to-noise ratios, children are able to benefit from semantic context when masked by both two-talker babble and speech-shaped noise. Overall, these findings highlight the importance of researching complex auditory environments where children spend much of their time learning and listening to spoken speech.

2. Introduction

2.1. Word Recognition

Speech recognition is a complex process that occurs over time, resulting in momentary ambiguities in the perception of the signal. For example, when listeners hear the sounds /bʊ…/, they may consider multiple possibilities for the upcoming words that are consistent with the perceived phonemes, such as *bush* or *book* (Allopenna et al., 1998; Marslen-Wilson, 1987). Listeners incrementally process incoming phonemes to activate similar sounding lexical candidates and inhibit competitors that are no longer consistent with the phonemic input. Cohorts, or lexical possibilities that share the same initial phonemes as the target, contribute to some of the competition listeners experience when listening to spoken words. As non-target words, such as cohorts, are suppressed, the target word is activated and mapped to its lexical representation (Allopenna et al., 1998). Young children also demonstrate similar processes of cohort activation and inhibition when listening to spoken words (Blomquist et al., 2023; Swingley et al., 1999), but do so less efficiently than adults (Sekerina & Brooks, 2007). Younger elementary aged children take longer to activate target lexical items and sustain activation of phonological competitors compared with older children, yet improve in efficiency with development (Jeppsen et al., 2024). For typical listeners, this process of activation and inhibition allows for flexibility and accuracy to overcome momentary signal ambiguities and ensure efficient access of lexical targets in spoken word recognition.

2.2. Semantic Context

Listeners also integrate semantic sentence context over time to resolve speech signal ambiguities and efficiently access potential upcoming lexical targets. Semantic prediction involves using meanings of words earlier in a sentence to restrict upcoming lexical possibilities, allowing for quicker lexical access and competitor suppression (Altmann & Kamide, 1999; Blomquist et al., 2023, Mani & Huettig, 2012). For example, upon hearing the sentence *"The dad reads the green book"*, listeners are able to narrow down upcoming lexical possibilities based on existing knowledge of the verb *reads*, allowing faster activation of the target word *"book"* and suppression of phonological competitors, such as *"bush"*. Children listening in ideal conditions have demonstrated these prediction abilities with verbs (Mani & Huettig, 2012; Nation et al., 2003), which develop across the course of development. In Blomquist et al. (2023), children participated in the visual world paradigm, an experimental task in which eye gazes are tracked while participants view images and listen to a spoken utterance. Participants were presented with informative or uninformative sentences. The stimuli contained a subject, a semantically informative or uninformative verb, a determiner, a neutral adjective, and a noun likely to follow the informative verb to create sentences such as *"The dad reads/finds the green book"*. Upon hearing the spoken sentences, children were instructed to select one of four images on a computer screen that matched the final word from the auditory stimuli. Eye gazes revealed that the semantically informative sentence contexts facilitated activation of the target word and suppression of the cohort competitor in children as young as 5 years old. Even in complex visual environments, children and adults exhibit these incremental prediction processes that aid efficient word recognition (Reuter et al., 2020).

2.3. Complex Auditory Environments

Even in ideal listening conditions, spoken word recognition is a complex process that occurs incrementally over the course of the speech stimuli. Many listeners, including young children, spend a significant amount of time in less than ideal listening environments. In these complex auditory environments, listeners must effectively perceive and comprehend target speech signals amidst competing sounds, commonly referred to as the "cocktail party problem" (Cherry, 1953). Although children do not attend cocktail parties, most children spend time in daycares, classrooms, or playgrounds, which can quickly become noisy environments (Manlove et al., 2001) that affect children's academic performance (Shield & Dockrell, 2008). In a classroom, children are tasked with listening to a teacher's voice in the presence of many competing sounds, such as other children's speech, outside traffic noise, air conditioning, and more.

In contexts with multiple signals, the cochlea transmits information to the central auditory system, where listeners must perceptually group sounds that belong to the same source and segregate the groups from other sources within the "auditory scene" (Bregman, 1990). By using acoustic information, such as spectral, temporal, and spatial cues, as well as learned world knowledge, listeners are often able to effectively separate sound sources from one another. When the presence of a competitor signal raises the threshold needed to detect the target speech stream, this is referred to as masking (Pollack, 1975). Energetic masking is primarily attributed to physical overlap and distortion that occurs in the cochlea when multiple sounds have overlapping spectral energy (Brungart, 2001). Even when limiting the effects of energetic masking, when the target and competing signals should be audible, listeners can still experience increased thresholds beyond what is expected, broadly termed informational masking (Brungart et al., 2006). Although the classifications of masking are interrelated and difficult to separate in real-life stimuli, speech maskers composed of a small number of talkers, such as two-talker babble (TTB), are believed to produce informational masking effects in addition to energetic effects, while speech shaped noise maskers (SSN) produce primarily energetic masking effects (Brungart et al., 2006). In comparison to adults, children typically perform poorer overall in masked speech recognition tasks, which becomes more pronounced in TTB conditions when compared to SSN conditions (Corbin et al., 2016; Hall et al., 2002; McCreery et al., 2020; Wightman & Kistler, 2005).

Perceptual segregation difficulties that listeners may experience in informational masking tasks are influenced by many cognitive factors, including linguistic knowledge, selective attention, and working memory (Buss et al., 2016; McCreery et al., 2020; Wightman & Kistler, 2005). McCreery et al. (2020) investigated the relationship between the high level cognitive processes of working memory and selective attention, language ability, and speech masking in developmental groups of children aged 5 to 6 and 9 to 10 years. Participants completed a variety of assessments that measured receptive language, short-term and working memory, attention, inhibitory control, and cognitive flexibility. Children were instructed to repeat target sentences that were presented in TTB or SSN, and signal-to-noise ratio (SNR) was adjusted based on response accuracy. Researchers determined that in the TTB masker condition, robust abilities in language, working memory, and selective attention skills were associated with better sentence recognition. Improved speech recognition in the SSN condition was only associated with vocabulary ability for 5 to 6 year old children. Overall, this study suggests that higher level cognitive and linguistic abilities interact with speech recognition in masked conditions, but vary based on masker type and individual child age.

In degraded auditory environments, children demonstrate the ability to use available semantic information to facilitate sentence recognition (Fallon et al., 2002), but do so less efficiently than adults (Nittrouer & Boothroyd, 1990). Previous work on children's word recognition with a closed set choice task suggests that children may benefit less from context information in the presence of speech maskers (Buss et al., 2016). Additionally, Buss et al. (2019) explored the effects of TTB and SSN maskers on sentence contexts across the lifespan. Speech reception thresholds (SRTs) were measured for high and low semantic contexts, demonstrating that children aged 5 to 16 years benefit less from semantic context when masked by TTB relative to SSN. Researchers suggested that this finding reflects the cognitive demands of parsing the target speech signal from a speech masker, an ability which may be limited in children.

2.4. Current Study

Previous research revealed that children take advantage of informative semantic context to narrow down possible upcoming lexical targets and inhibit competitor words to improve sentence recognition (Blomquist et al., 2023; Mani & Huettig, 2012). In addition, children exhibit poorer performance in tasks with TTB maskers in comparison to those with SSN maskers (Buss et al., 2016; Buss et al., 2019; Corbin et al., 2016; Hall et al., 2002). However, it remains unclear how these findings interact over time. It is possible that across semantic conditions, children would be slower to fixate on the target image and sustain consideration of the cohort in the TTB noise condition in comparison to the SSN condition, which would align with previous research indicating that children experience more interference from TTB conditions. In the TTB condition, the difference between the informative and neutral conditions may be less than the difference in the SSN condition, as the TTB masker may interfere with the benefit of semantic context to a greater extent. These possibilities would support previous research that masked speech can create ambiguous moments of uncertainty in which children may perform poorer in speech recognition tasks.

Therefore, the current study investigated the influence of different maskers and semantic sentence context on children's word recognition. Specifically, this study questioned if there is a difference between the time-course of children's word recognition in a sentence context in a TTB condition and a SSN condition. We hypothesized that in both masked conditions, children would still use semantic context in the informative condition to facilitate lexical access of the target and suppress cohort competitors. Additionally, we hypothesized that over time, children would be slower to fixate on the target image and would sustain fixations to the cohort image for longer in the TTB condition. If children experience more interference from a TTB masker, then the difference between the proportions of eye gazes between sentence conditions may be smaller than the difference found in the SSN condition. This is based on previous research suggesting that speech maskers require cognitive and linguistic resources that may not be as well developed in children, resulting in poorer performance (Buss et al., 2016; McCreery et al., 2020; Wightman & Kistler, 2005).

3. Methods

3.1. Participants

Participants were 16 children aged 5 to 7 years old (mean: 6;3, range: 5;4-7;2, 10 males) with typical hearing abilities. An additional three children participated but were not included in data analysis due to inability to complete the task (n=1), parent reported developmental disability (n=1), and out of age range (n=1).This age range was selected in consideration of kindergarten level language abilities and previous findings that suggest children with typical development in this age range are capable of efficiently using semantic prediction to inform prediction of possible upcoming lexical targets (Blomquist et al., 2023). Hearing screenings were conducted in a sound booth using a PC controlled audiometer with behavioral responses to tones at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz at 20dB. For participants unable to complete the hearing screening due to technological issues (n=6), typical hearing abilities were confirmed by parent report. All participants were learning English as their dominant language and had no reported intellectual or communicative disability based on parent reports. Institutional review board approval was obtained from the University of Maryland.

3.2. Materials

The auditory stimuli contain a primary speech stream and competing background noise. The primary stimuli comprises 18 sentence pairs (the same as Blomquist et al., 2021; Blomquist et al., 2023) in which one sentence contains a semantically informative verb (*reads*) and the other contains a semantically uninformative verb (*finds*). A native female speaker of Mainstream American English recorded the sentences in the following structure: "The [subject] [semantically informative or uninformative verb] [determiner] [neutral adjective] [noun likely to follow the informative verb]". For example, a stimuli pair may consist of the sentences "The dad reads the green book" and "The dad finds the green book". All of the sentence components had an age of acquisition less than 6 years (Blomquist et al., 2023; Kuperman et al., 2012). An additional 18 filler trials containing the uninformative verb followed by a cohort competitor noun (e.g., tractor), which is a word beginning with the same phonemes as the target word (e.g., trophy), are included.

The prerecorded target sentences were masked with two noise conditions using Audacity. The first masker is male TTB noise from Calandruccio et al. (2014) which was composed from two speakers reading passages from *Jack and the Beanstalk*. Before blending the audios, the two male masker recordings were edited to remove pauses greater than 300 ms. The second condition maximizes energetic masking through a SSN masker which is based on the male two-talker babble speech. The target sentences were root mean square (RMS) normalized and calibrated for presentation at 70 dB SPL. The target and masker were presented at an SNR of 10dB. This SNR was chosen to maximize accuracy in both conditions, as children aged 5 to 7 have been found to have lower average detection thresholds of 5.4 dB SNR for TTB and -0.5 dB SNR for SSN (Corbin et al., 2016). For each sentence stimulus, the noise masker gradually increased in intensity for 100 ms and began 400 ms before the target sentence. Following the target sentence, the noise masker continued for 400 ms and decreased in intensity for 100 ms. The stimuli were presented from loudspeakers that were calibrated with a sound level meter.

Visual stimuli consisted of four images in each trial: a target image matching the noun likely to follow the informative verb, a cohort competitor, and two unrelated images. For example, for the sentences “The child likes the grey cat” and “The child pets the grey cat”, the images included an image of a cat (target), keys (cohort competitor), a bike, and a bag (unrelated).

3.3. Procedure

Participants were seated in front of a computer monitor to begin the task. After the external computer and eye-tracking computer was turned on, children were then invited to follow along with game characters as they "capture items" throughout the task. The eye-tracking device was calibrated to each child's eye by using a five point cross setup. Following calibration, the experimenter ran a validation task before beginning the trials. Children saw four images on a computer monitor for 1200 ms before an attention-getting bullseye appeared in the center of the screen. After looking at the center of the bullseye for 800 ms, the trial began. Auditory stimuli played through speakers and participants pointed to the image that matched the last word of the sentence. Experimenters selected the image chosen by the child with a mouse-click. All eye movements were tracked on an Arm-Mounted EyeLink 1000 Plus (SR Research, Ottawa, Ontario, Canada). Children completed the task in two blocks, with each block containing 36 target sentences and 9 filler sentences. Experimental blocks were counterbalanced, as one contained sentences masked by two-talker babble noise, while the other block contained the speech-shaped noise condition. The language and hearing assessments were conducted following the eye-tracking task.

Parents completed two speech, language, and hearing questionnaires for their child that asked open-ended or numerical rating questions about the child's social and linguistic development in comparison to their peers (Hadley & Rice, 1993), as well as questions about the child's history and noise exposure. Participants completed a semantic verbal fluency task to measure semantic memory, in which children were instructed to verbally list as many animals as possible in one minute. Following this task, children participated in the hearing screening to ensure hearing abilities were within normal limits.

3.4. Data Cleaning

Eye gazes were sampled every 4 ms from the beginning of the trial following fixation at the center of the screen. Data cleaning was done in RStudio using the eyetrackingR package (Dink & Ferguson, 2015). Areas of interest (AOIs) were defined as 450 by 450 pixel squares where the images were displayed on the screen, which extended the boundaries of objects by 30 pixels to account for variability in eye movements. Critical trials were isolated by filtering filler trials and incorrect responses. Only trials with correct responses were included in analyses. Critical trials that exceeded a 0.5 track loss threshold, or moments where the eye tracking system failed to record eye gaze, were also removed. Of the critical trials, approximately 2% were removed due to track loss. All data cleaning, including filler trial filtering, resulted in a loss of 322 trials, resulting in 1073 critical trials remaining.

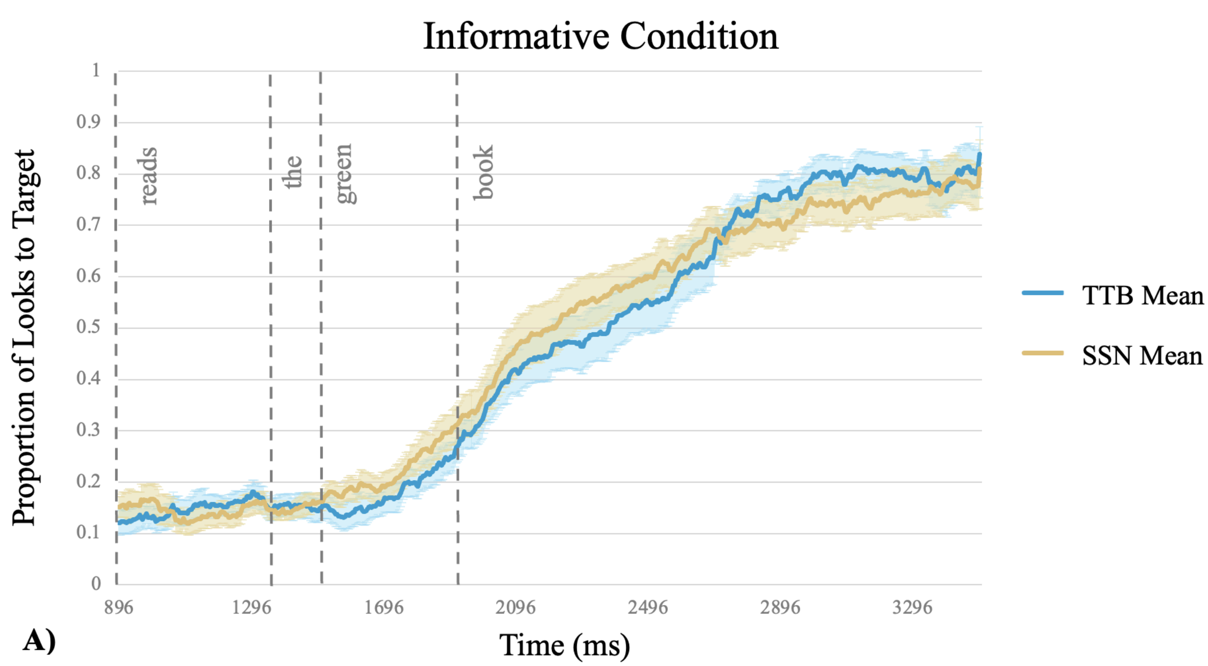
4. Results

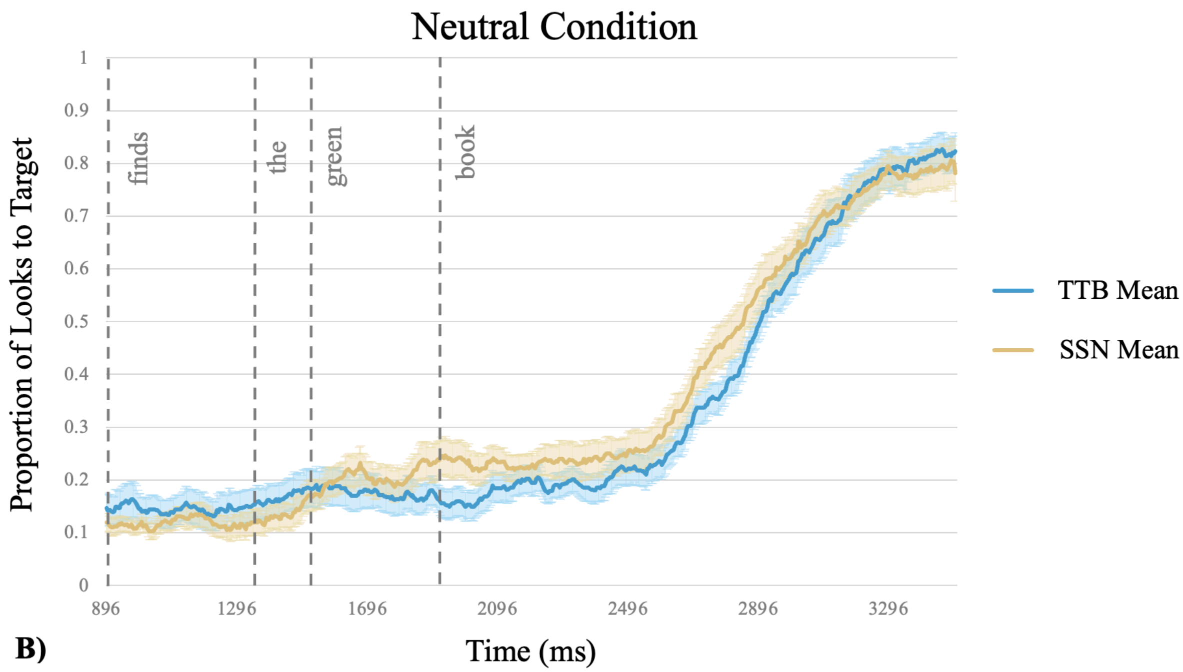
4.1. Response Accuracy

All participants achieved above 85% accuracy for every block and condition, although only four blocks across all participants contained accuracy lower than 95%. There was no significant difference in accuracy between noise conditions ($p > 0.1$).

4.2. Eye Gaze

Target fixations (Figure 1) were analyzed from 200 ms after the verb onset to approximately 1290 ms after the target noun onset. Since cohort competition depends on the phonetic input from the target noun, cohort fixations (Figure 2) were analyzed within a smaller time window from the target word onset, which occurred at approximately 1900 ms, to 1290 ms after the onset, or roughly 3200 ms. Within the time windows, proportion of looks were calculated as looks to the AOI over total looks to AOIs across the screen (Figure 3). Cohort eye gazes were isolated by subtracting the proportion of fixations toward the unrelated AOIs. A multiple linear regression model was performed in Excel using the Analysis ToolPak for all of the following analyses. The regression model predicted average eye gaze within the time windows of analyses with predictor variables of noise condition, semantic context, and verbal fluency. For categorical variables of noise condition and semantic context, numerical dummy coding was used. The reference level for semantic context was the informative condition and the TTB noise was used as a reference for the noise condition. Models with the interaction revealed no significant interactions between conditions ($p > 0.1$). The following results are from the Excel analyses.



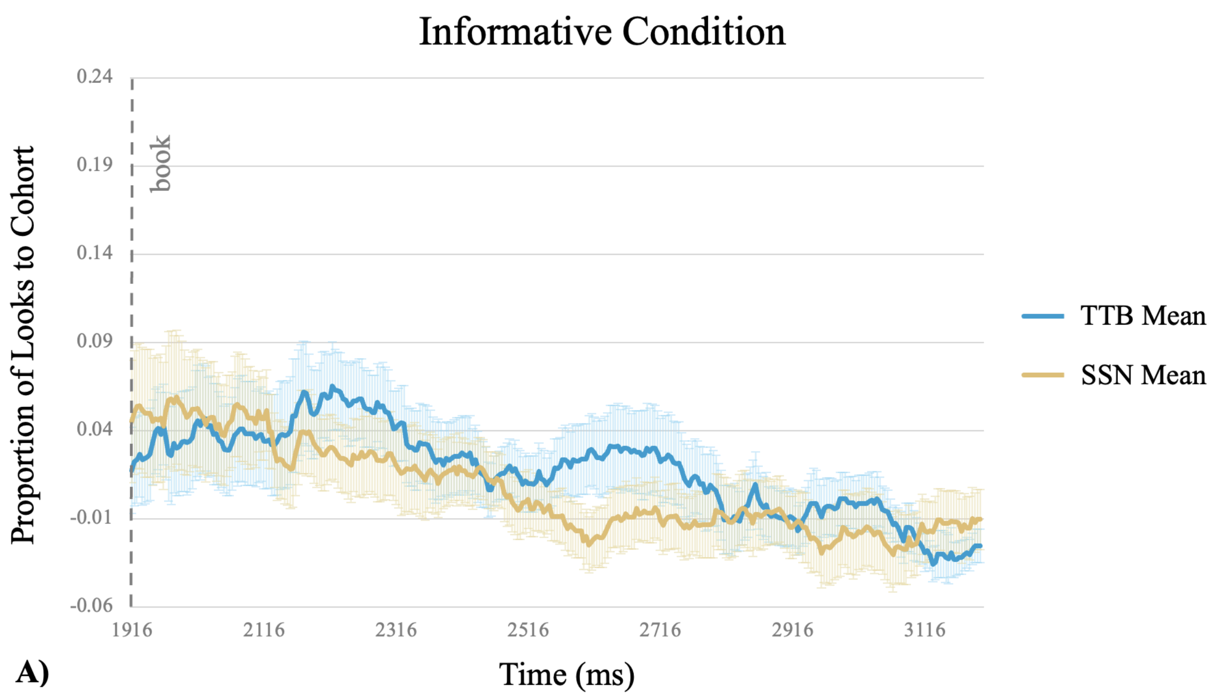


B)

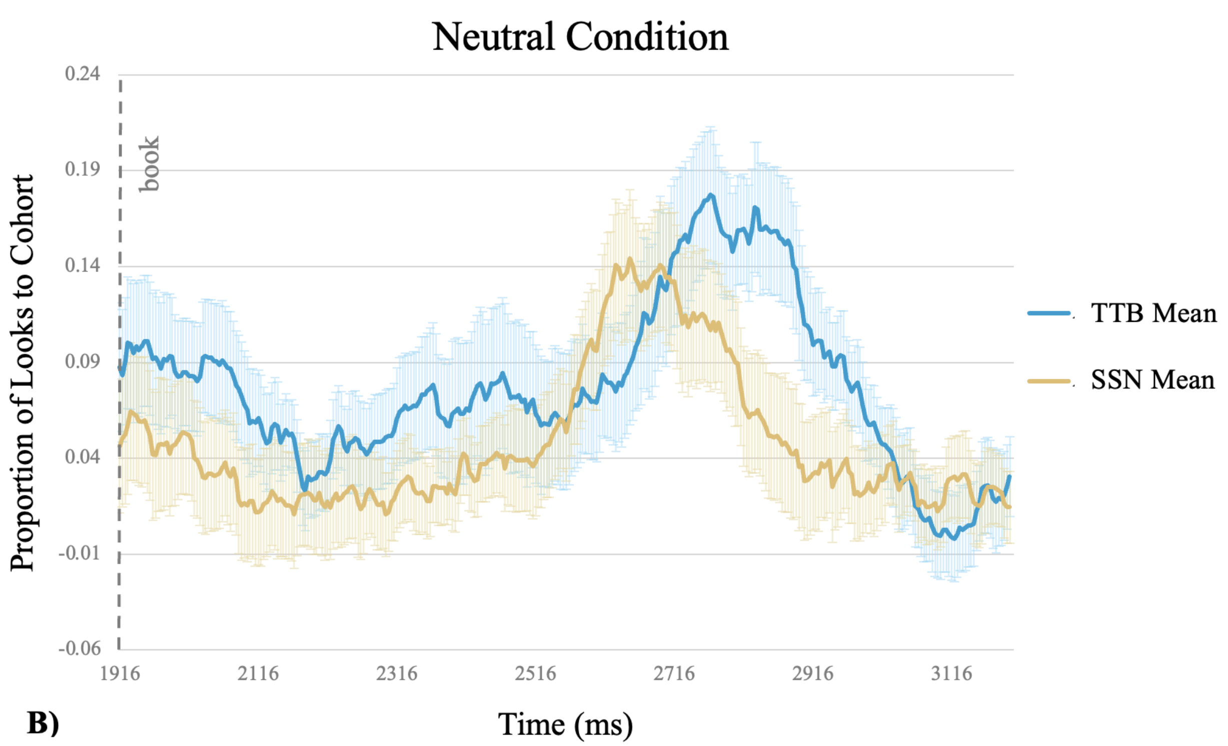
Fig. 1. Time window analysis of target eye gaze data for two-talker babble and speech-shaped noise conditions. **A)** Time window graph of proportion of looks to target in the informative verb condition. **B)** Time window graph of proportion of looks to target in the neutral verb condition. Dotted error lines represent standard error of the mean. Dashed lines indicate average onset of each word with a 200 ms delay.

4.3. Target looks

The model of target fixations revealed a significant semantic condition effect, with both noise conditions demonstrating more looks toward the target image in the informative condition compared to the neutral condition (β = -0.14, SE = 0.02), t(58) = -6.58, p < .001). In contrast to previous research (Buss et al., 2019*)*, there was no significant effect of TTB condition compared to the SSN condition (β = 0.02, SE = 0.02), t(58) = 0.93, p > .1).


Informative Condition
0.24
0.19
0.14
0.09
0.04
-0.01
-0.06
book
Proportion of Looks to Cohort
1916
2116
2316
2516
2716
2916
3116
Time (ms)
TTB Mean
SSN Mean

A)

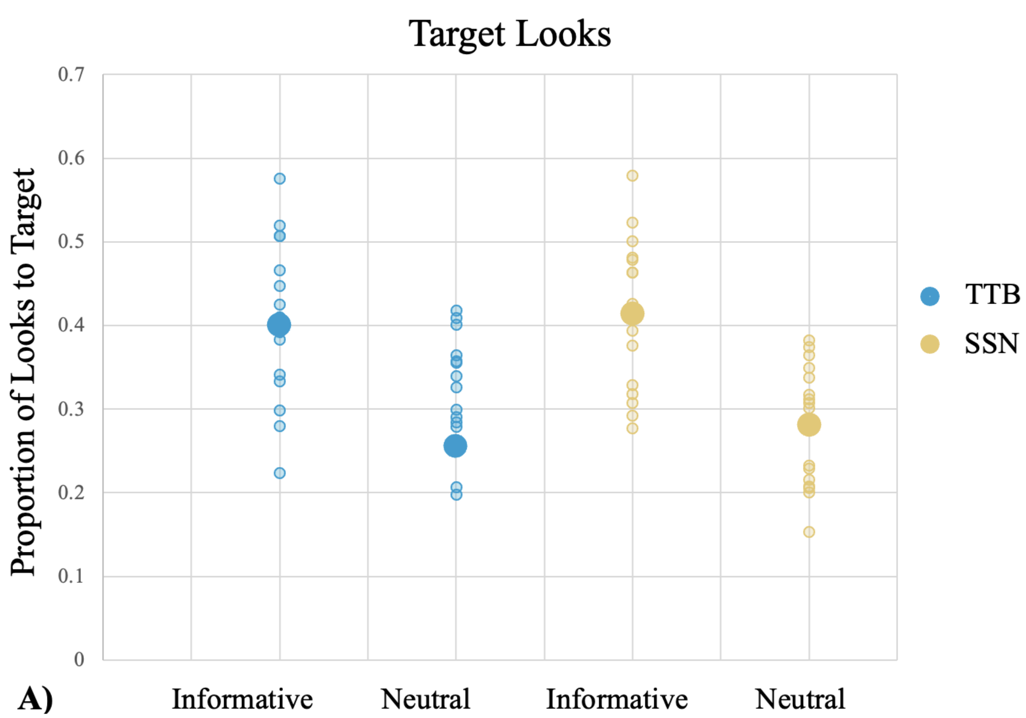

Neutral Condition
0.24
0.19
0.14
0.09
0.04
-0.01
-0.06
book
Proportion of Looks to Cohort
1916
2116
2316
2516
2716
2916
3116
Time (ms)
TTB Mean
SSN Mean

B)

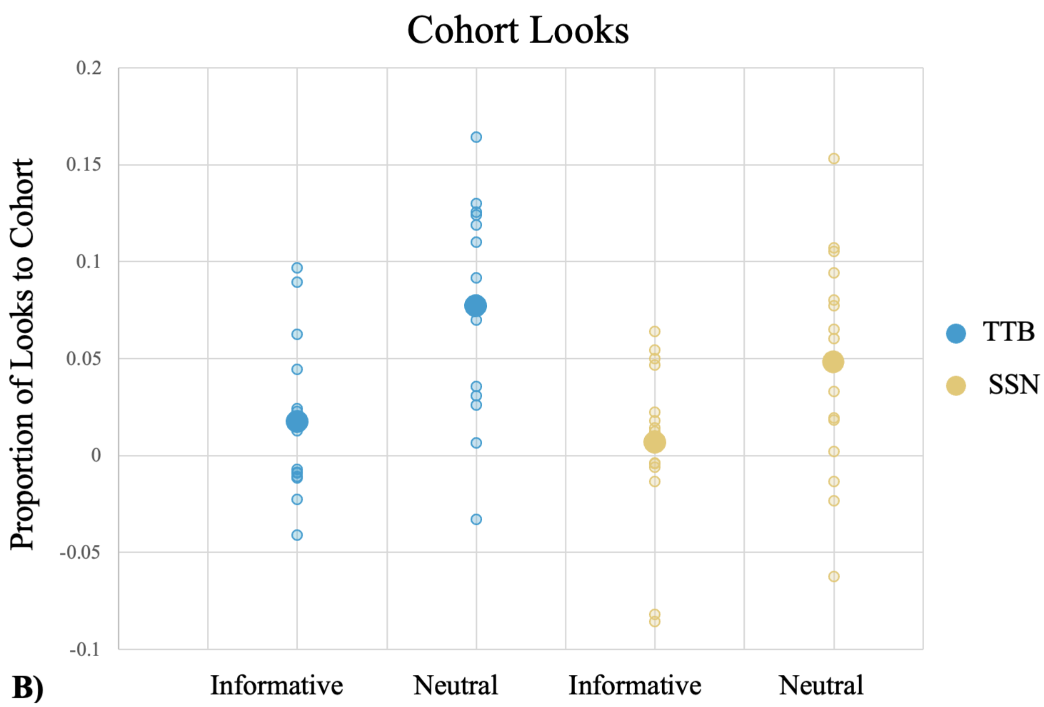
Fig. 2. Time window analysis of cohort eye gaze data for two-talker babble and speech-shaped noise conditions from 1914 ms to 3200 ms. **A)** Time window graph of proportion of looks to cohort in the informative verb condition. **B)** Time window graph of proportion of cohort to target in the neutral verb condition. Dotted error lines represent standard error of the mean. Dashed line indicates the average onset of target word with a 200 ms delay.

4.4. Cohort looks

The model of cohort fixations also revealed a significant semantic condition effect of less cohort competition in the informative condition compared to the neutral condition (β = 0.05, SE = 0.01), t(58) = 4.04, p < .001). In contrast to target looks, children displayed a nonsignificant trend more looks toward the cohort competitor in the TTB condition compared to the SSN condition (β = -0.02, SE = 0.01), t(58) = -1.62, p = 0.1). This difference across noise conditions suggests that in sentences lacking semantic context, children may experience more cohort competition when the masker is TTB, as displayed in Figure 2b and Figure 3b.



A)



B)

Fig 3. (A) Plot of proportion of target looks over the analyzed time window of 896 ms to 3200 ms. **(B)** Plot of proportion of cohort looks of the analyzed time window of 1914 ms to 3200 ms. Color indicates noise condition groups and horizontal axis labels indicate sentence condition. Average proportion of eye gazes for each participant is represented by smaller points whereas average proportion across conditions is indicated by larger points.

4.5. Verbal Fluency Scores

All children participated in a categorical verbal fluency assessment. The total number of unique words they produced were recorded to assess performance on the task. Verbal fluency scores were not found to be significantly related to children's performance in the eye-tracking task ($\beta = 0.002$, SE = 0.002), $t(58) = 1.14$, $p > .1$).

5. Discussion

5.1. Effect of Noise Maskers on Semantic Context

In this study, we investigated the effect of different noise maskers and semantic sentence context on children's word recognition. We hypothesized that across noise conditions children would still utilize semantic cues in the informative condition to facilitate earlier looks toward the target and suppress looks to the cohort when compared to the neutral condition. Additionally, we predicted that children would be slower to fixate on the target image in the TTB condition when compared to the SSN condition. We hypothesized that in the neutral sentence context, children would sustain consideration of the cohort for a longer duration in the TTB condition relative to the SSN condition.

In the presence of TTB and SSN maskers, children aged 5 to 7 years demonstrated use of informative semantic cues to facilitate access of an upcoming target word and suppress cohort competition. In comparison to the neutral condition, children looked more toward the target image and less toward the cohort image when provided with a semantically informative verb. Figure 1 exhibits the time windows of analyses for the informative and neutral sentence conditions, in which children display a greater proportion of earlier looks to the target when provided with an informative verb, suggesting more efficient lexical access of the target. Further, when children are not given informative semantic context, they display a greater proportion of looks to the cohort competitor over more time, indicating more uncertainty and less efficient access of the target. The finding that semantically informative sentence context allows for faster access of the target and suppression of the cohort competitor throughout childhood is supported by previous research (Blomquist et al., 2023), even in masked conditions (Buss et al., 2019).

We hypothesized that in the presence of a speech masker, children would benefit less from the informative sentence context and would be slower to fixate on the target when compared to a noise masker condition. Contrary to this prediction and previous research, there was no significant effect of noise condition for target looks. For example, Buss et al. (2019) similarly investigated the influence of different maskers on informative and uninformative sentence contexts with a different paradigm that measured participants SRTs with sentence repetition. This study determined that the SRT benefit gained from an informative sentence context masked by TTB was less than the benefit from an informative context masked by SSN, suggesting that speech maskers interfere with children's ability to use semantic context to facilitate word recognition. A consideration for the contrasting findings between the previous study and the current study is the task difficulty. Buss et al. (2019) analyzed the influence of different maskers on semantic context advantage using stimuli which may have been easily confused by young children ("Tough guys sound mean"). In the current study, all sentence stimuli had an age of acquisition under 6 years old (Blomquist et al., 2023; Kuperman et al., 2012), ensuring that participants were familiar with the verb, subject, adjective, and noun.

In Buss et al. (2019), the stimuli was created using a female talker masked by female TTB. This stimuli increases the perceptual similarity between the signals which creates more informational masking, and thus increases the difficulty of the task (Wightman & Kistler, 2005). In the present study, the stimuli included a female target speaker that was masked by an opposite sex masker, which provides a release from masking in both children and adults (Wightman & Kistler, 2005).

Finally, the current study presented the target speech at 10 dB SNR in both noise masker conditions, which falls well above the reported thresholds for word recognition of 5.4 dB SNR for TTB and -0.5 dB SNR for SSN in 5 to 7 year old children (Corbin et al., 2016). Overall, the auditory stimuli in the current study aimed to achieve high accuracy, resulting in only four out of 32 blocks with less than 95% accuracy. In masked speech recognition tasks, influences of executive functioning skills, such as selective attention and working memory, have been found to be associated with performance with a TTB masker (McCreery et al., 2020), suggesting that there may be cognitive limitations related to children's ability to perceptually isolate the target speech from the masker. Children's reduced ability to use semantic context in the TTB condition in Buss et al. (2019) may reflect the cognitive demands associated with a more difficult task. The task in the current study may not have exerted the same cognitive and linguistic load, resulting in a non-significant difference between the looks toward the target in the TTB condition compared to the SSN condition. Another consideration for this finding is the relatively small sample size, which led to variation in the data as shown by the error bars in Figure 1.

When considering cohort fixations, there was a marginal negative trend toward a difference between noise maskers, suggesting that children demonstrated a decrease in looks toward the cohort image in the SSN condition when compared to the TTB condition. In contrast to the finding of target looks, this trend aligns with our prediction that in the neutral TTB condition, children would sustain fixations toward the cohort image for a longer time than in the SSN condition. One possibility may be that suppression of the cohort competitor requires more cognitive and linguistic resources than target selection (Zhang & Samuel, 2018). In the presence of a TTB masker, which places a higher cognitive load on listeners than SSN, children may allocate more resources toward parsing the target speech stream from the background speech. When processing demand in these degraded conditions increases, children may be less able to suppress cohort competitors. Although these findings may be attributed to increased ambiguity in the target signal from the cognitive and linguistic demands of the TTB masker, it is also important to consider the significant variation and small sample size that makes up the data.

5.2. Limitations

Some limitations of this study include the need to increase the number of participants to reduce the standard error and increase generalization to the broader population. Additionally, future analyses may benefit from using more precise statistical measures, such as the Bootstrapped Differences of Timeseries (BDOTS), which has been used in similar studies analyzing eye gazes with the visual world paradigm over a specific time window (Blomquist et al., 2023). Including standardized measures of vocabulary, working memory, or attention may allow for further exploration of the role of cognitive and linguistic influences that might explain some variability in future tasks. Finally, since this task required children to remain still yet engaged over the course of an hour, child behavior and fatigue may have affected performance.

5.3. Future Directions and Implications

This study determined that children are able to use semantic context at advantageous SNRs in both TTB and SSN, future research may benefit from exploring the effects of a lower SNR with a similarly appropriate eye-tracking task for children. It is possible that with a lower SNR, where the target speech becomes more degraded, the benefit from semantic context is reduced, as in Buss et al., 2019. It may also be of interest to explore stimuli of increased linguistic complexity to reflect learning that occurs in a classroom context.

Overall, this study revealed that in the presence of both TTB and SSN maskers, children aged 5 to 7 years old are able to benefit from informative semantic sentence context.

Additionally, this study revealed that at an advantageous SNR, children did not experience significant interference from the TTB masker relative to the SSN masker when considering the target image, but may have experienced interference in suppressing the cohort competitor. Although study limitations should be considered alongside these findings, the results of this study may support the importance of high SNRs in classroom environments for effective spoken speech recognition. Future research may benefit from further exploration of these findings.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Blomquist, C., Newman, R. S., Huang, Y. T., & Edwards, J. (2021). Children with cochlear implants use semantic prediction to facilitate spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1636–1649. https://doi.org/10.1044/2021_jslhr-20-00319

Blomquist, C. M., Newman, R. S., & Edwards, J. (2023). The development of spoken word recognition in informative and uninformative sentence contexts. *Journal of Experimental Child Psychology*, *227*. https://doi.org/10.1016/j.jecp.2022.105581

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

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The*

Journal of the Acoustical Society of America, *120*(6), 4007–4018. https://doi.org/10.1121/1.2363929

Buss, E., Leibold, L. J., & Hall, J. W. (2016). Effect of response context and masker type on word recognition in school-age children and adults. *Journal of the Acoustical Society of America*, *140*(2), 968–977. https://doi.org/10.1121/1.4960587

Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked Sentence Recognition in Children, Young Adults, and Older Adults. *Ear and Hearing*, *40*(5), 1117–1126. https://doi.org/10.1097/aud.0000000000000692

Calandruccio, L., Bradlow, A. R., & Dhar, S. (2014). Speech-on-speech masking with variable access to the linguistic content of the masker speech for native and nonnative english speakers. *Journal of the American Academy of Audiology*, *25*(04), 355–366. https://doi.org/10.3766/jaaa.25.4.7

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children. *Ear and Hearing*, *37*(1), 55–63. https://doi.org/10.1097/aud.0000000000000201

Dink, J.W., & Ferguson, B. (2015). *eyetrackingR: An R library for Eye-tracking Data Analysis*. http://www.eyetracking-r.com

Fallon, M., Sandra, T., & Bruce, S. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America*, *111*(5), 2242–2249. https://doi.org/10.1121/1.1466873

Jeppsen, C., Baxelbaum, K., Tomblin, B., Klein, K., & McMurray, B. (2024). The development of lexical processing: Real-time phonological competition and semantic activation in school age children. *Quarterly Journal of Experimental Psychology*, *78*(3), 437–458. https://doi.org/10.1177/17470218241244799

Hadley, P., & Rice, M. (1993). Parental judgments of preschoolers' speech and language development: A resource for assessment and IEP planning. *Seminars in Speech and Language*, *14*(04), 278–288. https://doi.org/10.1055/s-2008-1064177

Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, *23*(2), 159–165. https://doi.org/10.1097/00003446-200204000-00008

Kuperman, V., Stadthagen-Gonazalez, H., & Brysbaert, H. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. https://doi.org/10.1037/a0029284

Manlove, E. E., Frank, T., & Vernon‐Feagans, L. (2001). Why should we care about noise in classrooms and child care settings? *Child and Youth Care Forum*, 30(1), 55–64. https://doi.org/10.1023/a:1016663520205

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

McCreery, R. W., Miller, M. K., Buss, E., & Leibold, L. J. (2020). Cognitive and linguistic contributions to masked speech recognition in children. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3525–3538. https://doi.org/10.1044/2020_jslhr-20-00030

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329. https://doi.org/10.1016/j.jecp.2003.09.001

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, *87*(6), 2705–2715. https://doi.org/10.1121/1.399061

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5–S5. https://doi.org/10.1121/1.1995329

Reuter, T., Dalawella, K., & Lew-Williams, C. (2020). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience*, *36*(4), 474–490. https://doi.org/10.1080/23273798.2020.1839665

Sekerina, I. A., & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*, *98*(1), 20–45. https://doi.org/10.1016/j.jecp.2007.04.005

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Shield, B. M., & Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *The Journal of the Acoustical Society of America*, *123*(1), 133–144. https://doi.org/10.1121/1.2812596

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. https://doi.org/10.1016/s0010-0277(99)00021-9

Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Zhang, X., & Samuel, A. G. (2018). Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. *Journal of Memory and Language*, *100*, 32–50. https://doi.org/10.1016/j.jml.2018.01.002