Fall 2019

The Role of Cognitive Control in Understanding Speech in Noise

Sarah Plock

THE ROLE OF COGNITIVE CONTROL IN UNDERSTANDING SPEECH IN NOISE

by

Sarah Plock

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Speech Pathology and Audiology

____
Dr. Bob McMurray
Thesis Mentor

Fall 2019

All requirements for graduation with Honors in the
Speech Pathology and Audiology have been completed.

____
Yu-Hsiang Wu
Speech Pathology and Audiology Honors Advisor

College of Liberal Arts and Sciences
The University of Iowa
Iowa City, Iowa

THE ROLE OF COGNITIVE CONTROL IN UNDERSTANDING SPEECH IN NOISE

By

Sarah Plock

A thesis submitted in partial fulfillment of the requirements for graduation with Honors in the Department of Communication Sciences and Disorders

Bob McMurray

Thesis Mentor

Fall 2019

Abstract

Purpose: During speech perception, lexical candidates compete for word recognition. Incorrect candidates are briefly activated and then become suppressed. An aspect of word recognition is the ability to suppress these incorrect candidates especially in noisy environments. Previous work examining lexical inhibition and domain-general cognitive control found no correlation, suggesting cognitive control was not involved in spoken word recognition. Few studies have examined individual characteristics that impact a listener's ability to process speech in noise. This study aims to understand the role cognitive control when speech is presented in noise.

Methods: We utilized the visual world paradigm (VWP) to measure lexical competition over the time course of word recognition. In the VWP, listeners heard words and clicked on the referent from a screen containing targets, cohorts, rhymes, and unrelated items, while eye-movements are monitored as a measure of lexical competition. Two classic cognitive controls tasks, the Flanker and Simon, were used to measure inhibition, a domain general cognitive mechanism. An experimental task, Temporal Flanker, was used because it simulates how speech unfolding over time.

Results: In the noise condition, listeners waited around 400 ms after the onset of the word to launch eye-movements. They showed slower and reduced activation of the target and increased competition. A significant interaction between Temporal Flanker score and timing of target fixations suggests that individuals who were better at the Temporal Flanker task were quicker to activate the target.

Discussion: The study showed a link between spoken word recognition and cognitive control. It has been well documented that the development of cognitive control is slow in childhood. This could have potential implications for children with Developmental Language Disorder or who

use a cochlear implant. The development of cognitive control may be a potential avenue for intervention for language and hearing disorders.

Introduction

A critical part of speech perception is lexical access – the ability to look up the meaning for a word that is heard. The mechanisms of this are well understood in typical adults listening to speech in quiet. It is also well documented that these mechanisms may differ among individuals, and that they differ in adverse listening (McMurray, Farris-Trimble, Seedorff, & Rigler, 2015; McMurray, Samelson, Lee, & Tomblin, 2010). However, less research has been done to understand the characteristics of an individual that affect their ability to performance lexical access in noise.

The present study aims to further understand the underlying cognitive processes involved in speech recognition in noise, specifically the role of cognitive control. A critical part of lexical access is suppressing competitors that are not consistent with the input. For example, when hearing the word *whistle*, the listener must suppress the competitor *wizard*. The question is whether cognitive control is involved in this process of competitor suppression. This will help uncover if the suppression process is specific to speech or if it involves more general cognitive mechanisms involved with decision-making and inhibition. The implications of this research can be applied to individuals with hearing impairments, as these finding may lead to cognitive training to improve the ability to process speech in noise.

Real-Time Spoken Word Recognition

Under ideal listening conditions speech perception is nearly effortless. However, this is accomplished despite the fact that speech unfolds over time, creating temporal ambiguity. This ambiguity is seen at early points in the input at which the listener cannot know what the target word will be (Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014; Marslen-Wilson, 1987).

For example, when a listener hears *san-*, there are multiple lexical items that may match the beginning of this word such as *sandal* and *sandwich* because there is not yet enough information that has been provided to identify the target word. This ambiguity can be heightened in adverse listening conditions such as background noise or degraded speech

To deal with this ambiguity, listeners initially activate multiple lexical candidates that may match the target word, and later they gradually suppress items that no longer match the input (Marslen-Wilson, 1987). In ideal listening conditions, the result of this process is that listeners activate mental representations of multiple words (lexical candidates) before they receive all the acoustic information to identify a single word (Farris-Trimble et al., 2014).

Visual World Paradigm

This theoretical model of speech perception has been validated by work using eye-tracking in the (VWP). The first use of eye-tracking to study real-time spoken word recognition was by Allopenna, Magnuson and Tanenhaus, (1998). In this study, listeners heard a target word and chose a picture on a computer screen while eye-movements were monitored. Each set of words was chosen to include a target (e.g. *beaker*), a cohort: a word that matched the target at the onset (e.g. *beetle*), a rhyme: a word that matched the target at the offset (e.g. *speaker*), and a phonologically unrelated word (e.g. *carriage*). Around approximately 200 ms after the start of the word, eye-movements were launched to the target and cohort items. Results showed that listeners launched eye-movements to the cohort earlier, and these fixations peaked higher and earlier than the fixations to rhymes. Although rhyme items had fewer fixations, their fixations persisted longer than cohort items. As more auditory information became available, the fixations to the cohort became suppressed and the listener fixated the target (Allopenna et al., 1998).

Methods like these have led researchers to propose a set of core cognitive mechanisms involved in real-time spoken word recognition. Upon receiving any auditory input, listeners activate multiple lexical items immediately. Lexical items are activated in parallel and update continuously as more auditory information becomes available to the listener. The lexical items compete with one another for word recognition at the local level, better known as lexical inhibition (Allopenna et al., 1998; Marslen-Wilson, 1987; McMurray et al., 2010).

Adverse Listening Conditions

These principles have been well worked for lexical access under ideal conditions (e.g. no background noise). Unfortunately, these ideal listening conditions rarely exist. Every day, listeners must overcome some form of adverse listening situation in order to effectively process auditory input. Adverse listening conditions can be created by background noise, inaudible speech, and through the use of a hearing aid. Brouwer and Bradlow (2016) showed that background noise caused increased activation of lexical competitors during a word recognition task. Results for this experiment showed that the cognitive mechanisms crucial for word recognition can be adjusted depending on listening conditions.

McQueen and Huettig (2012) asked how real-time spoken word recognition changes when the listener is presented with degraded speech. Their form of difficult listening differed from Brouwer and Bradlow (2016). Instead of background noise through the entire speech signal, they manipulated the words around the target word. They used varying types of competitor words that overlapped with the target word at the onset or offset. Noise was placed in varies parts of the sentences except for the critical words, the target, which has no noise. This allowed them to create the expectation of noise even though the signal itself (for the target word) was not noisy. When speech was presented in quiet, results were comparable to Allopenna et al. (1998) in which listeners looked more often to cohort than rhyme items. In noise, however, this pattern was weakened with listeners looking more often to rhyme items than they would have in quiet. This finding suggests that adverse listening conditions reduces confidence in what they listener believed the target word was, leading to more consideration of rhyme items and less consideration of cohort items (McQueen & Huettig, 2012), even when the target word itself was clear.

Differences in spoken word recognition also stem from an individual's hearing status. Farris-Trimble et al. (2014) examined post-lingually deafened cochlear implant (CI) users and normal hearing (NH) individuals and how they differed in fine-grained aspects of real-time spoken word recognition with a degraded signal. This study utilized the traditional paradigm used by Allopenna et al. (1998) with each set of words containing a target, cohort, rhyme, and an unrelated. This study compared two participant groups: CI-users, and NH listeners. In the first experiment CI users showed decreased target activation at the end of the time course as compared to NH listeners (Figure 1A). The CI users also showed increased rhyme activation as compare to NH listeners. In the second experiment NH listeners heard simulated CI speech. CI-simulated speech is a type of noise that simulates what it is like to have a cochlear implant. These listeners showed similar delays in activation as the CI group (Figure 1B). Their results showed a moderate delay in targets and increased activation for competitors which shows similar findings to Brouwer and Bradlow (2016). From this study, it is evident that post-lingually deafened CI-users behave similarly to the NH listeners who hear CI-simulated speech. This suggests that the degradation of a signal, whether that be due to difficult conditions or poor input, impacts how listeners are considering lexical candidates.

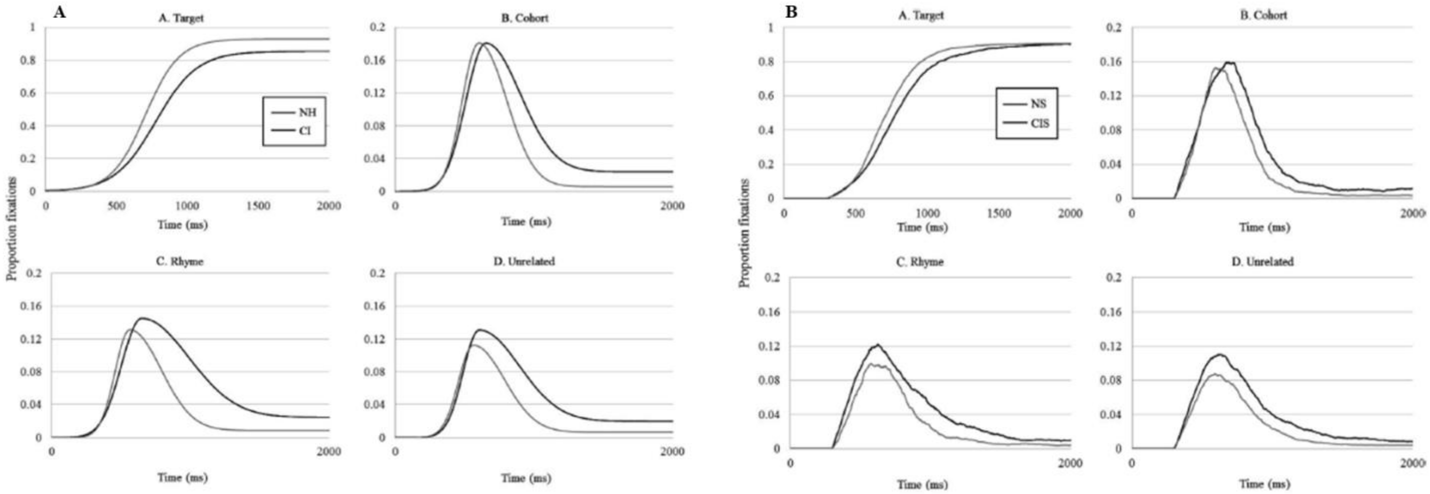

Fig. 1: Results from Farris-Trimble et al. (2014). A, the proportion of fixations to each of the items comparing NH listeners to CI-users. B, the proportion fixations to each of the items comparing normal speech (NS) to CI-simulated speech (CIS).

McMurray, Farris-Trimble, and Rigler, (2017) examined the differences in spoken word recognition between prelingually deaf cochlear implant (CI) users and NH individuals. In contrast to the post-lingually deaf adult CI users, prelingually deaf cochlear implant users were slower to fixate on items in the VWP, showed less competition from cohort items, and more competition from rhyme items (compared to age-matched controls). NH adults who heard highly degraded speech showed similar patterns as the CI users[1]. Thus, when listening is difficult enough, listeners use a “wait and see” approach, meaning that individuals wait to activate lexical items until more informational is made available.

It is evident from previous research that post-lingually deafened CI users do not appear to “wait-and-see”. However, it appears that pre-lingually deafened CI users do this. This suggests

[1] McMurray et al. (2017) used twice as much degradation of speech as in the study by McMurray et al. (2015). The differences observed in results may be accounted for by the difference in noise exposure.

that under highly difficult listening (highly degraded speech or developing with a CI) listeners "wait -and-see". But under moderately difficult listening (post-lingually deafened CI-user or moderate CI simulation) listeners show enhanced competition. Similar patterns between post-lingually deafened CI-users and NH listeners who did not develop with a CI suggest the degraded input causes the varied word recognition patterns.

Adverse listening conditions can also stem from individual differences such as age-related cognitive decline. Ben-David, Chambers, Daneman, Pichora-Fuller, Reingold, and Schneider, (2011) investigated how noise and aging affect real-time spoken word recognition. Older adults, even those who have low levels of hearing loss, typically require a higher signal-to-noise (SNR) than younger adults especially in noisy conditions. To control for age-related differences in the accuracy of word recognition, an SNR of -4 was used for the young participants and an SNR of 0 was used with the older participants. Ben-David et al. (2011) controlled for the number of overlapping syllables between competitor words by using sets of monosyllabic pairs (e.g. *house* – *mouse*) and disyllabic pairs (e.g. *candle* – *sandal*). Similar fixations patterns were present in both the younger and older participant groups. Similar to Brouwer and Bradlow (2016), enhanced competition was present in the noise condition. Differences between the younger and older participants were only observed in trials containing the target and a rhyme competitor in the noise condition. Older adults experienced increased difficulty in differentiating target words from rhymes. They found age-related differences only in the noise condition. This suggests that some age-related decline, possibly cognitive control, may be affecting their ability to recruit these general cognitive mechanisms when speech becomes difficult for local processes.

Summary. Studies examining how word recognition is impacted by noise or degraded speech have implications for everyday listening. Background noise causes increased competitor activation (Brouwer and Bradlow, 2016). McQueen and Huettig (2012) took this background noise one step further by making the words around the critical word noisy. This provided the illusion that the signal was noisy when in reality it was the only clear part of the input. They showed that a difficult listening situation reduces the listener's confidence in the target as they considered competitors more often. These studies showed that adverse listening conditions impact the listener's ability to confidently process the auditory input.

Studies conducted on pre-lingually and post-lingually deafened CI users showed differing patterns of spoken word recognition. Pre-lingually deafened CI users displayed a "wait-and-see" approach in which they waited longer in the time course to launch eye-movements (McMurray et al., 2017). However, post-lingually deafened CI users did not display this "wait-and-see" approach which leads to the idea that the "wait-and-see" approach may be a developmental pattern due to having a CI pre-lingually. Yet in the same study, post-lingually deafened CI users showed similar eye-movements to the NH listeners who heard CI-simulated speech. Lastly, work on aging and noise by Ben-David et al. (2011) suggest that some age-related decline, possibly cognitive control, may be contributing to the older adults' difficulty in processes speech in noise if they are unable to efficiently recruit general cognitive mechanisms.

Coping with Enhanced Competition

As the foregoing discussion suggests, listeners adopt a "wait and see" approach to speech perception when the degradation of speech becomes too difficult, as this can partially reduce competition (McMurray et al., 2017). In contrast, when speech is only moderately degraded, listeners show enhanced competition as a mechanism of dealing with this degradation (Farris-Trimble et al., 2014). The question in either case is how are these competitors suppressed?

One possibility is lexical inhibition. This form of inhibition is largely automatic and an unconscious process local to the word recognition system. When an auditory input comes in, an activated word inhibits other words that are similar sounding in the lexicon (e.g., between words) (Blomquist and McMurray, 2017). This created competition for word recognition. Lexical inhibition works in way that each word has an inhibitory connection to its neighbors. So that when a word like *sandal* becomes activated, it inhibits words like *sandwich* and *candle.* This inhibition causes them to be less active.

Dahan, Magnuson and Tanenhaus, (2001) documented this form of inhibition by using cross-splicing to manipulate the auditory stimuli to temporarily enhance competition between words. This method took the onset of one word (e.g., the *ne* from *net*) and put it with the offset of a competing word (the *k* from *neck*). This created temporary competition because the coarticulation of the onset of the word leads the listener to predict the competing word (*net*). The word *net* inhibits *neck* so that when the listener finally receives -*ck* from *neck*, the listeners is slower to activate *neck* because it was previously inhibited. This cross-splicing manipulation is a useful tool because it can detect that lexical inhibition is present. This is observed through the decreased activation of the target word because of temporary competition as the participant is suppressing the target word because of the misleading onset.

Dahan et al. (2001) used three splicing conditions: match-splice, word-splice, and nonword-splice. For example, the word-splice condition consisted of a competitor word (*net*) cross-spliced onto the target word (*neck*) to create (*ne*$_{t}$*ck*). The match-splice condition consisted of a target word (*net*) cross-spliced onto the same word (*ne*$_{t}$*t*). The nonword-splice condition consisted of a nonword cross-spliced (*nep*) onto the target word ($ne_{p}t$). Results from this study showed that the frequency of a word influences lexical access early in the auditory input. Listeners were slower in the word-splice condition. This shows that these partially activated lexical candidates inhibit each other for word recognition.

Kapnoula and McMurray (2016) showed that experience can affect the level of lexical inhibition and can be modified by training to fit the listener’s needs. In this study, individuals who experienced high-competition training were better at inhibiting competitors in the VWP. This high-competition training consisted of taking the listeners through a series of tasks that involved similar sound words (e.g., *net/neck* and *cat/cap*) prior to the main experiment. The idea that inhibition has plasticity can have real-world application. Increased inhibition helps in adverse listening conditions (i.e. degraded speech or background noise) as well as allowing the process used during speech perception to update and change depending on a listener’s current listening conditions (Kapnoula & McMurray, 2016).

In these studies, inhibition is primarily seen as within the lexicon. However, inhibition can also come from general cognitive processes and is a key part of cognitive control. Cognitive control is the conscious and effortful use of executive function skills such as attention, inhibition, and focus. This inhibition can be applied to any type of stimulus, not just speech, and can be applied to novel tasks. However, this inhibitory component of cognitive control could be helpful in noise as an individual must suppress incoming noise and irrelevant speech to focus on the target input.

There is little work on domain general inhibition and spoken word recognition. However, neuroscience offers hints. The prefrontal cortex (PFC) of the brain has been associated with cognitive control as referred to as ‘dynamic filtering’ by being able to hone in on appropriate information and inhibit irrelevant and inappropriate information (Shimamura, 2000; Novick, Trueswell, & Thompson-Schill, 2010). Previous work done on the neuroscience of speech perception suggests that in noise, listeners engage more of the frontal areas of brain which is the likely seat of cognitive control (Du, Buchsbaum, Grady, & Alain, 2014).

Research on cognitive control has focused on identifying the location of these general cognitive mechanisms and how they develop. Bunge, Dudukovic, Thomason, Vaidya, and Gabrieli (2012) examined differences in frontal lobe activation with children and adults. They used fMRI and classic cognitive control tasks. The neuroimaging tests (fMRI) to see what activates in the brain during specific tasks. Results showed the activation of PFC was associated with greater inhibition. However, children differed. In the same cognitive control tasks, the children recruited different brain regions than adults. This suggests that children go through a developmental phase between childhood and adulthood that may last an extended period of time before the frontal lobes are fully matured. This developmental period of the frontal lobes has special implications for children with CIs. If cognitive control is directly involved in understanding speech in noise, this developmental period could be a critical time for intervention as the brain is still learning how to recruit the areas in the frontal lobe associated with cognitive control.

There are many classic measures of cognitive control that are well-established in research. A prominent measure of inhibition is the flanker task. This task aims to measure the degree to which an individual can suppress a specific action (Wöstmann, Aichert, Costa, Rubia, Möller, & Ettinger, 2013). In the flanker task, the individual is presented with a row of arrows and must respond as quickly and as accurately to the direction of just the middle arrow. This task uses two types of trials: congruent and incongruent. In congruent trials, all the arrows are facing the same direction. In incongruent trials, the flanking arrows face a different direction than the middle arrow. It has been well-established that individuals are slower in the incongruent trials than congruent trials. Inhibition is at play in this task because the participant must consciously suppress some form of irrelevant information or inappropriate action.

Prior Work on Lexical Inhibition and Domain General Inhibition

Previous work has examined the relationship between lexical inhibition and performance on a nonlinguistic cognitive control task. Blomquist and McMurray (2017) used a modified version of the VWP from an earlier study by Dahan et al. (2001) that used cross-splicing to target local inhibition. Lexical inhibition was measured with the cross-splicing in the VWP; cognitive control was measured with a spatial Stroop task. They examined typically developing children ages 7-12 years of age. They found no correlation between the participant's performance on the spatial Stroop task and performance in the VWP (Figure 2). The lack of correlation suggests that cognitive control and lexical are not the same. It also suggests that domain general control may not be relevant to spoken word recognition. However, as this task being tailored to lexical inhibition, it is unknown if cognitive control is involved in word recognition more broadly. Moreover, the first major limitation of this study was that the task was conducted in quiet. Effortful processing is may not engaged in quiet but is engaged in adverse listening conditions (Peng & Wang, 2019), thus it is possible that cognitive

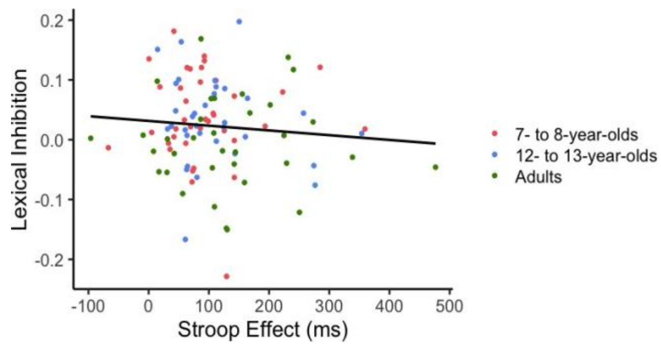

Fig. 2: No correlation found between lexical inhibition and performance on a nonlinguistic task.

control would be needed in noise. Another major limitation of this study was that the task did not have competitors on the screen. This was due to the study only examining lexical inhibition. They were not trying to measure the time course of competition.

The Present Study

The present study aimed to examine the relationship between cognitive control and real-time spoken word recognition to further understand if listeners recruit general cognitive inhibition mechanisms during speech perception in noise. Similar to the paradigm of Allopenna et al. (1998), Brouwer and Bradlow (2016), and Farris-Trimble et al. (2014), each trial consisted of a target (e.g. *sandal*), a cohort (e.g. *sandwich*), a rhyme (e.g. *candle*) and an unrelated item (e.g. *necklace*). The auditory stimuli were manipulated in half of the trials to be embedded in a multi-talker babble noise to simulate real-world listening conditions. The other half of the trials were presented in quiet. The timing of the onset of each word was randomly assigned in both the noise and quiet conditions so the listener could not anticipate the onset.

Participants also completed three nonlinguistic tasks that measure general inhibition skills. This included two classic cognitive control tasks, the Simon and Flanker tasks, along with a task that has never been used before, the Temporal Flanker task. The Temporal Flanker task displays the participant with a row of five arrows. Each arrow is only on the screen for 250 ms and the participant must respond to the direction of just the middle arrow after all the arrows were shown. This new task was included because it mirrors how speech unfolds over time in which the arrows only appear on the screen for a given amount of time and then they are gone. These tasks were used to determine if greater cognitive control abilities aided listeners in the noise condition of the word recognition task. The inclusion of tasks measuring general inhibition allow for the investigation of whether individuals recruit more general cognitive mechanisms when speech is presented in adverse listening conditions.

Methods

Participants

Thirty adults participated in this study. Participants were recruited through two methods: SONA, a university research credit portal, and via emails to the university community. Participants who were recruited through email were run at the University of Iowa Hospitals and Clinics and compensated $30 (N=15, 11 female). Participants who were recruited through SONA were run at Spence Laboratories of Psychology and received research credit for a psychological course through the University of Iowa (N=15, 9 female). All participants were native monolingual English speakers with normal hearing. Normal hearing was confirmed through a hearing screening conducted at the beginning of the experiment session. Abiding by IRB protocol, a written informed consent was collected from paid participants and a verbal informed consent was collected from Sona participants.

Design

This experiment used a traditional VWP task to investigate how real-time spoken word recognition was impacted when speech was presented in noise. The VWP task followed the design in Farris-Trimble et al. (2014). Each set of words consisted of a target (e.g. *sandal*), a cohort (e.g. *sandwich*), a rhyme (e.g. *candle*), and an unrelated word (e.g. *necklace*) for each trial. There were 60 such sets. On each trial, the participant heard an auditory input and chose a referent from the computer screen containing four referents. Words were selected that were easily visually representable (see Appendix A for a list of all the words used). Each word was presented in quiet and in noise to make a total of 480 trials (60 sets x 4 words/set x quiet vs. noise). Six words were excluded from analysis, *cavern*, *crone*, *mud*, *zip*, *socket*, and *cleaver*. This exclusion was decided based on a word's accuracy being lower than 90% across the participant pool.

Stimuli

Auditory stimuli. Auditory stimuli were recorded by a female with a Midwestern dialect in a sound attenuated room. Each stimulus was recorded in a carrier phrase (e.g. "She said *candle*") and given a variable amount of silence and noise at the onset for each condition in order to try and prevent participants from anticipating when the onset of the word will begin. This onset of silence or noise ranged from 200 to 1000 ms.

It was important to identify a signal-to-noise (SNR) that was challenging enough to slow participants down, but not too hard that accuracy was greatly impacted. We needed to maintain a reasonably high accuracy in the noise condition because only correct trials were used in the analysis. If the noise caused listeners to have a low accuracy in the task, then an insufficient number of correct trials would remain useable for the analysis. Thus, we conducted a pilot study on 6 participants to determine an optimal SNR. After it was confirmed that an SNR of +2.0 was appropriate, the experiment was then conducted with twenty-four more participants.

Visual stimuli. Visual stimuli were chosen following standard lab procedures in order to find appropriate and representative images for the auditory stimuli. For each word, approximately 10 images were chosen from a commercial clip art database. Members of the lab including undergraduate students, graduate students, and lab coordinators collaborated to discuss each prospective image. The selected images underwent edits if modifications were needed to ensure that all images had consistent levels of coloring and contained minimal distracting items. The final set of images needed final approval from the principal investigator (McMurray et al., 2010).

Procedure

After informed consent was given, participants were guided into a sound booth and seated in front of a 19” LCD display monitor with 1280 x 1024-pixel dimensions. The computer had a standard mouse and keyboard. For the paid participants in the University of Iowa Hospitals Clinics, auditory stimuli were presented at 60 dB through loudspeakers placed on both sides of the computer. SONA participants in Spence Laboratories of Psychology, auditory stimuli were presented through over the ear headphones. In both sound booths, a padded chin rest was placed directly in front of the monitor and was adjusted to comfort when necessary for each participant. After the participant was set in place for the experiment, the research assistant proceeded to the calibration process for the eye-tracker. The participant was then provided verbal instructions on the task. The research assistant informed the participant when breaks were allowed and addressed any questions or concerns regarding the experiment.

On each trial of the VWP experiment, a red dot was presented at the beginning of the trial. When the dot turned blue the participant clicked on the dot to begin the trial. They saw the four pictures on the monitor and heard an auditory input. This auditory input was played through headphones or over loudspeakers. The participant chose the referent that best represents the word they heard.

The participant was encouraged to take breaks when needed and give their best effort on each trial. Eye-movements and mouse clicks were recorded throughout the entire experiment. After completion of the VWP task, the participants completed four additional computer tasks examining cognitive control.

Eye-tracking recording and analysis. Eye-movements were recorded using a SR Research Eyelink 1000 eye-tracker. This desktop-mounted eye-tracker utilized a 9-point calibration system with corneal reflection and pupil used concurrently whenever possible to track. Drift corrections were done every 30 trials making a total of 16 drift corrections throughout the entire experiment. The recording of eye-movements was similar to McMurray et al. (2010) with data sampled every 4 ms beginning at the onset of each trial and ending when the participants clicked on a referent. These raw eye-position data were automatically converted into blinks, saccades, and fixations. Saccades and fixations were combined into a single event – a "look". A look began at the onset of a saccade and ended at the offset of a fixation. We used standard lab boundaries for the objects in the VWP task, 300 x 300 pixels.

Cognitive Control Tasks

Three cognitive control tasks were used. First, the Flanker task is a classic cognitive control measures used to observe a domain-general cognitive mechanism, inhibition. In the Flanker task, the participant is presented with a row of five arrows. The task is to respond as quickly and as accurately to the direction of the middle arrow. There are two conditions. In congruent trials, all of the arrows are facing the same direction. In incongruent trials, the middle arrow is opposite of the rest of the arrows. Individuals with better inhibition should show less interference with the flanking arrows in the incongruent trials. These trial types were randomized for a total of 120 trials. Participants placed both pointer fingers on the left and right arrow keys on the keyboard to respond.

Second, the Simon task is another classic cognitive control measure of domain-general inhibition. In the Simon task, the participant is presented with a dot. This dot can appear on either side of a fixation cross. The dot changes from a red to blue color randomly. The participant places both pointer fingers on the left and right arrows keys. The participant presses the left arrow key when the blue dot appears and the right arrow key when the red dot appears. On congruent trials, the blue dot appears on the left side of the screen or when the red dot appears on the right side. On incongruent trials, the blue dot appears on the right side of the screen or the red appears on the left side of the screen. Individuals with better inhibition should show less interference when the dot appeared on the incongruent side of the screen in the incongruent trials. The task consisted of 56 trials.

Third, the Temporal Flanker is a new cognitive control task attempting to measure domain-general inhibition. We used an adapted version of the task used by Hazeltine, Lightman, Schwarb, and Schumacher (2011). The participant is instructed to ignore the direction of the first two arrows and the last two arrows that appear on the screen in succession. The participant is to respond to the direction of just the central arrow. The type of flanker task mirrors how speech unfolds over time because each arrow is presented at distinct points in time as opposed to the traditional flanker in which all arrows are presented together for a given amount of time. Each arrow was present on the screen for 250 ms. After the appearance of all five arrows, a blank screen appeared until the participant responded, then a fixation cross appeared for 1000 ms to begin the next trial. The task consisted of 56 trials.

Preliminary analyses found no correlation between these measures, suggesting they could not be averaged together (as was intended). Instead, we estimated split-half reliability test on reach of the three cognitive control measures in order to choose the single most reliable measure. A split-half reliability test correlates the interference effect estimated from the odd and even trials separately. Interference is the difference between the mean averages of the incongruent and congruent trials. The split-half reliability of the Flanker task was 0.035. The Simon task had a split-half reliability of -0.135. The Temporal flanker had the highest split-half reliability of 0.414. Our analysis was to focus on the Temporal Flanker task. 38 trials were excluded due to reaction time (RT) not being within the specified range (+/- 2.5 standard deviations).

After the completion of the VWP task, the participant completed the following cognitive control tasks in order of the Flanker task, the Simon task, and then finally the Temporal Flanker task. At the end of the experiment, participants were debriefed on the purpose of the experiment. This time was also provided an opportunity to ask questions about the tasks they completed.

Results

The analysis begins with an overall test for accuracy during the word recognition task. Next the analysis of eye-movement data sheds light on the fine-grained differences of the process of real-time spoken word recognition during each listening condition. The purpose of investigating both accuracy and eye movements was to determine how successful each participant was in both listening conditions for the word recognition task. After eye-movement analyses, we computed reaction times in the Temporal Flanker task. We used the overall Temporal Flanker score to split subjects by the degree of interference into good and poor cognitive control. This grouping was done to analyze how cognitive control impacted a participant's ability to complete the word recognition task in both listening conditions.

Accuracy Analysis

We analyzed the participants' overall accuracy in the word recognition task to determine if anyone need to be excluded, and also to see if noise had a systematic effect on accuracy. Participants' mean accuracy in the quiet condition was 99.7%. In the noise condition, mean accuracy was 87.9%. Results from a t-test comparing the quiet and noise conditions concluded that there was a significant difference between performance in the quiet and noise conditions (t (29) = 23.9, $p < 0.001$). Although we found a significant difference in accuracy between the two listening conditions. This establishes that noise had the intended effect. However, the drop in accuracy may suggests that the noise level chosen was potentially problematic as it impacted participants' performance.

We also evaluated accuracy of individual words. Several words were excluded from the analysis (*cavern, crone, mud, zip, socket,* and *cleaver*) due to accuracy being lower than 90% across participants in both listening conditions.

Eye-movement Analysis

For the analysis of the eye-movements, only trials on which the target word was selected were included in the analysis. 1,001 out of 14,400 trials were excluded because participants did not correctly click on the target object.

We took the point of gaze every four ms and computed the proportion of the participants' looks to each object on the screen. For the analysis, time was adjusted (for the variable length of silence or noise) using Equation 1. We rounded the durations to a multiple of four because of the point of gaze data being collected every four ms.

Adjusted time = [time] – Round ([duration] – 4) x 4 (1)

As observed in previous work of spoken word recognition in quiet (Allopenna et al., 1998), Figure 3A shows that listeners began eye-movements around 200 ms after the onset of the word. As more information became available, the listener quickly suppressed the unrelated and rhyme items. The cohort item remained active longer, until more information for the target caused the listener to fixate on the target.

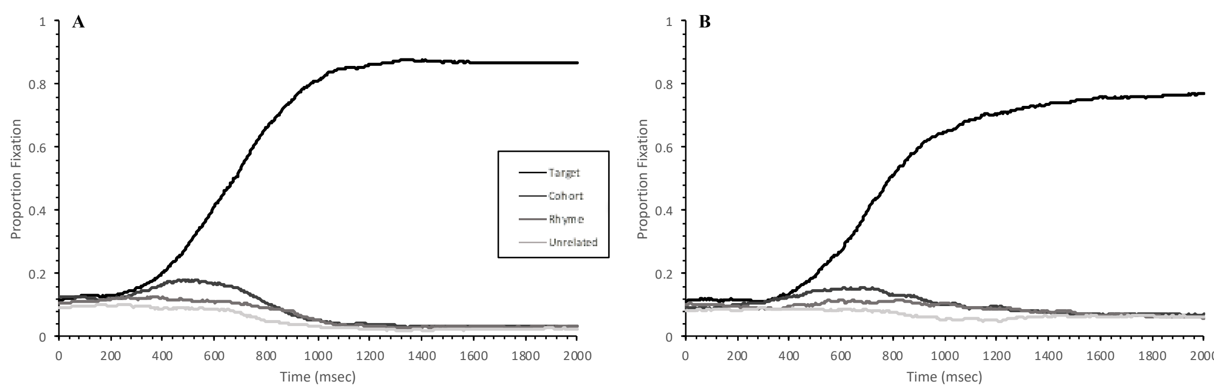

Fig. 3: A, average fixations in the quiet condition. B, average fixations in the noise condition.

Figure 3B shows eye-movements when the auditory stimuli were presented in noise. In this condition, listeners waited to launch eye-movements until around 400 ms after stimulus onset. Competitor words were suppressed more slowly and remained more active than in quiet. The peak activation of the target was reduced as compared to the quiet condition as listeners were less certain of what they heard.

We next examined the effect of noise on looks to each type of competitor. Figure 4 shows the proportion of looks to each item on the screen in both listening conditions. Figure 4A shows

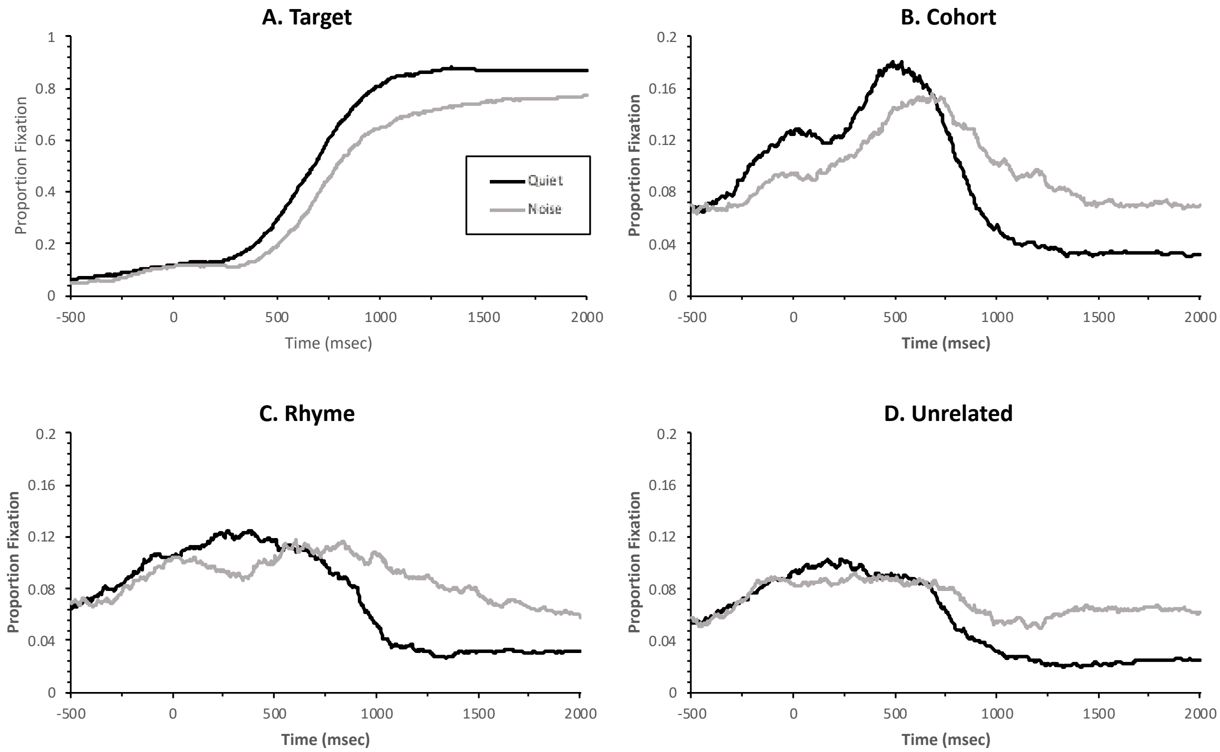

Fig. 4: A, proportion of fixations to the target item in noise versus quiet. B, proportion fixations to the cohort item in noise, versus quiet. C, proportion fixations to the rhyme item in noise versus quiet. D, proportion fixations to the unrelated item in noise versus quiet.

looks to the target item. In the noise condition, participants committed less to the target than in the quiet condition. Figure 4B shows looks to the cohort item. Participants activated the cohort item later and took longer to suppress it than in the quiet condition. Figure 4C shows the looks to the rhyme item. Similar to looks to the cohort item, participants activate the rhyme item later and took longer to suppress it. Figure 4D shows the looks to the unrelated item. Similar to patterns seen with the cohort and rhyme items, participants activated the unrelated item later in the time course. They also took longer to suppress the item and never fully suppressed compared to the quiet condition.

In order to statistically examine these fixations, we used a non-linear curvefitting approach. For target fixations, we fit a four-parameter logistic function to each subject for each condition. Fits were done with a constrained gradient descent method implemented in McMurray (2019). In this analysis, we only used a portion of the parameters that are used in the logistic function. The reason for excluding certain parameters is because they did not pertain to the theoretical research question at hand. For the target item, we focused on the following parameters: maximum height of the curve (max), crossover point, and slope. Max is a measure of peak activation. Crossover is the point on the middle of the slope. And finally, the slope is the degree of angle from the minimum to maximum asymptotes.

Table 1 reports the results of a repeated-measures ANOVA on the parameters of looks to the target comparing the quiet and noise conditions. There was a significant impact of noise on the max between the noise condition (max = 0.782) and the quiet condition (max = 0.876). A significant difference in slope was observed in the noise condition (slope= 0.0014) as compared to the quiet condition (slope = 0.0016). Finally, the crossover point varied significantly in the noise condition.

Table 1. For ANOVA measure conducted on the target parameters.

Parameter	*F*	*p*-value
Crossover	F (1,29) = 29.8	< 0.001
Slope	F (1,29) = 11.2	0.002
Max	F (1,29) = 78.6	< 0.001

Cognitive Control Analysis

Figure 5 shows the RT for congruent and incongruent trials in the Temporal Flanker task. On average, participants were almost 130 ms slower to respond when the flanking arrows did not match the central arrow (Figure 5). To verify the significant difference between congruent and incongruent trials, a simple t-

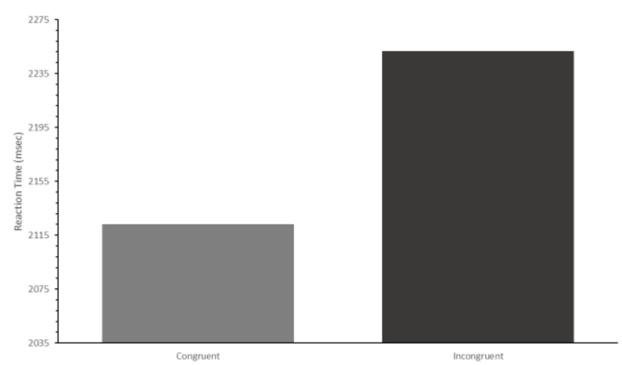

Fig. 5: The interference of congruent and incongruent trials on the Temporal Flanker task.

test was conducted (t (58) = -3.37, p = 0.001). This shows that the congruent trials had a significantly faster RT than the incongruent trials.

The Effect of Noise on Real-Time Spoken Word Recognition

A repeated-measures ANCOVA was conducted to determine how cognitive control (measured by the Temporal Flanker scores) influence a participant's performance as a function of noise. This analysis examined the effect of noise, the effect of Temporal Flanker, and the interaction between noise and Temporal Flanker. Separate ANCOVAs were done with each of the parameters of the curvefitting analysis. These are summarized in Table 2. We discuss the overall effect of noise here, and turn to the effects of cognitive control in the next section.

In order to minimize the number of statistical tests, we combined the crossover point and the slope of the target activation into a single measure of timing. To compute this measure, we took the log of the slope and converted it into a z-score. We then converted crossover to a z-score and multiplied it by -1. The final step was to take the average of the two z-scores (McMurray, Ellis, and Apfelbaum, 2018). We analyzed the cohorts and rhymes in a similar way. For cohorts and rhymes we fit a different function by using a double-gauss curve. This has six parameters as compared to the four parameters used to fit the targets. We used the same constrained gradient descent method and software for targets, cohorts, and rhymes.

Effects of noise were observed on various parameters of the target, cohort, and rhyme suggesting that listeners employ a "wait-and-see" approach to word recognition in noise. Table 2 shows there was a significant main effect of noise on the max parameter (F (1, 28) = 52.3, $p <$ 0.001). This suggests that the presence of noise impacted listeners' confidence in selecting the target. The timing of peak activation of the cohort (Mu) was significantly affected by noise (F (1,

28) = 11.7, *p* = 0.002). This shows that listeners were slower to activate the cohort in the noise condition (Mu = 763.3) as compared the quiet condition (Mu = 588.2). The height of cohort activation was also significant impacted by noise (*F* (1, 28) = 11.7, *p* = 0.002). This shows reduced activation of the cohort in the noise condition (H = 0.186) as compared to the quiet condition (H = 0.219). These findings support the notion that when speech is presented in noise, listeners engage a "wait-and-see" approach. This is evident from reduced max fixation to the target and reduced competition from the cohort and rhyme items.

Table 2. The reported measures for each of the item's parameters. F-test and p-values were observed for each of the following: main effect of noise, main effect of TempFlanker, and the interaction between noise and TempFlanker.

		Noise		Cognitive Control		Interaction	
Item	Parameter	*F*	*p*-value	*F*	*p*-value	*F*	*p*-value
Target	Timing	F (1,28) = 2.44	p = 0.129	F (1,28) = 0.004	p = 0.95	F (1,28) = 6.55	p = 0.016
	Max	F (1,28) = 52.3	p < 0.001	F (1,28) = 10.4	p = 0.003	F (1,28) = 0.112	p = 0.74
Cohort	Mu	F (1,28) = 11.7	p = 0.002	F (1,28) = 0.095	p = 0.76	F (1,28) = 1.09	p = 0.306
	H	F (1,28) = 10.2	p = 0.003	F (1,28) = 0.027	p = 0.87	F (1,28) = 0.26	p = 0.614
	S2	F (1,28) = 0.134	p = 0.717	F (1,28) = 3.66	p = 0.066	F (1,28) = 3.27	p = 0.081
	B2	F (1,28) = 17.6	p < 0.001	F (1,28) = 4.10	p = 0.053	F (1,28) = 0.063	p = 0.804
Rhyme	Mu	F (1,28) = 7.28	p = 0.012	F (1,28) = 2.27	p = 0.144	F (1,28) = 1.15	p = 0.294
	H	F (1,28) = 1.06	p = 0.312	F (1,28) = 0.200	p = 0.658	F (1,28) = 0.130	p = 0.721
	S2	F (1,28) = 0.696	p = 0.411	F (1,28) = 0.008	p = 0.927	F (1,28) = 0.044	p = 0.836
	B2	F (1,28) = 6.61	p = 0.016	F (1,28) = 4.23	p = 0.049	F (1,28) = 1.73	p = 0.199

The Effect of Cognitive Control on Real-Time Spoken Word Recognition

To analyze the interaction of cognitive control and noise, participants were divided into two groups based on a median split of Temporal Flanker scores. Participants were denoted as having good cognitive control if their Temporal Flanker score was less than 146.5 ms. Participants were denoted as having poor cognitive control if they had a Temporal Flanker score greater than 146.5 ms.

As noted previously, the max parameter of the target was affected by noise by itself. Table 2 shows there was also a significant main effect of cognitive control on the max parameter (*F* (1, 28) = 10.4, *p* = 0.016). This suggests that participants who obtained a better Temporal

Flanker score were made more fixations on the target (individuals with good cognitive control: mean = 0.86; individuals with poor cognitive control: mean = 0.79). There was a marginal effect of cognitive control on the S2 parameter of the cohort (F (1, 28) = 3.66, p = 0.066) This parameter tells us how long it participants to suppress competitors (individuals with good cognitive control: mean = 163.8; individuals with poor cognitive control: mean = 226.2) as denoted by the median split. There was also a marginal effect of cognitive control on the B2 parameter for the cohort (F (1, 28) = 4.10, p = 0.053). This parameter tells us how much participants suppressed competitors (individuals with good cognitive control: mean = 0.043; individuals with poor cognitive control: mean = 0.048). A significant effect of cognitive control was also observed for the B2 of the rhyme (F (1, 28) = 4.23, p = 0.049) (individuals with good cognitive control: mean = 1.61; individuals with poor cognitive control: mean = 0.88). These findings suggest that cognitive control may not necessarily be affecting the activation of competitors but rather the suppression of the competitors and the eventual degree of commitment to the target.

The Interaction of Cognitive Control and Noise on Real-Time Spoken Word Recognition

Finally, we asked if there was an interaction between cognitive control and noise. A significant interaction was observed between noise and Temporal Flanker score (F (1, 28) = 6.55, p = 0.016). This interaction arose because participants who performed better on the Temporal Flanker task were quicker to fixate on the target than individuals who performed worse in the Temporal Flanker task (Figure 6A). This finding suggests that cognitive control impacts real-time spoken word recognition. Figure 6A shows looks to the target. Participants who had poor cognitive control had lower peak activation of the cohort than participants with good cognitive control in quiet. Figure 6B shows that participants who had poor cognitive control took longer to suppress the cohort item than participants who had good cognitive control in noise.

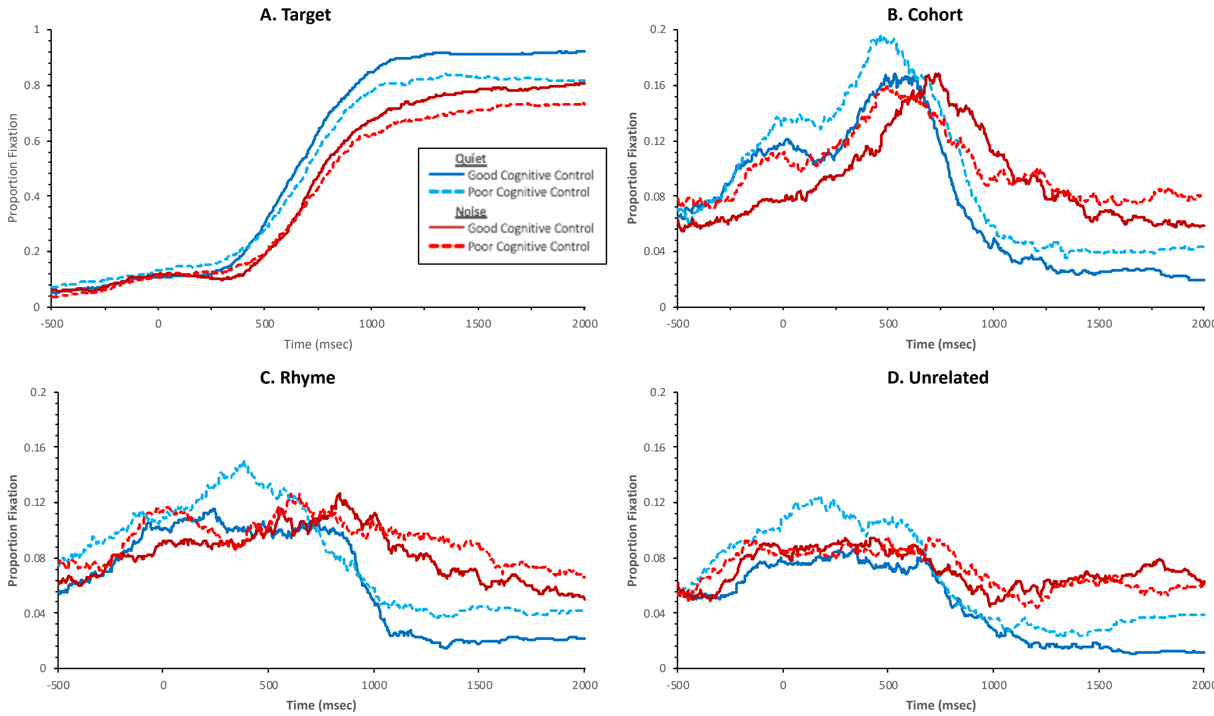

Fig. 6: Participants were grouped based off Temporal Flanker score. A, proportion of fixations to the target item in noise versus quiet. B, proportion fixations to the cohort item in noise, versus quiet. C, proportion fixations to the rhyme item in noise versus quiet. D, proportion fixations to the unrelated item in noise versus quiet.

Discussion

The study aimed to determine the relationship between cognitive control skills and understanding speech in noise. This question is critical for everyday listening as we live in a world filled with noise. The listener must be able to inhibit irrelevant lexical candidates to efficiently process a speech input. Further analysis of the interaction of cognitive control and noise showed that better cognitive control skills caused the participant to fixate to the target quicker.

To address this question, we used the VWP task to understand what participants consider before fixating on the target. This word recognition task was presented in two listening conditions: quiet and multi-talker babble. Looks to the target, cohort, rhyme, and unrelated items were recorded. Enhanced competition was observed as participants showed weaker suppression of competitors in the noise condition. Analyses of the different effects of noise and cognitive control on the different parameters suggest that noise may have an effect early in consideration of competitors words. Cognitive control may have an effect later in the time course on the suppression of the competitors as evident from marginal effects of cognitive control as opposed to an early effect on the activation of the competitors. Further analysis is needed to compare the unrelated items to the findings found in this study.

The significance of the interaction between timing and Temporal Flanker score suggest cognitive control skills could be advantageous for speech in noise. This indicates that individuals with better cognitive control skills may be getting to the target quicker than individuals with poor cognitive control skills. Previous work by Blomquist and McMurray (2017) found no correlation between lexical inhibition and cognitive control. This conclusion from this study was that cognitive control was not involved in spoken word recognition. Perhaps lexical inhibition was the only route to suppressing competitors. This study documents that there is an interaction with cognitive control and spoken word recognition. Results from the present study suggest that the architecture of spoken word recognition is not isolated to local regions. When speech becomes difficult, some type of recruiting of higher-level functioning, cognitive control, may become necessary.

The plasticity of the spoken word recognition task shown by Kapnoula and McMurray (2016) may have implications for cognitive control skills, if this same plasticity is applied to general cognitive mechanisms. This bodes well for individuals (e.g. CI users) who struggle to understand speech in noise as cognitive training could be beneficial for difficult listening situations. The implication of this study could be applied to individuals who are CI users. Pre-lingually deafened CI users show a "wait-and-see" approach to speech processing when the signal is degraded. This may due their immature frontal lobes not recruiting the right brain regions (Bunge et al, 2002). This neuroscience finding would account for the fact that post-lingually deafened CI-users behave more like NH listeners put through CIS.

In recent years, researchers have linked cognitive deficits or markers to developmental language disorders (DLD). It has been shown that children with DLD show significant deficits interference inhibition and attention (Evans, Gillam, and Montgomery, 2018). The present study's finding suggests that there is a link between spoken word recognition and cognitive control. These cognitive deficits in children with DLD may be the part of the explanation of the language disorder. Although the participants in this study were NH listeners, the implications for such findings can have an impact on multiple populations of people that struggle to understand speech in noise. This study shows that more research is needed to examine the link between spoken word recognition and cognitive control.

Acknowledgements

We would like to thank the National Institute of Health (NIH grant# P50DC000242 awarded to Bruce Gantz and Bob McMurray) and the Iowa Center for Research by Undergraduates (ICRU) for funding this study. We would also like to thank Francis Smith, Keith Apfelbaum, Jamie Klein-Packard, and Kristin Rooff for their contribution to this study.

Appendix A

List of Words used in Visual World Paradigm

Target	Cohort	Rhyme	Unrelated
bath	bass	path	couch
beak	beet	sneak	map
bear	base	pear	jet
boot	boom	suit	fox
cage	cave	gauge	hip
coat	cone	vote	ram
crown	crowd	drown	soup
dent	desk	tent	brush
gum	gut	drum	whale
hole	hose	goal	cap
horn	horse	corn	bib
lab	lamb	crab	tire
lips	list	chips	tape
mouse	mouth	house	chain
mug	mud	pug	cool
night	knife	bite	jar
pick	pit	kick	deer
porch	port	torch	milk
rip	rib	ship	dog
rose	robe	nose	pool
sick	sip	wick	door
sock	sod	dock	lap
throne	throat	crone	dish
trap	trash	nap	fist
tug	tub	rug	mic
type	tile	pipe	coach
wheat	wheel	seat	gem
wig	wind	fig	buzz
yarn	yard	barn	safe
zit	zip	sit	coal
batter	baggage	ladder	peacock
berry	barrel	fairy	rapids
carrot	carriage	parrot	tadpole
cavern	cashew	tavern	banner
coffee	coffin	toffee	knuckle
dollar	dolphin	collar	hammock
letter	lettuce	sweater	cannon
money	mother	honey	beagle
mountain	mousetrap	fountain	target
mustard	mustache	custard	penguin
paddle	package	saddle	monkey
pickle	picture	nickel	donkey
rocket	rocker	pocket	bubble
sandal	sandwich	candle	building
socket	soccer	locket	filling
tailor	table	sailor	candy
tower	towel	shower	hamster
turtle	turkey	hurdle	banjo
wizard	whistle	lizard	bottle
dragon	dragster	wagon	butter
funnel	fungus	tunnel	blanket
hockey	hotdog	jockey	button
windmill	window	treadmill	badger
robber	robin	bobber	necklace
magnet	magic	bonnet	putter
powder	power	chowder	billboard
pillow	pillar	willow	sunrise
beaver	beehive	cleaver	children
castle	cabin	tassel	water
campus	camera	camel	peanut

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.

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., & Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 54,* 243-262.

Blomquist, K., & McMurray, B. (2017). War of the words: development of the inter-lexical inhibition in typical children. Retrieved from: https://ir.uiowa.edu/honors_theses/.

Brouwer, S. & Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research, 45* (5), 1151-1160. doi: 10.1007/s10936-015-9396-9.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D.E. (2002). Immature frontal lobe contribution to cognitive control in children: evidence from fMRI. *Neuron, 33,* 301-311.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology, 42,* 317-367.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems.

Evans. J. L., Gillam, R. B., & Montgomery, J. W. (2018). Cognitive predictors of spoken word recognition in children with and without developmental language disorders. *Journal of Speech, Language, and Hearing Research, 61,* 1409-1425.

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The Process of Spoken Word Recognition in the Face of Signal Degradation. *Journal of Experimental Psychology: Human Perception and Performance, 40* (1), 308-327.

Kapnoula, E. C., & McMurray, B. (2016). Training alters the resolution of lexical interference: Evidence for plasticity of competition and inhibition. *Journal of Experimental Psychology: General, 145* (1), 8-30. http://dx.doi.org/10.1037//xge0000123.

McQueen, J. M. & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America, 131*, 509-517.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71-102. http://dx.doi.org/10.1016/00100277(87)90005-9.

McMurray, B. (2019). Nonlinear Curvefitting for Psycholinguistic (and other) Data. Retrieved from osf.io/4atgv.

McMurray, B., Ellis, T. P., & Apfelbaum, K. S. (2018). How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear and Hearing.*

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition, 169,* 147-164.

McMurray, B., Farris-Trimble, A., Seedorff, M., & Rigler, H. (2015). The effect of residual acoustic hearing and adaptation to uncertainty on speech perception in cochlear implant users: evidence from eye-tracking. *Ear and Hearing.*

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60 (1),* 1-39.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: evidence for the cognitive control connection. *Language and Linguistics Compass*, 906-924.

Peng, Z. E., & Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. *Journal of Speech, Language, and Hearing Research, 62*, 1068-1081.

Shimamura, A. P. 2000. The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28 (2)*. 207–18.

Wöstmann, N. M., Aichert, D. S., Costa, A., Rubia, K., Möller, H. J., & Ettinger, U. (2013). Reliaiblity and plasticity of response inhibition and interference control. *Brain and Cognition, 81*, 82-94.