The role of inhibitory control in spoken word recognition: Evidence from cochlear implant users

Sarah E. Colby
Department of Psychological & Brain Sciences
Department of Otolaryngology – Head and Neck Surgery
University of Iowa

Francis X. Smith
Department of Communication Sciences & Disorders
Department of Otolaryngology – Head and Neck Surgery
University of Iowa

Bob McMurray
Department of Psychological & Brain Sciences
Department of Communication Sciences & Disorders
Department of Otolaryngology – Head and Neck Surgery
Department of Linguistics
University of Iowa

Corresponding Author:
Sarah E. Colby, Ph.D.
sarah-colby@uiowa.edu
G60, Psychological & Brain Sciences Building
340 Iowa Ave., Iowa City, IA, USA, 52242

Abstract

During word recognition, listeners must quickly map sounds to meaning, while suppressing similar sounding competitors. It remains an open question whether domain-general inhibitory control is recruited for resolving lexical competition. Cochlear implant (CI) users present a unique population for addressing this question because they are consistently confronted with degraded auditory input, and therefore may need to rely on domain-general mechanisms to compensate. We examined spoken word recognition in CI users who were prelingually deaf (N=21), postlingually deaf (N=50), and normal hearing controls (NH; N=71). Participants recognized words while their eyes were tracked and completed an inhibitory control task. CI users were slower to recognize target words and did not resolve competition as fully as NH controls. Better inhibitory control predicted faster word activation in NH controls and postlingual, but not prelingual, CI users. Prolonged experience with acoustic language may thus influence how domain-general mechanisms are recruited for language processing.

Keywords: Spoken word recognition; inhibition; lexical competition; cochlear implant

Introduction

Spoken word recognition is a critical hub in language processing. It lies at the intersection of hearing, perception, and meaning, linking incoming speech to ongoing discourse. A critical question is the extent to which word recognition recruits domain-general mechanisms like inhibitory control, or whether it is entirely managed by processes internal to the language system. Three lines of work make a circumstantial case for a role for inhibitory control, but there is little direct evidence.

Word Recognition is Served by Competition

Word recognition is served by competition mechanisms. Under ideal conditions, word recognition begins immediately as the speech signal unfolds and proceeds incrementally as information accrues (Allopenna et al., 1998; McClelland & Elman, 1986). From the earliest moments, listeners consider multiple lexical candidates that match the partial incoming signal. For example, after hearing the "wi-" in *wizard* they may consider *wizard, window,* and *whistle.* Most of these candidates must then be ruled out for successful recognition. This can be visualized using eye movements in the Visual World Paradigm (VWP; Allopenna et al., 1998). In this task, participants match a spoken word to its referent from an array of pictures which includes the target word and potential candidates (e.g., for *wizard: window* and *lizard*). Eye movements to each competitor are monitored to index the degree to which different words compete over time; this shows strong evidence for partial activation and competition.

The competition posited in word recognition is superficially similar to the kind of cue- and response-conflict paradigms commonly invoked in work on domain-general inhibitory control. Inhibition – the ability to suppress a dominant or prepotent response – is a core executive function posited by the unity/diversity framework (Miyake et al., 2000; Miyake & Friedman, 2012), and it is distinct from other functions like updating or shifting. Inhibition can be assessed in tasks like the Stroop task, where participants suppress the impulse to read a conflicting written word (e.g., the word *blue* printed in *red*) and respond with the color of the text. This is similar to the need for suppressing lexical candidates during spoken word recognition.

Despite this similarity, most major mechanistic theories of word recognition (TISK; Hannagan et al., 2013; TRACE; McClelland & Elman, 1986) do not posit a role for domain-general inhibition, instead, proposing inhibitory connections *within* the lexicon, such that partially active words directly inhibit each other via lateral connections (Dahan et al., 2001; Luce & Pisoni, 1998). Unlike domain-general inhibition, this form of inhibition is precisely targeted to specific words (rather than across the board). These kind of processes can be targeted with specialized versions of the VWP and these lexical-inhibitory effects are not correlated to individual differences in domain-general cognitive control (Blomquist & McMurray, 2023; Kapnoula & McMurray, 2021).

Even if competition is managed by inhibitory processes within the lexicon, this does not rule out the possibility that domain-general inhibitory control is also involved. There is evidence that individuals with better inhibitory control are less distracted by orthographic competitors in a mouse-tracking paradigm (Zhao et al., 2022), suggesting that domain-general inhibitory control affects lexical competition, though at the level of visual forms not phonological candidates. Zhang and Samuel (2018) also document changes in lexical competition with resource depletion (in a dual task paradigm), suggesting some form of domain general involvement. Thus, the present study fills these gaps by a) directly assessing lexical competition using a standard variant of the visual world paradigm and b) relating this to a standard index of domain-general inhibitory control.

Inhibitory Control Covaries with Development and Communication Disorders

The second rationale for investigating the link between word recognition and inhibitory control is that word recognition varies across populations. Studies examining the real-time dynamics of lexical competition across the lifespan suggest it is slow to develop, with changes in efficiency well into adolescence and early adulthood (Rigler et al., 2015), and that it declines with age (Colby & McMurray, 2023). The dynamics of lexical processing are also disrupted by language disorders (developmental language disorder, aphasia; McMurray et al., 2010; Mirman et al., 2011; Yee et al., 2008). These changes take distinct forms with development, disorders, and aging affecting different aspects of competition (e.g., initial activation vs. late resolution of competition; McMurray et al., 2022). Specifically, development during childhood and adolescence primarily affects what has been termed “activation rate” – the speed of activating the target and suppressing the competitor (Figure 1B); language disorders affect the ultimate resolution of competition (at asymptote, Figure 1C); and aging leads to both slower activation rate and poorer resolution.

Critically, inhibitory control has a similar lifespan trajectory (Williams et al., 1999), with growth through adolescence and declines in older adulthood, and inhibitory control deficits have been linked to language disorders (Lukács et al., 2016). Consequently, understanding the degree to which domain-general inhibition is relevant to word recognition can reveal if differences in inhibition may account for these developmental and individual differences in word recognition. Thus, the present study included a wide age range (from 18 to 73) to capture

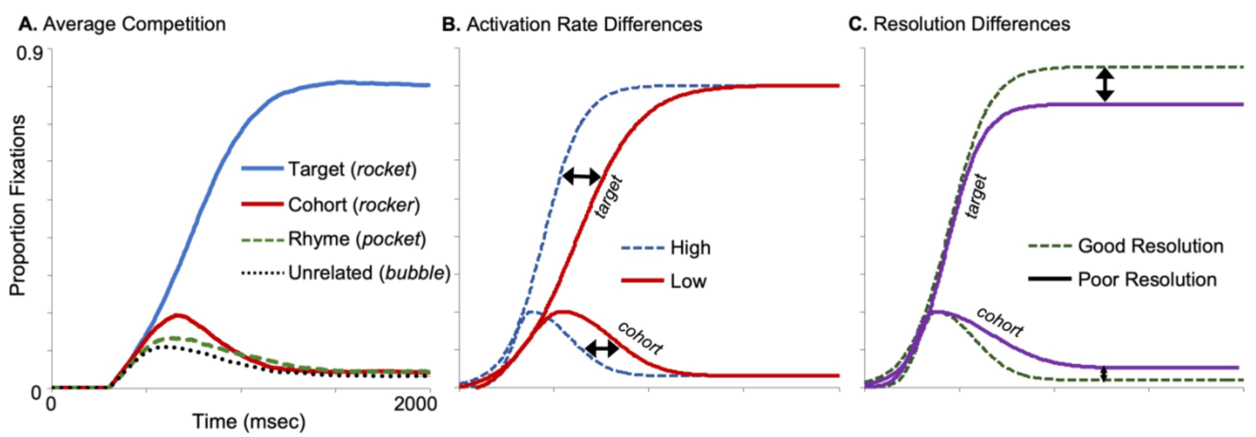


Figure 1. Average time course of fixations to different image types. A) Proportion fixations to target, cohort, rhyme, and unrelated images averaged across all participants; B) Differences in the activation rate appear across multiple components of the curves and are associated with typical development; C) Differences in resolution affect the asymptotes and have been linked to disrupted language or challenging listening conditions.

natural lifespan variation due to aging, as well as a population with known differences in language.

Is Inhibitory Control Used in Challenging Listening?

Finally, it has been posited that inhibitory control is recruited to manage difficult perceptual situations (as in the Framework for Understanding Effortful Listening [FUEL]; Pichora-Fuller et al., 2016). Ample work suggests that challenging listening may recruit some domain general resources; this is seen in both work on listening effort and studies showing recruitment of frontal regions during speech in noise (Du et al., 2014; Wild et al., 2012).

These findings have not been related to specifically to domain-general inhibitory control, or to the dynamics of lexical processing. Nonetheless, the dynamics of lexical competition are altered under various types of stimulus degradation (see Mattys et al., 2012 for review). Adverse conditions like the presence of background noise affect both the rate at which words are initially activated and the and ultimate resolution of competition (Brouwer & Bradlow, 2016; Hendrickson et al., 2020). It is not clear whether these adaptations are a natural result of noisy input, or if broader domain-general processes are involved.

We thus investigated this question in the context of Cochlear Implant (CI) users. CIs directly stimulate of the auditory nerve to replace typical (but now lost) acoustic hearing. However, due to inherent physical limitations of the electrical stimulation, CI users receive a fairly degraded input. While the majority of CI users lose their hearing in adulthood after developing language (postlingually deaf CI users), people who lose their hearing in childhood (prelingually deaf) face the added challenge of learning speech perception and language from a degraded signal.

Studies of CI users suggest a mixed role for classic domain-general cognitive variables in overall accuracy of speech perception. Skidmore (2020), for example, shows a moderate correlation between cognitive control and sentence recognition accuracy; while Heinrich et al.

(2016) show correlations of cognitive control only with sentence tasks, not with word recognition.

Moreover, the profile of lexical competition is altered in CI users. Postlingually deaf adults generally show slight delays in word recognition, but also fail to fully resolve competition (Farris-Trimble et al., 2014), what has been termed *Sustained Activation*. In contrast, prelingually deaf CI users show what has been termed *Wait and See* (Klein et al., 2021; McMurray et al., 2017), typified by a much longer delay in activating target words and *reduced* competition (since by the time listeners begin lexical access more of the word has unfolded). Normal hearing listeners exposed to spectrally-degraded speech can exhibit both these profiles, depending on the degree of degradation (Farris-Trimble et al., 2014; Hendrickson et al., 2020; McMurray et al., 2017).

It is not clear *why* these processing profiles arise. On one hand, this could simply be what happens when listeners are confronted with signal degradation. On the other hand, these profiles might arise from cognitive adaptations deployed to deal with signal degradation. Sustained activation could be useful, for example, for keeping competitors around in case an early decision must be revised; and wait and see could help listeners avoid making an early mistake by waiting for more information. Under this hypothesis, domain-general resources like inhibitory control might play a role in engaging these strategies. The two distinct profiles, however, raise a second question: if inhibitory control is involved in word recognition in challenging situations, does it only alter word recognition in one prescribed way, or can listeners use inhibitory control to flexibly modify different aspects of lexical competition to achieve their own listening goals?

The Present Study

We examined these possibilities in a group of varied listeners with CIs or normal hearing (NH). We used a standard variant of the Visual World Paradigm to track online spoken word recognition and a spatial Stroop task as a measure of inhibitory control. We examined three theoretically motivated indices of different components of lexical processing (activation rate, competition resolution, and peak competitor activation), with the hypothesis that if an effect of inhibitory control was present, it would primarily affect later aspects of word recognition (competitor resolution). We tested group membership (NH, prelingually deaf CI, postlingually deaf CI) as a potential moderator of these relationships. If inhibitory control plays a global and fixed role in word recognition, it should affect word recognition equally across all groups. Alternatively, if it is selectively invoked in challenging listening, we might expect larger effects for CI users. Finally, if the direction of the effect differs between pre- and postlingually deaf CI users, this suggests it can be deployed flexibly based on listeners needs. This in turn would support the claim the processing profiles shown by each group are actual cognitive adaptations and not just the result of the system dealing with degraded input.

Methods

Participants

Participants were monolingual adults with no neurological or developmental disorders (other than hearing loss) and normal or corrected-to-normal vision. CI users were recruited from the University of Iowa Hospitals and Clinics' Cochlear Implant Research Center. Seventy-one CI users completed all the experimental tasks. This group includes listeners with a range of hearing configurations (e.g., one or two CIs, with and without hearing aids) and device experience (M = 8.8 years, SD = 7.2) (see Supplementary Table S1). Of these, 21 participants were diagnosed with profound hearing loss before age 8 and are categorized as prelingual CI users. The remaining postlingual CI users developed hearing loss after age 18 and received their first implant later in life (no earlier than age 25).

Seventy-one normal hearing participants completed the study. They were age-matched +/- 2 years to the CI users (M_{age} = 52.7 years, SD = 14.8). These participants are a subset of the sample reported in Colby and McMurray (2023). These participants passed a hearing screening at octave frequencies from 0.25 - 8 kHz. All recruitment and experimental protocols were approved by the Institutional Review Board at the University of Iowa.

Power. This study was part of an ongoing clinical study and used a convenience sample. Sample size was limited by the number of available CI users which was difficult to predict. Thus, we did not conduct a traditional power analysis. Instead, sampling was conducted over a two-year period and stopped at a fixed time. Age-matched NH controls were recruited during this period with the total number matched to the CI users. Power was then calculated after the fact as a minimal detectable effect (MDE) size, given the obtained sample size. These analyses were based on a regression/correlation model which assumed an index from the VWP as a dependent variable, and with independent variables (IV) including age, inhibitory control, and processing speed crossed with two contrast codes for listener group (5 total IVs). Assuming N=142, α=0.05 and 1-β=0.80, the MDE of an individual variable was r^2>.053 so we had sufficient power to detect small effects.

Procedure

Participants were seated in a sound-attenuated booth with a 19” computer screen and two loudspeakers in front of them. The VWP was completed first, and the eye-tracker was calibrated with a 9-point calibration. Auditory stimuli were presented at 60 dB SPL. Following the VWP, participants completed the spatial Stroop task. Participants completed these tasks as part of a longer 2-hour visit to the lab.

Visual World Paradigm

Design. The VWP experiment used the same items and materials as reported in Colby and McMurray (2023). There were 60 item sets, 30 comprised of monosyllabic words, and 30 of bisyllabic words. Each item set included four words: a target, cohort competitor, rhyme competitor, and unrelated item (e.g., *rocket*, *rocker*, *pocket*, and *necklace*). These sets were chosen from an original list of 120 sets. After piloting the larger set of materials with 68 NH young adults, we selected the 60 items with the most prototypical pattern of competition. We obtained test-retest reliability on the final 60 sets from 29 participants who completed the VWP task twice with a week delay. Test-retest correlations were moderate-to-strong for our indices of interest (target timing: r = 0.75; competitor resolution: r = 0.62; competitor peak: r = 0.44).

Each item served as the target word once, with one additional item from each set chosen at random to serve as the target on an additional trial. This repeated target means that participants could not eliminate potential targets simply because one of the displayed competitors was the target on a previous trial (i.e., *pocket* could be the target again even if it was heard previously in the experiment). This resulted in 300 total trials (60 sets x 4 items x 1.25 repetitions).

On each trial, the four images of a set were presented in the four corners of the display, with a blue circle in the middle. Image placement was pseudo-random, such that every item type (target, cohort, rhyme, or unrelated) was equally likely to appear in any one quadrant. After a 500 msec preview of the display, the circle turned red, at which point participants clicked on the circle to play the target word. Participants then clicked on the image that best matched the auditory word.

Auditory Stimuli. Words were recorded by a female speaker of English in a quiet room at 44.1 kHz. Tokens with a consistent speaking style were chosen and edited to remove background noise and clicks. Final tokens were amplitude normalized to 70 dB.

Visual Stimuli. Stimuli consisted of colour clipart-style images. For each word, several candidate images were downloaded from a commercial clipart database. A small group of lab members then convened to choose a prototypical image and recommend changes to ensure a more prototypical depiction. These were then edited to have a cohesive style, ensure more prototypical orientations, colours (etc.), and remove distracting elements. Final images were scaled to 300 x 300 pixels relative to a 1024 x 1280 pixel screen.

Data Processing. Fixations and saccades recorded by the eyetracker were combined into looks using EyelinkAnalysis (ver. 4.12; McMurray, 2019). Regions of interest were defined as the 300 × 300 area covered by an image and extended by 100 pixels in each direction to account for any noise in the eye-track. Looks were categorized to one of the four images, or to nothing if it fell outside of the regions of interest. Only trials where the correct image was selected were included and any looks launched before the onset of the target word were ignored.

Each subject's mean timecourse of looks was calculated by averaging looks to each image type (target, cohort, rhyme, unrelated). Nonlinear curves were then fit to these averages to extract key indices that can be correlated with inhibitory control. A four-parameter logistic function was fit to the target looks with parameters for the two asymptotes the time of the crossover and the slope at the crossover. An asymmetric gaussian function (Seedorff et al., 2018) was fit to the cohort, rhyme, and unrelated looks. It has six parameters for the onset and offset asymptotes, the onset and offset slope, the height and time of the peak. These functions were fit using a constrained nonlinear curvefitter that minimized the least squared distance between the function and the data while ensuring that the function remained within reasonable bounds (ver. 29; McMurray, 2020).

We combined these parameters into theoretically motivated indices of word recognition. The *activation rate* of words was indexed by *target timing*, a composite score of the crossover and slope of looks to the target image. To construct this index, slope was log scaled (as it is zero bounded). Next, both parameters were converted to z-scores (relative to the entire sample), and crossover was multiple by -1 (so that larger values of both are related to faster activation rates). These were then averaged. *Competitor resolution* was indexed by the difference between the asymptote of target looks and the average of the asymptotic looks to the cohort (or rhyme) competitor and the unrelated image. *Peak activation* was indexed by the difference between the peak of cohort or rhyme activation and the looks to the unrelated image at the time of the cohort peak.

Spatial Stroop

On every trial, a fixation cross appeared for 200 msec and then a large arrow appeared on either the left or right half of the screen, pointing either to the left or to the right. The participant's task was to respond with the arrow keys on a keyboard to the direction that the arrow was pointing (ignoring the side of the presentation). On congruent trials, the arrow pointed in the same direction as the side on which it was presented (e.g., a left-pointing arrow on the left side of the screen). On incongruent trials, there was a mismatch between the direction that the arrow was pointing and the presentation side (e.g., a right-pointing arrow on the left side of the screen). The incongruency between the arrow and presentation side was expected to slow reaction time. The arrow remained on the screen until participants made a response. Participants were instructed to respond as quickly and accurately as possible. There were 64 congruent trials and 32 incongruent trials presented in a random order.

Data Processing. Responses were coded as correct if the participant responded with the arrow key that matched the direction that the arrow was pointing (e.g., right arrow key for right-pointing arrow). Accuracy in this task was high for all subjects (M_{NH} = 97.5%; $M_{CI\text{-}Post}$ = 97.1%; $M_{CI\text{-}Pre}$ = 95.2%). Response time was measured as they delay between the appearance of the arrow and when the participant made a key press. Average response time was slower in incongruent trials compared to congruent trials for all groups (M_{CONG} = 582 msec; M_{INCONG} = 701 msec), confirming the presence of the Stroop effect.

To calculate an individual's Stroop effect, we subtracted the average response time in congruent trials from their average response time in incongruent trials. Someone with a larger Stroop score has a larger impact of incongruency (poorer inhibitory control), while someone with a smaller score has better inhibitory control. Processing speed was taken as each individual's average response time on the congruent trials.

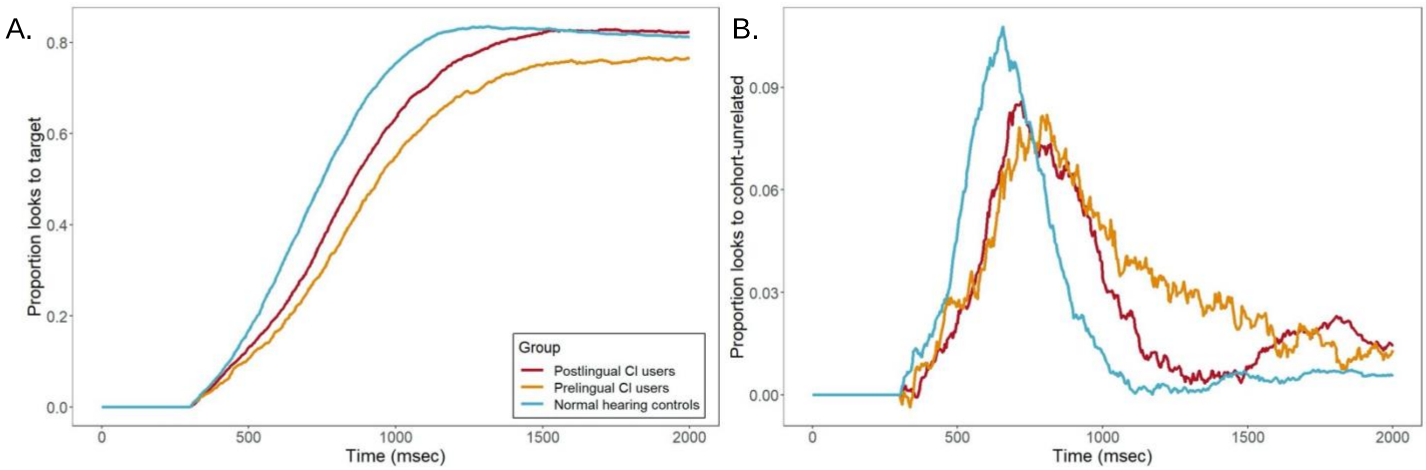
Analyses

We ran a series of linear regressions to determine the factors driving changes in spoken word recognition. The first set of models examined group differences in our three indices of interest (target timing, cohort and rhyme resolution, and cohort and rhyme peak) and the role of inhibitory control. Hearing group was contrast coded in two levels: 1) normal hearing controls (+1) versus all CI users (prelingual and postlingual CI users: -0.5); and 2) postlingual (+0.5) versus prelingual (-0.5) CI users (with NH controls set to 0). Predictors also included age (centered), processing speed (centered) and Stroop score (centered). This latter factor also interacted with hearing group. Standardized beta coefficients are reported throughout for better comparison of effect sizes.

To better understand the role of inhibitory control, we ran an additional regression investigating target timing. Hearing group, age, and processing speed were coded as described above, but Stroop score was split into three separate factors based on hearing group. That is, to isolate the role of inhibitory control as a main effect, we created one factor for each hearing group that corresponded to a subject's Stroop score if they were a member of that group (i.e., for Prelingual CI users, the Prelingual Stroop factor would be set to their Stroop score and the Postlingual Stroop factor and NH Stroop factor would be set to 0. For other participants, the Prelingual Stroop factor would be set to 0). Within these factors, Stroop scores were z-scored.

Results

Figure 1A shows fixations to the four image types (target, cohort, rhyme, unrelated) as a function of time, averaged across all listeners. We observed the typical pattern of fixations; there



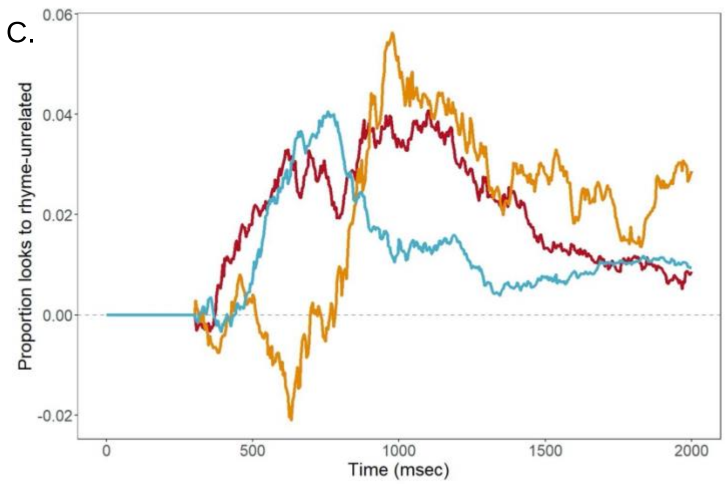


Figure 2. Proportion looks to (A) the target image, (B) the cohort image minus looks to the unrelated image, and (C) the rhyme image minus looks to the unrelated image, split by hearing group.

was an increase in looks to all image types around 200 msec after the onset of the stimulus (approximately the amount of time it takes to launch an eye movement). Looks to the target increased until reaching asymptote at around 1000 msec, while looks to the competitors quickly peaked and were suppressed at around 500 msec. Figure 2 breaks down fixations to each image type by hearing group. As expected from previous work (Farris-Trimble et al., 2014; McMurray et al., 2017), we observed the fastest timecourse of target fixations in the NH control group, followed by the postlingual CI users, then the prelingual CI users (Figure 2A). We also see reduced peak cohort fixations (Figure 2B) in both groups of CI users suggesting both are exhibiting a partial wait-and-see profile (though the prelinguals showed more delay; Figure 2A). Similarly, we observed the most complete competitor resolution in the NH controls, with slightly reduced resolution in postlingually deaf CI users, and prelingually deaf CI users showing incomplete resolution (later part of Figures 2B and C).

To statistically characterize these group differences and determine whether inhibitory control moderates these effects, our first regressions predicted one of the indices of word recognition from hearing group and Stroop congruency, with processing speed and age as additional factors.

Table 1 summarizes the linear regression predicting target timing. The NH group was faster to activate target words than the CI users (β = 0.81, t(133) = 8.65, $p < .001$). There was no significant difference between the CI users (Post- vs. Prelingual: β = 0.35, t(133) = 1.45, p = .15). Regardless of hearing group, listeners with faster processing speed were also faster to activate target words (β = -0.15, t(133) = -2.25, p = .03). The negative relationship is expected given how the variables are scaled: a higher processing speed corresponds to a slower average response time, while target timing was scaled such that higher values meant faster target fixations. Therefore, individuals with slower processing speeds are those who are slower to fixate targets. While there was no main effect of Stroop congruency (β = -0.06, t(133) = -0.59, p = .56), there was an interaction with post- versus prelingual hearing group (β = -0.55, t(133) = -2.02, p = .04). This suggests that the role of domain-general inhibitory control differed for the two CI groups (Figure 3A).

Table 1. Summary of a linear regression predicting target timing from Hearing group (NH vs. All CIs, and Post CI vs. Pre CI), Stroop congruency, Processing Speed, and Age. P values > .2 are not shown.

Factor	**Beta**	**SE**	**t(133)**	**p**
Hearing group (NH vs. All CIs)	0.81	0.09	8.65	< 0.001
Hearing group (Post CI vs. Pre CI)	0.35	0.24	1.45	0.15
Stroop Congruency	-0.06	0.1	-0.59	-
Processing Speed	-0.15	0.07	-2.25	0.03
Age	0.06	0.07	0.77	-
Hearing group (NH vs. All CIs) x Stroop	-0.21	0.11	-1.85	0.07
Hearing group (Post CI vs. Pre CI) x Stroop	-0.55	0.27	-2.02	0.04

The results for the resolution indices (cohort and rhyme; Figure 3B and D) are presented in Table 2. For both competitor types, CI users did not resolve competition as fully as the NH group (Cohort: $\beta = 0.40$, $t(133) = 3.15$, $p = .002$; Rhyme: $\beta = 0.41$, $t(133) = 3.20$, $p = .002$). Within the CI users, postlingual CI users resolved competition more fully than prelingual users (Cohort: $\beta = 0.80$, $t(133) = 2.46$, $p = .02$; Rhyme: $\beta = 0.84$, $t(133) = 2.59$, $p = .01$). There was no significant effect of any of the other factors—including inhibitory control (Stroop congruency) on cohort or rhyme resolution.

Results of regressions investigating the peak indices (cohort and rhyme; Figure 3C and E) are presented in Table 3. There were no significant effects of any of the investigated factors for either the cohort or rhyme peak index.

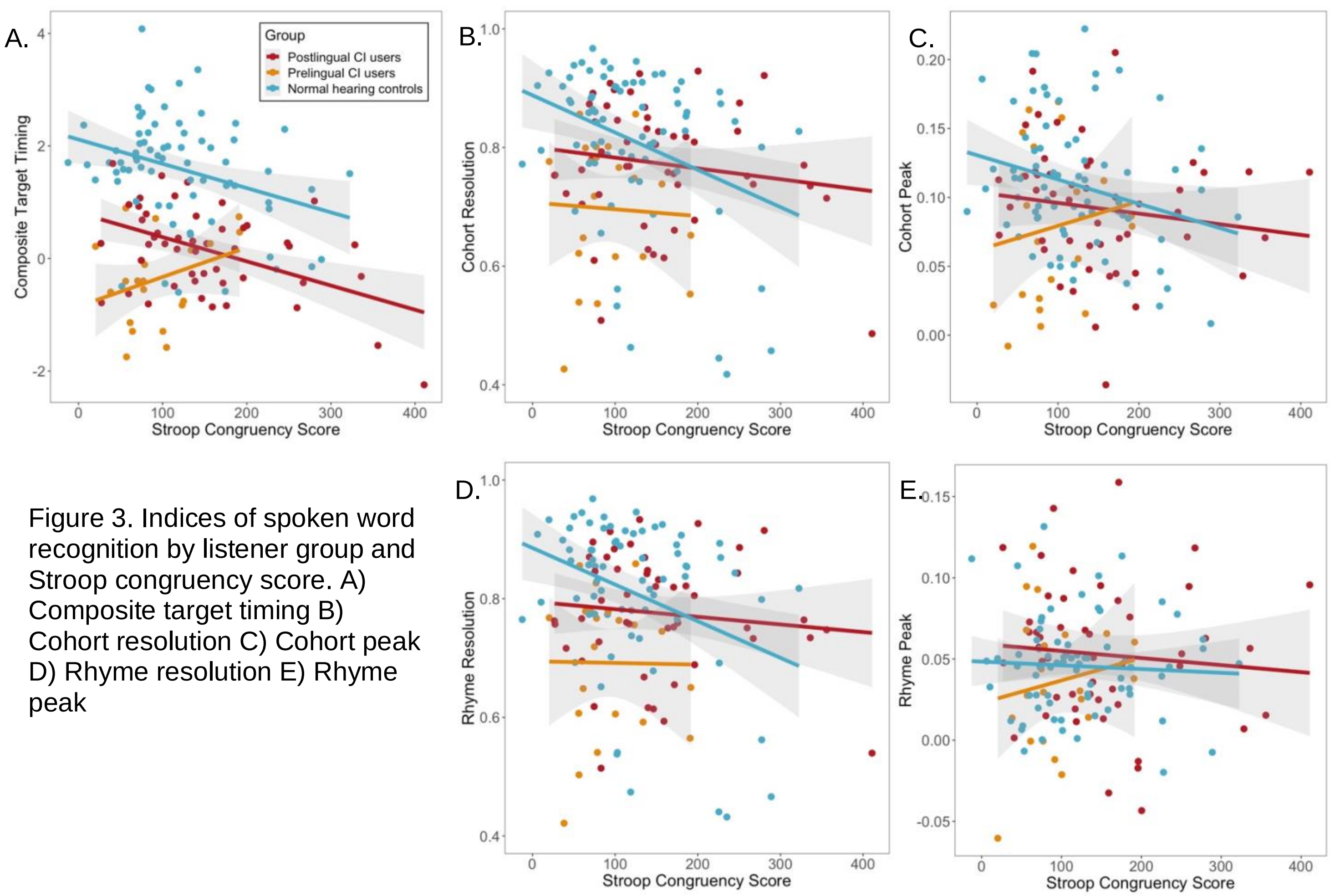


Figure 3. Indices of spoken word recognition by listener group and Stroop congruency score. A) Composite target timing B) Cohort resolution C) Cohort peak D) Rhyme resolution E) Rhyme peak

Table 2. Summary of linear regressions predicting cohort and rhyme resolution from Hearing group (NH vs. All CIs, and Post CI vs. Pre CI), Stroop congruency, Processing speed, and Age. P values > .2 are not shown.

Analysis	Factor	Beta	SE	t	p
Cohort Resolution	Hearing group (NH vs. All CIs)	0.40	0.127	3.15	0.002
	Hearing group (Post CI vs. Pre CI)	0.80	0.326	2.46	0.02
	Stroop Congruency	-0.18	0.128	-1.42	0.16
	Processing Speed	-0.04	0.091	-0.45	-
	Age	-0.07	0.099	-0.66	-
	Hearing group (NH vs. All CIs) x Stroop	-0.16	0.151	-1.09	-
	Hearing group (Post CI vs. Pre CI) x Stroop	0.01	0.369	0.04	-
Rhyme Resolution	Hearing group (NH vs. All CIs)	0.41	0.127	3.20	0.002
	Hearing group (Post CI vs. Pre CI)	0.84	0.326	2.59	0.01
	Stroop Congruency	-0.15	0.128	-1.20	-
	Processing Speed	-0.016	0.091	-0.18	-
	Age	-0.085	0.099	-0.86	-
	Hearing group (NH vs. All CIs) x Stroop	-0.184	0.150	-1.22	-
	Hearing group (Post CI vs. Pre CI) x Stroop	-0.012	0.369	-0.03	-

Table 3. Summary of linear regressions predicting cohort and rhyme peak from Hearing group (NH vs. All CIs, and Post CI vs. Pre CI), Stroop congruency, Processing Speed, and Age. P values > .2 are not shown.

Analysis	Factor	Beta	SE	t(133)	p
Cohort Peak	Hearing group (NH vs. All CIs)	0.23	0.13	1.72	0.09
	Hearing group (Post CI vs. Pre CI)	0.18	0.34	0.53	-
	Stroop Congruency	-0.03	0.13	-0.21	-
	Processing Speed	-0.1	0.1	-1.05	-
	Age	0.04	0.10	0.34	-
	Hearing group (NH vs. All CIs) x Stroop	-0.23	0.16	-1.49	0.14
	Hearing group (Post CI vs. Pre CI) x Stroop	-0.35	0.39	-0.92	-
Rhyme Peak	Hearing group (NH vs. All CIs)	-0.02	0.14	-0.14	-
	Hearing group (Post CI vs. Pre CI)	0.47	0.35	1.34	0.18
	Stroop Congruency	0.06	0.14	0.42	-
	Processing Speed	-0.04	0.1	-0.45	-
	Age	-0.07	0.11	-0.61	-
	Hearing group (NH vs. All CIs) x Stroop	-0.06	0.16	-0.39	-
	Hearing group (Post CI vs. Pre CI) x Stroop	-0.32	0.4	-0.80	-

Given the interaction between inhibitory control and the CI user group for target timing, we next conducted a follow-up analysis on to determine the nature of the moderation (Table 4). We again found that NH listeners were faster to activate targets (β = 0.89, t(133) = 10.60, p < .001). Within CI users, prelingual CI users were slower to activate targets than postlingual (β =

Table 4. Summary of a linear regression predicting target timing from Hearing group (NH vs. All CIs, and Post CI vs. Pre CI), Stroop congruency, Processing Speed, and Age. P values > .2 are not shown.

Factor	**Beta**	**SE**	**t(133)**	**p**
Hearing group (NH vs. All CIs)	0.89	0.08	10.60	< 0.001
Hearing group (Post CI vs. Pre CI)	0.42	0.20	2.12	0.04
Processing Speed	-0.18	0.07	-2.65	0.009
Age	0.06	0.07	0.89	-
NH Stroop Congruency	-0.17	0.06	-2.77	0.006
Post CI Stroop Congruency	-0.15	0.06	-2.47	0.02
Pre CI Stroop Congruency	0.07	0.06	1.25	-

0.42, t(133) = 2.12, p = .04). Listeners with slower domain-general processing speed were slower to activate targets (β = -0.18, t(133) = -2.65, p = .009) [i]. For the split Stroop congruency scores, the normal hearing and postlingual CI factors were significant (NH: β = -0.17, t(133) = -2.77, p = .006; Post CI: β = -0.15, t(133) = -2.47, p = .02), suggesting that NH listeners and postlingual CI users with poorer inhibitory control are slower to recognize words. While the prelingual CI Stroop congruency factor was not significant (β = 0.07, t(133) = 1.25, p = .21), it is worth noting that the direction of the relationship reverses for this group, suggesting that there might be the opposite relationship in prelingual CI users (and with only 21 listeners the MDE for this group in isolation was r^2>.29 – a very large effect).

Discussion

Consistent with previous findings (Farris-Trimble et al., 2014; McMurray et al., 2017), we found that NH listeners were faster to activate words and suppressed competitors more than CI users, regardless of their onset of deafness. Within CI users, prelingually deaf CI users showed more extreme delays than postlingual CI users. With respect to competition, postlingual CI users did not suppress competitors as fully as NH listeners at the end of the timecourse of processing and the prelingually deaf CI users showed even less competitor resolution. However, unlike previous findings, there was no statistical difference in the peak activation of the competitors between the listeners groups. These results do not perfectly align with the previously identified word recognition profiles in CI users (Wait-and-see and Sustained Activation; Farris-Trimble et al., 2014; McMurray et al., 2017). This is consistent, however, with recent work highlighting these profiles as continuous dimensions along which word recognition can vary across many types of listeners (CI and NH), and not standalone cognitive strategies uniquely associated with one listener group (McMurray et al., 2023).

We consistently found an influence of processing speed on the timing of word activation (about 22.4% of the explained variance). While this is not surprising, it is the first demonstration of such an effect (to our knowledge) and may reflect the influence of domain-general properties. Because of the nature of the VWP, where the cognitive processes underlying word recognition are inferred from eye movements, it is possible that this effect explains some non-linguistic variation in the task (e.g., visual search, eye-movement control) and not differences in word recognition itself. That is, our measure of processing speed is likely explaining differences in a breadth of abilities, from eye movement control and visual search to decision making. Some of these abilities are relevant for spoken word recognition, and some are necessary for the VWP (but not language processing).

With respect to the role of inhibitory control during word recognition, we found evidence that inhibitory control has an effect in certain listener groups. Contrary to our prediction, inhibitory control played an early role in activating target words. Normal hearing listeners and postlingual CI users with better inhibitory control were faster to recognize words. While it was not significant, the direction of this effect flipped for prelingual CI users, suggesting that they may be engaging inhibitory control differently than the other listener groups, to slow down rather than speed up word recognition. Further work, ideally with a larger sample size, should investigate the nature of the relationship between inhibition and spoken word recognition in prelingual CI users.

Previous work has inconsistently found a relationship between inhibitory control and word recognition (Dey & Sommers, 2015; Kapnoula & McMurray, 2021; Zhang & Samuel, 2018; Zhao et al., 2022). When a relationship has been found, it was with respect to inhibitory ability and resolution of competition (Dey & Sommers, 2015; Zhao et al., 2022). We do not find the same relationship here, but rather between inhibition and the earlier metric of target activation. Our results suggest that domain-general inhibitory control may be used to improve the efficiency of recognizing a word in listeners who have experience with acoustic language (NH listeners and postlingual CI users).

However, we also note that the effect of inhibitory control was a small contributor to the overall explained variance. For example, Stroop congruency accounted for 21% of the explained variance in postlingual CI users and for 23.3% in NH listeners. In contrast, the dichotomous NH versus CI contrast code accounted for 67.7% of the variance across the whole sample. This suggests that domain-general inhibitory control is not a requirement for achieving efficiency, but rather may play a more supportive role.

In this regard, there were hints that the role of inhibitory control may break down differently in different groups. We found a significant interaction between inhibitory control and language status. Postlingually deaf CI users showed gains in efficiency with better inhibitory control, while prelingual listeners showed either no effect or even a reversal effect. While our sample of prelingually deaf listeners was too small for a definitive picture, the fact that this group has been shown to generally exhibit a distinct profile of word recognition (Wait and See; McMurray et al, 2023) raises the possibility that they are using inhibitory control differently (and more flexibly) to achieve this goal, even as postlingually deaf listeners use inhibitory control to become more NH-like.

We set out to investigate whether domain-general resources are recruited for word recognition in challenging listening situations. Our results suggest that inhibitory control is engaged differently by listeners with varying experiences of hearing loss and language development, with normal hearing listeners and postlingual CI users engaging inhibitory control to improve efficiency in word recognition. We add to a growing body of work that suggests that there are not discrete strategies for spoken word recognition in challenging listening situations, but rather flexible dimensions along which listeners can vary.

Acknowledgements

The authors would like to thank Kristin Rooff and Marissa Huffman for their help with data collection. This work was supported by National Institutes of Health Grant P50 000242 to B. Gantz and B. McMurray and Grant DC0008089 to B. McMurray.

Declaration of Interest

The authors have no competing interests to report.

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

Blomquist, C., & McMurray, B. (2023). The Development of Lexical Inhibition in Spoken Word Recognition. *Developmental Psychology*, *59*(1), 186–206.

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. https://doi.org/10.1007/s10936-015-9396-9

Colby, S. E., & McMurray, B. (2023). Efficiency of spoken word recognition slows across the adult lifespan. *Cognition*. https://doi.org/10.31234/osf.io/gcj76

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*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychology and Aging*, *30*(3), 634–646. https://doi.org/10.1037/pag0000033

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences*, *111*(19), 7126–7131. https://doi.org/10.1073/pnas.1318738111

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. https://doi.org/10.1037/a0034353

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00563

Heinrich, A., Henshaw, H., & Ferguson, M. A. (2016). Only Behavioral But Not Self-Report Measures of Speech Perception Correlate with Cognitive Abilities. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00576

Hendrickson, K., Spinelli, J., & Walker, E. (2020). Cognitive processes underlying spoken word recognition during soft speech. *Cognition*, *198*, 104196. https://doi.org/10.1016/j.cognition.2020.104196

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, *223*, 105031. https://doi.org/10.1016/j.bandl.2021.105031

Klein, K., Walker, E., & McMurray, B. (2021). *Delayed lexical access and cascading effects on semantic activation during spoken word recognition in children with hearing aids and cochlear implants: Evidence from eye-tracking* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/mdzn7

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing*, *19*(1), 1–36.

Lukács, Á., Ladányi, E., Fazekas, K., & Kemény, F. (2016). Executive functions and the contribution of short-term memory span in children with specific language impairment. *Neuropsychology*, *30*(3), 296–303. https://doi.org/10.1037/neu0000232

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McMurray, B. (2019). *Eyelink Analysis* (4.12) [Computer software]. https://osf.io/c35tg/

McMurray, B. (2020). *Nonlinear Curvefitting for Psycholinguistic (and other) Data* (Version 30) [Computer software]. https://osf.io/4atgv/

McMurray, B., Apfelbaum, K. S., & Tomblin, J. B. (2022). *The slow development of real-time processing: Spoken word recognition as a crucible for new thinking about language acquisition and disorders*. https://psyarxiv.com/uebfc/

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. https://doi.org/10.1016/j.cognition.2017.08.013

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McMurray, B., Smith, F. X., Huffman, M., Muegge, J., Jeppsen, C., Kutlu, E., & Colby, S. E. (2023). *Fundamental Dimensions of Real-Time Word Recognition in Challenging Conditions: Insights from Cochlear Implant Users*. PsyArXiv. https://doi.org/10.31234/osf.io/fmy2a

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in Aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68. https://doi.org/10.1016/j.bandl.2011.01.004

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear & Hearing*, *37*(1), 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690–1703. https://doi.org/10.1037/dev0000044

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language*, *102*, 55–67. https://doi.org/10.1016/j.jml.2018.05.004

Skidmore, J. A., Vasil, K. J., He, S., & Moberly, A. C. (2020). Explaining Speech Recognition and Quality of Life Outcomes in Adult Cochlear Implant Users: Complementary Contributions of Demographic, Sensory, and Cognitive Factors. *Otology & Neurotology*, *41*(7), e795–e803. https://doi.org/10.1097/MAO.0000000000002682

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *The Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of Inhibitory Control Across the Life Span. *Developmental Psychology*, *35*(1), 205–213.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-Semantic Activation in Broca's and Wernicke's Aphasia: Evidence from Eye Movements. *Journal of Cognitive Neuroscience*, *20*(4), 592–612. https://doi.org/10.1162/jocn.2008.20056

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

Zhao, L., Yuan, S., Guo, Y., Wang, S., Chen, C., & Zhang, S. (2022). Inhibitory control is associated with the activation of output-driven competitors in a spoken word recognition task. *The Journal of General Psychology*, *149*(1), 1–28. https://doi.org/10.1080/00221309.2020.1771675

---

[i] Age was moderately correlated to processing speed (r = .38). We ran the same model after residualizing the effect of age from processing speed, and the results did not change. This analysis can be found on the OSF repository associated with this project (https://osf.io/4nkpx/).