



HHS Public Access

Author manuscript

Ear Hear. Author manuscript; available in PMC 2023 January 21.

Published in final edited form as:

Ear Hear. 2022 ; 43(5): 1487–1501. doi:10.1097/AUD.0000000000001203.

Lexical Access Changes Based on Listener Needs: Real-Time Word Recognition in Continuous Speech in Cochlear Implant Users

Francis X. Smith¹, Bob McMurray^{1,2}

¹Dept. of Psychological and Brain Sciences, University of Iowa

²Dept. of Otolaryngology, University of Iowa

Abstract

Objectives.—A key challenge in word recognition is the temporary ambiguity created by the fact that speech unfolds over time. In normal hearing (NH) listeners, this temporary ambiguity is resolved through incremental processing and competition among lexical candidates. Post-lingually deafened cochlear implant (CI) users show similar incremental processing and competition but with slight delays. However, even brief delays could lead to drastic changes when compounded across multiple words in a phrase. This study asks whether words presented in non-informative continuous speech (a carrier phrase) are processed differently than in isolation and whether NH listeners and CI users exhibit different effects of a carrier phrase.

Design.—In a Visual World Paradigm experiment, listeners heard words either in isolation or in non-informative carrier phrases (e.g., “*click on the...*”). Listeners selected the picture corresponding to the target word from among four items including the target word (e.g., *mustard*), a cohort competitor (e.g., *mustache*), a rhyme competitor (e.g., *custard*), and an unrelated item (e.g., *penguin*). Eye movements were tracked as an index of the relative activation of each lexical candidate as competition unfolds over the course of word recognition. Participants included 21 post-lingually deafened CI users and 21 NH controls. A replication experiment presented in the supplementary materials included an additional 22 post-lingually deafened CI users and 18 NH controls.

Results.—Both CI users and the NH controls were accurate at recognizing the words both in continuous speech and in isolation. The time course of lexical activation (indexed by the fixations) differed substantially between groups. CI users were delayed in fixating the target relative to NH controls. Additionally, CI users showed less competition from cohorts than NH controls (even as previous studies have often report increased competition). However, CI users took longer to suppress the cohort and suppressed it less fully than the NH controls. For both CI users and NH controls, embedding words in carrier phrases led to more immediacy in lexical access as observed by increases in cohort competition relative to when words were presented in isolation. However, CI users were not differentially affected by the carriers.

All correspondence should be addressed to: Francis X. Smith, Department of Psychological and Brain Sciences, 255 PBSB, University of Iowa, Iowa City, IA 52242, USA. francis-smith@uiowa.edu.

Conflicts of Interest

There are no conflicts of interest, financial or otherwise.

Conclusions.—Unlike prior work, CI users appeared to exhibit “wait-and-see” profile, in which lexical access is delayed minimizing early competition. However, CI users simultaneously sustained competitor activation late in the trial, possibly to preserve flexibility. This hybrid profile has not been observed previously. When target words are heard in continuous speech, both CI users and NH controls more heavily weight early information. However, CI users (but not NH listeners) also commit less fully to the target, potentially keeping options open if they need to recover from a misperception. This mix of patterns reflects a lexical system that is extremely flexible and adapts to fit the needs of a listener.

Introduction

Speech recognition is challenging to explain even in ideal listening situations. The speech signal is variable between talkers, and even between productions from a single talker (Liberman & Cooper, 1967; McMurray & Jongman, 2011; Newman et al., 2001), and it varies due to speaking rate (Miller et al., 1986) and coarticulatory context (e.g., Daniloff & Moll, 1968). In addition, the signal unfolds over time, leading to periods of temporary ambiguity about the intended word. For example, as a listener hears a word like *sandal*, at onset (e.g., after hearing *san-*) the input is briefly consistent with many possible words like *sandal*, *Santa* and *sandwich*.

These challenges are more daunting when the signal is unreliable, as in cochlear implant (CI) users. Due to inherent limitations in the electrical properties of the CI, frequency resolution is profoundly reduced. While CI users generally fare well with speech perception, there is considerable variability in outcomes (Blamey et al., 2013; Gomma et al., 2003). Variability in speech recognition performance in CI users is not fully explained by the auditory periphery. As a result, researchers have begun to examine cognitive processes that are harnessed to cope with ambiguity in every day (clear) speech in NH listeners that could be adapted to deal with increased ambiguity in CI users. This study focuses on the process of selecting words from a range of candidates as speech unfolds – lexical access. Experience with the CIs may alter this process because of the poorer reliability of their input (Farris-Trimble et al., 2014; McMurray et al., 2017), altering the competition mechanisms that underlie word-recognition.

Prior work has largely focused on words in isolation (Farris-Trimble et al., 2014; McMurray et al., 2017). But this process may change in several ways when words are embedded in phrases. Thus, the goal of this study is to better understand how this competition is altered in CI users when processing words in simple sentences.

Spoken Word Recognition

In normal hearing (NH) listeners, word recognition has been framed in terms of the problem of temporary ambiguity created by the fact that speech unfolds over time. It is generally agreed that NH listeners resolve this ambiguity via competition (Dahan & Magnuson, 2006; Hannagan et al., 2013; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994; Weber & Scharenborg, 2012; Zwitserlood, 1989). As soon as any input arrives, listeners partially activate multiple lexical items consistent with the signal. For example, while hearing the onset of *sandal* (*sa-*), they may consider *sandwich* and *sandbox*. Listeners

remove items from consideration when they are no longer supported by the input (e.g., at *-al*, *sandbox* drops out), and may briefly (but partially) consider words that do not overlap at onset (e.g., *handle*, *shandy*) (Alloppenna et al., 1998; Toscano et al., 2013). Thus, lexical processing begins immediately after any input, proceeds incrementally, and responds to new information immediately as it arrives.

One of the primary methods for investigating online word recognition is the Visual World Paradigm (VWP) (Tanenhaus et al., 1995). In a typical VWP experiment, participants hear a target word and see four pictures on the screen. The task is to click on the matching picture. The objects represent words that may be considered during processing (e.g., phonological competitors) or unrelated words. Participants' eye movements are tracked while they complete this task. To select a picture, participants must find it and plan a motor action, and this usually requires them to usually launch several eye movements. Because eye movements are launched while lexical processing is ongoing, the proportion of fixations to each object capture that word's degree of consideration at the point when each fixation is launched.

Typically, eye movements begin shortly after the onset of speech (~200 ms) and reflect the earliest information in the signal heard. At word onset, if a participant hears *sandal*, they will fixate both the target (*sandal*), and pictures of words that overlap at onset (cohorts like *sandwich*). Rhymes (e.g., *candle*) receive fixations later in the trial. By about one second after word onset, listeners have generally fully committed to the target.

Several lines of evidence point to the idea that the lexical system is geared toward managing uncertainty about the intended word, which could help listeners deal with challenging listening scenarios (such as spectral degradation). First, competition is biased by word frequency (Dahan, Magnuson, & Tanenhaus, 2001; Marslen-Wilson, 1987), which could serve as a prior probability of hearing a given word when the bottom-up input is insufficient. Second, words exert inhibition such that a partially active target word actively suppresses competitors (Dahan, Magnuson, Tanenhaus, et al., 2001; Luce & Pisoni, 1998). Inhibition could amplify weak differences in activation among candidates, which could be helpful in quickly suppressing residual activation of non-relevant lexical entries when the input is partially ambiguous. Third, the fact that both rhyme and anadrome competitors are active (Alloppenna et al., 1998; Connine et al., 1993; Gregg et al., 2019; Toscano et al., 2013) indicates that the system considers words even after a disconfirming segment is heard. When input is unreliable this type of system could help recover from errors when one or more phonemes were misperceived. In this situation, one of the competitors (for the mispronounced input) are likely to be the target and could be quickly reactivated.

Thus, the dynamics of lexical access do not passively reflect the unfolding input. Rather, this competition process is flexible (and may even be tuned with training, Kapnoula & McMurray, 2016) and may be well suited to dealing with uncertainty. These dynamics might also adapt to help listeners in challenging situations, such as individuals with cochlear implants who receive spectrally degraded input that may be harder to recognize.

Cochlear Implants

CI users face additional forms of uncertainty beyond temporary ambiguity, as the auditory input is characterized by reduced information. The input that CI users receive preserves the temporal envelope of the speech signal, but otherwise is spectrally degraded because 1) the thousands of frequencies discriminable on the cochlea are collapsed to a small number of electrodes (with limited benefits for adding additional electrodes: Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001); and 2) current processing strategies do not preserve temporal fine structure and may ignore some frequencies entirely.

There is substantial variability in CI users' outcomes (Blamey et al., 2013; Gomaa et al., 2003), and even experienced CI users struggle in challenging listening situations (Armstrong et al., 1997; Hochberg et al., 1992). While much of this variation is due to differences in the fidelity of the signal (Blamey et al., 2013; Henry et al., 2005; Litvak et al., 2007), CI users may also differ in how they calibrate the way lexical competition unfolds to manage uncertainty.

Supporting this, studies with NH listeners have shown that lexical competition is affected by the reliability of the signal. Brouwer and Bradlow (2016) found that adding broadband noise increases competition of both cohorts and rhymes (see also Ben-David et al., 2011). This could in part be due to the fact that the cues discriminating the target and competitors are no longer as informative (e.g., spectral detail is lost). However, enhanced activation for competitors could help cope with greater uncertainty. That is, it could be beneficial to keep competing lexical candidates available to help recover from a misperception. In these studies, the fact that the target was noise-masked makes it difficult to isolate the effects of adaptation from the fact that poor input simply contains less information contrasting the target from competitors. Thus, McQueen and Huettig (2012) added noise to a carrier sentences but the target word was always clear. When the carrier sentence contained noise, listeners looked less to the cohort and looked more to a rhyme (implying that they were processing speech less immediately). This suggests that the expectation of difficulty alone can alter the dynamics of lexical competition¹.

The changes in the profile of competition differ between these studies: Huettig and McQueen show reduced cohort (but increased rhyme) fixations while Brouwer and Bradlow show across the board increases in competitor fixations. These differences may be due to differences in tasks, the type of noise and the level of uncertainty. Still, at the broadest level, these studies suggest the lexical system can be adapted when input is unreliable or uncertain.

CI users' long-term experience with challenging input and may lead them to adapt lexical competition. Farris-Trimble et al. (2014) examined post-lingually deafened CI users in a VWP task. Relative to NH controls, CI users showed a slight delay in processing (about 70 milliseconds). The delay is likely a processing cost resulting from the decrease in signal quality, as it was also observed for NH controls hearing noise-vocoded speech. However, CI

¹It is not clear whether this is beneficial to listeners or not. McQueen and Huettig (2012) did not include a target on the screen to gauge listeners efficiency of recognizing the word. However, we now know that this profile of reduced cohort/increased rhyme fixations is consistent with a "wait and see" approach in which listeners delay lexical access until more information arrives (McMurray et al., 2017).

users also showed evidence of *sustained competitor activation*: cohorts and rhymes remained active longer (Farris-Trimble et al., 2014). This may be adaptive for coping with unreliable input quality (as in Brouwer and Bradlow, 2016), allowing greater flexibility for recovering after a misperception.

Since the signal quality is necessarily degraded in CI users, these changes may just be an automatic consequence of the poor input. McMurray et al. (2016) investigated this in a phoneme categorization task along with eye-tracking to gauge lexical competition. CI users showed more competitor fixations across all continua steps, which is consistent with the sustained competitor activation observed by Farris-Trimble et al. (2014). Importantly, after accounting for the ultimate categorization decision, both CI and NH listeners' fixations were equally sensitive to the degree of phonetic ambiguity of the auditory stimulus (indicated by the slope of the fixation response as a function of continuum step). This suggests that sustained competitor activation is unlikely to derive from the lack of ability to categorize acoustic input, as both groups were similarly sensitive to changes in VOT and frication spectrum.

Evidence that changes in lexical competition may be adaptive comes from a study examining the effects of mispronunciation. McMurray, Ellis, and Apfelbaum (2019) presented either onset or offset mispronunciations in a VWP experiment (e.g., *dog* mispronounced as *tog*) and examined the speed of fixating the target. CI users were *less* disrupted by onset mispronunciations than NH listeners. This is consistent with a strategy by which CI users slow down and keep competitors active so that they can recover when they discover a mistake (e.g., when they hear /g/, realize that *tog* is not a word, and must reactivate *dog*).

This sustained competitor activation profile is not the only possibility. McMurray et al. (2017) examined pre-lingually deafened CI users in a standard VWP task. They found substantially larger delays (>200 ms). This led to differences in the overall pattern of competition with *reduced* consideration of cohort competitors but increased consideration for rhymes (as in McQueen and Huettig, 2012). The fact that cohorts receive less consideration suggests that the delay in target fixations does not derive from increased competition. Instead, listeners appear to delay lexical processing; then when lexical access begins, they may have already heard information that rules out cohorts—a profile we termed “wait and see”. A similar pattern was also observed in NH listeners hearing severely degraded speech (4-channel vocoding), and Hendrickson et al. (2020) showed it for NH listeners hearing quiet speech. Thus, this is a general change and not unique to prelingually deaf listeners. To be clear, the term “wait and see” is meant solely as a description of the empirical phenomenon and is not meant to imply any specific mechanism – a point we return to in the general discussion.

To summarize, when listening is somewhat challenging, CI users show moderate slowing and sustain competitor activation (Farris-Trimble et al., 2014). This is at least in part a lexical phenomenon (McMurray et al., 2016) and may be adaptive (McMurray et al., 2019). However, when listening is more difficult, they “wait-and-see”, with much larger delays

in lexical access and, as a result show reduced competition from cohorts and increased competition for rhymes (McMurray et al., 2017; Hendrickson et al., 2020).

Sentence Context

Most of the above studies used words in isolation. However, the dynamics of competition may differ when words are heard in sentences (as they typically are). The obvious factor introduced in sentences is syntactic and semantic expectations (which could constrain lexical selection). There is evidence that listeners will use as much information as they have available to anticipate upcoming information in discourse (Altmann & Mirkovi, 2009; Kamide et al., 2003). While listeners could use these anticipatory mechanisms to help cope with delays in processing brought on by difficult listening scenarios, we are interested in the basic effect of embedding single words in continuous speech (without yet involving reliable anticipatory cues to the target word). Even a simple carrier phrase that is completely non-informative with respect to the target word could have at least four impacts on lexical competition.

First, a delay in lexical access of 70 milliseconds—the delay shown by post-lingually deaf adult CI users—may seem small. However, if such a delay occurred for each word in even a short phrase, it could compound, and listeners may fall far behind by the end.

Second, even semantically neutral sentences impose cognitive demands. Listeners must construct syntactic and semantic interpretations which could use resources that may be needed for lexical access (Zhang & Samuel, 2018). For example, syntactic complexity may require working memory (Just & Carpenter, 1992), or domain-general cognitive control, especially when misinterpretations are likely to occur (Novick et al., 2005, 2014). Critically, Zhang & Samuel's (2018) study of lexical access suggests that any effect of resource reduction should not be seen in the speed of initial lexical access, but rather later, in the resolution of lexical competition.

Third, prosodic cues could allow the auditory system to predict when a word is likely to begin, making it easier to initiate lexical access (Dilley & Pitt, 2010). This more predictable timing could speed lexical access as opposed to the delays predicted above.

Finally, processing speech in sentences requires a listener to segment the incoming speech signal. This makes competing predictions. One prediction would be that segmentation could require resources that are needed for lexical competition, impeding word recognition. Alternatively, in continuous mapping models (McClelland & Elman, 1986), segmentation is not distinct from lexical access. Rather segmentation occurs because listeners always activate many partial matches to the input (including items with different segmentation such as *car#go* and *cargo*), and competition sorts it out. Thus, the need to segment may lead listeners to adopt faster and more incremental lexical access (which facilitates this segmentation by access approach).

These factors could either facilitate or impede word recognition in sentence context. Thus, the first step is to identify the global effects of a word simply being in a surrounding sentence or phrase. We thus examined word recognition in carrier phrases

without constraining semantic context. Critically, while these phrases were uninformative, the location of the word in the sentence was unpredictable so listeners had to actually process the phrase to identify which word was the target. We asked whether the delays shown by CI users in carriers are similar to those of words in isolation, and if lexical processes change in continuous speech. If CI users show greater delays in sentence contexts (relative to NH), or if they show dramatically impaired accuracy this would suggest the cognitive demands of sentence processing can impact word recognition. Alternatively, CI users could show more immediate processing; this would suggest the demands of sentence processing encourage speeded activation and competition (over and above any cost imposed by the need for more cognitive resources).

While our goals are basic-science, the impact of carrier phrases on word recognition is important clinically because of ongoing debates about the use isolated words or sentences to as a measure of outcomes in hearing remediation (Geller et al., submitted; Mueller, 2001; Taylor, 2003). Understanding how word recognition differs in the context of a carrier could help understand how these tasks relate. This cannot be done simply by comparing existing sentence and word recognition tasks as words, talkers and other factors may differ between tasks. Thus, investigating the effect of simple carrier phrases on lexical processing is a starting point to examine the overall differences between sentences (more broadly) and words in isolation.

The Present Study

The present study uses the VWP to ask how lexical competition changes in continuous speech (relative to isolated words) in both CI users and NH listeners. Target words were acoustically identical across conditions to isolate the impact of the carrier phrases. Phrases were semantically unrelated to the critical word, in order to eliminate the effect of semantic/syntactic expectations to see potential effects of compounding delays, general resource utilization and segmentation from. To ensure that listeners could not just ignore the carrier phrases, we used a large number of frames, both short (*now choose the X*) and long (*in this display, find the picture of the X*), and target words could appear at several positions (*now find the X in this display*). Of course, syntactic and prosodic information offer clues to the position of the target word, but these cannot be useful unless the listener actually processes the phrase. We assessed lexical competition similarly to Farris-Trimble et al. (2014) in a VWP task with pictures corresponding to the target, cohort, rhyme, and unrelated words.

In developing hypotheses, we considered two aspects of lexical competition: the initial activation (of targets and cohorts) and later resolution of competition. If the small delay in individual words compounds across multiple words, we would expect to see delays in initial activation in continuous speech, and a potential decrease in accuracy. Alternatively, if the demands of processing words in continuous speech encourage greater incremental processing (e.g., for segmentation), we would predict faster initial activation.

Effects on later competitor resolution may come from other sources. CI users may show differences from NH listeners in asymptotic target and competitor looks, for example showing reduced target and increased competitor fixations in carrier phrases. This pattern is consistent with two explanations. First, this profile may be more useful in continuous

speech: listeners may need to revise after a misrecognition, so they do not fully commit to the target in case they need to reactivate a competitor. Second, this pattern also fits with Zhang and Samuel's (2018) finding that cognitive load impacts late competition resolution and would reflect the increased cognitive demands of sentence processing. The present study cannot disentangle these two explanations.

Finally, a null result is also possible – if carrier phrases exert no effect on recognition, this might imply that the time course of recognition is entirely driven by the signal quality of the target word itself (which is acoustically identical in both contexts).

There was a choice about how to handle the words in isolation. If the word was heard immediately at trial onset, pre-target fixations would generally be at screen center at that moment. However, in the carrier phrase condition, the delay before the target word (~1 second) would give time for fixations to be directed to the objects, creating differences at baseline. This could be fixed by delaying the presentation of the target word, eliminating the differences at baseline. However, this makes the onset of the word (in isolation) unpredictable, and this contrasts with the more predictable carrier phrase. We dealt with this issue in two ways. First, the primary experiment reported here adopts the latter, delayed onset, approach as we felt the baselining issues created significant statistical issues. However, we also conducted a replication with the immediate onset word in isolation; this shows nearly identical results and is reported completely in Supplement S1 and summarized after the experiment. Second, to deal with the baseline issues in the replication, we use a procedure described in the methods. This was also used in the supplemental experiment for continuity and to eliminate any further between subject-variance in these pre-target fixations.

Finally note that our study used a convenience sample of CI users with a wide array of device types and profiles (though all experienced post-lingually deaf adults). We did not have the sample size for a detailed individual analysis. Thus, we focus on CI users as a group of listeners with long term experience with difficult speech, a point we return to in the discussion.

Materials and Methods

Participants

Twenty-one (10 women) experienced CI patients (greater than 1-year device use) were tested. Participants were recruited from a patient registry maintained by the University of Iowa Dept. of Otolaryngology and tested when they came for their annual device tuning. Participants were required to be between 18 and 75 years, have no known neurological disorders, not have single-sided deafness, and have post-lingual hearing loss. CI users had a mean age of 59.9 years ($SD = 7.8$). Four were unilateral CI users, three were bilateral CI users, three were bimodal CI users, and the remaining eleven used a hybrid CI with aided acoustic hearing on the ipsilateral ear (some also wore a contralateral hearing aid). See Supplement S2 for more detail.

Twenty-one NH controls (10 women) were age-matched to the CI users and recruited through community advertisement. The NH controls had a mean age of 59.6 years ($SD =$

9.5). All NH controls underwent a hearing screening to ensure that their thresholds were no higher than 25 dB in the ranges of 250–4000 Hz. NH control participants were paid \$30.

All participants provided informed consent in accordance with university Institutional Review Board approved protocols for this study.

Design

This experiment used 20 item sets (see Supplement S3) of four words. Each item set consisted of a base word (target, e.g., *mustard*), a cohort (e.g., *mustache*), a rhyme (e.g., *custard*), and an unrelated word (e.g., *penguin*). This resulted in a total of 80 words. In half of the item sets all four words had one syllable, and in the other half they had two. Pictures of each of the four words in a set always occurred together during a trial. Each word in a set was the stimulus on different trials. This created four trial types. When the base word (*mustard*) was the target, each word played their assigned role (e.g., *mustache* was the cohort, *custard* was the rhyme) to create a target/cohort/rhyme/unrelated trial (TCRU). When *mustache* was the stimulus, there was a cohort (*mustard*), but *custard* and *penguin* are unrelated, a target/cohort/unrelated trial (TCUU). When *custard* was the stimulus, there was a rhyme (*mustard*), but the other words (*mustache*, *penguin*), were unrelated, a target/rhyme/unrelated trial (TRUU). Finally, when *penguin* was the stimulus, all other items were unrelated (TUUU).

The primary experimental condition was *context*: whether words were heard in isolation or in a carrier phrase, manipulated within-subject. In the carrier phrase condition, words were embedded in 40 carrier phrases (Supplement S3). The phrases crossed two factors. First, phrases were evenly divided between short (3–6 words) and long phrases (8–11 words). Second, half of the phrases ended with the target word and half had a post-phrase after the target. These factors were not intended for analysis and were only intended to prevent participants from ignoring the phrase. For each participant, phrases were randomly assigned to target words such that each phrase occurred twice within each set of 80 words.

This context condition was consistent within a block of 80 trials. Blocks alternated between words in isolation or in carrier phrases across a total of 6 blocks, and the initial block type was counterbalanced across subjects. Within a block, we did not restrict target words or phrase (the same context or word could repeat). This was done to prevent participants from discerning the correct response on the basis of which item had not yet been used in that block.

In the carrier phrase condition, listeners were expected to launch several fixations during the phrase, but prior to word onset; thus, the baseline level of fixations would be non-zero. To equate this for the words-in-isolation condition the word was delayed by the duration of a randomly selected carrier phrase (see Supplement S1 for a replication with no delay).

Each word within a set was used as an auditory stimulus 3 times in each context condition (isolation vs. carrier phrases). The experiment thus had 20 sets \times 4 words/set \times 2 contexts \times 3 repetitions for a total of 480 trials across 6 blocks.

Stimuli

Auditory stimuli were recorded by a male speaker with a Midwestern dialect. Words and phrases were recorded on a head mounted microphone in a sound attenuated room and digitized at 44,100 hz by a Kay 4300B A/D board. Carrier phrases were recorded with the nonword *tud* as the target². Five recordings of each carrier phrase were made, and the clearest was selected and the *tud* was excised. Fifty ms of silence were added to the beginning of each phrase. A small amount of silence was added to the end of each phrase to ensure a more natural prosody. Phrases had an average length of 1580 ms ($SD=543$).

The target words were recorded in the carrier phrase *choose the picture of the x*. Words in isolation were identical to the words spliced into the carrier phrases with the exception that 50 ms of silence was added to the onset of the words in isolation. There were 160 target stimuli (80 words, 2 exemplars per word). Words had a mean duration of 437 ms (range: 262–732).

Every carrier phrase was combined with every target exemplar to create 6400 possible stimuli (40 contexts \times 80 words \times 2 exemplars), of which any participant heard 240.

Visual stimuli were clip art images created using a lab procedure designed to ensure that each image is a prototypical depiction of the word (McMurray et al., 2010). Most were drawn from a lab database consisting of images previously developed with this procedure. For each image, first, multiple candidates were downloaded from a commercial clipart database. Next, a focus group of lab members reviewed the images and choose the most canonical image and determine what, if any, editing is needed. Next, images were edited to ensure a prototypical color, orientation and size, and to remove any distracting background elements. Final images were approved by a senior lab member with experience with the VWP.

Procedure

The experiment was implemented in Experiment Builder (SR Research Ltd., Ontario, Canada). Testing was carried out in a sound-attenuated booth. Participants were seated in front of a 17" LCD monitor at 1280 \times 1024 pixels. Auditory stimuli were presented in a sound field via JBL Professional 3 Series amplified speakers at $\pm 26^\circ$ from center. Listeners were permitted to set the volume of presentation to a volume comfortable for them.

On each trial, four images (300 \times 300 pixels each) were displayed on the screen at each corner, 50 pixels from the screen edge. Participants saw a red fixation circle intended to draw their gaze back to center after each response. After 500ms the circle turned blue, and participants clicked on it to hear the stimulus. In the carrier phrase condition, this began immediately; in the words-in-isolation condition there was a delay corresponding to one of the randomly selected carrier phrases. Participants were not explicitly told that there would

² *Tud* was used because 1) it is a coronal consonant and a central vowel and was expected to exert less coarticulation on the sentence context than other choices; and 2) because the stop consonant provided closure between the sentence and the target word that offered a clear point for cutting.

be delays in the word in isolation condition. After hearing the stimulus, participants clicked on the corresponding picture.

Eye-tracking recording and processing

Fixations were monitored with an Eyelink 1000 desktop-mounted eye tracker. A chin rest was used to stabilize the participant's head. The eye-tracker was calibrated using the standard nine-point procedure. A drift-correct was performed every 20 trials to correct for drifts in the eye-track over the experiment. In the event of a failed drift correction the eye tracker was recalibrated. The Eyelink 1000 uses the location of the pupil and corneal reflection to determine point of gaze every 4 milliseconds. The eye-movement record was automatically parsed into blinks, saccades, and fixations. As with previous studies (McMurray et al., 2010), each saccade was combined with the following fixation to form a "look" which lasted from the initiation of the saccade to the end of the following fixation. When identifying the object associated with each look, image boundaries were extended by 100 pixels to allow for error in the eye track. This extension did not create overlap between any regions of interest.

For analysis, we computed the proportion of trials on which the participant was fixating the target, cohort, rhyme (etc) at each 4 msec window (e.g., Figure 1). Curves were computed separately for each listener and for each condition. Analysis of the fixations used only correct trials, as this allows us to assess identify the timecourse of competition when we know that the participant eventually settled on the correct target. In contrast, including all trials could be problematic as the CI users are more likely to occasionally misperceive the word. On these trials, they would be more likely to fixate the competitors (since they would be clicking on that word), creating an artifact in the averaged time-course of fixations. Fixations for each analysis were drawn from all relevant trial types; Target fixations were estimated from all trials, Cohort from TCRU and TCUU trials, and rhyme from TCRU and TRUU trials. This meant that competitor effects were effectively counterbalanced against any differences in the properties of the two words in the set (e.g., the cohort effect averaged looks to both *mustard* [when *mustache* was the target] and *mustache* [when *mustard* was the target]).

We were concerned that differences in pre-target looking could vary among participants, creating noise due to basic oculomotor strategies (not word recognition). This was anticipated to be more of a problem in our replication experiment in which the target words were presented without a delay. To eliminate these differences, we subtracted the proportion of looks to the unrelated item at each time point from the looks to the item of interest. This is analogous to baselining the data although it is done at each time (e.g., target at time 200 msec minus unrelated at time 200) and is not relative to a specific pre-stimulus window (as would be done in EEG, for example). Raw (not baseline corrected) data are shown in Supplement S4.

Fixation curves were analyzed using a curvefitting approach based on prior work (Farris-Trimble & McMurray, 2013; McMurray et al., 2019) in which nonlinear functions are fit to each participants' data, and parameters like the slope and asymptote are used as dependent variables.

Target fixations were modeled with a logistic function which predicted the proportion of fixations to the target as a function of time (in milliseconds) after the onset of the target word. We used a 4-parameter logistic function that captured the minimum asymptote (baseline), maximum asymptote (maximum looks), the slope at the transition (rate of change at the crossover point) and the crossover (timing measure). This logistic function was fit using a constrained gradient descent algorithm that minimizes the least squared error while keeping the function between 0 and 1 and constraining parameters in reasonable ways (e.g., the baseline parameter must always fall below the maximum parameter) (McMurray, 2017).

Cohort curves did not conform to the expected shape for all individual subjects (even as the group means did). To overcome this, we jackknifed³ cohort fixation data before curve fitting (Apfelbaum et al., 2011; Ulrich & Miller, 2001), and adjusted the F statistic by dividing it by $(n-1)^2$ where n is the number of observations in each cell of the design (F_{jk}). Jackknifing was done separately for each group (NH/CI). Jackknifed competitor fixations were then modeled with an asymmetrical Gaussian function made up of two Gaussians that meet at the midpoint. This allows for separate parameter estimates for the asymptote and slope on either side of the peak of the curve as well as the peak height, and peak timing.

In mapping these parameters to theoretical constructs, we note that changes in the initial activation of targets and competitors should appear in a number of places. When activation is more rapid, we would expect steeper target slopes and earlier crossovers. These will be coupled with higher cohort peaks (since people make more rapid commitments to the cohorts). Conversely, late-stage competition resolution should appear as differences in the asymptotes of the targets and cohorts, and in the offset slope of the cohort.

Results

Accuracy

First, we examined mouse-clicks with a 2×2 ANOVA (listener group \times carrier phrase) to ensure that listeners were accurate enough in the task to justify the fixation analysis. Both listener groups were generally accurate, though CI users were significantly less so ($M = 93.10\%$ correct, $SD = 5.86\%$) than NH listeners ($M = 99.77\%$ correct, $SD = 0.55\%$; $F(1,40) = 27.58$, $p < .001$). Accuracy was also similar between words in isolation and in carrier phrases, though there were significantly more errors in the carrier phrase condition ($M = 96.00\%$, $SD = 3.63\%$; $F(1,40) = 14.24$, $p < .001$) than in the word in isolation condition ($M = 96.87\%$, $SD = 2.77\%$). There was a significant interaction, $F(1,40) = 7.95$, $p = .007$. This was driven by the fact that the effect of context was only significant for CI users, $t(20) = 3.58$, $p = .002$, but not for NH listeners, $t(20) = 1.22$, $p = .236$. This suggests that carrier phrases had a detrimental (not facilitatory) effect on CI users' performance. However, even with these differences, accuracy was sufficiently high across both listener groups and conditions to justify analysis of the fixations, and the differences were numerically small and unlikely to substantially bias the fixation analysis.

³To jackknife the data, we computed average fixation curves within each carrier condition and listener group as a function of time across all participants, but with one excluded. We next fit curves to this jackknifed data. This procedure is repeated, excluding a different participant each time, resulting in a data set with the same number of subjects as the original data set. These estimates are then compared using test statistics that are adjusted for the fact that there is less variance between "subjects" after jackknifing.

Fixations

Figure 1 shows the fixations to the target, cohort and rhyme (after baseline correction) as a function of time. Both groups initiated looks to the target faster when words occurred in carrier phrases. In addition, both groups fixated cohorts earlier and to a slightly higher degree in carrier phrases. CI users still show the previously observed delays relative to NH listeners (Farris-Trimble et al., 2014), but the effect of carrier phrases was similar to that of the NH.

Target fixations.

Figure 2 shows the timecourse of target fixations. Logistic fits were good in all conditions, average $R = .998$ across the four cells of the design (CI/NH [listener group] \times Word in Isolation / Carrier Phrase [context]). There were small differences among conditions (Supplement S5), but fits were highly accurate, and all groups were near ceiling. No subjects were excluded for poor fits.

The *baseline* parameter showed no significant effect of listener group or context, all $p > .25$. There was a significant interaction, $F(1,40) = 4.44$, $p = .041$. However, paired t -tests revealed no significant difference between carrier phrases and words-in-isolation for either CI users, $t(20) = 1.67$, $p = .113$, or NH listeners, $t(20) = 1.29$, $p = .211$.

Analysis of the *crossover* (Figure 2B) revealed a significant effect of listener group, $F(1,40) = 32.67$, $p < .001$, as CI users had significantly later crossovers than NH listeners (by 127 ms). There was a marginally significant effect of context, $F(1,40) = 3.91$, $p = .055$. However, there was also a significant interaction, $F(1,40) = 6.46$, $p = .010$. Follow-up paired t -tests reveal that NH listeners showed earlier crossover points in carrier phrases compared to words-in-isolation, $t(20) = 3.60$, $p = .002$, while CI users did not, $t(20) = 0.36$, $p = .722$.

The target *slope* (Figure 2C) patterned differently than the crossover. We found a significant effect of listener group, $F(1,40) = 5.59$, $p = .023$. NH listeners showed steeper slopes than CI users. There was also a main effect of context, $F(1,40) = 8.54$, $p = .006$, with no significant interaction, $F(1,40) = 2.93$, $p = .095$. Both groups showed steeper slopes when words were in isolation compared to occurring in carrier phrases.

Maximum looks to the target (Figure 2D) were scaled with the empirical logit prior to analysis to attain a more Gaussian distribution. This showed no significant effect of listener group, $F(1,40) = 0.50$, $p = .485$. However, there was a significant effect of context, $F(1,40) = 6.93$, $p = .012$, and there was a significant listener group \times context interaction, $F(1,40) = 8.16$, $p = .007$. We followed up on this with paired t -tests for each group (CI and NH). These tests revealed that CI users showed higher maximum target looks for words-in-isolation than in carrier phrases, $t(20) = 4.84$, $p < .001$. In contrast, NH listeners showed no significant difference between contexts, $t(20) = 0.14$, $p = .893$.

Cohort Fixations.—Cohort fixations in each condition are shown in Figure 3. The fits of the jackknifed data were good across conditions and for both CI users and NH listeners, mean $R = .989$, with no significant differences (Supplement S5). No subjects were excluded for poor fits.

Onset baseline showed no significant differences between listener group or context and there was no significant interaction, all $p > .65$. As with the target, this suggests that subtracting unrelated items was a useful baseline method. The *onset slope* similarly showed no significant effects of group or context and no significant interaction between the two, all $p > .35$.

Peak height (Figure 3B) showed significant differences between listener groups, with CI users showing reduced peak cohort fixations compared to NH, $F_{jk}(1,40) = 9.99, p = .003$. There was a significant effect of context with increased peak looks in carrier phrases compared to words-in-isolation, $F_{jk}(1,40) = 12.32, p = .001$. There was no significant interaction between listener group and context, $F_{jk}(1,40) = 0.16, p = .686$. *Peak timing* (Figure 3C) did not differ significantly by group or context, and there was no significant interaction, all $p > .4$.

Offset baseline (Figure 3D) showed a marginally significant effect of group, $F_{jk}(1,40) = 3.33, p = .076$. CI users showed slightly higher baselines at the end of the trial ($M = .006$), relative to NH ($M = .002$). There was no significant effect of context, $F_{jk}(1,40) = 2.72, p = .107$, and no significant interaction, $F_{jk}(1,40) = 0.54, p = .465$. *Offset slope* (Figure 3E), however, showed a significant effect of listener group with CI users taking more time to suppress cohorts than NH listeners, $F_{jk} = 7.52, p = .009$. There was no effect of context, $F_{jk}(1,40) = 0.58, p = .450$, and no interaction, $F_{jk}(1,40) = 0.32, p = .570$.

Rhyme Fixations.—Figure 4 shows rhyme fixations in each condition over time. Even the group-level looking showed non-canonical patterns, particularly in CI users. This was driven by the fact that many individual CI users fixated the rhymes more than the unrelated baseline even before word onset. Because of these early looks, rhyme looks could not be fit with the asymmetrical Gaussian. This difference in baseline looking (unique to CI users) may suggest a meta-linguistic strategy on these trials (as rhyme trial data can only be taken from TCRU and TRUU trials). Rhymes in TRUU trials may simply be more salient and CI users may take advantage of this to cope with difficulty processing the input.

When modeling the CI data, a variety of curve-fitting procedures were exhausted without success. Supplement S6 presents several coarser analyses of rhyme fixations across listener group and context. In these analyses, rhymes are active later for CI users than for NH listeners, stay active for a greater extent of the trial, and reach marginally higher peak fixations. We are hesitant to draw strong conclusions regarding rhymes, however, given the relatively weak effects and the unexplained difference at baseline.

Replication Summary

As we described, one concern with this experiment was that in the words-in-isolation condition onset of the target word was unpredictable, whereas if listeners were following the carrier phrase the onset was predictable in this condition. While this was necessary to achieve a comparable baseline, some of the differences between the two context conditions could be due to predictability. We thus replicated this experiment using identical methods, except that the words presented in isolation were always presented 50 msec after clicking

on the fixation circle. Twenty-two CI users with a similar profile to the primary experiment were tested.

A full report is available in Supplement S1 (Figure 5, 6 for targets and cohorts). The results are strikingly similar across the two experiments. In this experiment, CI users showed a delay in target fixations relative to NH, a reduction in peak cohort fixations, and a reduced asymptote for the target. The effect of context was also consistent across experiments with a faster commitment to the target in both groups in carrier phrases. The effects on maximum target looks differed between experiments, but the replication was consistent with the broader finding of a larger effect of sentence context on the asymptote in CI users.

General Discussion

The primary goal of this study was to understand the impact of non-informative carrier phrases on the dynamics of word recognition. This was motivated by specific issues raised by CI users, but as the role of uninformative carrier phrases on the dynamics of word recognition has not been widely studied in the basic science of language processing this study adds to our understanding of these impacts as well. We start by addressing the overall difference between CI and NH listeners. We then turn to the influence of carrier phrases on both the earliest moments of processing, and the later resolution of competition.

CI vs. NH listeners

Across two experiments, CI users showed somewhat larger delays ($M = 135$ ms) (in both context conditions) than have been previously observed (prior delays are typically around 75 ms: Farris-Trimble et al., 2014). CI users also showed decreased peak cohort fixations, whereas Farris-Trimble et al. (2014) found increased peak cohort fixations for similar CI users. CI users show greater rhyme consideration than NH listeners (Supplement S6). These three findings look a lot like the “wait-and-see” pattern described earlier for prelingually deaf children (McMurray et al., 2017), and for NH listeners hearing substantially degraded input or soft speech (Hendrickson et al., 2020; McMurray et al., 2017). This is surprising given that the population was post-lingually deaf and is a fairly close match to that of Farris-Trimble et al. (2014).

One of our hypotheses was that the delay observed previously in isolated words might compound across a sentence to create even larger delays. However, even in this population where we expected the delay in isolated words to be small, we found a pattern of lexical competition changes normally associated with listeners who encounter much greater difficulty. It could be that we happened to test a sample of CI users who are simply experiencing greater difficulty than would be expected with their CI. However, as Supplement S2 indicates, CI users were experienced, and nearly all showed good speech perception outcomes. Moreover, this was observed in a sample of 43 CI users across both experiments. Unfortunately, our sample sizes in each experiment were not sufficient to attempt a serious individual differences analyses (e.g., as a function of device type or performance), but this is an important area for future work.

There are two possibilities for these differences. First, Farris-Trimble et al. (2014) may have sample unrepresentatively good performers. If so, our data suggest that far more CI users exhibit this profile than previously observed. Perhaps this profile is much more common, and perhaps helpful, even for CI users with relatively good outcomes. Alternatively, the structure of the experiments in this study may have led CI users to adopt the “wait-and-see” profile even if they would not normally exhibit this in a typical experiment with only words in isolation. Specifically, listeners could be adapting to the fact that on many trials a sentence may be present (even if that adaptation is not helpful on all trials). That is, listeners may flexibly adapt lexical processing to task-demands across the experiment, even if they cannot do so at a finer timescale (between 10-minute blocks). It is not clear whether this adaptation is helpful.

Critically, it is important to note that wait-and-see is meant only as a description of the empirical pattern of data. The underlying mechanisms are unclear. They could derive from a delay in lexical access, reduced signal fidelity or both (c.f., McMurray, Apfelbaum and Tomblin, submitted). Understanding which of these mechanisms is at play is critical for understanding how CI users achieve this adaptation and is an important area for future work.

Surprisingly, CI users not only showed a “wait-and-see” profile here, but also showed evidence of sustained competitor activation. CI users took longer to suppress the competitor than NH. Moreover, they show a reduced asymptote to the target, and the analyses across both experiments (Supplement S1) suggest they suppress the competitor less fully than NH listeners. This suggests that these two profiles are not mutually exclusive—CI users can both wait and see and withhold a complete commitment to the target word. These profiles may derive from different causes and solve different problems. This combination of “wait-and-see” and sustained competitor activation may be ideal when CI users process sentences.

Effect of Continuous Speech: Timing and Activation Rate

Overall effects of carrier phrases on the *timing* of target fixations were small. NH listeners showed earlier crossover points in our main experiment when listening to words in continuous speech. However, the benefit was relatively small, and it was not found in our supplemental follow-up. In contrast, both groups showed a steeper slope in words in isolation than in carrier phrases. Normally crossovers are anti-correlated with slope: when listeners show a delayed crossover, it is because they have a shallower slope (Farris-Trimble & McMurray, 2013; McMurray et al., 2010, 2019). Here we see the reverse in our primary experiment (a later crossover, but steeper slope in words-in-isolation) or no correlation in the replication (no change in crossover, but a steeper slope in words-in-isolation).

Given this, steeper slopes suggest that when words occur in isolation it takes longer to initiate lexical access even if processing “ramps up” to make up for this delay. In contrast, in carrier phrases, target fixations numerically begin earlier (though this is not directly captured in any of our parameters) and show similar crossover times, even as they start earlier. Notably, competitor activation was also significantly altered by the presence of carrier phrases in both groups. Even for CI users, who showed lower overall cohort activation

than NH listeners, embedding acoustically identical target words into a carrier phrase led to higher peak cohort fixations. Thus, this too appears to be general difference due to context.

One explanation for this pattern of results would be the need for immediate processing. Carrier phrases increase the need for immediate and incremental processing because auditory input precedes and often follows the target word. Listeners do not have the luxury of time, and this could lead them to adopt this more immediate processing mode. Critically, more immediate processing also leads to information early in the word being weighted more heavily. Consequently, target fixations may get above baseline faster, and they may also show increased cohort fixations, since the earliest bits of information are consistent with both. This is exactly what was observed across both experiments and is consistent with what we have termed increased *activation rate* (McMurray et al., submitted).

It may be the need for segmentation that drives listeners to more heavily weight early information during processing. Continuous mapping models (such as TRACE: McClelland & Elman, 1986) rely on competition to handle word segmentation. These models activate a range of possible segmentations and let competition arrive at the optimal parse for the input. This allows these models to accomplish segmentation without any dedicated segmentation processes. If segmentation is accomplished via competition and listeners need to process speech in phrases, heavily weighting early information would allow the competition to begin earlier; the need for this may be enhanced in uninformative carrier phrases where semantic context is not available to guide segmentation. This could lead to overall increases in the speed of both lexical access and segmentation as word boundaries may be discovered more quickly.

Even as listeners initiated lexical access faster in continuous speech, it remains that overall slope was shallower in this condition. This suggests a slower activation rate by the end of the trial. Listeners may be attempting to utilize semantic and syntactic information from the carrier phrase to constrain lexical access and rule out competitors. However, as the phrases in this study were non-informative, this may impose processing costs that slow lexical processing after initial access. In normal discourse, such a process would likely speed activation by ruling out competitors inconsistent with the semantic context (Dahan & Tanenhaus, 2004), but in this study this process is likely costly as it offers little to no constraint on lexical access. This again speaks to the flexibility of lexical access dynamics.

It is important to note that this effect was much stronger in the primary experiment than in the replication. In the primary experiment it appeared as a significant effect on slope, but only as a marginal main effect on crossover (which was larger in NH listeners). In contrast, in the replication study it only appeared in the slope. The larger effect in NH in the primary experiment may derive from the fact that this effect would likely be harder to observe in CI users because of their substantial delays in processing. The differences between experiments may derive from the predictability in word onsets. Carrier phrases provide timing cues as to when each word is about to begin which could help listeners prepare to initiate lexical access or to launch an eye-movement (Dilley & Pitt, 2010). In the word-in-isolation condition of the main experiment, the seemingly random delays introduced meant that only the sentence context condition had a predictable target-onset. As a result, the crossover

difference observed in NH listeners could be due to either preparation (in carrier phrases) or surprisal (in the word-in-isolation condition). However, in the replication, the word was predictable in both cases, which may explain why the effect was weaker. However, the significant effect on slope – even in the replication – suggests this cannot entirely explain the effect.

However, at the broadest level, CI users appear to show qualitatively similar effects of carrier phrases on lexical processing to NH listeners, even as CI users still experience delays in fixating the target. The impact of carrier phrase on crossover was minimal for both groups of listeners; both groups seem to weight early information more heavily for words in phrases than words in isolation; and both groups show overall shallower target slopes in sentence contexts. This is not consistent with our hypothesis that continuous speech could lead to catastrophic delays in processing for CI users. Perhaps this is unsurprising given CI users' general success at everyday communication. The phrases in this study, however, were intentionally non-informative so it is encouraging that even in the least helpful contexts, sentence processing does not impose significant costs in word recognition for CI users. Rather, the effect of carrier phrases suggests that CI users can harness similar segmentation strategies to those that NH listeners use.

Competition Resolution in Continuous Speech

Across both experiments some results suggest CI users did not resolve competition as fully as NH listeners. Our primary experiment found competitor fixations took significantly longer to reach baseline in CI users (though this was not significant in the replication); in both experiments, the offset baseline for the cohort approached significance, and a combined analyses (Supplement S1) showed robust evidence for this. We also found indirect evidence for this in target fixations, where the maximum looks in carrier phrases were reduced in CI users relative to NH listeners (in both experiments). This—in addition to the increases in cohort fixations—is a strong indicator (and perhaps statistically more reliable) marker that CI users are not fully committing to the target—the sustained activation profile. This suggests CI users keep competitors active and do not fully commit to the target even at the end of processing. This may be adaptive: when mistakes in recognition are more common it would make sense to hold onto competitors, even if only to a small degree, as it would make it easier to reactivate these items to correct for potential misrecognitions (McMurray et al., 2009). We note that the phrase that follows the target word (in half the carrier phrases) was uninformative; thus the decision to do this is likely more of a heuristic applied to any sentence, and not tied to the specific content (and see Sarrett et al., 2020 for converging evidence for the maintenance of phonetic ambiguity).

These findings support a role for cognitive adaptation of lexical processing that may be intended to maximize the potential to recover from misrecognitions. That is, by withholding a complete commitment, CI users may be more flexible in recovering from a misperception. Indeed, the slower growth of target fixations later in the trial (in carrier phrases) may also support this as another marker that listeners avoid making strong commitments too rapidly. Critically, it is in sentence context where this approach would be most useful (since later words may provide information or it may take time to integrate semantic or syntactic

expectations). This suggests CI users exhibit this sustained activation strategy specifically in contexts where it would be most beneficial.

Clinical Implications

For the purposes of clinical measures, we found that accuracy was lower (particularly for CI users) when words were embedded in continuous speech. Our task is unlike any of the clinically common sentences recognition tasks: it was a 4AFC closed set task, carrier phrases were uninformative, and the response was picture selection, not speech production. Nonetheless, even without these critical aspects of clinical sentence-based tasks, embedding words in carrier phrases led to more errors. It is not clear if these increased errors reflect more challenging perceptual processes, or something about the higher-level demands of sentence processing (e.g., if syntactic parsing is resource demanding). However, two reasons favor the latter. First, the auditory input was identical across the conditions. Second, the eye-movement results did not show large delays in processing the specific target word between the conditions. This adds to the evidence sentence level tasks invoke more than just auditory/perceptual processes. While such demands may increase their real-world predictability, they may also make them less diagnostic of specific perceptual issues (Geller et al., submitted). However, it is encouraging to note that the lack of compounding delay suggests that CI users may be relatively unaffected by the short sentences used in many clinical assessments. In fact, embedding words in sentences may be ideal for encouraging variance in word recognition tasks if participants are performing near ceiling (making it difficult to detect differences).

Limitations

There are several limitations to this study. First, we used non-informative carrier phrases which are uncommon in everyday language use. This makes it impossible to tell if these effects (like increased competitor competition) would be observed with more natural sentences. However, had we used more natural sentence contexts the semantic constraints alone may have obscured any differences caused simply by embedding words in continuous speech.

Second, the patterns of rhyme fixations in both experiments suggest that CI users may be exhibiting some meta-linguistic strategy. It could be that CI users pick up on the target-rhyme relationship more easily given people's wide awareness of rhymes. They may then exploit this by launching fixations to the rhyme above baseline. Even discounting the early looking, the rhyme effects in both experiments were weak, making it hard to detect effects of listener group or context condition. More careful construction of rhyme sets may eliminate both of these issues in future studies. However, this strategy did not appear to affect the cohort or target fixations (which were at baseline), so it is unlikely to affect our primary conclusions.

Third, neither the primary experiment nor the replication was able to perfectly match the words-in-isolation to the carrier phrase condition in terms of the predictability of the word. The primary experiment added "carrier phrase duration" quantity of silence to the words in isolation. This equated the time for visually scanning the display between the conditions,

but as a result the onset of the words in isolation could not be predicted (as it could in the phrases). The replication controlled for this with a very predictable word onset (always 50 ms after mouse click). However, this allowed more time to scan the scene in the carrier phrase condition. Even as neither is perfect, the fact that the pattern of results was broadly similar across the two experiments seem to make this a small drawback.

Finally, our study compares CI users *as a group* to NH listeners. Our convenience sample of CI users was necessarily broad, with substantial heterogeneity in device type, age, and etiology and length of deafness (Supplement S2). Any of which could influence results. Our sample sizes were insufficient for any serious individual differences analyses (e.g., a correlation or sub-group approach), and though the VWP shows decent test/retest reliability (Farris-Trimble & McMurray, 2013), it is not generally used as a single subject measure. Thus, we focused on group differences, which treats a particularly heterogenous population as one group. There may be important differences among CI users. For example, hybrid and bimodal listeners with access to F0 may be able to better segment words from running speech than electric only listeners (Spitzer et al., 2009). These are important avenues for future work. At the same time, we note that at the level of lexical processing (as studied with the VWP), the pattern of lexical competition shown by CI users often matches that of NH listeners with vocoded speech - a mediocre approximation of real CI input (Farris-Trimble et al., 2014; McMurray et al., 2017), and also with background noise (Brouwer & Bradlow, 2016), and even in quiet speech (Hendrickson et al., 2020). Thus, at the level we are studying, the specifics of the signal processing may not matter substantially. Instead, we would argue that in the context of this experiment, our CI users should be viewed as a fairly representative sample of individuals with long term post-lingual experience with degraded speech input.

Conclusions

While CI users showed substantial delays in target fixations, they showed no evidence of catastrophic delays when words were presented in non-informative carrier phrases. In both carrier phrases and isolated word conditions, CI users show decreased cohort competitor fixations during processing compared to NH listeners, though they also sustain competitor activation later into the trial as well as well. This suggests that real-time processing is altered in multiple ways depending on the task demands. Moreover, carrier phrases lead listeners (both NH and CI) to weight early information more heavily than when words are presented in isolation as evidenced by increased activation of cohort competitors. Finally, late in the trial, CI users, particularly in carrier phrases, show less complete commitment to the target, and increased competitor activation. These results suggest that the lexical system is flexible and can adapt to fit the needs of the listener.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Thanks to Kristin Rooff and Sarah Colby for assistance in data collection as well as feedback on the design and interpretation of this project. Thanks to Susan Wagner Cook for early feedback leading to the creation of supplemental experiment to control for timing issues. This research was supported by NIH Grants DC008089 awarded to Bob McMurray, and DC000242 awarded to Bob McMurray and Bruce Gantz. Portions of this article were presented at the 42nd Annual MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, Maryland, February 9, 2019. F.X.S designed and conducted experiments, analyzed data, and wrote the paper; B.M. designed experiments, offered critical revisions, and provided guidance on statistical analyses. Both authors discussed the results and implications and reviewed the manuscript at all stages. This article has appeared as a preprint (<https://psyarxiv.com/wyaxd/>).

Source of Funding:

This research was supported by NIH Grants DC008089 (to B.M.) and DC000242 (to B.M. and B. Gantz).

References

- Allopenna PD, Magnuson JS, & Tanenhaus MK (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. 10.1006/jmla.1997.2558
- Altmann GTM, & Mirkovi J (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609. 10.1111/j.1551-6709.2009.01022.x [PubMed: 20396405]
- Apfelbaum KS, Blumstein SE, & McMurray B (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review*, 18(1), 141–149. 10.3758/s13423-010-0039-8 [PubMed: 21327343]
- Armstrong M, Pegg P, James C, & Blamey P (1997). Speech perception in noise with implant and hearing aid. *The American Journal of Otology*, 18(6 Suppl), S140–1. <http://www.ncbi.nlm.nih.gov/pubmed/9391635> [PubMed: 9391635]
- Ben-David BM, Chambers CG, Daneman M, Pichora-Fuller MK, Reingold EM, & 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(1), 243–262. 10.1044/1092-4388(2010/09-0233)
- Blamey P, Artieres F, Baskent D, Bergeron F, Beynon A, Burke E, Dillier N, Dowell R, Fraysse B, Gallégo S, Govaerts PJ, Green K, Huber AM, Kleine-Punte A, Maat B, Marx M, Mawman D, Mosnier I, O'Connor AF, ... Lazard DS (2013). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients. *Audiology and Neurotology*, 18(1), 36–47. 10.1159/000343189 [PubMed: 23095305]
- Brouwer S, & Bradlow AR (2016). The Temporal Dynamics of Spoken Word Recognition in Adverse Listening Conditions. *Journal of Psycholinguistic Research*, 45(5), 1151–1160. 10.1007/s10936-015-9396-9 [PubMed: 26420754]
- Connine CM, Blasko DG, & Titone D (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32(2), 193–210.
- Dahan D, & Magnuson JS (2006). Spoken Word Recognition. In *Handbook of Psycholinguistics* 10.1016/B978-012369374-7/50009-2
- Dahan D, Magnuson JS, & Tanenhaus MK (2001). Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements. *Cognitive Psychology*, 42(4), 317–367. 10.1006/cogp.2001.0750 [PubMed: 11368527]
- Dahan D, Magnuson JS, Tanenhaus MK, & Hogan EM (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5–6), 507–534. 10.1080/01690960143000074
- Dahan D, & Tanenhaus MK (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 30(2), 498–513. 10.1037/0278-7393.30.2.498 [PubMed: 14979820]

- Daniloff R, & Moll K (1968). Coarticulation of Lip Rounding. *Journal of Speech and Hearing Research*, 11(4), 707–721. 10.1044/jshr.1104.707 [PubMed: 5719225]
- Dilley LC, & Pitt MA (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. 10.1177/0956797610384743 [PubMed: 20876883]
- Farris-Trimble A, & McMurray B (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, 56(4), 1328–1345. 10.1044/1092-4388(2012/12-0145)
- Farris-Trimble A, McMurray B, Cigrand N, & Tomblin JB (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. 10.1037/a0034353 [PubMed: 24041330]
- Fishman KE, Shannon RV, & Slattery WH (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language, and Hearing Research*, 40(5), 1201–1215. 10.1044/jslhr.4005.1201
- Friesen LM, Shannon RV, Baskent D, & Wang X (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, 110(2), 1150–1163. 10.1121/1.1381538 [PubMed: 11519582]
- Geller J, Holmes A, Schwalje A, Berger J, Gander P, Choi I, & McMurray B (n.d.). Validating the Iowa Test of Consonant Perception. *Journal of the Acoustical Society of America*
- Gomaa NA, Rubinstein JT, Lowder MW, Tyler RS, & Gantz BJ (2003). Residual Speech Perception and Cochlear Implant Performance in Postlingually Deafened Adults. *Ear and Hearing*, 24(6), 539–544. 10.1097/01.AUD.0000100208.26628.2D [PubMed: 14663353]
- Gregg J, Inhoff AW, & Connine CM (2019). Re-reconsidering the role of temporal order in spoken word recognition. *Quarterly Journal of Experimental Psychology*, 72(11), 2574–2583. 10.1177/1747021819849512
- Hannagan T, Magnuson JS, & Grainger J (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, 4(SEP), 563. 10.3389/fpsyg.2013.00563 [PubMed: 24058349]
- Hendrickson K, Spinelli J, & Walker E (2020). Cognitive processes underlying spoken word recognition during soft speech. *Cognition*, 198(January), 104196. 10.1016/j.cognition.2020.104196 [PubMed: 32004934]
- Henry BA, Turner CW, & Behrens A (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America*, 118(2), 1111–1121. 10.1121/1.1944567 [PubMed: 16158665]
- Hochberg I, Boothroyd A, Weiss M, & Hellman S (1992). Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear and Hearing*, 13(4), 263–271. 10.1097/00003446-199208000-00008 [PubMed: 1397769]
- Just MA, & Carpenter PA (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. 10.1037/0033-295X.99.1.122 [PubMed: 1546114]
- Kamide Y, Altmann GTM, & Haywood SL (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. 10.1016/S0749-596X(03)00023-8
- Kapnola EC, & 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. 10.1037/xge0000123 [PubMed: 26709587]
- Liberman A, & Cooper F (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461. <http://psycnet.apa.org/journals/rev/74/6/431/> [PubMed: 4170865]
- Litvak LM, Spahr AJ, Saoji AA, & Fridman GY (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *The Journal of the Acoustical Society of America*, 122(2), 982–991. 10.1121/1.2749413 [PubMed: 17672646]
- Luce PA, & Pisoni D (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36. http://journals.lww.com/ear-hearing/Abstract/1998/02000/Recognizing_Spoken_Words__The_Neighborhood.1.aspx [PubMed: 9504270]

- Marslen-Wilson W (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. <http://www.ncbi.nlm.nih.gov/pubmed/18480284> [PubMed: 3581730]
- McClelland JL, & Elman JL (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, 18, 1–86. [PubMed: 3753912]
- McMurray B (2017). Nonlinear Curvefitting for Psycholinguistics (Version 24) <https://osf.io/4atgv/>
- McMurray B, Apfelbaum KS, & Tomblin JB (n.d.). The slow development of real-time processing: Spoken Word Recognition as a crucible for new thinking about language acquisition and disorders. *Current Directions in Psychological Science* <https://psyarxiv.com/uebfc/>
- McMurray B, Ellis TP, & Apfelbaum KS (2019). How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear and Hearing*, 40(4), 961–980. 10.1097/AUD.0000000000000681 [PubMed: 30531260]
- 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(August), 147–164. 10.1016/j.cognition.2017.08.013 [PubMed: 28917133]
- McMurray B, Farris-Trimble A, Seedorff M, & Rigler H (2016). The Effect of Residual Acoustic Hearing and Adaptation to Uncertainty on Speech Perception in Cochlear Implant Users. *Ear and Hearing*, 37(1), e37–e51. 10.1097/AUD.0000000000000207 [PubMed: 26317298]
- McMurray B, & Jongman A (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219–246. 10.1037/a0022325 [PubMed: 21417542]
- McMurray B, Samelson VM, Lee SH, & Tomblin JB (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60(1), 1–39. 10.1016/j.cogpsych.2009.06.003 [PubMed: 19836014]
- McMurray B, Tanenhaus MK, & Aslin RN (2009). Within-category VOT affects recovery from “lexical” garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1), 65–91. 10.1016/j.jml.2008.07.002 [PubMed: 20046217]
- McQueen J, & Huettig F (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, 131(1), 509–517. 10.1121/1.3664087 [PubMed: 22280612]
- Miller J, Green K, & Reeves A (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43, 106–115. <http://content.karger.com/ProdukteDB/produkte.asp?Doi=261764>
- Mueller HG (2001). Speech audiometry and hearing aid fittings: Going steady or casual acquaintances? *Hearing Journal*, 54(10). 10.1097/01.hj.0000294535.51460.0c
- Newman RS, Clouse SA, & Burnham JL (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 109(3), 1181–1196. 10.1121/1.1348009 [PubMed: 11303932]
- Norris D (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. 10.1016/0010-0277(94)90043-4
- Novick JM, Hussey E, Teubner-Rhodes S, Harbison JJ, & Bunting MF (2014). Clearing the garden-path: improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186–217. 10.1080/01690965.2012.758297
- Novick JM, Trueswell JC, & Thompson-Schill SL (2005). Cognitive control and parsing: Reexamining the role of Broca’s area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263–281. 10.3758/CABN.5.3.263
- Sarrett ME, McMurray B, & Kapnoula EC (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain and Language*, 211(December), 104875. 10.1016/j.bandl.2020.104875 [PubMed: 33086178]
- Spitzer S, Liss J, Spahr T, Dorman MF, & Lansford K (2009). The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, 125(6), EL236–EL241. 10.1121/1.3129304 [PubMed: 19507928]

- Tanenhaus M, Spivey-Knowlton M, Eberhard K, & Sedivy J (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. 10.1126/science.7777863 [PubMed: 7777863]
- Taylor B (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *Hearing Journal*, 56(1), 40–46. 10.1097/01.HJ.0000293000.76300.ff
- Toscano JC, Anderson ND, & McMurray B (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review* 10.3758/s13423-013-0417-0
- Ulrich R, & Miller J (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38(5), 816–827. 10.1017/S0048577201000610 [PubMed: 11577905]
- Weber A, & Scharenborg O (2012). Models of spoken-word recognition. *WIREs Cogn Sci*, 3, 387–401. 10.1002/wcs.1178
- Zhang X, & Samuel AG (2018). Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. *Journal of Memory and Language*, 100(December 2015), 32–50. 10.1016/j.jml.2018.01.002
- Zwitserslood P (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64. [PubMed: 2752705]

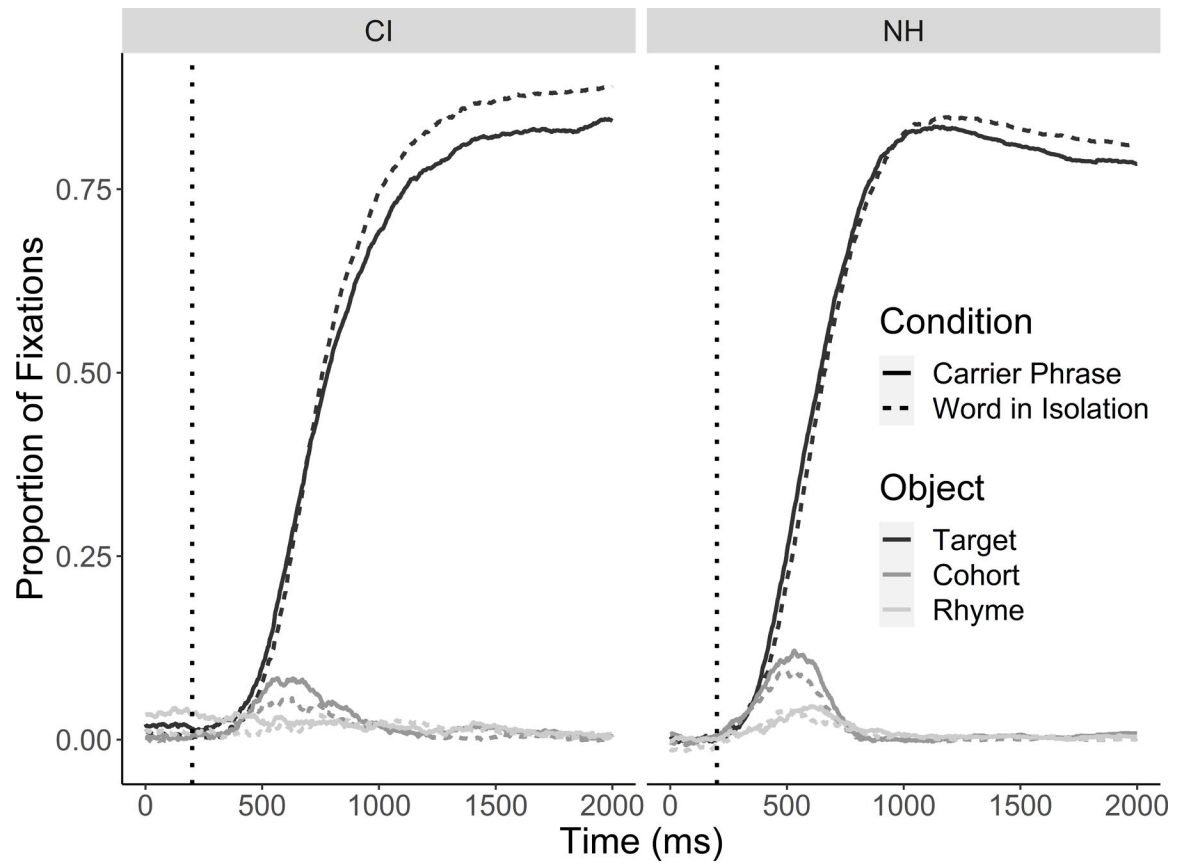


Figure 1. Looks to the target and competitors after subtracting looks to the unrelated at each time point for both CI users and NH listeners.

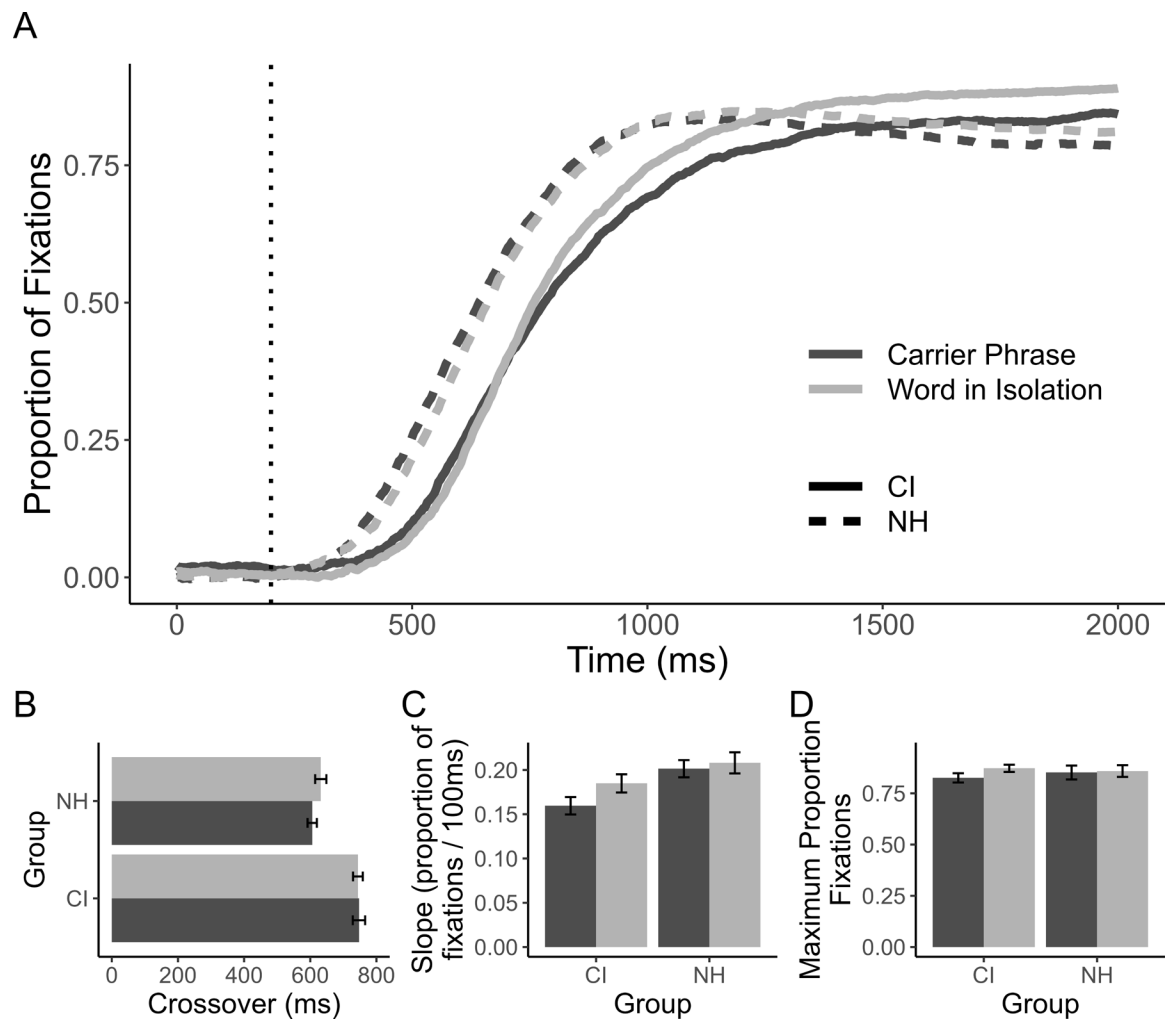


Figure 2.

A) Target fixations as a function of time after subtracting looks to the unrelated item for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters; error bars indicate SEM. B) the crossover; C) the slope; D) peak looks (at asymptote).

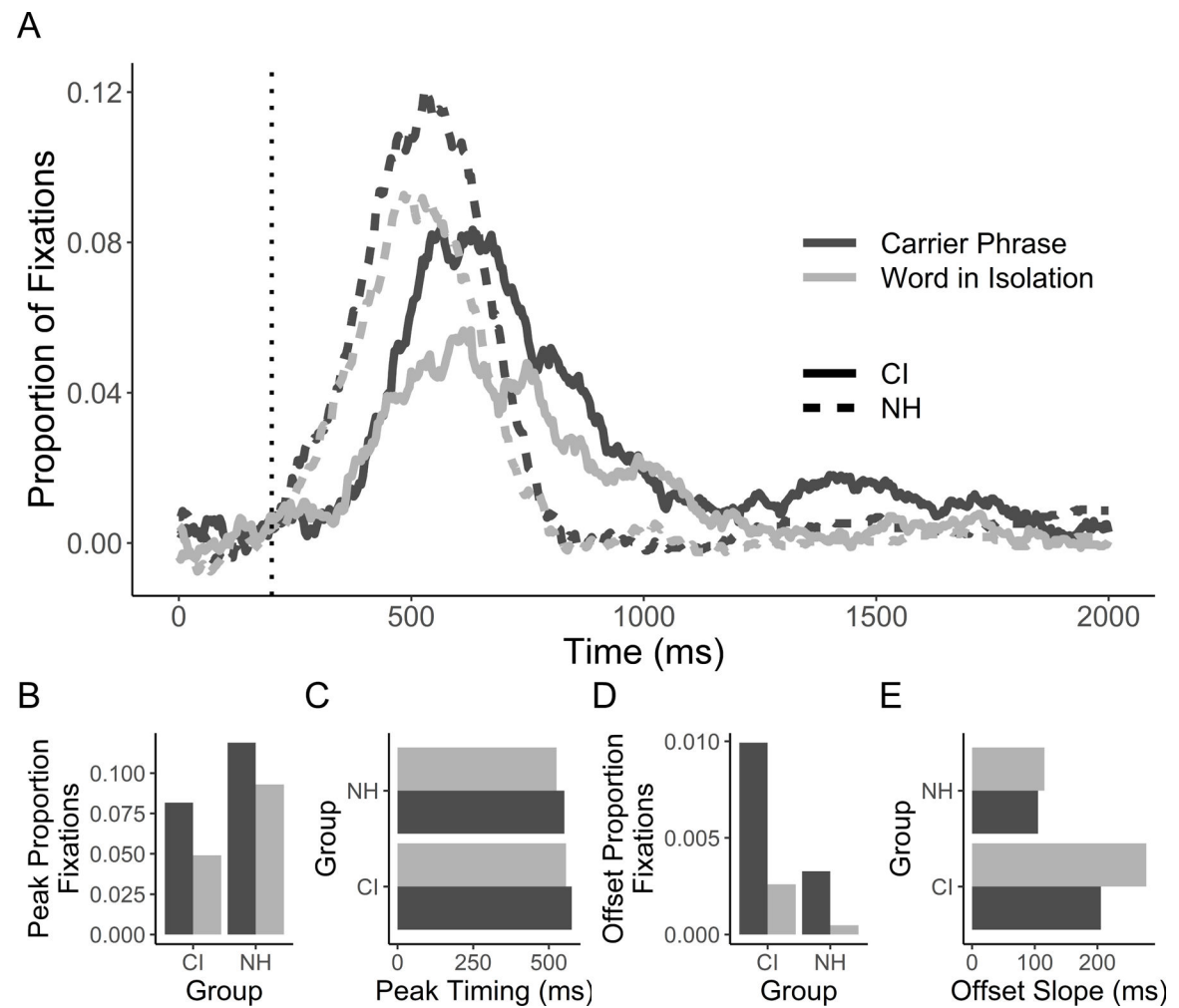


Figure 3.

A) Cohort fixations as a function of time after subtracting looks to the unrelated, for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters. B) Peak height; C) peak timing; D) offset baseline; E) and offset slope. Note that in B-E, no error bars are included because data were jackknifed prior to curvefitting and there is no clear way to estimate standard error for jackknifed data.

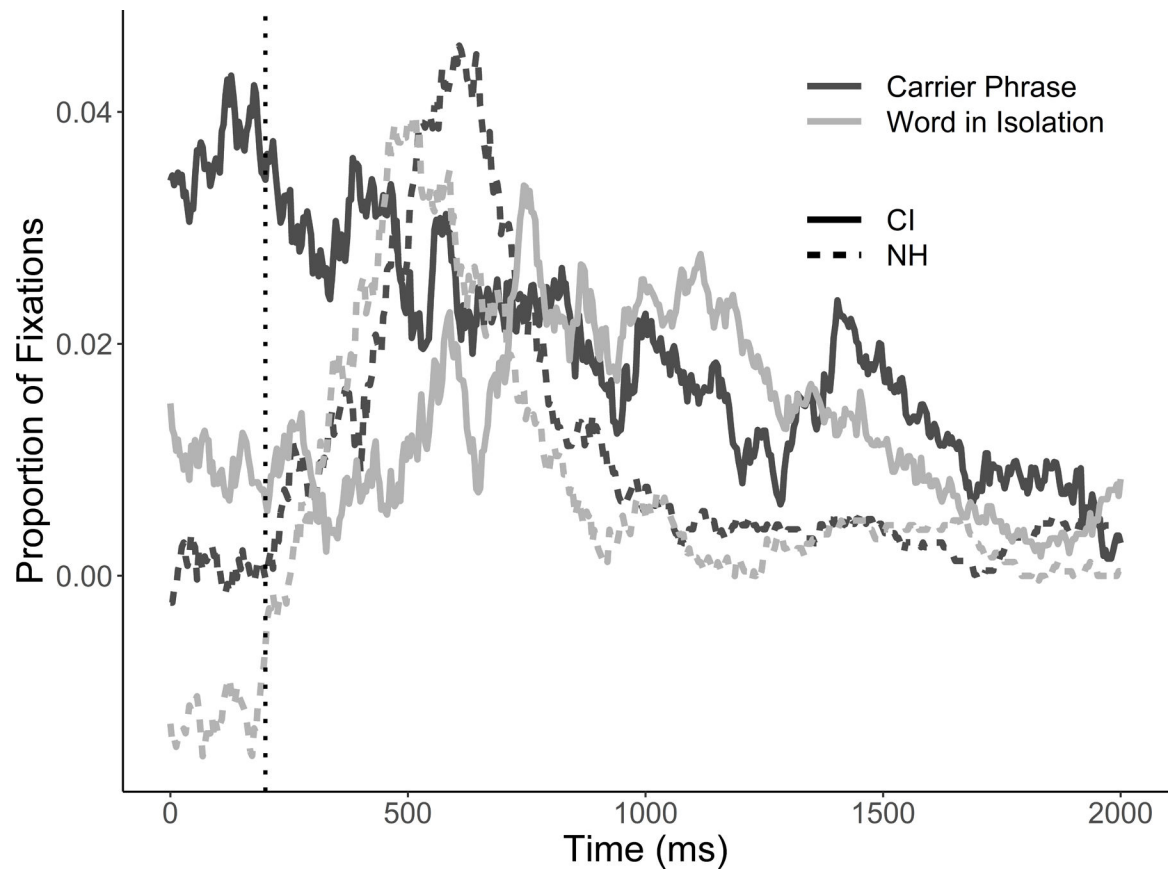


Figure 4. Looks to the rhyme after subtracting looks to the unrelated item at each time point for both listener groups and sentence conditions.

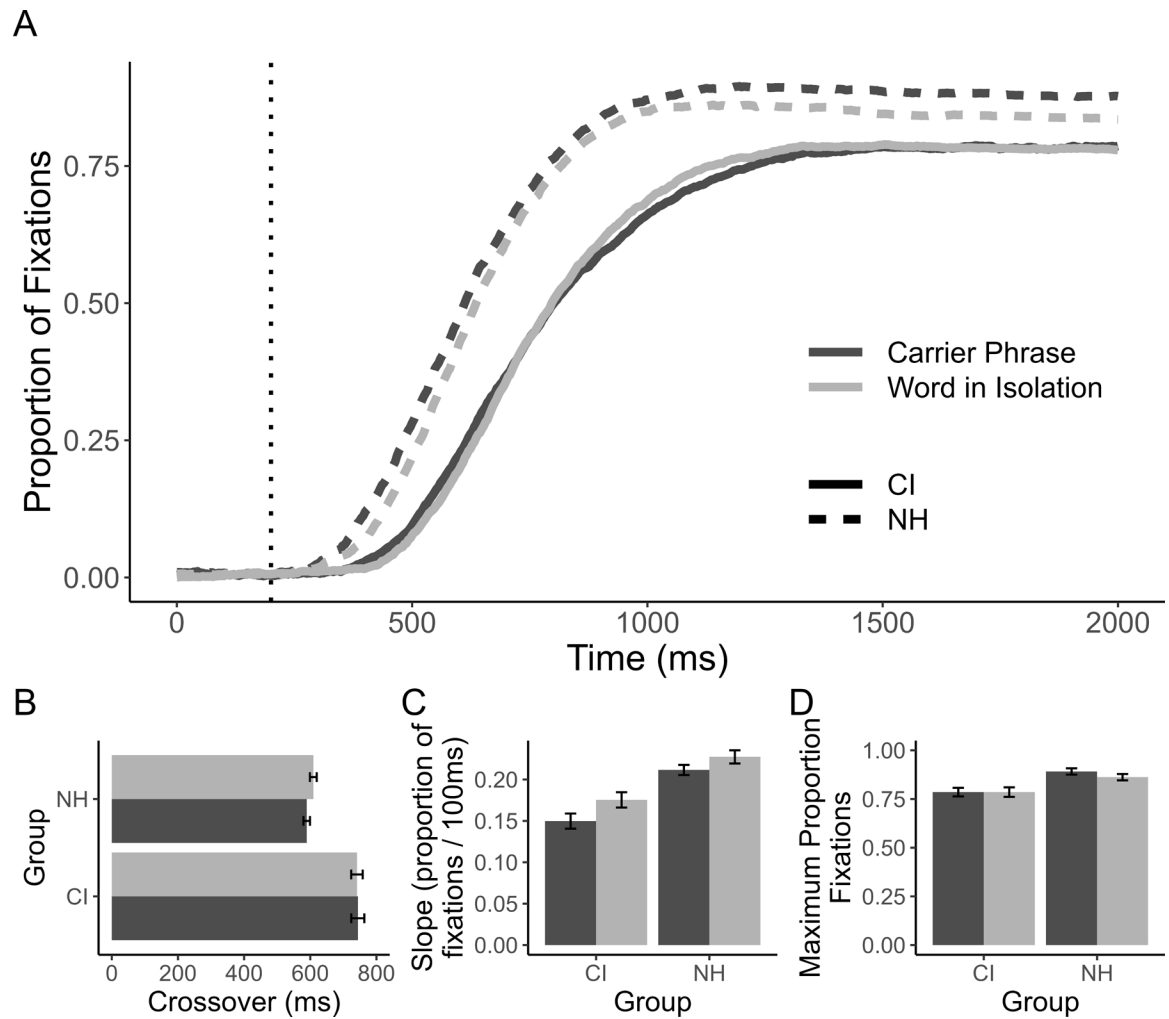


Figure 5.

A) Target fixations as a function of time after subtracting looks to the unrelated item for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters; error bars indicate SEM. B) the crossover; C) the slope; D) peak looks (at asymptote).

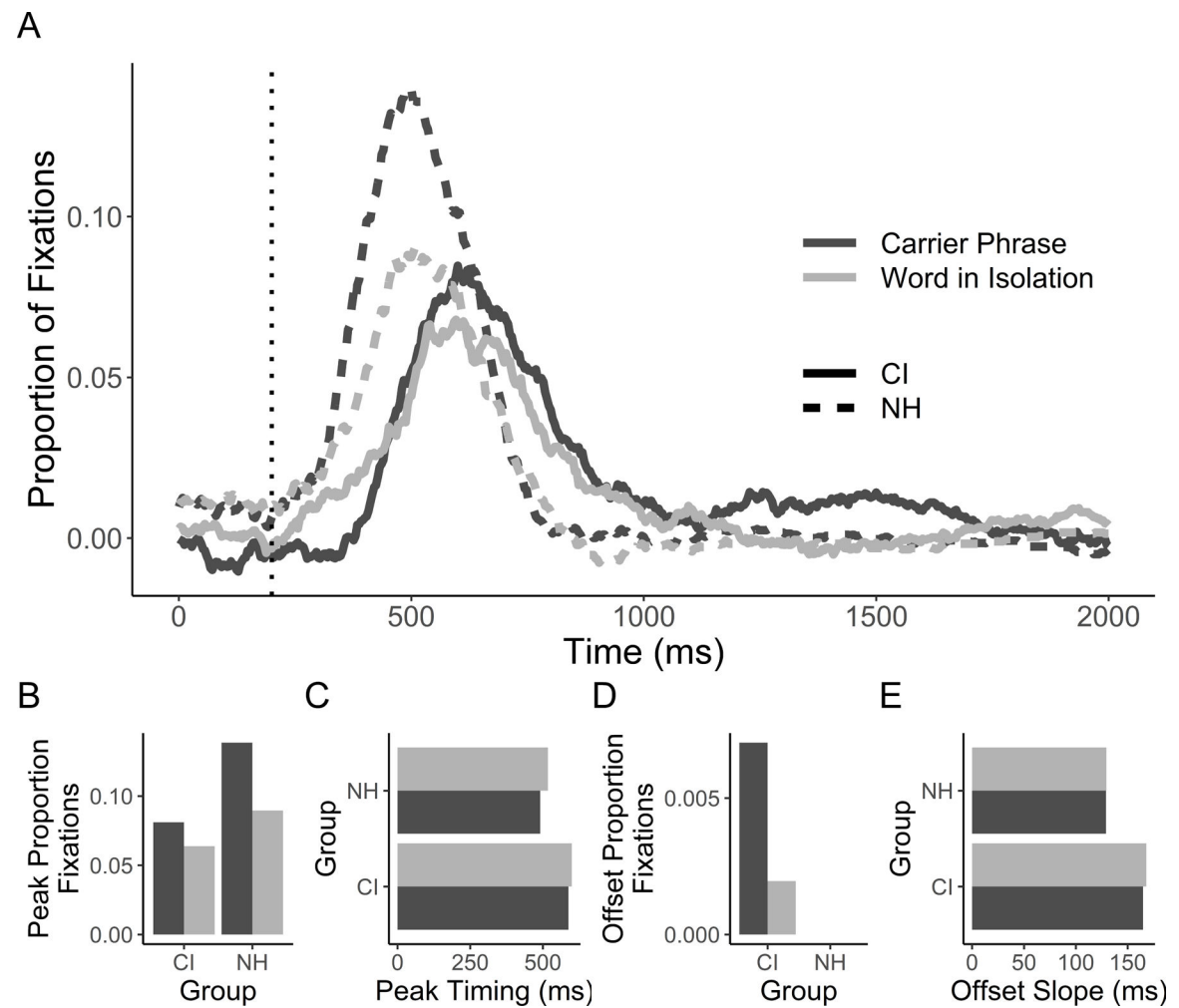


Figure 6.

A) Cohort fixations as a function of time after subtracting looks to the unrelated, for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters. B) Peak height; C) peak timing; D) offset baseline; E) and offset slope. Note that in B-E, no error bars are included because data were jackknifed prior to curvefitting and there is no clear way to estimate standard error for jackknifed data.