SUPPLEMENTAL DIGITAL CONTENT

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

Running Head: LEXICAL ACCESS IN CONTINOUS SPEECH

Francis X. Smith
And
Bob McMurray

Dept. of Psychological and Brain Sciences and
Dept. of Otolaryngology
University of Iowa

Corresponding Author
Francis X. Smith
Dept. of Psychological and Brain Sciences
255 PBS
University of Iowa
Iowa City, IA 52245
Francis-smith@uiowa.edu
319-335-0692

S1. Replication Study

The primary experiment controlled the timing of the target words by adding an amount of silence identical to the length of a randomly chosen carrier phrase to the onset of each word in isolation. This ensures that in both context conditions, listeners have similar time to scan the scene and leads to similar amounts of baseline looking. However, it also creates a predictability confound: in carrier phrases the target word is fairly predictable (if the listener is actually processing the phrases) while in the word-in-isolation condition it is not. We thus conducted a replication in which the word-in-isolation condition had no delay between the onset of the trial and the stimulus – making the word perfectly predictable in both cases (though creating issues in the baseline fixations).

Materials and Methods

Participants. Twenty-two experienced CI users (with more than 1-year of device use; 11 women) were recruited for this study from the same patient registry as in the primary experiment. Eligibility criteria were the same as in the primary experiment. No participant was allowed to participate in both. Listeners had a mean age of 63.7 years (SD = 6.2). Three participants were unilateral CI users, four were bilateral CI users, four 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 details on the participants.

NH controls were recruited through community advertisement. This resulted in 18 NH controls (13 women). Their average age was slightly lower than that of the CI participants (M = 54.0, SD = 15.3). 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. The design of the replication experiment was similar to the primary experiment with several exceptions. First, we removed the phrase-matched delay in the word-in-isolation condition to enhance the predictability of word onset in this condition. Second, this reduced time spent during each trial in the word-in-isolation condition allowing for more trials within an experimental session. As a result, each auditory stimulus was presented 4 times / context condition (as opposed to the 3 / condition in the main experiment). This resulted in 640 total trials across 8 blocks.

Stimuli. The stimuli were the same as those described in the primary experiment (see Supplement S3 for a complete list).

Procedure. The procedures were the same the primary experiment.

Eye-movement recording and analysis. Eye-movements were recorded, processed and analyzing using a similar technique to the primary experiment. Critically, fixations were baselined by subtracting the unrelated fixations from the target/cohort and rhyme fixations prior to analysis (see Supplement S4 for visualizations of this data prior to baselining).

Results

Accuracy. Both groups were generally accurate in ultimately recognizing the word (selecting the correct picture), though CI users were less accurate (*M* = 93.56% correct, *SD* = 4.91%) than NH listeners (*M* = 99.67% correct, *SD* = 0.50%). Accuracy was also similar between context conditions, though there were more errors in the carrier phrase condition (*M* =

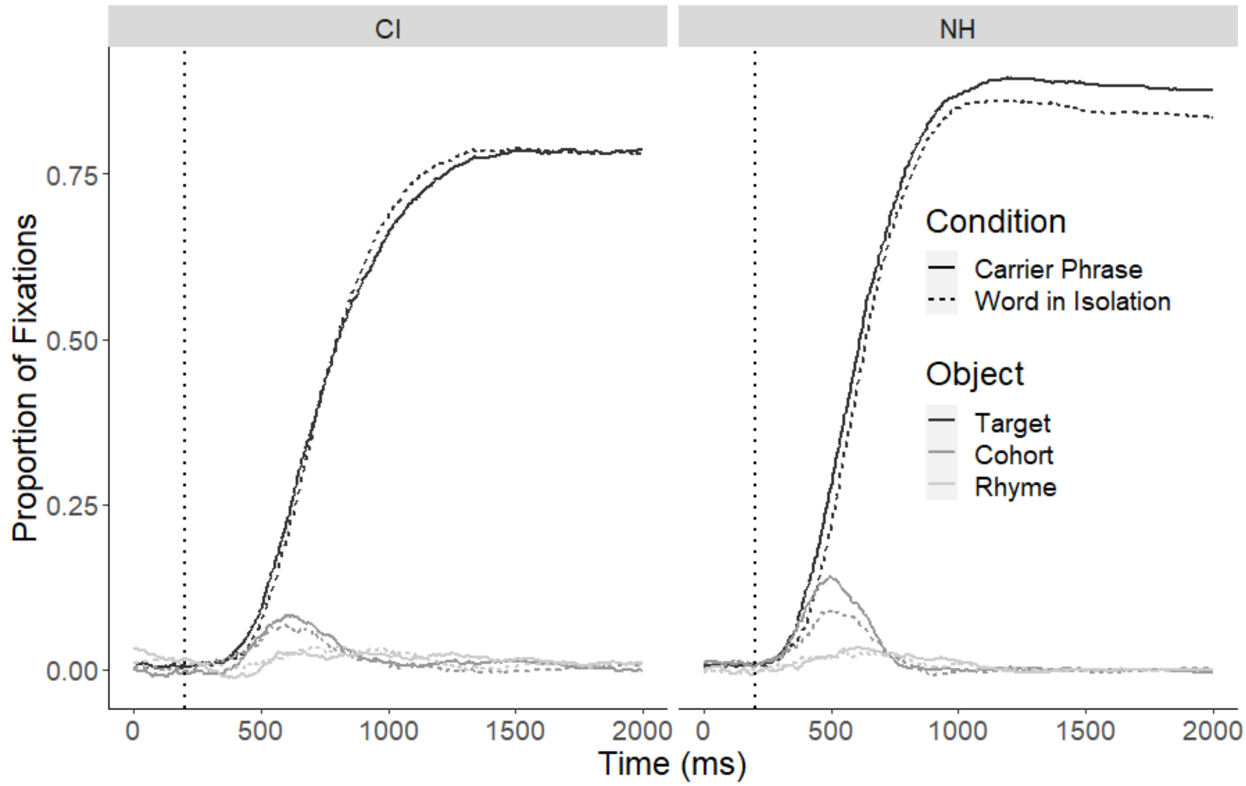

Figure S1.1. Looks to the target and competitors after subtracting looks to the unrelated at each time point for both CI users and NH listeners. The vertical dashed lines represent the 200ms oculomotor delay.

96.33% correct, *SD* = 3.17%) than with words in isolation (*M* = 96.98% correct, *SD* = 2.24%). Both of these differences were significant in a 2×2 ANOVA on accuracy as a function of listener group and context condition. There was a significant main effect of listener group, $F(1,38) = 29.42$, $p < .001$, and a significant effect of context, $F(1,38) = 5.39$, $p = .026$. There was no interaction, $F(1,38) = 2.49$, $p = .123$.

Fixations. Figure S1.1 shows looks to the various items on the screen as a function of time after target onset after baselining (Supplement S4 for visualizations without baselining). Early on, both groups looked more at the cohort than unrelated items (cohort fixations were above 0), suggesting they exhibited incremental processing. NH listeners showed slightly earlier looks to both the target and cohort than CI users and more cohort fixations. CI users also showed slightly lower overall looks to the target, and competitor fixations remained above baseline longer than NH listeners. Both of these patterns are consistent with Farris-Trimble et al. (2014). Further, both groups look slightly earlier to target words when words are presented in carrier phrases rather than in isolation (as was observed in the primary experiment). To analyze these results, we examined each candidate item type individually (Target, Cohort, Rhyme) on the appropriate trial types using the same process as the main experiment.

Target fixations. Fixations to the target as a function of time, listener group and context condition are shown in Figure S1.2. Logistic fits were good across both groups and both conditions, with an average *R* of .998 across all four cells of the design. While CI users (M = .*998*) showed a significantly poorer fit than NH listeners (*M* = .999) (see Supplement S5),

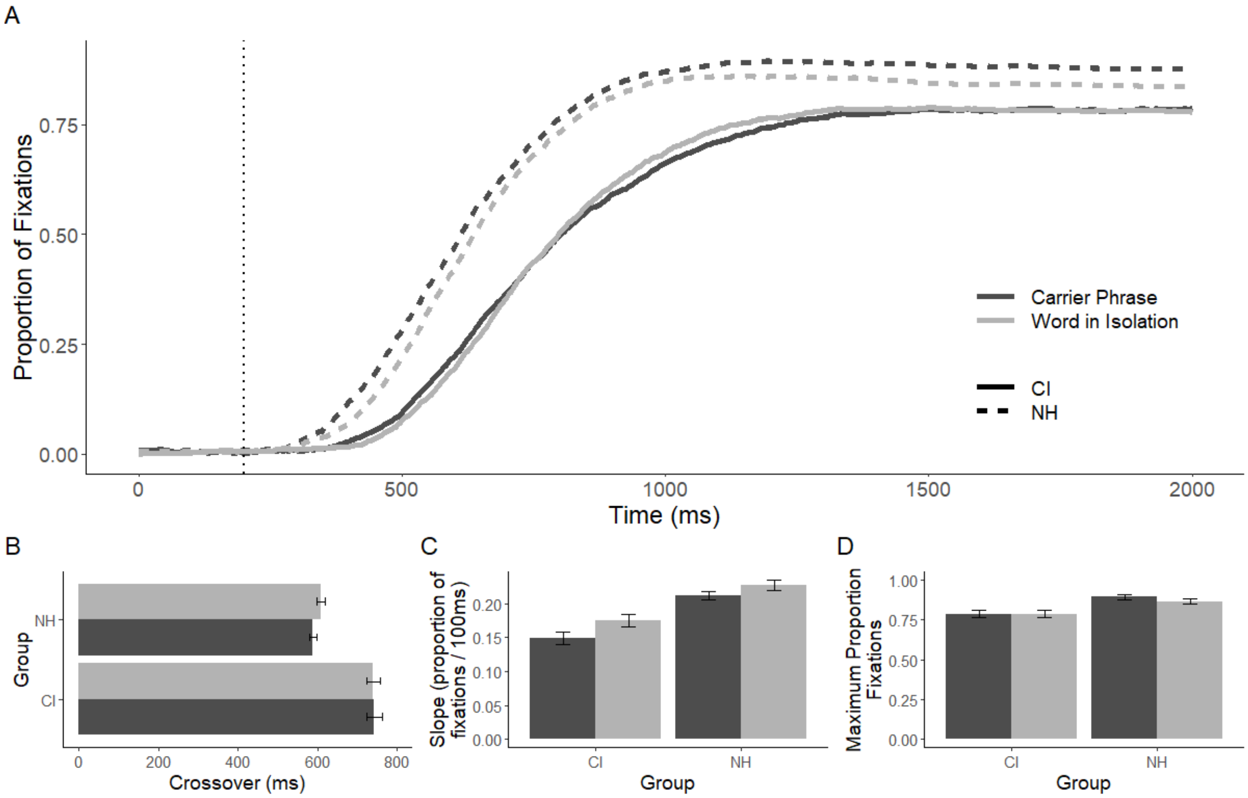

Figure S1.2. A) Target fixations as a function of time after subtracting looks to the unrelated item for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters; error bars indicate SEM. B) the crossover (in milliseconds); C) the slope (change in proportion looks per 100ms); D) maximum looks (at asymptote).

differences were numerically negligible. Each of the four parameters for each subject was then compared across condition and group with 2×2 mixed ANOVAs assessing the effect of context (carrier phrase vs. word-in-isolation) and group (CI vs. NH).

The ***baseline*** parameter showed no significant differences between group or context and no significant interaction, all $p > .20$. This suggests that the subtraction of looks to unrelated items appropriately corrected looks to the target, as it does not cause differences in the baseline before the onset of the target word.

Crossover (Fig. S1.2B) showed a significant effect of listener group, $F(1,38) = 45.89$, $p < .001$, as CI users had significantly later crossovers than NH listeners (by on average 143 ms). There was no significant effect of context, $F(1,38) = 1.22$, $p = .276$, and no significant interaction, $F(1,38) = 2.16$, $p = .150$.

The ***slope*** of target fixations patterned similarly to the primary experiment (Figure S1.2C). We found a significant main effect of listener group, $F(1,38) = 26.32$, $p < .001$, and context, $F(1,38) = 19.29$, $p < .001$, with no significant interaction, $F(1,38) = 1.09$, $p = .302$. NH listeners showed steeper slopes than CI users, and both groups showed steeper slopes when words were in isolation than in carrier phrases.

For analysis, ***maximum looks*** to the target (Figure S1.2D) were first scaled with the empirical logit. There was a significant effect of listener group, $F(1,38) = 10.44$, $p = .003$. This was due to the fact that CI users showed reduced maximum looks relative to NH listeners. There

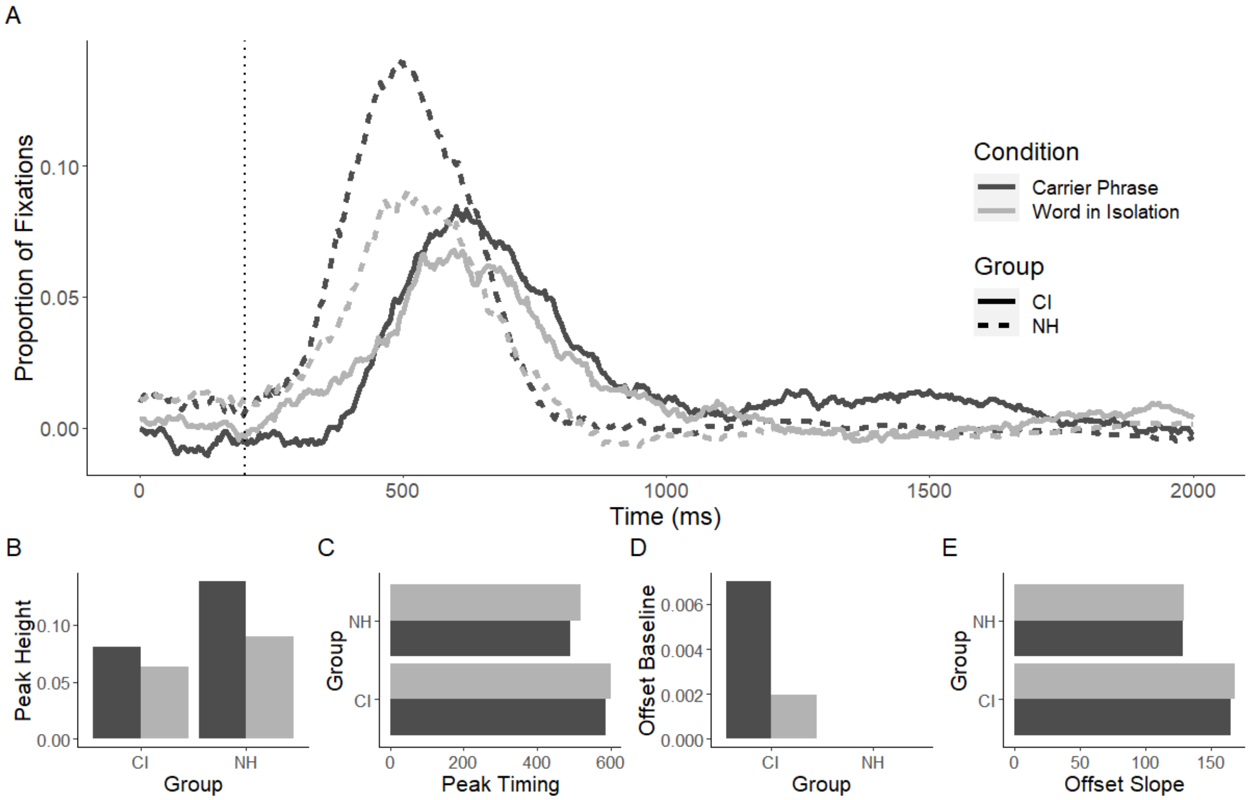

Figure S1.3. A) Cohort fixations as a function of time after subtracting looks to the unrelated, for each listener group and sentence context condition (the dotted vertical line represents the 200ms oculomotor delay). Panels below indicate individual curvefit parameters. B) Peak height; C) peak timing (ms); 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.

was also a main effect of context, $F(1,38) = 4.21$, $p = .047$ and a significant interaction, $F(1,38) = 6.27$, $p = .017$. Paired-sample t-tests revealed that there was no difference in maximum looks between contexts for CI users, $t(21) = 0.43$, $p = .669$, while NH listeners showed higher maximum looks when words were embedded in carrier phrases, $t(17) = 2.52$, $p = .022$.

Cohort fixations. The timecourse of cohort fixations as a function of listener group and context is shown in Figure S1.3. As in the primary experiment, Cohort fixations were jacknifed prior to curvefitting, and F statistics were adjusted to account for jackknifing. Fits were good across conditions and for both CI users and NH listeners, mean $R = .989$ with no significant differences (Supplement S5). Each of the six parameters was then examined in a 2×2 ANOVA assessing the effect of context (sentence-context vs. word-in-isolation) and group (CI vs. NH).

Onset baseline showed no significant differences between group or context and there was no significant interaction, all $p > .10$. As with the target baseline, this suggests that subtracting unrelated items serves as a useful baselining method. The ***onset slope*** parameter similarly showed no significant differences between group or context and no significant interaction between the two, all $p > .20$.

Peak height (Figure S1.3B) showed significant differences between listener groups, with CI users showing reduced peak fixations compared to NH listeners, $F_{jk}(1,38) = 9.11$, $p = .005$. There was also a significant effect of context ($F_{jk}(1,38) = 15.29$, $p < .001$): when the target word was presented in a carrier phrase, peak looks increased. The listener group × context interaction was marginally significant, $F_{jk}(1,38) = 3.52$, $p = .068$, driven by the fact that the increase in competitor peak was numerically larger in NH listeners.

Peak timing (Figure S1.3C) similarly showed significant differences between groups, with CI users showing a delayed cohort peak relative to NH listeners (by an average of 90 ms), $F_{jk}(1,38) = 14.29$, $p < .001$. There was no significant effect of context, $F_{jk}(1,38) = 1.03$, $p = .316$. There was no significant group × context interaction, $F_{jk}(1,38) = 0.16$, $p = .693$.

Offset baseline (Figure S1.3D) showed a marginally significant effect of listener group, $F_{jk}(1,38) = 4.09$, $p = .050$. CI users had numerically higher baselines (M= .0045) than NH listeners (M=.0000003). There was no effect of context, $F_{jk}(1,38) = 1.50$, $p = .228$, and no interaction, $F_{jk}(1,38) = 1.50$, $p = .228$. The ***offset slope*** (Figure S1.3E) showed no significant effect of listener group or context and there was also no significant interaction, all $p > .10$.

Rhyme Fixations. As in the primary experiment, CI users fixated the rhyme above baseline even before target word onset (Figure S1.4). Because of this we could not adequately fit curves to the rhyme data for CI users and thus could not rely on the statistical approach we used for cohort fixations. As in the primary experiment, we report coarser analyses of rhyme fixations, to draw tentative conclusions (Supplement S6). They find, as in the main experiment, that rhymes were fixated for a greater extent of the trial by CI users than by NH listeners, and they showed higher peak fixations, though the timing of rhyme activations does not differ by group.

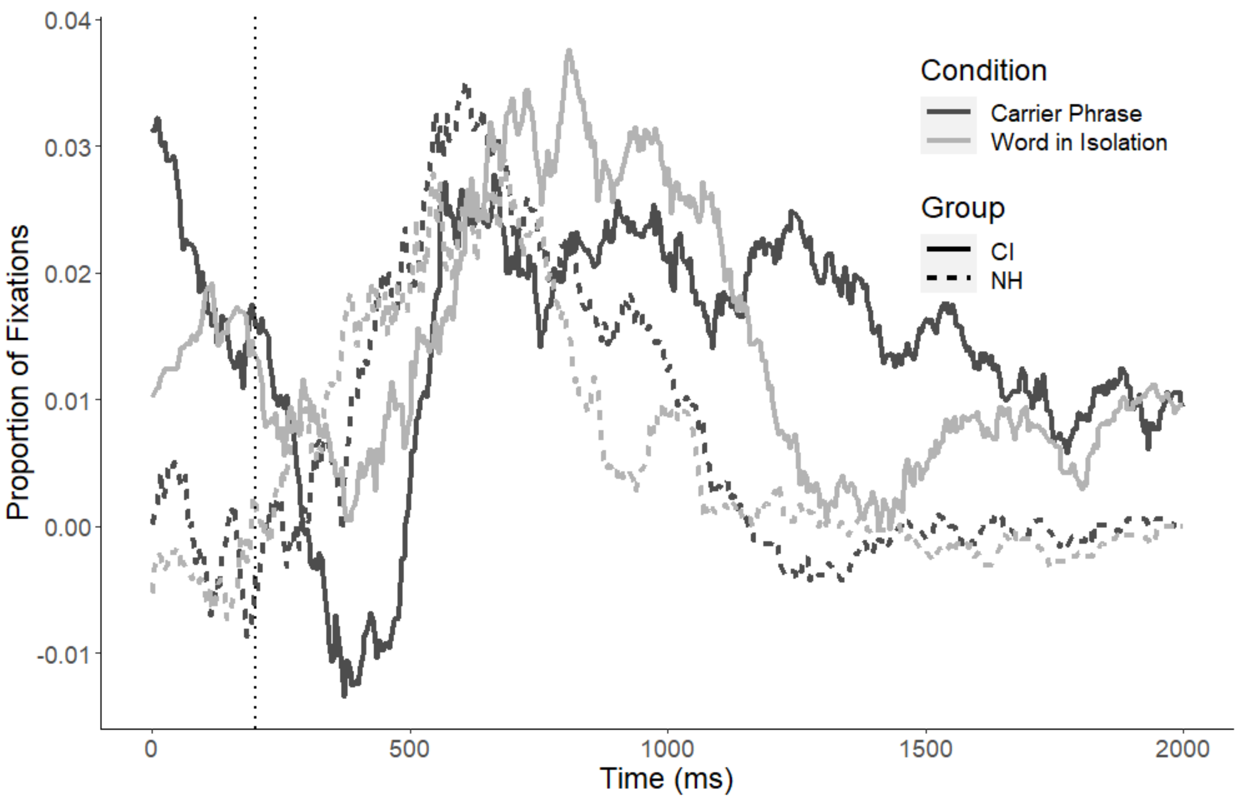

Figure S1.4. Looks to the rhyme after subtracting looks to the unrelated item at each time point for both listener groups and sentence conditions (the dotted vertical line represents the 200ms oculomotor delay).

Discussion

With respect to the effect of listener group, CI users showed delayed lexical activation relative to NH listeners. CI users were about 143 ms slower to fixate the target than NH listeners (target crossover, Figure S1.2B). Peak looks to cohort competitors were also about 90 ms later than NH listeners (Figure S1.3C). These delays were also observed in carrier phrases, extending prior work which focused on words in isolation (Farris-Trimble et al., 2014).

The delays observed in the present study however were larger than previously reported among post-lingually deafened CI users (Farris-Trimble et al., 2014). We also saw a *decrease* in peak cohort looking in CI users (Figure S1.3B) which was not observed earlier. The larger delays in lexical activation combined with the decrease in peak target looks and cohort competitor activation levels suggest that CI users in this experiment were exhibiting something closer to the "wait-and-see" pattern of lexical access, which is not predicted for this population by prior work (Farris-Trimble et al., 2014). CI users also showed a slight increase in cohort competitor activation at offset of the trial which, while only marginally significant, suggests they may have been preserving some competitor activation to correct for misperceptions.

The effects of carrier phrases on the speed of lexical access were quite similar across the listener groups. Both CI users and NH listeners showed no difference in the time of their crossover point between words presented in carrier phrases compared to words presented in isolation. The timing of peak looks to the cohort were similarly unaffected by context. Carrier phrases seem to present no difficulty for committing to the target for either group. However, both groups showed a significant increase in target slope when words were presented in isolation. This suggests that lexical access may initiate slightly earlier in continuous speech than in isolation. Conversely, even as the crossover is reached at similar times in both conditions, in isolation, lexical access may begin slightly later and then "ramp up" faster.

In general, CI users showed a similar response to carries phrases as NH listeners. However, an interaction was observed in the resolution of competition late in the trial. Maximum target looks (at asymptote) showed a larger difference between CI users and NH listeners in carrier phrases than with words in isolation. These larger differences suggest that CI users may alter how strongly they resolve competition in sentence contexts.

Cross-Experiment Analyses

We originally hypothesized that CI users would show sustained competitor activation, as had been observed in previous research with post-lingually deafened adults (Farris-Trimble et al., 2014). The pattern of results across the two experiments revealed some signs of this profile. In particular, maximum target fixations were reduced in CI users, particularly in carries phrases. However, cohorts showed mixed results with a significant effect on the offset slope (but not the asymptote) in the primary experiment, and a marginally significant effect on asymptote here.

One explanation may be that each experiment lacked the power to detect the effects of sustained competitor activation due to the jackknifing procedure used for cohorts. This procedure is quite conservative in its corrected F-statistic. Given this pattern, and the fact that the two experiments generally showed a similar pattern of results, we combined data from both to determine if there was evidence for sustained activation in analysis with greater power. We were hesitant to apply this approach to all of the parameter estimates, as the experiments differed in critical aspects of timing that may affect the earliest fixations. However, late measures (such as offset slope and offset baseline of cohorts) would be unlikely to be affected by these timing differences. Moreover, this procedure was not necessary for target looks as they were not jackknifed. Thus, we performed a combined analysis of the offset slope and offset baseline using a 2×2×2 mixed ANOVA using listener group, context, and experiment as factors.

Offset slope showed a significant effect of listener group, with CI users taking longer to suppress cohorts than NH listeners, $F_{jk}(1,78) = 9.77$, $p = .003$. No other main effect was significant, all $p > .3$, and there were no significant interactions. The ***offset baseline*** similarly showed a significant effect of listener group, as CI users looked to cohorts more at the end of the trial than NH listeners, $F_{jk}(1,78) = 7.31$, $p = .010$. There was also a significant effect of context, as carrier phrases led to higher looks to cohorts at the end of the trial than words in isolation, $F_{jk}(1,78) = 4.12$, $p = .049$. Experiment did not have a significant effect and there were no significant interactions, all $p > .2$.

As this analysis was post-hoc, we should treat results with caution, particularly main effect of context on the offset baseline which was not observed in the analyses of each experiment. However, the pattern of results with respect to listener group are consistent with the marginal effects seen in each experiment, and they fit the profile of sustained competitor activation. This suggests that the individual experiments simply lacked the power to detect the effect. CI users do indeed seem to keep competitors active, and this may allow for easier corrections in response to misperceptions.

S2. Participants

Table S2.1. *Description of CI users in the primary experiment.*

Missing values in the Onset of Deafness column indicate a gradual hearing loss. AZ-Bio and CNC word scores are taken from the hearing configuration used by the patient on the day of the experiment.

Configuration	Gender	Age (yrs)	Age at Implantation	Onset of Deafness	Ear	Manufacturer	Implant Model	Etiology	AZBio%	CNC Word%
Bilateral	Male	69	60	44	B	Cochlear	Nucleus CI 512 (CA)	Unknown	92	93
Bilateral	Female	60	55	45	B	AB	Clarion HiRes 90k	Hereditary	97	88
Bilateral	Female	63	52	47	B	Cochlear	L - Nucleus CI 24 RE R – Nucleus CI 512 (CA)	Unknown	79	86
Bimodal	Female	54	48	47	R	Cochlear	Nucleus CI 422	Unknown	92	99
Bimodal	Female	54	51	30	R	AB	Clarion HiRes 90k	Unknown	75	42
Bimodal	Male	44	43	42	R	AB	Clarion HiRes Ultra Slim J	Unknown	62	30
Hybrid	Female	58	52	24	R	Cochlear	Nucleus Hybrid L24	Unknown	84	90
Hybrid	Female	41	36	---	R	Cochlear	Nucleus Hybrid L24	Unknown	82	91
Hybrid	Male	64	59	50	L	Cochlear	Nucleus Hybrid S12	Family History	73	60
Hybrid	Male	58	54	50	L	Cochlear	Nucleus Hybrid L24	Unknown	26	30
Hybrid	Male	62	57	45	L	Cochlear	Nucleus Hybrid L24	Unknown	71	86
Hybrid	Female	61	57	40	R	Cochlear	Nucleus Hybrid L24	Unknown	88	74
Hybrid	Male	55	52	35	R	Cochlear	Nucleus Hybrid L24	Unknown	92	87
Hybrid	Male	70	67	30	R	Cochlear	Nucleus Hybrid L24	Unknown	85	86
Hybrid	Male	71	67	60	R	Cochlear	Nucleus CI 422 + HA	Unknown	51	74
Hybrid	Female	61	60	45	L	Cochlear	Nucleus S12	Family History	92	91
Hybrid	Male	70	68	64	R	Cochlear	Nucleus Hybrid L24	Unknown	38	86
Unilateral	Female	58	29	22	L	Cochlear	Nucleus CI 24M	Hereditary	83	63
Unilateral	Male	60	56	55	L	Cochlear	Nucleus CI 422	Unknown	95	96
Unilateral	Female	62	36	36	L	AB	Clarion Radial Bipolar/Standard 1.0	Otosclerosis	90	79
Unilateral	Male	53	31	27	R	Cochlear	Nucleus CI 24M	Unknown	82	57

Table S2.2. *Description of CI users in the supplemental experiment.*

AZ-Bio and CNC word scores are taken from the hearing configuration used by the patient on the day of the experiment.

Configuration	**Gender**	**Age (yrs)**	**Age at Implantation**	**Onset of Deafness**	**Ear**	**Manufacturer**	**Implant Model**	**Etiology**	**AZBio%**	**CNC Word%**
Bilateral	Female	42	25	19	B	AB	Clarion HiFocus II-CII	Hereditary	43	78
Bilateral	Female	57	40	31	B	AB	Clarion HiFocus II-CII	Unknown	90	83
Bilateral	Female	59	50	25	B	Cochlear	L – Nucleus CI 512 R – Nucleus CI 24 RE	Unknown	92	92
Bilateral	Female	59	52	51	B	Cochlear	L – Nucleus CI 512 R – Nucleus CI 24 RE	Cochleitis	92	73
Bimodal	Male	65	46	45	R	Cochlear	Nucleus CI 24M	Unknown	91	83
Bimodal	Male	66	62	35	L	Med-El	Concerto Flex 28	Unknown	83	85
Bimodal	Male	67	64	33	L	Cochlear	Nucleus CI 532	Unknown	79	88
Bimodal	Male	69	67	66	R	Med-El	Synchrony Flex 24	Unknown	86	90
Hybrid	Male	56	54	40	L	Cochlear	Nucleus Hybrid L24	Unknown	86	96
Hybrid	Female	63	50	35	B	Cochlear	L – Nucleus Hybrid L24 R – Nucleus Hybrid S8	Autoimmune sensorinueral	65	89
Hybrid	Male	63	48	38	R	Cochlear	Nucleus EAS2	Hereditary	65	79
Hybrid	Female	65	53	50	B	Cochlear	L – Nucleus Hybrid S8 R – Nucleus Hybrid L24	Unknown	58	75
Hybrid	Female	65	62	29	L	Cochlear	Nucleus Hybrid L24	Unknown	75	66
Hybrid	Male	66	64	40	R	Cochlear	Nucleus Hybrid S12	Unknown	93	89
Hybrid	Male	67	65	50	R	Cochlear	Nucleus Hybrid L24	Unknown	53	71
Hybrid	Female	67	64	38	L	Cochlear	Nucleus Hybrid L24	Unknown	55	24
Hybrid	Female	67	55	43	R	Cochlear	Nucleus EAS3 + HA	Unknown	45	67
Hybrid	Female	70	67	65	R	Med-El	Concerto Flex 24	Hereditary	56	91
Hybrid	Male	71	60	42	R	Cochlear	Nucleus Hybrid S12	Unknown	34	72
Unilateral	Female	59	55	19	L	Cochlear	Nucleus CI 422	Unknown	21	19
Unilateral	Male	65	59	55	L	AB	Clarion HiRes 90k	Unknown	59	71
Unilateral	Male	65	58	50	R	AB	Clarion HiRes 90k	Unknown	66	56

S3. Stimuli

Table S3.1 *Words used in experiments.*

Target	Cohort	Rhyme	Unrelated
batter	battle	ladder	monkey
beak	beet	geek	wave
berry	barrel	fairy	jacket
carrot	carriage	parrot	beagle
chip	chin	whip	boat
coffee	coffin	toffee	badger
dollar	dolphin	collar	hammock
grill	grin	drill	foot
lap	lab	zap	chef
mail	maid	bale	jet
money	mother	honey	wagon
mustard	mustache	custard	penguin
paddle	package	saddle	waiter
pickle	picture	nickel	target
pin	pig	fin	mat
race	rake	face	soup
rain	rays	lane	mime
rope	rose	soap	mule
tower	towel	shower	magic
well	web	bell	ram

Table S3.1 *Carrier phrases used in experiments.*

Carrier Phrases	
here choose the	in this display, choose the image of the
here select the	in this display, find the image of the
here find the	in this display, select the picture of the
this time choose	on this screen, select the image of the
this time select	on this screen, choose the picture of the
this time find	on this screen, find the picture of the
now choose the	for this trial, choose the image of the
now select the	for this trial, find the image of the
now find the	for this trial, select the picture of the
click on the	to keep going, choose the picture of the
this time select ___ on this screen	in this display, select the image of the ___ to keep going
this time choose ___ on this screen	in this display, choose the picture of the ___ to proceed
this time select ___ in this display	in this display, find the picture of the ___ to continue
this time choose ___ in this display	on this screen, choose the image of the ___ to keep going
this time select ___ for this set	on this screen, find the image of the ___ to proceed
this time choose ___ for this set	on this screen, select the picture of the ___ to continue
now click the ___ on this screen	for this trial, select the image of the ___ to proceed
now find the ___ on this screen	for this trial, choose the picture of the ___ to continue
now find the ___ in this display	for this trial, find the picture of the ___ to keep going
now find the ___ for this set	to keep going, choose the image of the ___ on this screen

S4. Pre-Baseline Fixations

Figure S4.1 shows fixations as a function of time after onset of target word, averaged across all correct TCRU trials in the word in isolation condition in the primary experiment; Figure S4.2 shows the same for the replication reported in S1.

Prior to target onset, all items were fixated frequently but roughly equally, as there was no information to guide eye movements. After target onset, both groups fixated the target and cohort; rhymes became active later and to a lesser degree. Relative to NH listeners, CI users showed a slight delay in all fixations. They also had generally higher fixations before the onset of the target, but delayed all fixations after the onset of the target word (relative to NH listeners). The baselining procedure described in the main experiment helps to correct for differences in overall pre-target fixations between groups (and contexts). In both experiments, prior to 0 msec (the onset of the target word) cohort and target fixations were generally similar to the unrelated; however, the enhanced rhyme fixations for CI users are visible in both.

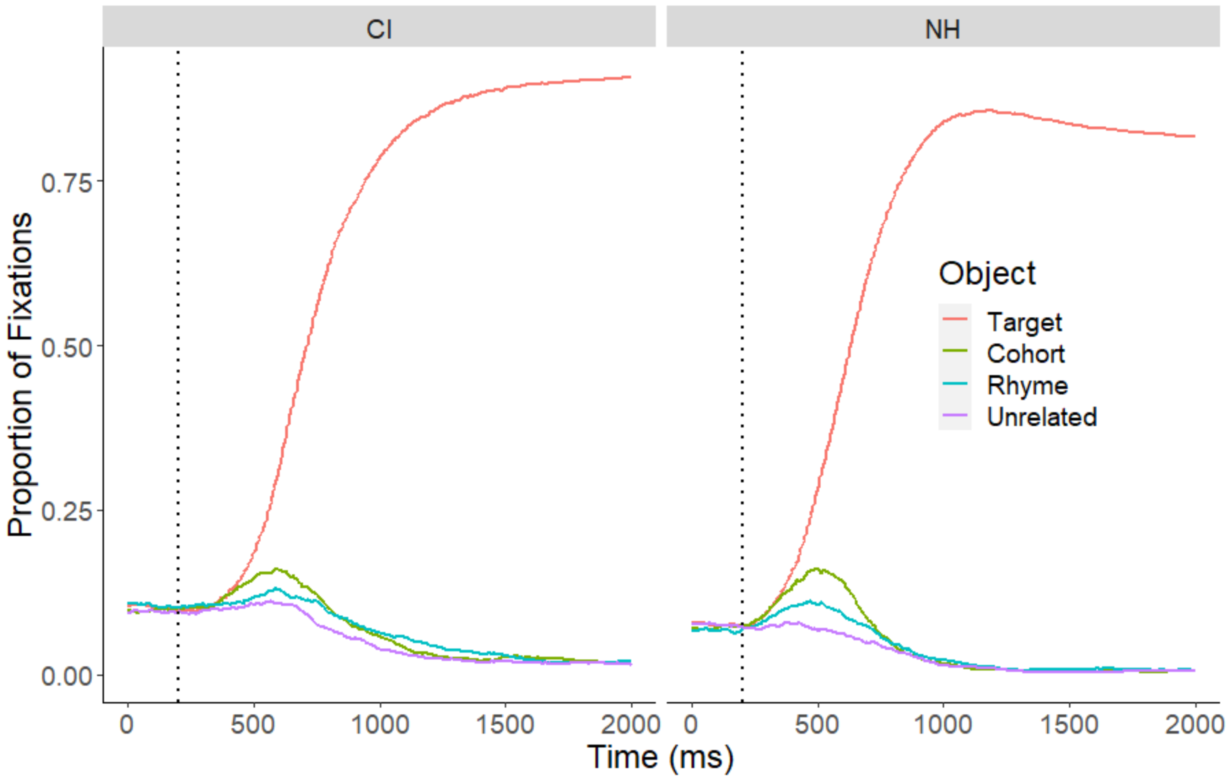

Figure S4.1 Looks to the four object types over time after onset of target word as a function of sentence context in the primary experiment. The dotted vertical line represents the 200ms oculomotor delay.

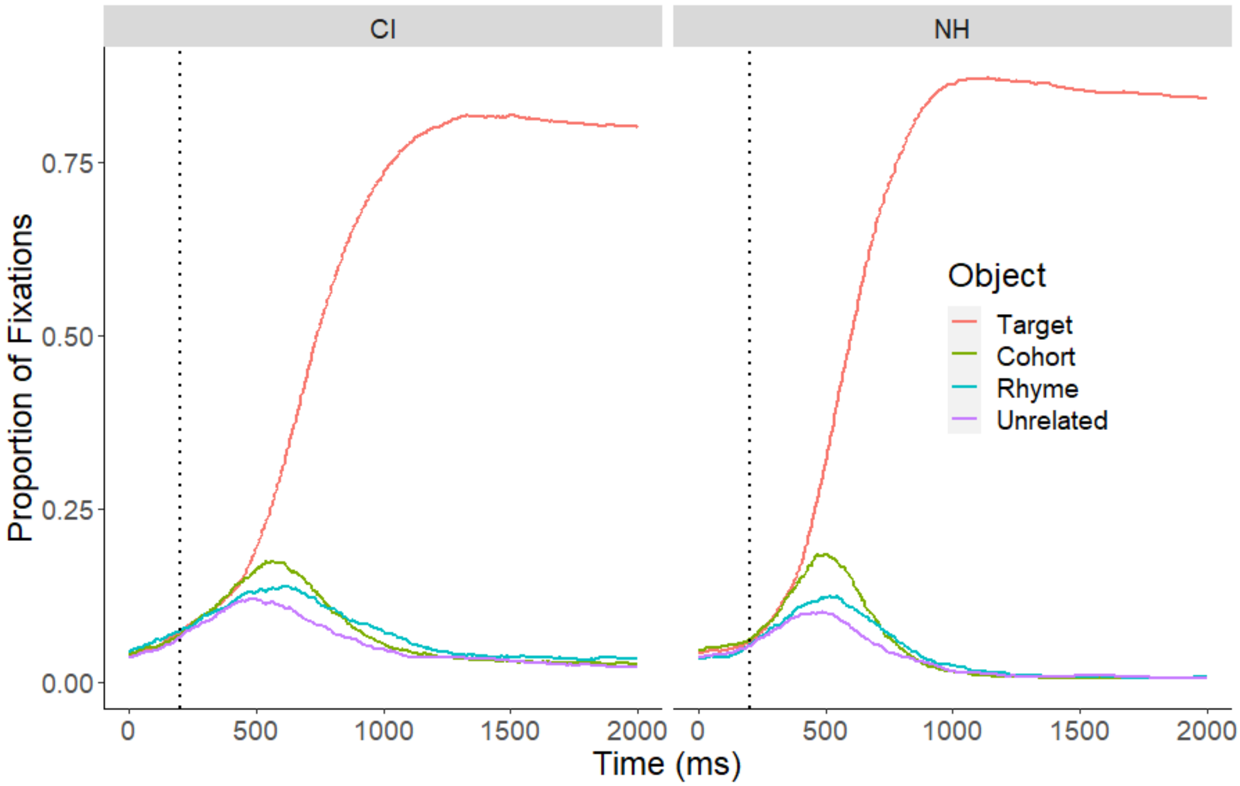

Figure S4.2 Looks to the four object types over time after onset of target word as a function of sentence context in the replication experiment. The dotted vertical line represents the 200ms oculomotor delay

S5. Analyses of Fit Quality

A critical assumption of the curvefitting approach is that fit quality does not differ substantially across conditions of the model. Here, we present a summary of this for each analysis.

Primary Experiment: Targets.
Logistic fits were good overall, average $R = .998$, across the four cells of the design (CI/NH × Word in Isolation/Carrier Phrase). A 2×2 mixed ANOVA of the R values showed a significant difference between listener groups, $F(1,40) = 13.83$, $p = .001$ with better fits for NH patients ($M = .9998$) than for CI patients ($M = .9973$). There was also a marginally significant main effect of context, $F(1, 40) = 3.79$, $p = .059$, with better fits, numerically, in the word-in-isolation condition ($M = .9982$) than in the carrier phrase condition ($M = .9977$). There was no significant interaction between the two, $F(1,40) = 2.01$, $p = .164$. Despite these significant differences, it is important to note that the mean differences between condition were less than .01, suggesting only minimal differences.

Primary Experiment: Cohorts
The fits of the jackknifed data were good across conditions and for both CI users and NH listeners, mean $R = .989$. To compare quality of fits across groups we used 2×2 mixed ANOVA of the R values across the four cells. Fit quality did not significantly differ across either group, $F_{jk}(1,40) = 2.67$, $p = .111$, or context, $F_{jk}(1,40) = 0.32$, $p = .572$. There was also no interaction between group and context, $F_{jk}(1,40) = 0.45$, $p = .505$.

Replication Experiment: Targets.
Goodness of fit (correlation) was analyzed with a 2×2 ANOVA examining the effect of listener group and context. This showed a main effect of listener group with CI users ($M = .998$) showing slightly lower quality of fit than NH listeners (.999), $F(1,38) = 14.84$, $p < .001$. There was no significant effect of context, $F(1,38) = 1.68$, $p = .202$. There was no significant interaction between group and context, $F(1,38) = 1.62$, $p = .211$. In all cases, fits were near ceiling.

Replication Experiment: Cohorts
To compare quality of fits across groups we used 2×2 mixed ANOVA of the R values across the four cells. Fit quality did not significantly differ across either listener group, $F_{jk}(1,38) = 3.59$, $p = .066$, or context, $F_{jk}(1,38) = 0.009$, $p = .926$. There was also no interaction between group and context, $F_{jk}(1,38) = 0.01$, $p = .905$.

S6. Analyses of Rhymes

As we described in the main text, CI user's fixations to rhymes could not be analyzed in the same manner as cohorts. This was for two reasons. First, the baseline procedure that had been applied successfully to target and cohort fixations led to rhyme fixations exhibiting a non-canonical pattern which could not be fit with the asymmetrical Gaussian commonly used to model competitor looks. Second, CI users—but not NH listeners—showed increased looking to the rhymes during the carrier phrase, before the target was heard. This created an unequal baseline that made it hard to compare groups. We are hesitant to make claims about the cause of this surprisingly increased baseline. Each item in the set was equally likely to be the target on throughout the experiment, we counterbalanced the location of the rhyme on the screen across trials. Nonetheless this peculiar pattern prevents us from implementing our intended analyses.

Instead, here we present several coarser analyses that examine differences in rhyme fixations between context conditions and listener groups. As these were not the planned analyses, they should be treated as exploratory.

For these analyses, we extracted three measures directly from the proportion of fixations over time. These measures were intended—much like curvefit parameters—to describe meaningful properties of the curves, without assuming any particular form (see Rigler et al., 2015). First, we computed: the peak height of fixations to rhymes after subtracting fixations to the unrelated item at each time point. Second, we computed the time at which this peak was reached. Finally, we computed the extent to which the rhyme was active over and above the unrelated competitor (e.g., the proportion of time points at which the rhyme was more active than the unrelated).

These measures do not rely on curvefitting procedures and did not require jackknifing as the cohort analyses in the main text did. Instead, we started by smoothing the proportion of fixations to both rhymes and unrelated items using a 48 ms triangular window to aid in the estimation of a single peak height and time. We then subtracted unrelated fixations from the rhyme fixations for that subject and condition.

Height was defined as the greatest value along this time series, and time as the point in time where this was reached. To calculate the extent of rhyme fixations, we computed the number of time points at which the rhyme item fixations minus unrelated item fixations were greater than .02 (and converted that to a proportion over the entire length of the trial). These measures were then analyzed in standard ANOVAs (similarly to the curvefit parameters). For all measures, the first 250 ms of the time course was ignored to avoid including the unexplained increase in rhyme looks in CI users described above and in the main text.

Primary Experiment Rhymes

Figure S5.1 A shows the smoothed timecourse of rhyme and unrelated fixations for the primary experiment presented in the main text.

Peak height (Figure S5.1B) showed a significant effect of listener group. CI users made more fixations to the rhyme item than NH listeners, $F(1,40) = 13.50$, $p = .001$. No other effect was significant, all $p > .15$. ***Peak timing*** (Figure S5.1C) showed a significant main effect of context with both listener groups looking to rhymes earlier in carrier phrases than in isolated words, $F(1,40) = 6.75$, $p = .013$. No other effects were significant, all $p > .15$. Rhyme ***extent*** (Figure S5.1D) also showed a significant main effect of group with CI users showing greater overall rhyme activation than NH listeners, $F(1,40) = 18.10$, $p < .001$, with no other significant effects, all $p > .10$.

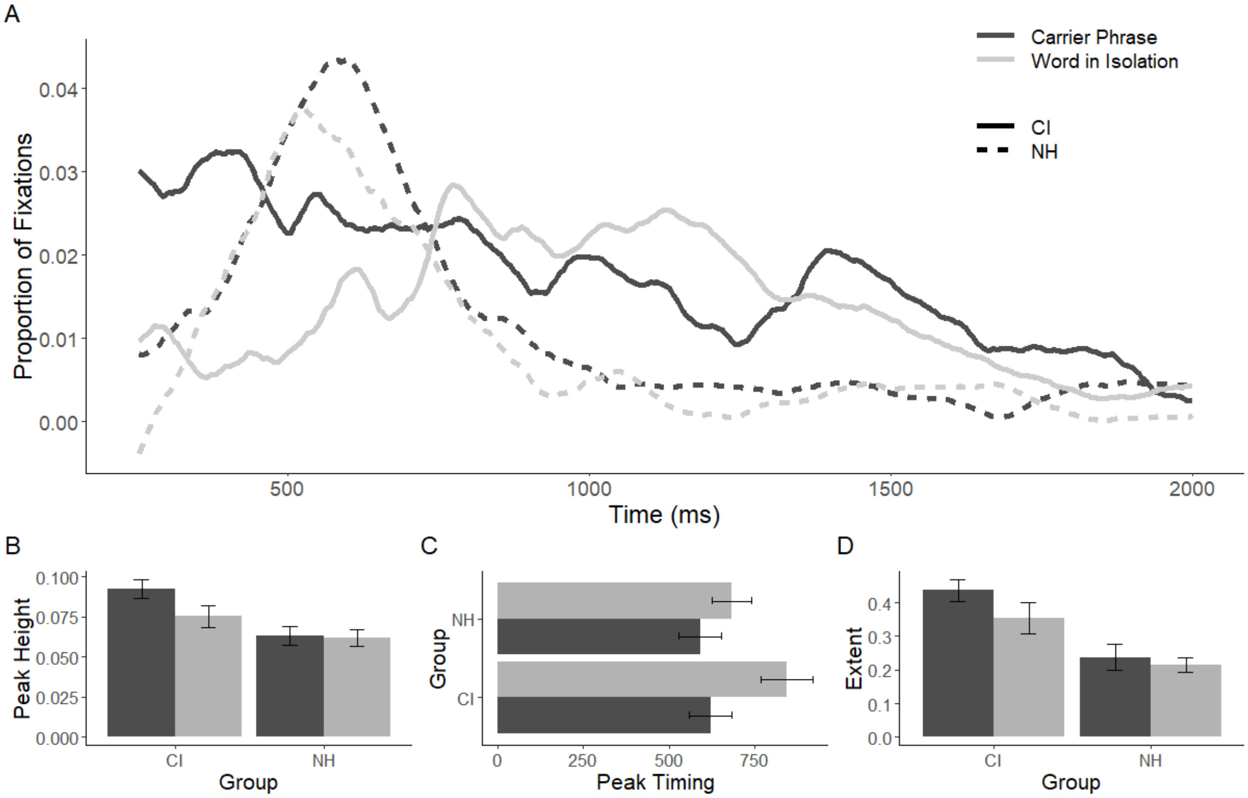

Figure S5.1. A) Rhyme fixations as a function of time after subtracting looks to the unrelated and smoothing, for each listener group and sentence context condition. Panels below indicate individual curvefit parameters; error bars indicated SEM. B) Peak height; C) peak timing (ms); D) extent (proportion of the trial during which fixations to rhymes minus unrelated were greater than .02.

Replication Experiment

We next repeated the same analysis for the replication (Supplement S1). Figure S5.2A shows the smoothed timecourse on which the estimates (S5.2B-D) were based.

Peak height (Figure S5.2B) showed a marginally significant effect of listener group, with CI users making numerically more rhyme fixations than NH listeners, $F(1,38) = 3.54$, $p = .067$. No other effect was significant, all $p > .4$. ***Peak timing*** (Figure S5.2C) showed a significant main effect of listener group, $F(1,38) = 10.66$, $p = .002$, with CI users reaching their peak rhyme fixations later than NH listeners. No other effect was significant, all $p > .5$. Rhyme ***extent*** (Figure S5.2D) also showed a significant main effect of group with CI users a longer period of increased rhyme fixations compared to NH listeners, $F(1,38) = 14.54$, $p < .001$. There were no other significant effects, all $p > .4$.

Summary

Across the two experiments, CI users showed evidence of increased rhyme activation, primarily seen in the measure of extent. This suggests that CI users are showing the sustained competitor activation profile that was also observed with cohorts (as described in the main text). Additionally, in the main experiment (and marginally in the replication, S1), CI users showed

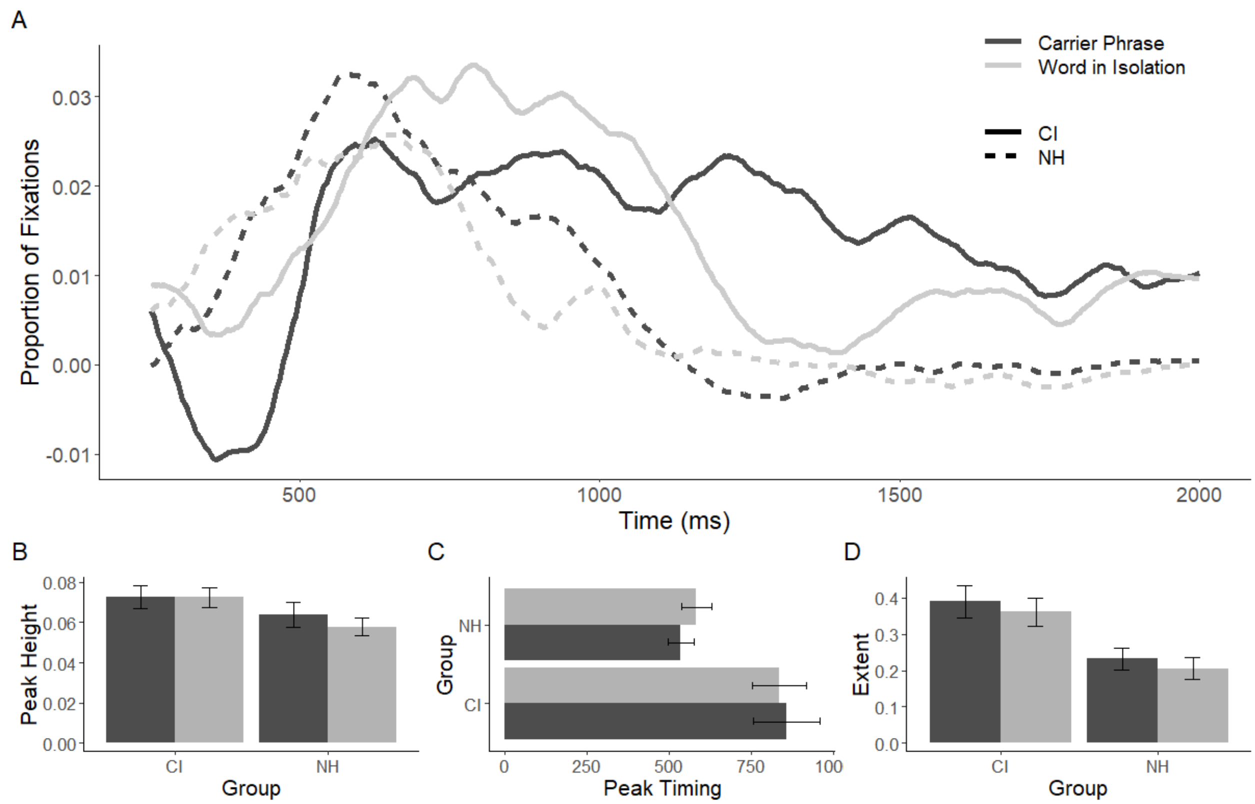

Figure S5.2. A) Rhyme fixations as a function of time after subtracting looks to the unrelated and smoothing, for each listener group and sentence context condition. Panels below indicate individual curvefit parameters; error bars indicated SEM. B) Peak height; C) peak timing (ms); D) extent (proportion of the trial during which fixations to rhymes minus unrelated were greater than .02.

evidence of increased peak activation of rhymes relative to NH listeners. This increased rhyme peak is also observed as part of the "wait and see" profile (McMurray et al., 2017). Thus, this adds support to the idea that the CI users exhibited a "wait-and-see" profile (either as part of their response to the experimental task as a whole, or as part of their general profile of listening). The overall pattern of rhyme results supports the conclusions from target and cohort analyses presented in the main text.

References

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

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

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