Published in final edited form as:
J Exp Psychol Hum Percept Perform. 2014 February ; 40(1): 308–327. doi:10.1037/a0034353.

The process of spoken word recognition in the face of signal degradation

Ashley Farris-Trimble, Ph.D.[*],
Dept. of Psychology and, Dept. of Communication Sciences and Disorders, Delta Center, University of Iowa

Bob McMurray, Ph.D.,
Dept. of Psychology and, Dept. of Communication Sciences and Disorders, Delta Center, University of Iowa

Nicole Cigrand, B.A., and
Dept. of Psychology and, Dept. of Communication Sciences and Disorders, Delta Center, University of Iowa

J. Bruce Tomblin, Ph.D.
Dept. of Communication Sciences and Disorders, Delta Center, University of Iowa

Abstract

Though much is known about how words are recognized, little research has focused on how a degraded signal affects the fine-grained temporal aspects of real-time word recognition. The perception of degraded speech was examined in two populations with the goal of describing the time course of word recognition and lexical competition. Thirty-three postlingually-deafened cochlear implant (CI) users and 57 normal hearing (NH) adults (16 in a CI-simulation condition) participated in a visual world paradigm eye-tracking task in which their fixations to a set of phonologically related items were monitored as they heard one item being named. Each degraded-speech group was compared to a set of age-matched NH participants listening to unfiltered speech. CI users and the simulation group showed a delay in activation relative to the NH listeners, and there is weak evidence that the CI users showed differences in the degree of peak and late competitor activation. In general, though, the degraded-speech groups behaved statistically similarly with respect to activation levels.

Keywords

cochlear implants; eye-tracking; online processing; word recognition

1.0 Introduction

A critical problem in language comprehension is mapping incoming acoustic material to words in a lexicon in which many word-forms are highly overlapping. At early points in the signal, multiple words may be consistent with the input that has been received, resulting in a temporary ambiguity that must be resolved by later input. Empirical psycholinguistic paradigms have measured the process of mapping the acoustic signal onto candidates in the mental lexicon (*lexical access*) and have revealed that from the earliest moments of the

Corresponding Author: Ashley Farris-Trimble, 8888 University Dr., Dept. of Linguistics, Simon Fraser University, Vancouver, BC V5A 1S6, 778-782-9708, ashley_trimble@sfu.ca.
[*]First author is now affiliated with Simon Fraser University but conducted this research at the University of Iowa.

input, listeners consider multiple lexical candidates in parallel, which compete over time until only one remains. Such work has yielded a rich set of temporal dynamics that reveal underlying competition mechanisms (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan & Gaskell, 2007; Dahan, Magnuson, & Tanenhaus, 2001; Luce & Cluff, 1998), and these dynamics have been instantiated in a number of models (e.g., Luce, Goldinger, Auer, & Vitevitch, 2000; Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008).

A limitation of this research, however, is that it has primarily examined normal-hearing listeners under ideal conditions. A wealth of studies have documented how degrading the input affects speech recognition accuracy (Dorman & Loizou, 1997; Dorman, Loizou, & Rainey, 1997; Duquesnoy & Plomp, 1980; Hawkins & Stevens, 1950; Kalikow, Stevens, & Elliott, 1977; Loizou, Dorman, & Tu, 1999; Nittrouer & Lowenstein, 2010; Nittrouer, Lowenstein, & Packer, 2009; Pichora-Fuller, Schneider, & Daneman, 1995; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Sommers, Kirk, & Pisoni, 1997). Similarly, a number of studies have looked at *how well* phonemic information can be assembled into words under adverse listening conditions by asking whether word recognition accuracy can be predicted from phoneme recognition accuracy (Boothroyd & Nittrouer, 1988; Bronkhorst, Bosman, & Smoorenburg, 1993). However, measurements of accuracy focus on the final product of word recognition, not on the temporal dynamics how listeners achieve it, leaving it an open question as to whether the real-time properties of lexical access may differ with a degraded signal.

The present study begins to examine real-time spoken word recognition and lexical access under degraded input by examining the time course of spoken word recognition in a particular and important form of signal degradation: cochlear implants. We examined both adult, postlingually-deafened cochlear implant (CI) users (Experiment 1) and normal hearing (NH) listeners hearing CI-simulated speech (Experiment 2). We used the visual world paradigm (VWP; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to measure the temporal unfolding of lexical access, examining both the overall time course of activation and also assessing specific markers in the eye-movement record for the timing and degree of activation. The obvious hypothesis was that differences in speech processing arise when speech is degraded; however, what was not known was which specific aspects of the time course of processing would differ. We also hypothesized that these differences would not be due to the nature of the input alone, but also to fundamental differences in lexical processing that result from long-term experience with a CI, which we examine by comparing the findings of the two experiments. These questions are important from a theoretical perspective, as current models of lexical access were conceived around the problem of mapping a clear input to lexical candidates.

1.1 Degraded input and cochlear implants

CI users must contend with a severely degraded input (see, e.g., Niparko, 2009 for more detailed information on CIs). CIs degrade frequency resolution, transmit temporal fine structure poorly, and often result in the loss of some low-frequency information (among many other degradations). CI users' long-term experience with this degradation raises the possibility that their word recognition processes reflect both the atypical input and any adaptations they have made to cope with it.

CI users are generally accurate (though variable) at recognizing speech in quiet, but recognition is not stable under difficult listening conditions, for instance in noise (Friesen, Shannon, Baskent, & Wang, 2001; Fu, Shannon, & Wang, 1998; Stickney, Zeng, Litovsky, & Assmann, 2004) or in open-set tasks (Balkany, Hodges, Menapace, et al., 2007; Helms, Müller, F., et al., 1997). However, it is likely that even when they successfully recognize a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

word, CI users arrive at the correct end-state choice via a different route, either in terms of the cognitive processes they deploy, or in terms of the types of preliminary decisions that are made during processing. Nittrouer and Lowenstein (2012) offer hints of this. They examined the effect of degraded speech on working memory and recovery of phonetic structure. NH adults accurately identified CI-simulated words, but in a memory task, recall was slower and less accurate than with unprocessed speech, implying that higher-order processing was affected even when recognition was not overtly impaired. Similarly, in a task that required attention to phonetic structure, degraded speech slowed response time without impairing accuracy, suggesting there may be differences in how processing unfolds over time.

Examining CI users alone may not completely address the broader issue of how degraded input affects word recognition for several reasons. First, the CI may offer a unique form of degradation. Second, and perhaps more importantly, differences between NH individuals and CI users may also derive from their long-term experience with that degradation (over and above the degradation itself). CI users may tune their word recognition systems in various ways to cope with this degraded input. For example, they may keep competing lexical candidates active because they are accustomed to being uncertain and having to revise their interpretations. Indeed, there is substantial evidence of significant adaptations in long-term CI users (Dorman & Ketten, 2003; Dorman & Loizou, 1997; Giraud, Price, Graham, Truy, & Frackowiak, 2001; Giraud, Truy, & Frackowiak, 2001; Gray, Quinn, Court, Vanat, & Baguley, 1995; Harnsberger et al., 2001; Pelizzone, Cosendai, & Tinembart, 1999; Perkell, Lane, Svirsky, & Webster, 1992; Svirsky, Silveira, Suarez, Neuberger, Lai, & Simmons, 2001; Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997; cf. Fu, Shannon, & Galvin, 2002), but this has typically been shown with either cognitive neuroscience measures or accuracy, and there as yet has been no characterization of how the time course of lexical access may change.

One way to eliminate the effects of long-term experience is to compare CI users to NH individuals recognizing CI-simulated speech, who have not had as much time to adapt. Though it is impossible to exactly replicate electric-only hearing with an acoustic signal, simulations have often been used to make inferences about CI users or to pinpoint differences specific to CI users (Cullington & Zeng, 2008; Friesen et al., 2001; Qin & Oxenham, 2003; Stickney et al., 2004; Throckmorton & Collins, 2002; Turner, Gantz, Vidal, Behrens & Henry, 2004). CI-simulated speech shows many of the same qualitative degradations, and isolated word recognition performance for CI simulations is comparable to that of actual CI users (Fu & Shannon, 1998, 1999; Fu, Shannon, & Wang, 1998; Nelson, Jin, Carney, & Nelson, 2003; Stickney et al., 2004). While some short-term adaptation to CI-simulated speech occurs (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Fu, Nogaki, & Galvin, 2005; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008; Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011; Rosen, Faulkner, & Wilkinson, 1999), many of these adaptations require implicit feedback, making them unlikely to occur in designs such as what was used here.

Examining both CI users and CI simulation allows us to consider the effects of experience over and above the degraded signal itself. Of course, this approach assumes that the simulated speech accurately approximates the degradation caused by the CI, a point to which we will return in the discussion section. However, to the extent it differs, this comparison can also be informative by asking which properties of degraded word recognition can be seen with multiple forms of degradation. Either way, if differences are observed it clearly points the way toward future work on different forms of degradation; however, where they are not observed, this points to perhaps stable, immediate responses of the system to degradation.

1.2 Real-time spoken word recognition

The goal of this study is to precisely characterize the time course of lexical access in CI and CI simulations; here we briefly review what is known about this in typical listeners. A fundamental issue in spoken word recognition and lexical access is the sequential nature of the input – the problem of integrating information over time, and the ambiguity created by the lack of complete information at early points in the word (e.g., Marslen-Wilson, 1987). Even assuming that the phonemes of a word are accurately encoded, the fact that those phonemes arrive over time raises questions about how and when the lexicon is accessed. For instance, do listeners wait until they have heard the entire word before accessing a single lexical entry, or do they access candidates earlier using partial information?

Under ideal listening conditions, listeners activate potential matches as soon as the earliest input is received and can often recognize a word before it is heard in its entirety (Grosjean, 1980; Marslen-Wilson, 1973; Marslen-Wilson & Tyler, 1980; McMurray, Clayards, Tanenhaus, & Aslin 2008). Because multiple words may match the signal at early points in time, multiple lexical candidates are initially activated (Allopenna et al., 1998; Dahan & Gaskell, 2007; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson, 1987; Marslen-Wilson, 1990; McMurray, Samelson, Lee, & Tomblin, 2010; Spivey, Grosjean, & Knoblich, 2005; Zwitserlood, 1989; Zwitserlood & Schriefers, 1995), and this set of competitors is continually evaluated and narrowed down as more input is received (Allopenna et al., 1998; Connine, Blasko, & Titone, 1993; Dahan & Gaskell, 2007; Dahan, Magnuson, Tanenhaus, et al., 2001; Marslen-Wilson & Zwitserlood, 1989; McMurray et al., 2010). Finally, some words (e.g., high-frequency words) are activated more strongly than others (Dahan & Gaskell, 2007; Dahan, Magnuson, & Tanenhaus, 2001; Grosjean, 1980; Marslen-Wilson, 1987; Tyler, 1984; Zwitserlood, 1989), and more active words inhibit less active words (Dahan, Magnuson, Tanenhaus, et al., 2001; Luce & Pisoni, 1998). Our understanding of these temporally unfolding aspects of word recognition derives from empirical paradigms that measure how strongly listeners consider multiple lexical candidates in real time, paradigms like cross-modal priming, gating and the visual world paradigm.

1.3 Eye tracking and the visual world paradigm

In adaptations of the VWP to study word recognition (e.g., Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; McMurray et al., 2010), four pictures appear on a computer screen. The listener hears the name of one of them and clicks on the referent. The names for the pictures are phonologically related to the auditory stimulus, allowing the experimenter to assess classes of lexical competitors. If, for instance, the auditory stimulus were *wizard*, the screen might contain a picture of a *wizard* (the target), a *lizard* (a rhyme), a *whistle* (a cohort, which overlaps at onset) and an unrelated item like *baggage.* As listeners perform this task, eye movements to each object are recorded. Averaged across trials, this offers a millisecond-by-millisecond measure of the strength by which competitors are considered during recognition. Indeed, the proportion of fixations to various lexical competitors corresponds closely to activation patterns generated by models like TRACE (McClelland & Elman, 1986) when simple linking functions are employed to map activation across the whole lexicon to the words pictured on the screen (Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; McMurray et al., 2010; McMurray, Tanenhaus, & Aslin, 2009).

The VWP is ideal for examining clinical populations. The task is simple and does not require metalinguistic judgments. Impaired listeners typically succeed in selecting the appropriate picture, and the task has been used effectively with aphasics (Yee, Blumstein, & Sedivy, 2008), dyslexics (Desroches, Joanisse, & Robertson, 2006), and children with specific language impairment (McMurray et al., 2010). It is also quite reliable, with test/re-test reliabilities above .7 for some components (Farris-Trimble & McMurray, in press). Crucially, the VWP can reveal processing differences across groups even when responses are equally accurate, because the measure of interest is the fixations to competitors *prior* to the overt response (McMurray et al., 2010). These fixations reflect ongoing processing and are typically not under the listener's conscious control. The simultaneous measures (trial-by-trial accuracy and fixations) also allow us to condition the analysis of eye-movements on whether or not the listener identified the correct word, so we can determine whether different populations reach the same correct end-state via a different route.

2.0 Experiment 1

2.1 Materials and Methods

2.1.1 Design—Twenty-nine sets of four words were used (Appendix A). Each set contained a base word (e.g., *wizard*), a cohort of the base (*whistle)*, a rhyme of the base (*lizard*), and a phonologically unrelated item (*baggage*). The words in each set were not semantically related and were equated for number of syllables and stress pattern (all two-syllable words carried strong-weak stress). Words were selected from a number of similar experiments (e.g., Allopenna et al., 1998; McMurray et al., 2010) to offer sufficient phonetic diversity (e.g., covering the range of vowels and consonants) that our measure captured the general dynamics of lexical processing, rather than the specifics of how particular phonemes are perceived.

The four pictures in a set always appeared together, and each member of the set served as the stimulus in an equal number of trials. Each of the 116 words (29 sets × 4 items) occurred as the auditory stimulus five times for a total of 580 trials. Because of the structure of the sets, there were four types of trials, with each defined by which word was the auditory stimulus (here termed *target*, as it was the item we expected the listener to click on). Table 1 illustrates the role of each word in the set *wizard*, *whistle*, *lizard*, *baggage*, as a function on which word was the auditory stimulus. The letters naming each trial-type refer to the relationship among the items on the screen to the auditory stimulus. For instance, in a TCRU (or Target-Cohort-Rhyme-Unrelated) trial, the base word *wizard* is the auditory stimulus and target; the other three items serve as *wizard*'s cohort (*whistle*), rhyme (*lizard*), and unrelated item (*baggage*). In a TCUU (or Target-Cohort) trial, *whistle* is the auditory stimulus/target item, while *wizard* serves as its cohort, and *lizard* and *baggage* are both unrelated to the target. These arrangements allow us to estimate the time course over which certain types of competitors are considered across multiple trial types: targets, for example, can be examined on all four trial-types, cohorts on TCRU and TCUU trials, and rhymes on TCRU and TRUU (Target-Rhyme) trials. Crucially, while sets were reused across multiple trials, each word in a set served as the auditory stimulus an equal number of times. As a result, the participant was could not guess the target item in advance, and all words played a role in the analysis as both a target item and a competitor.

2.1.2 Auditory stimuli—The 116 auditory stimuli were recorded by a female speaker with a standard American accent. They were recorded in a soundproof room at a sampling rate of 44100Hz using a Kay Elemetrics Computerized Speech Lab 4300B (Kay Elemetrics Corp., Lincoln Park, NJ). Stimuli were produced in the carrier phrase "He said ___" to ensure a declarative sentence intonation. Several tokens of each stimulus were obtained and the single best exemplar was isolated from the carrier phrase using both visual and auditory inspection. Stimuli were RMS amplitude-normalized and were low-pass filtered with an upper cut-off of 7.2 kHz because though the recordings were clear, at the somewhat higher volumes that most of the CI patients preferred, our playback equipment introduced a slight distortion in the high frequency ranges. Finally, 100 ms of silence were added before and after each word.

2.1.3 Visual stimuli—Visual stimuli came from a database of clipart pictures that underwent a rigorous selection, editing, and approval process used across multiple studies by McMurray and colleagues (Apfelbaum, Blumstein, & McMurray, 2011; McMurray et al., 2010). The pictures were similar in style and roughly equivalent in visual saliency.

2.1.4 Participants—CI users were recruited from an ongoing research project conducted at the University of Iowa, and NH participants were recruited from the community through advertisement. Thirty-three adult CI users and 26 NH individuals participated. An additional six CI users were recruited but not run because of difficulty calibrating the eye-tracker (4) or insufficient familiarity with a computer (2). Of the 33 CI users, four were excluded because their reaction times averaged longer than three seconds (RTs in this paradigm are usually on the order of 1200–1400 ms: McMurray et al., 2010). This resulted in a total of 29 CI users contributing to the analysis. None of the 26 NH listeners were excluded by these criteria.

The 29 CI users showed diverse device configurations (Table 2). Four were implanted unilaterally; five received bilateral implants sequentially (at different times); 11 received bilateral implants simultaneously; eight used a hybrid hearing-preservation implant (i.e. a short-electrode implant, which preserves some low frequency hearing, often used with a hearing aid in the same ear); and one used a hybrid implant in one ear and a standard implant in the other. All participants reported normal or corrected-to-normal vision in at least one eye.

The NH controls reported normal hearing and normal or corrected-to-normal vision, and all were native monolingual speakers of American English. Because we relied on self-report rather than hearing tests, it is possible that the older NH controls had some age-related hearing loss of which they were unaware. If anything, such hearing loss might make the NH group perform more like the CI users (since the signal would be degraded, albeit in different ways[1]). The two hearing-groups did not differ in age (CI: M = 59.7 years, range 35–81; NH: M = 59.8 years, range 35–89; $t < 1$). Most participants were paid $30 for participation; 13 CI users received audiology services instead as part of an earlier IRB protocol.

2.1.5 Procedure—The experiment was implemented with Experiment Builder software (SR Research Ltd., Ontario, Canada). To control for background noise, the experiment was administered in an open-door sound-attenuated booth. Extra noise from the testing computers was minimized by replacing the fans in the eye-tracking control computer with low-noise fans. Auditory stimuli were presented through a Sound Blaster X-Fi soundcard over two front-mounted Bose loudspeakers (each at approximately a 45° angle from the midline of the participant) amplified by a Sony STR-DE197 amplifier/receiver. Volume was initially set to 65 dB, and each participant was allowed to adjust it to a comfortable level using a knob on the speaker amplifier during the practice trials[2] (described below). Stimuli were presented at a user-optimized (rather than fixed) level and participants chose the implant program that they would normally use to listen to speech in a quiet room. Both choices were made because the focus of this study is on the process of lexical access, not the

[1]The specific frequencies lost in normal-age related hearing loss and CI processing differ for many listeners: age-related loss typically degrades higher frequencies, while CIs typically allows for better perception of high-frequency sounds than low-frequency sounds. However, our experiment was designed to look at overall differences in word recognition, rather than the recognition of particular sounds, using a stimulus set that included many types of sounds were in all positions. Thus, we do not expect such effects to play a large role.

[2]Though each participant could adjust the volume levels from the initial level, few NH participants did. Some NH participants reduced the volume to approximately 60dB, while some CI users increased the volume up to 70 dB.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

end-state accuracy, and as a result, we preferred participants to listen under conditions optimal to their performance and comfort.

Participants were seated in front of a 1280×1024 LCD computer screen and a desktop-mounted eye-tracker. They read the instructions, and the experimenter explained the procedure and ensured that they understood the task. They performed eight practice trials before beginning the experiment. These familiarized the participants with the procedure using words and pictures that were not included in the test trials. At the beginning of each trial, the four pictures from a set appeared in the four corners of the screen. Each picture was 300×300 pixels and located 50 pixels from the screen edge. Picture location was randomized across trials; the order of trial presentation was randomized across participants. Along with the pictures, a red dot appeared in the middle of the screen; after 500 ms, it turned blue, and the participant clicked on it to initiate the auditory stimulus. Then, the blue circle disappeared and a word played over the speakers. The participants clicked on the picture that matched the word they heard. Participants were instructed to guess if they were not sure. There was no opportunity to replay the word.

2.1.6 Eye tracking—Eye movements were recorded with a desktop-mounted SR Research Eyelink 1000 eye-tracker (SR Research Ltd., Ontario, Canada); a chin rest was used to stabilize the head. The eye-tracker was calibrated with a 9-point procedure. To offset drift of the eye-track over time, a drift correction was performed every 29 trials. If a drift correct failed, the eye-tracker was recalibrated. The Eyelink 1000 uses the location of the pupil and corneal reflection to determine point of gaze in screen coordinates every four ms. The continuous record of gaze location was automatically classified into saccades, fixations and blinks. As in previous studies (McMurray et al., 2002; 2010), each saccade was combined with the subsequent fixation into a single unit, termed a "look." A look thus lasted from the beginning of a saccade to the end of the subsequent fixation. Fixations launched during the first 300 ms of each trial (100 ms of silence at the beginning of each sound file plus 200 ms required to plan and launch an eye movement) were not included in the analysis as they could not have been driven by the auditory stimulus. The trial was deemed to end 200 ms before the mouse-click as we found a subset of participants who tended to initiate the movement toward or fixate back to the center before the actual click. We compared the screen coordinates of the looks to those of the images to determine the object of fixation. Consistent with prior work, image boundaries were extended by 100 pixels in each direction to allow for a small amount of error in the eye-track, thus capturing looks that were intended for the item. This was not large enough create overlap between images.

2.2 Results

We conducted three sets of analyses. First, we examined the accuracy and reaction time of the mouse clicks to determine the overall performance profiles of both hearing-groups. Next, we examined the pattern of fixations over time to look for gross differences between the two groups in the time course of target and competitor fixations during recognition. Finally, we identified moments in the fixation record that assessed either the *timing* at which each word-type was considered or ruled out or the *degree* to which candidates were considered.

2.2.1 Mouse clicks (accuracy & reaction time)—Both groups were generally accurate, though the CI users had higher error rates on average: the CI users made errors in picture selection in 5.2% of trials (SD = 5.1%; range: 0.2 – 20.0%), and NH controls made significantly fewer errors in 0.6% of trials (SD = 1.3%; range: 0 – 6.6%; $t(53) = 4.93$, $p < .001$). The NH controls' highly accurate responses confirm that our stimuli were clear and audible. Adult CI users' average reaction time (RT) was 2134 ms (SD = 367; range: 1357–2751); with control participants' responding significantly faster at 1576 ms (SD = 233;

range: 1199–2177; $t(53) = 4.24$, $p < .001$). Importantly, even the CI users performed the task well. Only three (of 29) were less than 90% accurate (none of the NH listeners fell below this).[3] The analyses of the fixations include only correct trials, to focus on the subset of trials in which accuracy was equal – if there were still differences in the fixations, this constitutes strong evidence for underlying differences in processing. Of 31,900 total trials across both groups, 950 were excluded (CI: 867; NH: 83).

2.2.2 Gross time course of fixations—As a starting point, Figure 1 plots the proportion of trials in which NH participants fixated each item as a function of time on correct TCRU trials, a typical way to plot VWP data. At about 400 ms, looks to the target and cohort, both of which were consistent with the stimulus up to that point, began to diverge from the other competitors; by 500ms, cohort looks peaked and began to decline as the cohort was disambiguated from the target. Also around this time, the rhyme began to receive fixations as its similarity with the target began to play a role. Both cohort and rhyme competitors received more looks than the unrelated item (as the analyses below demonstrate statistically), suggesting that these words were being considered. Figure 2 compares the CI users and the NH group for each of the four competitors averaged across trial-types in which that type was present on the screen. It suggests that CI users' initial looks to the target and cohort (driven by the early portion of the auditory input) diverged from unrelated looks somewhat later than the NH listeners', around 500 ms; that CI users did not suppress looks to cohorts and rhymes as quickly as NH listeners; and that they were less likely to fixate the target toward the end of the trial and more likely to fixate other objects.

One problem with this description is that it is averaged across individuals. Nonlinear curves like this may not reflect the average performance of any individual. For example, if CI users had steep slopes for their target fixations (Figure 2A), but were more variable in when they transitioned, this would average to a shallow slope, which does not reflect any individual. Figures S1–S4 in Note S2 of the online supplement, which show individual fixation functions, illustrate the relationship between the average function and the individual measures. To evaluate group differences statistically, we needed an approach that is less sensitive to artifacts of averaging across participants. Thus, we fit non-linear functions (Figure 3) to each participant's averaged looks to each of the four objects as a function of time (using techniques previously applied in McMurray et al., 2010; Farris-Trimble & McMurray, in press) and compared this parametric description of the time course of fixations across groups. It is a common strategy for participants to move their mouse to the target and then look back to the center dot while they are clicking in anticipation of the next trial. Indeed, visual inspection of fixation curves for individual participants revealed many instances of this strategy, as characterized by looks to the target item peaking and then declining. To eliminate the effects of this meaningless decline in the curve-fit analyses below, each trial ended 200 ms before the participant's response.

2.2.2.1 Target fixations: Target fixations were modeled by a logistic function predicting fixations to the target as a function of time, t (Figure 3A). Individual differences are captured by four parameters: the minimum asymptote b (representing base fixations), the maximum asymptote p (maximum fixations), the slope at the transition s (rate of increase in fixation), and a crossover point c (timing of fixations), as shown in Equation 1.

[3]Removing these three subjects from subsequent analyses did not change any of the results at an α-level of .05.

$$P(target) = \frac{p-b}{1+\exp\left(4 \cdot \frac{s}{p-b} \cdot (c-t)\right)} + b \quad (1)$$

We achieved good fits for both groups of subjects (the NH group had slightly better fits, but both groups were at ceiling: average R^2 for both hearing-groups = .999; $t(53) = 2.19$, p = .03). Figure 4A plots the logistic function for each group based on the average parameters within that group, a representation that is an arguably more accurate depiction of the typical participant. This can help determine if any differences observed in Figure 2A reflect differences among individuals, and which may be artifacts of averaging. Comparison of Figure 4A and 2A suggests that the average logistic reflects largely the same differences observed in the average fixations.

The parameters for each subject were compared across hearing-groups with independent sample t-tests (Table 3). Because the first 300 ms of each trial were fixed at zero, we did not analyze the minimum asymptote. As the maximum asymptote is bounded by 0 and 1, we transformed this parameter with the empirical logit prior to analysis to yield a more normal distribution (but report the more meaningful raw curve values). As Table 3 suggests, CI users had significantly fewer maximum target fixations, a significantly later crossover point and significantly shallower slopes than NH participants.

2.2.2.2 Competitor fixations: As in McMurray et al. (2010), competitor fixations were modeled with an asymmetrical Gaussian function (Figure 3B, Equation 2) made up of two Gaussians that meet at the midpoint, and permit different minimum value and slope on each side.

$$P(competitor) = \begin{cases} \exp\left(\frac{(t-\mu)^2}{-2\sigma_1{}^2}\right)(h-b_1)+b_1 & \text{if } t \le \mu \\ \exp\left(\frac{(t-\mu)^2}{-2\sigma_2{}^2}\right)(h-b_2)+b_2 & \text{if } t > \mu \end{cases} \quad (2)$$

The upper and lower functions describe the time course before and after peak fixation. The peak's location in milliseconds (μ) and its height in proportion of fixations (*h*) are the same across both Gaussians. The onset and offset slopes (σ_1, σ_2) and lower minima (b_1, b_2) are specified for each independently. The minima are affected by starting and ending levels of fixation, while the slopes represent the rate of increase or decrease in fixations over time.

We estimated these parameters separately for each participant and for each competitor. Cohort fixations were derived from TCRU and TCUU trials; rhyme fixations from TCRU and TRUU trials; and unrelated fixations from all four trial-types. Fits were good, though somewhat better for the NH participants (Cohort: average R^2 for CI = .989, for NH = .995, $t(53) = 3.61$, p = .001; Rhyme: R^2 for CI = .983, for NH = .989, $t(53) = 2.50$, p = .02; Unrelated: R^2 for CI = .988, for NH = .994, $t(53) = 2.6$, p = .01). Curves constructed from average parameters are shown in Figure 4B–D, again showing a pattern similar to group averages. As before, the two groups were compared on each parameter, and we did not analyze the initial lower minimum. Empirical logit transformations were applied to the peak and offset baseline parameters.

Results are shown in Table 4. For looks to the **cohort**, CI users had a significantly slower (shallower) onset slope, a later midpoint, and a slower offset slope. They also had significantly higher offset baselines. The two groups did not differ significantly in peak height. Looks to the **rhyme** showed a similar pattern: CI users were slower than NH listeners to initially fixate the rhyme (onset slope), had a delayed midpoint, and were

NIH-PA Author Manuscript

NIH-PA Author Manuscript

marginally slower to stop fixating it (offset slope). They also fixated the rhyme at offset more than the NH listeners (offset baseline). Finally, looks to the **unrelated** item showed that CI users had later midpoints and more offset fixations than the NH group.

2.2.2.3 Device Factors: We next considered the variety of CI configurations in our sample, particularly the hybrid configuration. Because hybrids preserve some low-frequency information, their users may receive more useful information. We performed three one-way ANOVAs on each of our curve fit measures. In the first, we compared four groups of users: unilateral users, simultaneously-implanted bilateral users, sequentially-implanted bilateral users, and hybrid users. There was a marginal effect of group for the unrelated offset slope parameter ($F(3,29) = 2.8$, $p = .06$), but no other group differences (for all other parameters, $F < 1$). A second ANOVA grouped the two bilateral configurations and compared them to the two unilateral configurations. This found a significant effect of group on the unrelated offset slope parameter ($F(2,29) = 4.2$, $p = .025$) and again, no other group differences ($F < 1.5$, $p > .25$ for all others). Finally, we collapsed the CI users into two groups: hybrid and other. We performed a third ANOVA and found no group differences ($F < 2.6$, $p > .12$). In general, it appears that in this sample, the type of CI used does not greatly impact our measures of lexical processing.

2.2.2.4 Summary: These analyses suggest that CI users differed from NH listeners in two primary ways. First, CI users showed an overall slowing in the time course of lexical processing, both in the rate that fixations built, and in the absolute timing of those fixations. CI users considered words at a slower rate, as is evidenced by the shallower target, cohort and rhyme onset slopes (Figure 2A–C), and they were slower to suppress competitors (shallower offset slopes; Figure 2B–D). The absolute timing of both target and competitor functions was also delayed, indicated by later target crossover (difference of 80 ms, Table 3) and competitor midpoints (average difference of 61 ms; Table 4). CI users waited to get more information before they begin to commit to any interpretations, and they updated their candidate set more slowly. These results will be examined in more depth shortly.

Second, later in the time course, CI users fixated the target less and the competitors more than their NH counterparts. This was illustrated by lower target maxima and higher competitor offset baselines (Figure 2, Tables 3 and 4). It is tempting to interpret this finding as weaker sustained commitment to the target and less suppression of competitors by the CI users. However, to show that we must show that competitors received more fixations (or less suppression) *relative to other items*, and it is possible that CI users simply fixate everything more than their NH counterparts. In fact, CI users' incomplete suppression of competitors extended to the unrelated item (Figure 2D), even though it is not an expected phonological competitor. Given this, when interpreting CI users' increased competitor fixations, we must be careful to account for their increased fixations to the unrelated object. If their heightened cohort fixations at the end of the time course (offset baseline) are not greater than those to the unrelated object, this may reflect greater overall uncertainty rather than cohort activation specifically. We will return to this when we examine the degree of consideration of each item below.

2.2.3 Markers of lexical processing—Until this point, we have been comparing the timing or degree of fixations within each class of competitors to broadly characterize group differences. However, the process of lexical access is really one of considering multiple competitors simultaneously, ruling some out, and committing to others. To accurately characterize this we need to understand how competitors relate to each other. Thus, our third analysis examined specific moments in the fixation record that relate to key components of this process. We focused on three important time-points, given our understanding of lexical access. First, we determined the earliest point at which participants made more looks to the target or cohort than the unrelated item. This assesses how soon listeners begin to activate items that are consistent with the input. Next, we identified the earliest point at which participants made more fixations to the target than to the cohort. This represents the point at which information in the signal is leading listeners to begin to rule out the cohort in favor of the target. Finally, to examine whether the two groups of listeners were considering competitors to different *degrees*, we examined the fixations to the cohort and rhyme relative to the unrelated item at particular points in the time course, namely at the point of peak competitor fixation and at the participants' average reaction times.

2.2.3.1 Early sensitivity to the signal: Our first measure asked how early listeners can make use of information in the signal to begin accessing lexical candidates. At early portions of the input, target and cohort cannot typically be distinguished, so greater fixations to either of them relative to the unrelated item would constitute evidence for signal-driven commitment to a set of lexical competitors. That is, for each participant, we estimated the time at which the sum of looks to the target and cohort items differed from the sum of looks to the two unrelated items by an absolute value of at least 0.10 for at least 50ms consecutively.[4] We analyzed only TCUU trials as looks to the rhyme typically follow a different time course. CI users disambiguated the target/cohort from the unrelated items later than the normal hearing controls (287 vs. 203 ms[5]; $t(40) = 4.2$, $p < .001$), even though the two groups did not differ significantly in the timing of their first fixation (123 vs. 113 ms; $t(53) = .27$, $p = .79$). That is, while both groups launched eye movements at about the same time, the CI users took longer to fixate an item that was consistent with the auditory signal, indicating that they were slower to activate the relevant lexical item.

2.2.3.2 Competitor suppression: We next asked when participants stopped fixating items that were previously candidates. This assesses listeners' sensitivity to new information in the speech signal and how quickly they can use this information to suppress competitors. We estimated the time at which looks to the target diverged from looks to the cohort. This reflects the point when phonetic information that is inconsistent with the cohort has been processed, the candidate set has been updated, and the listener is now suppressing the cohort. We employed the same 0.1 difference criterion used above and used only the TCUU trials.

By this measure, CI users ceased fixating the cohort competitor later than normal hearing listeners (347 vs. 295 ms; $t(52) = 3.48$, $p = .001$). This later suppression could take two possible forms. First, since CI users were slower to begin using the signal, it is possible that the duration of time over which CI users consider cohort competitors may be similar to the NH group, and it is simply shifted in time due to this delay (e.g., cohorts are active for 200 ms in both groups, but that activation begins later in CI users). Alternatively, this slower suppression may reflect an overall slowing, such that the amount of time over which cohorts are active is longer (e.g., the overall timespan during which cohorts for active is longer in CI users). To disentangle these possibilities, we determined whether CI users were delayed in suppressing competitors *relative to* their delayed onset of activation by measuring the length of time between suppression of the unrelated items relative to the target+cohort and suppression of the cohort (*cohort-duration*).

The two hearing-groups did not differ in cohort-duration (CI users: 60ms, NH listeners: 92ms; $t(52) = 1.42$, $p = .16$). This suggests that CI users were delayed in overall activation, but did not consider cohorts for a longer period of time – they were simply later to activate

[4]These analyses were also conducted using different thresholds (0.25 as well as 0.1) and with both absolute thresholds and thresholds relative to each participant's maximum fixation. Results were unchanged.

[5]These mean values take into account the 100ms silence preceding each word and the 200ms oculomotor delay.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and later to suppress cohorts than NH listeners. The fact that cohort-duration was shorter for the CI users than NH listeners (though not significant) hints at the possibility that while CI users were slower to activate competitors, they may "catch up" over the course of recognition.

2.2.3.3 Degree of Consideration: Finally, we assessed the apparently increased competitor fixations observed in CI users by asking how much competitors were fixated relative to unrelated items. Because fixations to the unrelated item fixations were presumably not phonologically motivated, they serves as a baseline measure of uncertainty or inattention; conversely cohort or rhyme fixations that exceed unrelated fixations indicate these items are competing for lexical activation. To examine this, we first determined specific markers in the time course for each participant (the times when cohort and rhyme fixations reached peak and the time when participants made their selection). Next we computed the proportion of fixations to the competitor and unrelated object at those time points, to establish whether participants fixated the competitor more than the unrelated item at that time.

To estimate the time of peak fixations, we first smoothed the proportion of fixations over time for each participant with a symmetrical 80 ms triangular window separately for cohorts and rhymes. We then found the maximum fixations to those objects, and determined the earliest time at which that peak occurred. Finally, we computed the proportion of fixations to that competitor and the unrelated item at that time. For the end-of-time course analysis, we found each participant's average RT and measured fixations to each competitor item at that point. We used only TCRU trials for this analysis to ensure equal opportunity for looks to each item type.

To examine the degree of consideration of each phonological competitor, the proportions of fixations to cohort and unrelated items *at the time of each participant's peak fixation* (Figure 5A) were submitted to a 2×2 ANOVA examining hearing-group (between-subjects) and word-type (within). Proportions were transformed via the empirical logit. We found a significant effect of word-type ($F(1,53) = 140$, $p < .001$) with more cohorts fixations than unrelated. There was no main effect of hearing-group ($F < 1$), and no interaction ($F < 1$). Thus, at the point of peak cohort fixations, both groups fixated the cohort more than the unrelated item.

A slightly different pattern was observe for rhymes (Figure 5B). A 2×2 ANOVA found a significant effect of word-type on looks ($F(1,53) = 56$, $p < .001$), no effect of hearing-group ($F<1$), but a significant word-type × hearing-group interaction ($F(1,53) = 9.8$, $p = .003$). Follow-up tests showed that both groups made more fixations to the rhyme than to the unrelated item (CI: $t(28) = 6.6$, $p < .001$; NH: $t(25) = 3.9$, $p = .001$). However, CI users made more rhyme fixations than NH participants ($t(53) = 2.7$, $p = .009$), but unrelated fixations were not different ($t(47) = .9$, $p = .38$).

We used a similar approach to examine degree of consideration at the end of processing. Here, NH listeners should show almost no fixations to competitor objects (as they have been ruled out) and any looks by CI users would suggest difficulty ruling out certain classes of competitors. We computed proportions of fixations to the cohort, rhyme and unrelated objects at 200 ms before each subject's average RT. These were submitted to two mixed-design ANOVAs examining word-type (cohort/rhyme vs. unrelated) x hearing-group (CI vs. NH). Looks to the cohort and unrelated items showed a nearly significant effect of word-type ($F(1,53) = 3.9$, $p = .052$), an effect of hearing-group ($F(1,53) = 18.5$, $p < .001$), and no interaction ($F(1,53) = 1.6$ $p = .21$). Individual t-tests revealed that CI users looked more to the cohort than to the unrelated object at their average RT ($t(28) = 2.4$, $p = .03$), but NH participants did not ($t(25) = .5$, $p = .63$, Figure 5C). The ANOVA examining rhymes also showed an effect of word-type ($F(1,53) = 35.7$, $p < .001$), an effect of hearing-group ($F(1,53) = 45.4$, $p < .001$) and an interaction ($F(1,53) = 19.8$, $p < .001$). CI users looked more to the rhyme than to the unrelated object at the average RT ($t(28) = 7.3$, $p < .001$), but the NH group did not ($t(25) = 1.1$, $p = .29$, Figure 5D).

These results taken together provide a picture of lexical competition across the two groups of listeners. When the cohort was at its most active (peak fixations), both groups fixated it than the unrelated item, and did not differ from each other. In contrast, when the rhyme was at its most active, the difference in fixations was greater for CI users than for NH listeners. Moreover at the end of processing, the CI users fixated both the cohort and the rhyme more than the unrelated item, even while clicking on the target, while the NH group fixated neither,.

2.3 Discussion

Overall, adult CI users and NH listeners differed in both the time course and degree of fixations for phonological competitors. CI users were slower to begin fixating targets and to suppress fixations to competitors. At the end of the time course, they also showed fewer looks to the target and more looks to the competitor than NH listeners. Given that we only examined trials in which the listeners clicked the target, this means that on many trials, CI users were clicking the target while they looked at something else. However, at the moments of peak consideration of these lexical competitors, CI users did not show any more fixations to cohort than NH controls, although there was evidence that they considered rhymes more strongly.

One surprising result were small differences in fixations to the unrelated items. The fact that such differences were not observed throughout the whole time course rules out a simple oculomotor account which would predict heightened fixations either early (when there is maximal uncertainty about the target), or uniformly throughout the time course, but not specifically in the second half of the time course. This raises the possibility that CI users fixated non-target objects because they were less certain about the auditory stimulus overall, not because of fundamental differences in the time course of lexical access. However, these were accounted for by the lexical markers analyses so they cannot explain all of our results.

The overall differences between the two groups could be due to a number of factors. The poorer frequency resolution and loss of low-frequency information experienced by CI users may cause in-the-moment uncertainty about specific phonetic cues, resulting in later lexical activation and a somewhat more equal activation across competitors (including the target). In this case, the degraded input would be the proximal cause of quantitative difference in CI users' activation. On the other hand, the source of CI users' differences in lexical activation may lie deeper. CI users may have fundamentally reorganized their speech processing systems, an adaptation that would result in qualitatively different lexical access strategies. For example, they may actively delay their commitment to lexical items to ensure that they have sufficient input to make an accurate decision. Similarly, they may have learned that they are likely to mishear particular phonemes or phonemes at particular positions in a word, and therefore they keep competitors more active in case they need to revise their interpretations. This could account for differential effects on rhymes – if the onset phoneme was misheard, the rhyme and target would contain no further disambiguating material.

To start to tease these two possibilities apart, we conducted a follow-up experiment with NH listeners and CI-simulated speech. If the differences we saw in CI users are due solely to the degraded input, CI-simulated speech should produce a similar pattern of results. On the other hand, if the source of those differences is more fundamental, then we expect CI users and CI simulations to show different time courses of word recognition.

3.0 Experiment 2

3.1 Materials and Methods

The design, visual stimuli, procedure and analysis were identical to Experiment 1. The stimulus type (CI-simulated or normal speech) was a between-participants factor.

3.1.1 Auditory stimuli—The auditory stimuli for the normal-speech condition were identical to those used in Experiment 1. The CI-simulated speech was produced using the Tiger CIS software, which uses the continuous interleaved sampler processing strategy (http://www.tigerspeech.com, version 1.5.02). Though modern CIs have up to 22 electrodes, we simulated an 8-channel CI, as speech recognition performance in CI users seems to asymptote somewhere between 7 and 10 electrodes (Fishman, Shannon, & Slattery, 1997; Friesen et al., 2001). The Tiger CIS software splits the signal into bands and derives an amplitude envelope for each band, then replaces the spectral information in those bands with noise modulated by each band's amplitude envelope. Our 8 channels spanned a frequency range between 200 and 7000 Hz, with corner frequencies based on the Greenwood function (Greenwood, 1990) and a 24 dB/octave slope. The envelope detection used a 400 Hz cutoff frequency and a 24 dB/octave slope. The white noise carrier had the same frequency and slope as the band filter. As in Experiment 1, stimuli were presented over loudspeakers.

3.1.2 Subjects—There were 31 participants, 16 in the simulation (CIS) group and 15 in the normal speech (NS) group. Two additional subjects were excluded from analysis because they did not make sufficient eye movements. Participants were monolingual speakers of American English, reported normal hearing and normal or corrected-to-normal vision, and were between the ages of 18 and 30. Participants received course credit or were paid $30 for their participation.

3.2 Results

3.2.1 Mouse clicks (accuracy & reaction time)—CIS and NS participants made errors on 1.6% of trials (SD = 1.2%; range: 0.3%–4.1%) and 0.4% of trials (SD = 0.3%; range: 0%–1%), respectively ($t(29) = 3.7$, $p = .001$). CIS participants' average RT (on correct trials) was 1490ms (SD = 124; range: 1278–1688), while NS participants' averaged 1278ms (SD = 105; range: 1056–1447; $t(29) = 5.1$, $p < .001$). The degradation of the signal slowed response time and made recognition somewhat more difficult.

3.2.2 Gross time course of fixations—We evaluated group differences using the curve fitting techniques described above. For all object types, we obtained good fits ($R^2 = .981 – .999$), and the fits were equally good in both groups (all $p > .12$). Raw proportion of fixations are shown in Figure 6 and the average fits in Figure 7.

Results of t-tests on the logistic curve fits for the target are shown in Table 5. The CIS group had significantly later crossover points and significantly shallower slopes than the NS group; the two groups did not differ in their maximum levels of fixation.

Table 6 shows comparisons for the asymmetric Gaussian fits for the cohort and rhyme fixations. For looks to the **cohort,** CIS listeners had a significantly shallower onset slope, as well as marginally later midpoints compared to NS listeners. The two groups did not differ significantly in the other parameters. Looks to the **rhyme** and **unrelated** items showed another pattern. In both cases, CIS listeners had significantly higher offset baselines than NS listeners, but did not differ significantly elsewhere.

To summarize, CIS and NS participants differed in only a few parameters, most of which (target slope and crossover, cohort onset slope and midpoint) characterized the time course of considering and ruling out lexical candidates. The CIS group fixated the target and cohort at a slower rate (target and cohort onset slopes) and showed an overall delay (target crossover and cohort midpoint). The peak proportion of fixations to all items was equivalent between the two groups (as was observed in the CI users). However, unlike the CI users, the proportion of fixations to non-target items at the end of the recognition process differed in the rhyme and unrelated items, but not the cohort.

3.2.3 Markers of Lexical Processing—The three markers of lexical processing examined above were also analyzed for the CIS and NS groups.

3.2.3.1 Early sensitivity to the signal: CIS listeners disambiguated the target/cohort from the unrelated items later than the NS listeners (273 vs. 238 ms; $t(26) = 2.5$, $p = .02$). Like the CI analysis in Experiment 1, the two groups made their first fixations at approximately the same time (CIS: 91 vs. NS: 41 ms; $t(29) = 1.8$, $p = .08$). However, it is important to point out the magnitude of the effect (35 ms) was less than what was observed in CI users (84 ms).

3.2.3.2 Competitor suppression: CIS listeners also suppressed fixations to the cohort later than NS listeners (363 vs. 285 ms; $t(29) = 3.8$, $p = .001$), but again it is possible that this derives from an overall delay in the function, rather than a specific difference in suppressing fixations to the competitors. However, as in Experiment 1, there was no difference in cohort duration: for CIS listeners the time between fixating and suppressing the cohort averaged of 90ms, while for NS listeners this was 57ms ($t(29) = 1.6$, $p = .12$).

3.2.3.3 Degree of consideration: Finally, we measured the proportion of fixations to the non-target items at key landmarks in the time course. The degree of fixations to cohort and unrelated items at the time of peak cohort fixation (Figure 8A) was submitted to a 2×2 ANOVA examining hearing-group (between-subjects) and word-type (within). There was a significant effect of word-type ($F(1,29) = 67$, $p < .001$) as cohorts were fixated more than unrelated items. There was no main effect of hearing-group ($F(1,29) = 1.1$, $p = .31$), and no interaction ($F(1,29) = 2.9$, $p = .102$). Similarly, A 2×2 ANOVA comparing looks to the rhyme and unrelated items at peak (Figure 8B) showed an effect of word-type ($F(1,29) = 43$, $p < .001$) and hearing-group ($F(1,29) = 5.5$, $p = .03$), but no interaction ($F(1,29) = 1.0$, $p = .32$). Overall, at the point of maximum rhyme looks, participants looked more to the rhyme than to the unrelated item, and this was true in both groups, though the CIS listeners made more overall fixations to both items than the NS group (rhyme: $t(24) = 2.2$, $p = .04$; unrelated: $t(25) = 2.2$, $p = .04$).

Lastly, we computed the proportions of fixations to the cohort, rhyme and unrelated objects at each participant’s average RT (Figure 8C, D). These were submitted to two ANOVAs examining word-type and hearing-group. The cohort analysis found a significant effect of word-type ($F(1,29) = 12.2$, $p = .002$), but no effect of hearing-group ($F<1$) and no interaction ($F<1$). The rhyme analysis showed an effect of word-type ($F(1,29) = 17.4$, $p < .001$), no effect of hearing-group ($F(1,29) = 1.2$, $p = .29$), and a word-type × hearing-group interaction ($F(1,29) = 9.5$, $p = .005$). This interaction derived from an effect of word-type for the CIS group ($t(15) = 5.1$, $p < .001$), but not for the NS group ($t(14) = .77$, $p = .45$).

In sum, CIS listeners did not differ from controls in fixations at peak cohort activation (the same pattern that we saw in the CI users and their controls). Moreover, the CIS group exceeded the NS controls in all fixations at peak rhyme activation, suggesting increased overall looking by the CIS group but not necessarily greater rhyme activation; in comparison, only the CI users’ rhyme fixations (but not the unrelated fixations) exceeded those of their NH controls at the same point, suggesting greater rhyme activation (relative to the unrelated item). Finally, the CIS listeners fixated both the cohort and rhyme competitors more at RT than the unrelated item, a pattern also shown by the CI users. In contrast, the NS controls fixated the cohort more than the unrelated object at RT, while the NH controls did not fixate either item at that late point.

4. Cross-Experiment Analyses

Thus far, we have compared the CI users to their NH controls and the CIS group to their NS controls, without making direct quantitative comparisons between the two experiments. Because the two experiments were not designed to be compared statistically, there are several differences between them that make a balanced comparison difficult. The participants in Experiment 1 were much older than the participants in Experiment 2, a comparison we discuss further in the online supplement (Note S1). Moreover, the two experiments had different numbers of participants. Nevertheless, it seemed important to determine whether the two experiments had fundamentally different outcomes.

We first examined each of the curve fit parameters with a two-way hearing-group (degraded or normal) by experiment (1 or 2) ANOVA on each of the curve fit parameters (Table 7). Several important findings emerged. First, there is a significant effect of degradation for nearly every parameter. Only the peak height parameters are consistently non-significant or marginal. The rhyme midpoint and offset slope are also marginally significant, and the unrelated offset slope does not differ significantly by group. In general, then, it is clear that hearing degraded speech, whether from a CI or in a simulation, affects the process by which words are recognized in nearly every way, but the peak amount of competitor activation is least affected. Second, there is a significant effect of experiment on several parameters. This is presumably due to differences in age and sample size across the two experiments. Finally and most importantly, there are significant or marginally-significant interactions in only three parameters: target maximum, cohort offset baseline, and the rhyme midpoint. Specifically, the CI users differed from their NH controls these parameters, while the CIS group did not differ from the NS group.

We also examined how much phonological competitors (cohort and rhyme) were fixated relative to the unrelated item as a measure of competition. To determine how different types of degraded speech affected competition, we performed three-way ANOVAs on word-type (competitor vs. unrelated) × hearing-group (degraded vs. non-degraded) × experiment (1 vs. 2) on the degree of consideration data. The most important of these results is the three-way word-type × hearing-group × experiment interaction, which would indicate that the disparity between looks to a competitor and looks to the unrelated item differed based on the form of degradation (CI vs. simulation). Across these analyses this was significant only for rhyme fixations at peak ($F(1,82) = 7.3$, $p = .008$). This interaction is driven by the CI users: while all four groups of participants looked at the rhyme significantly more than the unrelated item at this point, the difference was much greater for the CI users than for the other groups.

The results of this cross-experiment analysis clearly support the assertion that degraded speech has a strong impact on lexical processing. Simply hearing a degraded signal, whether through a CI or through a simulation, influences both the timing of activation and the degree of late activation (Table 7). The only parameters that were consistently unaffected by degraded speech were those describing the peak competitor looks. Degrading the signal apparently has no effect on how much a word is fixated at its peak, though it may affect how much a competitor is activated *relative to the unrelated item*, as in the case of the rhyme.

The evidence that using a CI resulted in different or additional changes in word recognition that could not be explained by the degraded signal alone was more mixed. Within each experiment there were a number of findings that differed. However, in this cross-experiment analysis, we found support for differences in the cohort offset baseline (the retention of cohort fixations), and marginal evidence for a reduction in peak target fixations and a later rhyme peak. In sum, we can argue definitively that hearing a degraded signal slows the course of processing and may affect some competition dynamics, but at the higher standard of this cross-experiment analysis there is not a strong statistical case for differences based on the source of that degradation. However, as we describe next, the qualitative summary in Table 7 and inspection of the figures definitely suggests differences in how the CI users behave relative to their controls and how the CIS group behaves relative to its controls.

5. General discussion

The goals of these Experiments were 1) to characterize the effect of degraded speech on the time course of lexical access in CI users and 2) to tease apart whether these differences are specific to the CI (including the possibility of long-term adaptation) or a more general response to degradation. Table 7 summarizes the results from each experiment; results which differ across the two experiments are in bold. With respect to the first goal, we found that CI users are slower than NH listeners to activate all words, and at the end of processing, they activate target words less (as reflected in the time course of target fixations) and competitor words more (as reflected in the degree of fixation analyses). They show more uncertainty overall, as is evident from increased looks to even the unrelated item, but that alone does not explain their increased fixations to the competitors at the end of the time course.

Experiment 2 presented degraded speech to a group of listeners for whom this type of degradation was novel. NH listeners hearing CI-simulated speech resembled the CI users in some ways. Like the CI users, the CIS group was delayed relative to the NS controls in mapping the signal to lexical candidates, specifically the target and cohort. This result raises the possibility that the delay evidenced by both groups derives from early-stage perceptual processes that are impaired by the degraded stimulus. Interestingly, however, while these difference in timing as a response to degradation were clear across both experiments, signal degradation did not appear to yield an increase in how strongly candidates were active (early in processing). While the former finding seems fairly intuitive, the latter was unexpected. However, it does appear to be consistent with some models of word recognition: simulations with TRACE (see McMurray et al., 2010) manipulating input noise do not show increases in cohort activation in response to noise.

At the same time, a comparison of the results of Experiment 1 and 2 (Table 8) does suggest a number of differences that may be responses to the CI rather than to degradation in general. While these should be evaluated against the mixed statistical evidence in the cross-experiment analysis, they are worth noting. Unlike the CI users, the CIS group did not show decreased target fixations at the end of the time course, and while they revealed some increased late rhyme activation, it was not as pervasive as that of the CI users. There could be two sources for this. First, while simulations can yield similar profiles of behavioral performance (speech recognition) to CIs, they each impose somewhat different degradations to the signal. Second, we speculate that this could derive from a long-term adaptation to degraded speech, and indeed we could see how reducing commitments overall could be an adaptive as it would be less difficult for listeners to revise if subsequent information overrides an early choice (c.f., McMurray, Tanenhaus & Aslin, 2009 for discussion). Future work should strive to flesh out these differences and their causes.

More broadly however, the effects of degradation observed in both groups were not simply quantitative. CI users and the simulation listeners partially violate many of the assumptions about lexical processing from years of research and many experimental paradigms. Whereas it is typically found that NH individuals begin activating words at the earliest sound, both CI users and the CIS group were less immediate in mapping the signal to lexical candidates, suggesting that they needed more of the degraded information to reach an equivalent degree of activation. One component of this may be the somewhat harsh onset of the processed speech that could lead to a short lag before lexical access can begin. Similar studies using words in running speech may help clarify this. Both groups were also delayed in suppressing the cohort competitors, though the time between initially activating the target and suppressing the cohort was no greater for the degraded groups than for the normal controls. This indicates that while activation under a degraded input may have been less immediate, CI users and the CIS group nevertheless updated their candidate set incrementally as they received information in real time—there simply seems to have been a delay in how long it takes for that information to impact lexical access.

In terms of competition and suppression, while the CI users showed sustained activation even late in the time course for both the cohort and the rhyme, the CIS group showed only sustained rhyme activation. The CI users in particular thus appear to violate the assumption that competitors are inactive by the time of lexical access. Moreover, the fact that rhymes appear to be more susceptible to this is particularly interesting given that rhyme activation is typically weaker than cohort activation (Allopenna et al., 1998; Marslen-Wilson, Moss, & Van Halen, 1996) and is sometimes not found at all (e.g., Marslen-Wilson & Zwitserlood, 1989). Our explanation is that while cohorts eventually receive some disambiguating information (e.g., the –al in *sandal* rules out cohorts like *sandwich*), rhymes do not (once a listener has committed to *candle* after hearing *sandal,* there is never any information to rule it out). Thus, if some activation is built early in the time course (perhaps because the CI user mis-hears the first sound of the word, which is not uncommon), there may not be any suppressing information later. Thus, it is perhaps not surprising that both groups maintain rhyme activation to some extent, as the uncertainty created by the degraded input causes it to become more active and there is no further information in the signal to rule it out. However, the fact that CI listeners continue to maintain cohort activation at RT (despite disambiguating information in the signal) suggests that the lack of disambiguating information in the rhyme is not the only explanation.

From a theoretical perspective, these findings show that the theoretical models of lexical access developed to describe word recognition in ideal listening conditions may need to be adapted in systematic ways for characterizing the perception of degraded speech. Specifically, the mechanisms by which listeners activate lexical candidates in parallel may differ when the signal is harder to recognize: they are slower, they suffer from a delay, and the degree of commitment may be modulated by learning. The next step to understanding these differences is to determine what it is about CI-degraded speech that is relevant. Is it the specific way in which this signal is degraded, with the loss of frequency resolution, formant transitions, and low-frequency information? Or is it simply the intelligibility of the signal as a whole? Exploring the real-time parallel activation and competition processes that underlie lexical access in noise, in hard-of-hearing individuals, or in speech degraded in other ways (e.g., sinewave speech: Remez, Rubin, Pisoni, & Carrell, 1981; chimaeric speech: Smith, Delgutte, & Oxenham, 2002) may reveal how the loss of different types of information in the signal affects word recognition and lexical access.

Finally, the degraded signal could affect what types of information are most relevant during word recognition. It is well known that NH listeners use a number of higher-level sources of information such as talker information, visual cues, speech rate, as well as semantic, syntactic and discourse context, to help them recognize words. It seems likely that CI users rely on these higher-level sources of information even more than their NH counterparts, particularly as many of them do not rely on auditory input alone (e.g., Dorman, Loizou & Rainey, 1997; Kaiser, Kirk, Lachs & Pisoni, 2003; Loizou, 1998; Wong, Miyamoto, Pisoni, Sehgal & Hutchins, 1999). This is an important line of future study, and the visual world paradigm is in many ways ideal for this as it has a rich history of looking at the interaction of contextual constraints (e.g., the visual scene, syntax, semantics) on language understanding at multiple levels from speech perception to semantics and pragmatics (Altmann & Kamide, 1999; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004; Dahan & Tanenhaus, 2004; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus et al., 1995). Examining the degree to which CI users rely on information at these higher levels (prosodic, semantic or syntactic cues) relative to their NH counterparts may also help illuminate other compensation mechanisms that CI users develop. In sum, it is clear that people adapt to degraded speech from the neural level up to the level of speech perception—if this adaptation also continues to the level of lexical access and beyond, then studying this process could have both theoretical and clinical benefits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Many thanks to Camille Dunn, Dan McEchron and the members of the MACLab at the University of Iowa for assistance with data gathering, data analysis, and comments and suggestions. Thanks also to the CI Research Team at the University of Iowa as well as the participants and their families. This research was funded in part by grants from the National Institutes of Health to the University of Iowa: DC000242, DC008089, and DC011669.

References

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the timecourse of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language. 1998; 38:419–439.

Altmann GT, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition. 1999; 73(3):247–264. S0010027799000591 [pii]. [PubMed: 10585516]

Apfelbaum K, Blumstein S, McMurray B. Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. Psychonomic Bulletin and Review. 2011; 18:141–149. [PubMed: 21327343]

Balkany T, Hodges A, Menapace C, et al. Nucleus Freedom North American clinical trial. Otolaryngology Head and Neck Surgery. 2007; 136:757–762. [PubMed: 17478211]

Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. Journal of the Acoustical Society of America. 1988; 84(1):101–114. [PubMed: 3411038]

Bronkhorst AW, Bosman AJ, Smoorenburg GF. A model for context effects in speech recognition. Journal of the Acoustical Society of America. 1993; 93(1):499–509. [PubMed: 8423265]

Chambers CG, Tanenhaus MK, Eberhard KM, Filip H, Carlson GN. Circumscribing referential domains during real-time language comprehension. Journal of Memory and Language. 2002; 47:30–49.

Chambers CG, Tanenhaus MK, Magnuson JS. Actions and affordances in syntactic ambiguity resolution. Journal of Experimental Psychology: Learning, Memory and Cognition. 2004; 30:687–696.

Connine CM, Blasko DG, Titone D. Do the beginnngs of spoken words have a special status in auditory word recognition? Journal of Memory and Language. 1993; 32:193–210.

Cullington HE, Zeng FG. Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. Journal of the Acoustical Society of America. 2008; 123:450–461. [PubMed: 18177173]

Dahan D, Gaskell MG. The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. Journal of Memory and Language. 2007; 57:483–501. [PubMed: 18071581]

Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology. 2001; 42:317–367. [PubMed: 11368527]

Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes. 2001; 16:507–534.

Dahan D, Tanenhaus MK. Continuous mapping from sound tomeaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. Journal of Experimental Psychology: Learning, Memory and Cognition. 2004; 30:498–513.

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General. 2005; 134(2):222–241. [PubMed: 15869347]

Desroches AS, Joanisse MF, Robertson EK. Specific phonological impairments in dyslexia revealed by eyetracking. Cognition. 2006; 100:B32–B42. [PubMed: 16288732]

Dorman M, Ketten D. Adaptation by a cochlear-implant patient to upward shifts in the frequency resolution of speech. Ear and Hearing. 2003; 24:457–460. [PubMed: 14534415]

Dorman M, Loizou P. Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with continuous interleaved sampling (CIS) processors. Ear and Hearing. 1997; 18:147–155. [PubMed: 9099564]

Dorman M, Loizou P, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. Journal of the Acoustical Society of America. 1997; 102:2403–2411. [PubMed: 9348698]

Duquesnoy AJ, Plomp R. Effect of reverberation and noise on the intelligibility of sentences in case of presbyacusis. Journal of the Acoustical Society of America. 1980; 68:537–544. [PubMed: 7419810]

Farris-Trimble A, McMurray B. The reliability of eye tracking in the Visual World Paradigm for the study of individual differences in real-time spoken word recognition. Journal of Speech Language and Hearing Research. in press.

Fishman KE, Shannon RV, Slattery WH. 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. 1997; 40:1201–1215.

Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. Journal of the Acoustical Society of America. 2001; 110:1150–1163. [PubMed: 11519582]

Fu QJ, Nogaki G, Galvin JJ. Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. Journal of the Association for Research in Otolaryngology. 2005; 6:180–189. [PubMed: 15952053]

Fu QJ, Shannon RV. Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. Journal of the Acoustical Society of America. 1998; 104:2570–2577. [PubMed: 9821336]

Fu QJ, Shannon RV. Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant. Ear and Hearing. 1999; 20:332–344. [PubMed: 10466569]

Fu QJ, Shannon RV, Galvin JJ. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. Journal of the Acoustical Society of America. 2002; 112:1664–1674. [PubMed: 12398471]

Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. Journal of the Acoustical Society of America. 1998; 104:3586–3596. [PubMed: 9857517]

Giraud AL, Price CJ, Graham JM, et al. Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron. 2001; 30:657–663. [PubMed: 11430800]

Giraud AL, Truy E, Frackowiak R. Imaging plasticity in cochlear implant patients. Audiology and Neurotology. 2001; 6:381–393. [PubMed: 11847465]

Gray RF, Quinn SJ, Court I, Vanat Z, Baguley DM. Patient performance over eighteen months with the Ineraid intracochlear implant. Annals of Otology, Rhinology, and Laryngology Supplement. 1995; 166:272–277.

Greenwood DD. A cochlear implant frequency-position function for several species--29 years later. Journal of the Acoustical Society of America. 1990; 87:2592–2605. [PubMed: 2373794]

Grosjean F. Spoken word recognition processes and the gating paradigm. Perception and Psychophysics. 1980; 28:267–283. [PubMed: 7465310]

Harnsberger JD, Svirsky MA, Kaiser AR, Pisoni DB, Wright R, Meyer TA. Perceptual “vowel spaces” of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. Journal of the Acoustical Society of America. 2001; 109:2135–2145. [PubMed: 11386565]

Hawkins JE, Stevens SS. The masking of pure tones and of speech by white noise. Journal of the Acoustical Society of America. 1950; 22:6–13.

Helms J, Müller J, FS, et al. Evaluation performance with the COMBI 40 Cochlear Implant in adults: A multicentric clinical study. Journal for Oto-Rhino-Laryngology, Head and Neck Surgery. 1997; 59:23–35.

Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP. Perceptual learning of noise vocoded words: Effects of feedback and lexicality. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34(2):460–474. [PubMed: 18377182]

Hervais-Adelman A, Davis MH, Johnsrude IS, Taylor KJ, Carlyon RP. Generalization of perceptual learning of vocoded speech. Journal of Experimental Psychology: Human Perception and Performance. 2011; 37(1):283–295. [PubMed: 21077718]

Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. Journal of Speech, Language, and Hearing Research. 2003; 46:390–404.

Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America. 1977; 61:1337–1351. [PubMed: 881487]

Loizou P, Dorman M, Tu Z. On the number of channels needed to understand speech. Journal of the Acoustical Society of America. 1999; 106:2097–2103. [PubMed: 10530032]

Loizou P. Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants. IEEE Signal Processing Magazine. 1998; 15:101–130.

Luce PA, Cluff MS. Delayed communication in spoken word recognition: Evidence from cross-modal priming. Attention, Perception, & Psychophysics. 1998; 60(3):484–490.

Luce PA, Goldinger SD, Auer ET, Vitevitch MS. Phonetic priming, neighborhood activation, and PARSYN. Perception and Psychophysics. 2000; 62:615–625. [PubMed: 10909252]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. [PubMed: 9504270]

Marslen-Wilson W. Linguistic structure and speech shadowing at very short latencies. Nature. 1973; 244:522–523. [PubMed: 4621131]

Marslen-Wilson W. Functional parallelism in spoken word recognition. Cognition. 1987; 25:71–102. [PubMed: 3581730]

Marslen-Wilson, W. Activation, competition, and frequency in lexical access. In: Altmann, GTM., editor. Cognitive Models of Speech Processing. Cambridge, MA: MIT Press; 1990. p. 148-172.

Marslen-Wilson W, Moss HE, Van Halen S. Perceptual distance and competition in lexical access. Journal of Experimental Psychology: Human Perception and Performance. 1996; 22:1376–1392. [PubMed: 8953227]

Marslen-Wilson W, Tyler LK. The temporal structure of spoken language understanding. Cognition. 1980; 8:1–71. [PubMed: 7363578]

Marslen-Wilson W, Zwitserlood P. Accessing spoken words: The importance of word onsets. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:576–585.

McClelland JL, Elman J. The TRACE model of speech perception. Cognitive Psychology. 1986; 18:1–86. [PubMed: 3753912]

McMurray B, Clayards M, Tanenhaus MK, Aslin RN. Tracking the time course of phonetic cue integration during spoken word recognition. Psychonomic Bulletin and Review. 2008; 15:1064–1071. [PubMed: 19001568]

McMurray B, Samelson VM, Lee SH, Tomblin JB. Individual differences in online spoken word recognition: Implications for SLI. Cognitive Psychology. 2010; 60:1–39. [PubMed: 19836014]

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition. 2002; 86:B33–B42. [PubMed: 12435537]

McMurray B, Tanenhaus MK, Aslin RN. Within-category VOT affects recovery from “lexical” garden paths: Evidence against phoneme-level inhibition. Journal of Memory and Language. 2009; 60:65–91. [PubMed: 20046217]

Nelson PB, Jin SH, Carney AE, Nelson DA. Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. Journal of the Acoustical Society of America. 2003; 113:961–968. [PubMed: 12597189]

Niparko, J. Cochlear Implants: Principles and Practices. 2. Philadephia: Lippincott Williams & Wilkins; 2009.

Nittrouer S, Lowenstein JH. Learning to perceptually organize speech signals in native fashion. Journal of the Acoustical Society of America. 2010; 127:1624–1635. [PubMed: 20329861]

Nittrouer S, Lowenstein JH. Separating the effects of acoustic and phonetic factors in linguistic processing with impoverished signals by adults and children. Applied Psycholinguistics. 201210.1017/S0142716412000410

Nittrouer S, Lowenstein JH, Packer RR. Children discover the spectral skeletons in their native language before the amplitude envelopes. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35:1245–1253. [PubMed: 19653762]

Norris D. Shortlist: A connectionist model of continuous speech recognition. Cognition. 1994; 52:189–234.

Norris D, McQueen JM. Shortlist B: A Baysian model of continuous speech recognition. Psychological Review. 2008; 115:357–395. [PubMed: 18426294]

Pelizzone M, Cosendai G, Tinembart J. Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for Ineraid implanted subjects. Ear and Hearing. 1999; 20:228–237. [PubMed: 10386849]

Perkell J, Lane H, Svirsky MA, et al. Speechof cochlear implant patients: A longitudinal study of vowel production. Journal of the Acoustical Society of America. 1992; 91:2961–2978. [PubMed: 1629489]

Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. Journal of the Acoustical Society of America. 1995; 97:593–608. [PubMed: 7860836]

Qin MK, Oxenham AJ. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. Journal of the Acoustical Society of America. 2003; 114:446–454. [PubMed: 12880055]

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science. 1981; 212:947–950. [PubMed: 7233191]

Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. Journal of the Acoustical Society of America. 1999; 106:3629–3636. [PubMed: 10615701]

Sedivy JC, Tanenhaus MK, Chambers CG, Carlson GN. Achieving incremental processing through contextual representation: Evidence from the processing of adjectives. Cognition. 1999; 71:109–147. [PubMed: 10444906]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Smith Z, Delgutte B, Oxenham A. Chimaeric sounds reveal dichotomies in auditory perception. Nature. 2002; 416:87–90. [PubMed: 11882898]

Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal hearing and cochlear implant listeners I: The effects of response format. Ear and Hearing. 1997; 18:89–99. [PubMed: 9099558]

Spivey MJ, Grosjean M, Knoblich G. Continuous attraction toward phonological competitors. Proc Natl Acad Sci U S A. 2005; 102:10393–10398. [PubMed: 15985550]

Stickney GS, Zeng F-G, Litovsky R, Assmann P. Cochlear implant speech recognition with speech maskers. Journal of the Acoustical Society of America. 2004; 116:1081–1091. [PubMed: 15376674]

Svirsky MA, Silveira A, Suarez H, Neuburger H, Lai TT, Simmons PM. Auditory learning and adaptation after cochlear implantation: A preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. Acta Otolaryngol. 2001; 121:262–265. [PubMed: 11349792]

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634. [PubMed: 7777863]

Throckmorton CS, Collins LM. The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model. Journal of the Acoustical Society of America. 2002; 112:285–296. [PubMed: 12141354]

Tyler LK. The structure of the initial cohort: Evidence from gating. Perception and Psychophysics. 1984; 36:417–427. [PubMed: 6533566]

Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. Journal of the Acoustical Society of America. 2004; 115:1729–1735. [PubMed: 15101651]

Tyler RS, Parkinson AJ, Woodworth GG, Lowder MW, Gantz BJ. Performance over time of adult patients using the Ineraid or Nucleus cochlear implant. Journal of the Acoustical Society of America. 1997; 102:508–522. [PubMed: 9228814]

Wong D, Miyamoto RT, Pisoni DB, Sehgal M, Hutchins GD. PET imaging of cochlear-implant and normal-hearing subjects listening to speech and nonspeech. Hearing Research. 1999; 132:34–42. [PubMed: 10392545]

Yee E, Blumstein SE, Sedivy JC. Lexical-semantic activation in Broca's and Wernicke's Aphasia: Evidence from eye movements. Journal of Cognitive Neuroscience. 2008; 20:592–612. [PubMed: 18052783]

Zwitserlood P. The locus of the effects of sentential-semantic context in spoken-word processing. Cognition. 1989; 32:25–64. [PubMed: 2752705]

Zwitserlood P, Schriefers H. Effects of sensory information and processing time in spoken-word recognition. Language and Cognitive Processes. 1995; 10:121–136.

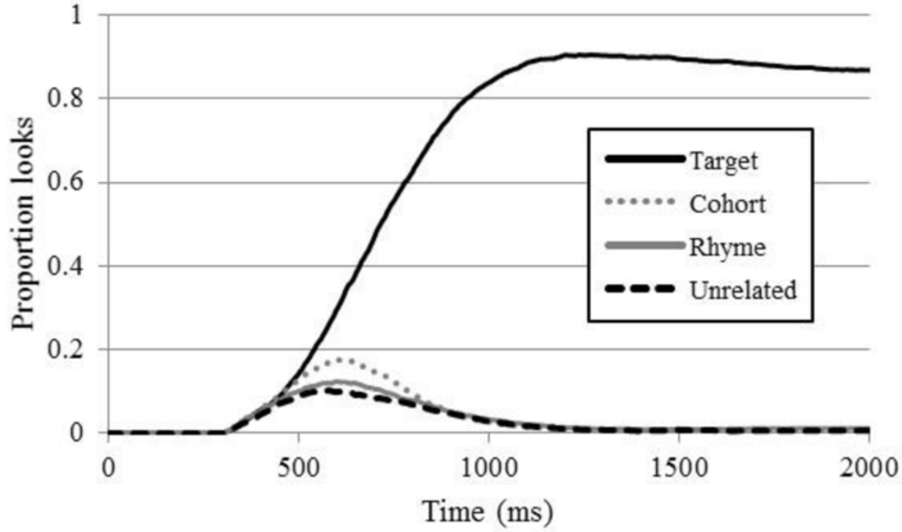

Figure 1.
Proportion fixations by normal hearing listeners to each word type.

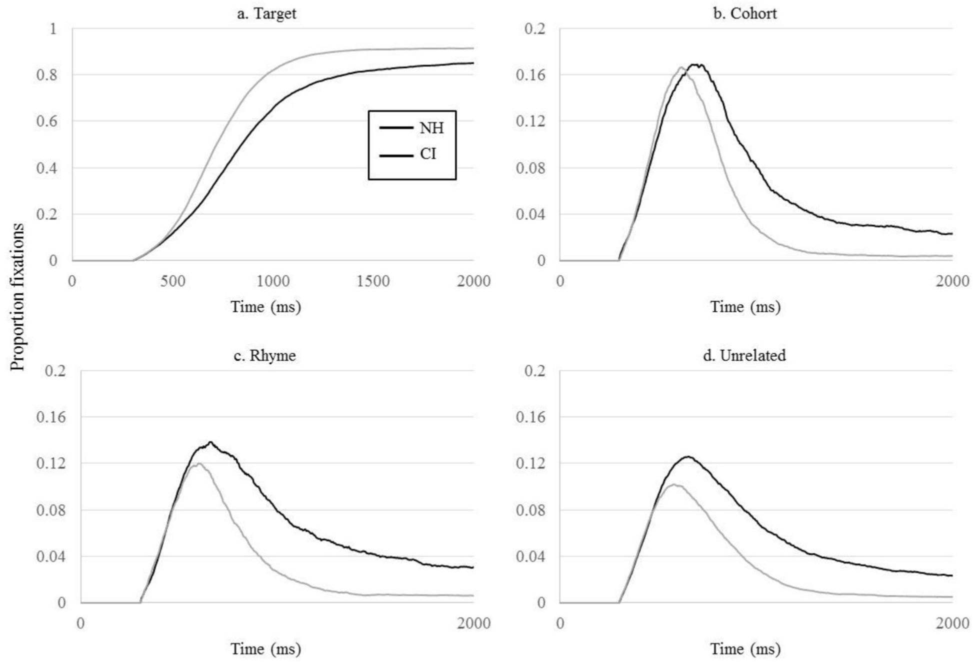

Figure 2.

Proportion fixations by the NH and CI groups as a function of time. A) Fixations to the target (in all trials); B) cohort (TCRU and TC trials); C) rhyme (TCRU and TR trials); D) unrelated (all trials). *N.B.* Y-axis for competitor activation is on a different scale than for target activation.

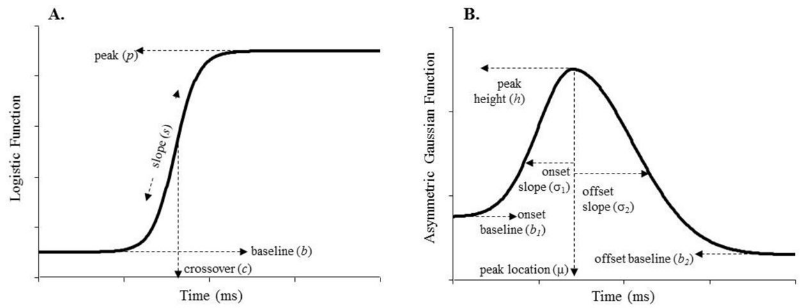

Figure 3.
Parameters of the logistic (A) and asymmetric Gaussian (B) functions used to fit target and competitor fixations (respectively).

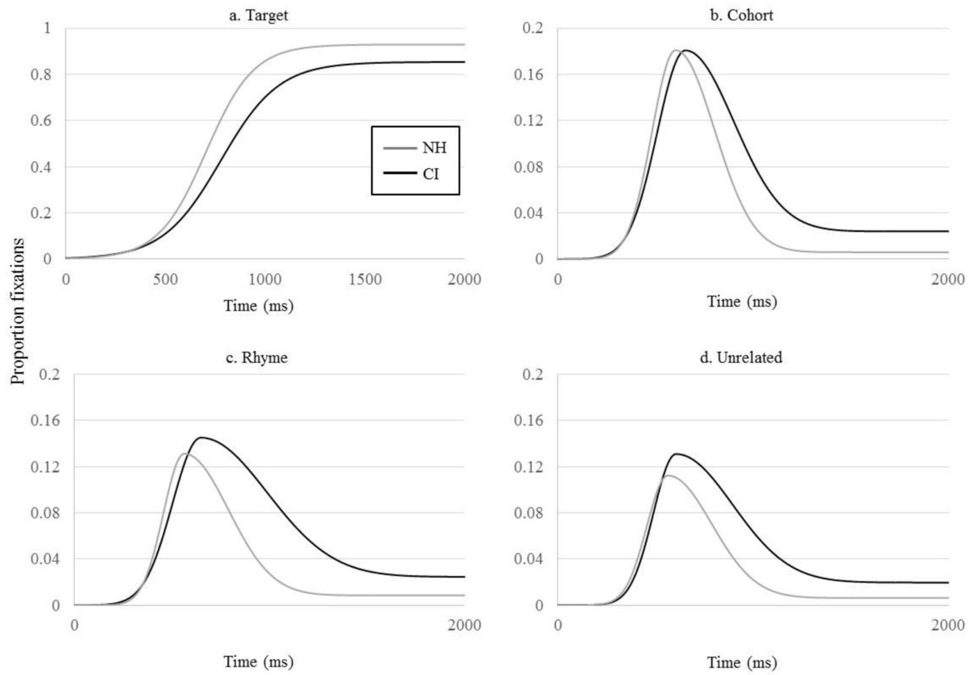

Figure 4.
Average curve fit for the NH and CI groups as a function of time. A) Fixations to the target; B) cohort; C) rhyme; D) unrelated. *N.B.* Y-axis for competitor activation is on a different scale than for target activation.

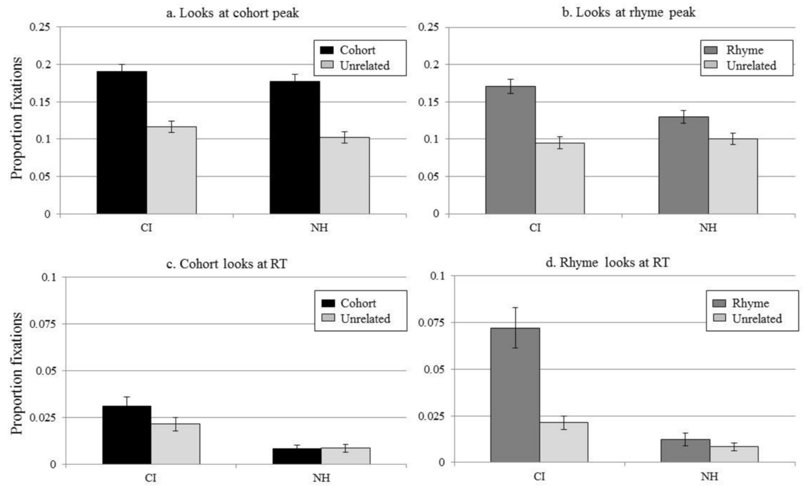

Figure 5.

Proportion fixations by the CI and NH groups to A) the cohort and unrelated items at time of peak cohort fixation. B) the rhyme and unrelated items at time of peak rhyme fixation. C) the cohort and unrelated items at average reaction time. D) the rhyme and unrelated items at average reaction time.

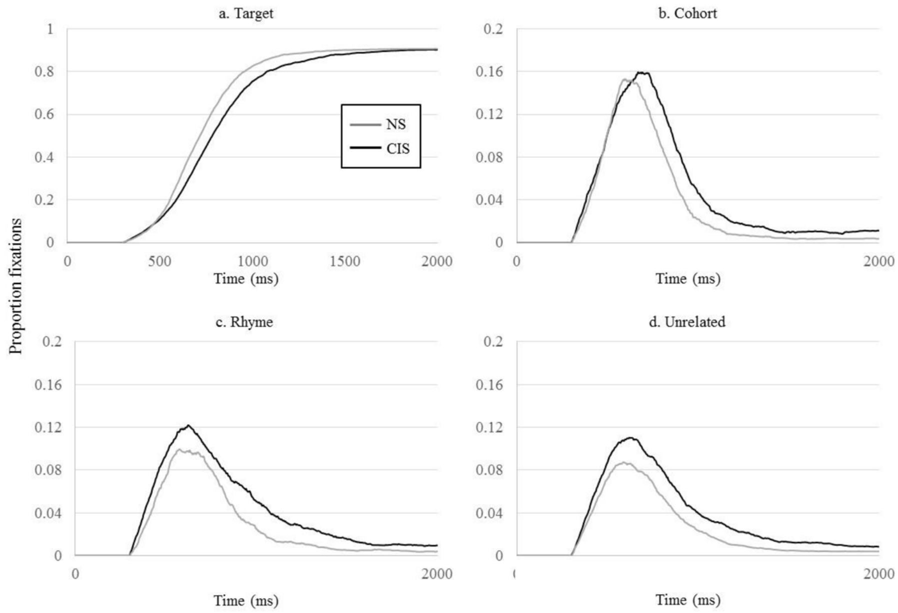

Figure 6.

Proportion fixations by the NS and CIS groups as a function of time. A) Fixations to the target (in all trials); B) cohort (TCRU and TC trials); C) rhyme (TCRU and TR trials); D) unrelated (all trials). *N.B.* Y-axis for competitor activation is on a different scale than for target activation.

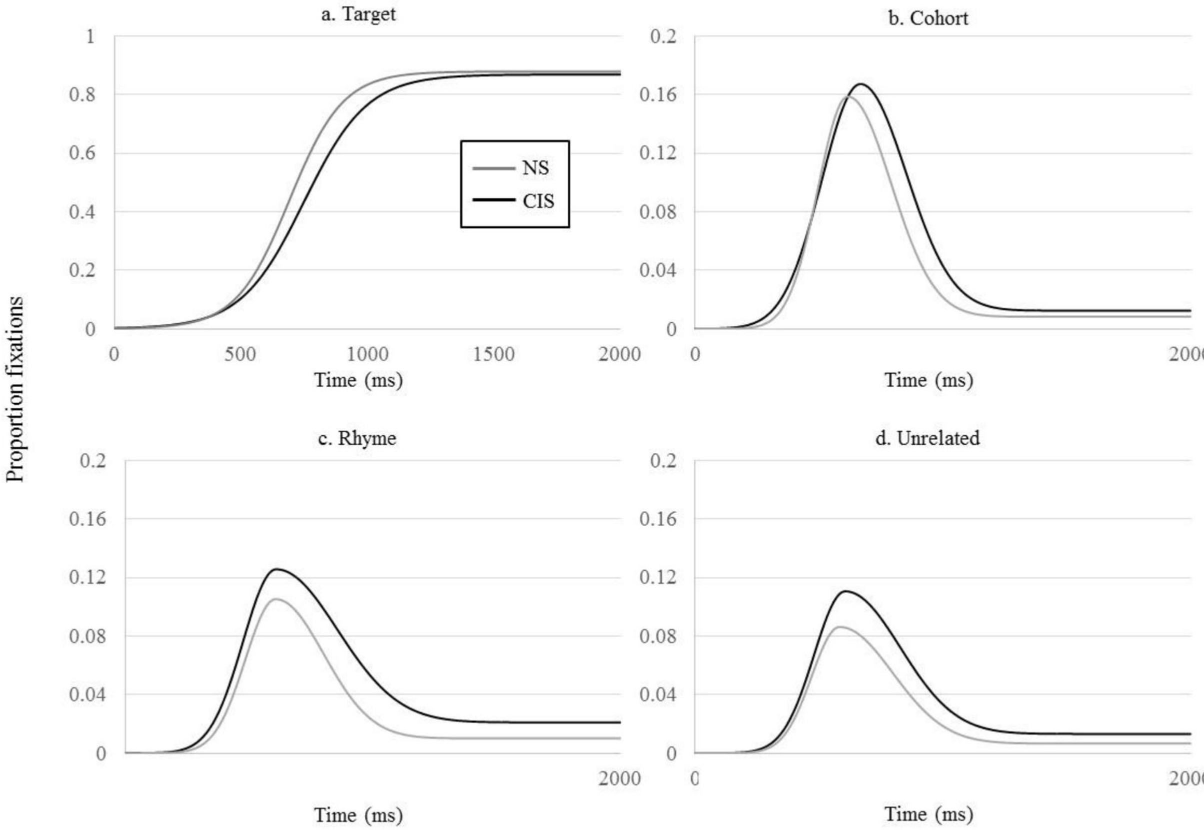

Figure 7.

Average curve fit for the NS and CIS groups as a function of time. A) Fixations to the target; B) cohort; C) rhyme; D) unrelated. *N.B.* Y-axis for competitor activation is on a different scale than for target activation.

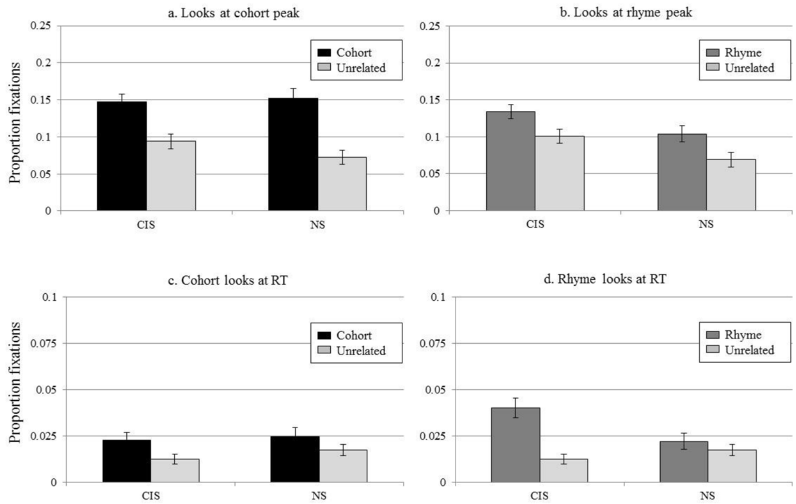

Figure 8.

Proportion fixations by the CIS and NS groups to A) the cohort and unrelated items at time of peak cohort fixation. B) the rhyme and unrelated items at time of peak rhyme fixation. C) the cohort and unrelated items at average reaction time. D) the rhyme and unrelated items at average reaction time.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Role of Word by Trial Type.

		Auditory stimulus	Visual stimuli			
			wizard	**whistle**	**lizard**	**baggage**
Trial Type	**TCRU**	**base (*wizard*)**	**target**	cohort	rhyme	unrelated
	TCUU	**cohort (*whistle*)**	cohort	**target**	unrelated	unrelated
	TRUU	**rhyme (*lizard*)**	rhyme	unrelated	**target**	unrelated
	TUUU	**unrelated (*baggage*)**	unrelated	unrelated	unrelated	**target**

T=target, C=cohort, R=rhyme, U=unrelated

Table 2

Demographic Characteristics of the CI Group

Subj. #	Age (years)	Etiology	Age at onset of deafness	Age at implantation	Implant type[a]	Implant manufacturer[b]	lPTA	rPTA
2	65	unknown	9	59	uni	AB	118	93
3	41	unknown	30	33	uni	AB	87	118
4	55	noise exposure	unknown	53	bi-sim	AB	73	72
6	63	unknown	52	57	bi-sim	AB	75	115
10	54	hereditary	1	51	hy-8	C	75	75
16	81	infection	59	69	uni	C	118	117
17	56	Meniere's disease	39	40	uni	AB	82	77
19	59	unknown	unknown	51	hy-8	C	85	83
22	70	unknown	51 (L); 47 (R)	69	bi-sim	AB	88	87
23	66	unknown	57	59	bi-seq	AB	65	95
24	49	unknown	40	43	bi-seq	AB	102	102
29	67	noise exposure	57	63	bi-sim	AB	77	68
30	57	other	38	39	bi-seq	AB	107	90
41	42	unknown	unknown	32	hy-8	C	80	88
48	76	infection (R); unknown (L)	73	74	bi-sim	AB	113	115
72	64	unknown	unknown	63	hy-12	C	70	68
74	64	unknown	unknown	63	hy-12	C	n/a	n/a
75	74	hereditary	61	70	bi-sim	AB	92	87
76	65	otosclerosis	unknown	58	bi-seq	C	118	97
77	35	hereditary	26	26	bi-sim	AB	92	98
79	50	unknown	20	48	bi-sim	C	118	118
81	61	hereditary	55	57	bi-sim	C	107	102
82	45	unknown	32	45	hy-12	C	n/a	n/a
83	67	hereditary	53	59	bi-sim	AB	102	103
84	73	unknown	66	67	bi-sim	AB	70	63
86	44	unknown	24	26	bi-seq	AB	118	103
90	63	unknown	unknown	61	hy-12	C	n/a	n/a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Subj. #	Age (years)	Etiology	Age at onset of deafness	Age at implantation	Implant type[a]	Implant manufacturer[b]	lPTA	rPTA
94	63	unknown	unknown	61	hy-12	C	n/a	n/a
99	65	Autoimmune Sensorineural Loss	unknown	57	hy-8 (L); uni (R)	C	70	70

Notes:

[a]Implant type: uni = unilateral; bi-sim = simultaneous bilateral; bi-seq = sequential bilateral; hy-12 = 10-electrode hybrid; hy-8 = 8-electrode hybrid;

[b]Implant manufacturer: AB = Advanced Bionics; C = Cochlear

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Comparison of Parameters Describing the Time Course of Target Fixations in the Cochlear Implant (CI) group and the Age-Matched Normal Hearing (NH) Group

	M (SD)				
	CI	**NH**	**df**	**t**[a]	**p**
Maximum (p[fix])[b,c]	.85 (.113)	.93 (.064)	53	4.1	<.001
Crossover (ms)[d]	777 (58)	703 (40)	50	5.6	<.001
Slope (Δp[fix]/ms)[e]	.0015 (.0004)	.0019 (.0003)	53	5.0	<.001

Note:

[a]T-tests assume unequal variances;

[b]Proportion of total fixations;

[c]Target maximum averages and standard deviations are not transformed so as to show meaningful measures. The statistics are performed on the transformed values;

[d]Milliseconds;

[e]Change in proportion of fixations over time in milliseconds

Table 4

Parameters Describing the Time Course of Competitor Fixations in the Cochlear Implant (CI) Group and the Age-Matched Normal Hearing (NH) Group

		M (SD)				
	Parameter	**CI**	**NH**	**df**	**t**[a]	**p**[b]
Cohort	Onset slope (ms)[c]	144 (27)	122 (20)	52	3.5	.001
	Midpoint (ms)	652 (72)	603 (50)	50	3.0	.004
	Peak height (p[fix])[d,e]	.18 (.05)	.18 (.05)	53	0.1	
	Offset slope (ms)	249 (70)	199 (54)	52	7.0	.005
	Offset baseline (p[fix])	.024 (.013)	.006 (.004)	52	8.7	<.001
Rhyme	Onset slope	146 (55)	107 (18)	34	3.6	.001
	Midpoint	649 (140)	563 (50)	36	3.1	.004
	Peak height	.15 (.04)	.13 (.04)	53	1.0	
	Offset slope	339 (309)	226 (42)	29	2.0	.060
	Offset baseline	.025 (.015)	.008 (.005)	53	4.6	<.001
Unrelated	Onset slope	122 (29)	110 (29)	52	1.5	
	Midpoint	610 (83)	562 (62)	51	2.5	.017
	Peak height	0.138 (0.040)	0.119 (0.035)	50	1.1	
	Offset slope	274 (103)	240 (58)	45	1.5	
	Offset baseline	0.023 (0.012)	0.006 (0.004)	39	6.3	< .001

Notes:

[a] T-tests assume unequal variances;

[b] P values not shown are $p > .1$;

[c] Milliseconds: the competitor onset and offset slope measurements correspond to the standard deviation (σ) or width of a Gaussian and are therefore represented in ms;

[d] Proportion of total fixations;

[e] Competitor peak height and offset baseline averages and standard deviations are not transformed so as to show meaningful measures. The statistics are performed on the transformed values.

Table 5

Comparison of Parameters Describing the Time Course of Target Fixations in the CI Simulation Group (CIS) and the Normal Speech Group (NS)

	M (SD)				
	CIS	**NS**	**df**	**t[a]**	**p[b]**
Maximum (p[fix])[c,d]	.87 (.08)	.88 (.08)	24	0.8	
Crossover (ms)[e]	750 (40)	692 (37)	29	4.2	<.001
Slope (Δp[fix]/ms)[f]	.0017 (.0004)	.0021 (.0004)	29	2.5	.02

Notes:

[a] T-tests assume unequal variances;

[b] P values not shown are $p > .2$;

[c] Proportion of total fixations;

[d] Target maximum averages and standard deviations are not transformed so as to show meaningful measures. The statistics are performed on the transformed values;

[e] Milliseconds;

[f] Change in proportion of fixations over time in milliseconds

Table 6

Parameters Describing the Time Course of Competitor Fixations in the CI Simulation Group (CIS) and the Normal Speech Group (NS)

		M (SD)				
	Parameter	**CIS**	**NS**	**df**	**t[a]**	**p[b]**
Cohort	Onset slope (ms)[c]	156 (50)	120 (21)	20	2.7	.015
	Midpoint (ms)	668 (89)	617 (53)	25	2.0	.062
	Peak height (p[fix])[d,e]	.167 (.034)	.159 (.049)	24	0.7	
	Offset slope (ms)	185 (61)	175 (30)	22	0.6	
	Offset baseline (p[fix])	.012 (.010)	.008 (.007)	25	1.7	
Rhyme	Onset slope	134 (52)	123 (37)	27	0.7	
	Midpoint	610 (111)	608 (67)	25	0.1	
	Peak height	.126 (.04)	.105 (.04)	27	1.4	
	Offset slope	247 (113)	194 (49)	21	1.7	
	Offset baseline	.021 (.013)	.010 (.009)	27	2.5	.018
Unrelated	Onset slope	133 (41)	116 (24)	25	1.4	
	Midpoint	605 (92)	584 (61)	26	0.7	
	Peak height	0.115 (0.040)	0.094 (0.039)	28	1.4	
	Offset slope	222 (46)	215 (78)	22	0.3	
	Offset baseline	0.019 (0.008)	0.007 (0.008)	14	3.2	.006

Notes:

[a]T-tests assume unequal variances;

[b]P values not shown are $p > .1$;

[c]Milliseconds: the competitor onset and offset slope measurements correspond to the standard deviation (σ) or width of a Gaussian and are therefore represented in ms;

[d]Proportion of total fixations;

[e]Competitor peak height and offset baseline averages and standard deviations are not transformed so as to show meaningful measures. The statistics are performed on the transformed values.

Table 7

The Effects of Group and Experiment on the Curve fit Parameters for Experiment 1 and Experiment 2

	Effect of group		Effect of experiment		Interaction	
	F (1,82)	p[a]	F (1,82)	p	F (1,82)	p
Target maximum[b]	9.9	.002	2.8	.096	3.4	.068
Target crossover	39.8	<.001	3.3	.072	<1	
Target slope	24.6	<.001	5.9	.017	<1	
Cohort onset slope	18.8	<.001	<1		1	
Cohort midpoint	11.2	0.001	<1		<1	
Cohort peak height	<1		2.1		<1	
Cohort offset slope	5.2	0.025	11.3	0.001	2.2	
Cohort offset baseline	35	<.001	5.1	0.026	6.3	0.014
Rhyme onset slope	6.6	.012	<1		2.1	
Rhyme midpoint	3.6	.061	<1		3.3	.074
Rhyme peak height	2.9	.094	4.5	.037	<1	
Rhyme offset slope	3.8	.054	2.1		<1	
Rhyme offset baseline	23.0	<.001	<1		<1	
Unrelated onset slope	4.4	.039	1.6		<1	
Unrelated midpoint	4.1	.046	<1		<1	
Unrelated peak height	3.1	.083	1.2	.028	<1	
Unrelated offset slope	<1		4.8	.031	<1	
Unrelated offset baseline	45.1	< .001	<1		<1	

Notes:

[a] P values not shown are $p > .1$;

[b] Statistics are performed on transformed values of the target maximum, the competitor peak heights and the competitor offset baselines.

Table 8

Summary of Qualitative Differences Between Experiment 1 and Experiment 2

	Experiment 1	Experiment 2
Curvefits	• CI users delayed relative to NH controls **in fixations to all items** • **CI users have fewer target fixations than NH controls** • CI users have more late fixations **to all competitors** than NH controls	• CIS group delayed relative to NS controls **only in target and cohort fixations** • **CIS group and NS controls have equal target fixations** • CIS group has more late **rhyme and unrelated fixations** than NS controls
Early sensitivity to the signal	• CI users slower to activate relevant item than NH controls	• CIS group slower to activate relevant item than NS controls
Competitor suppression	• CI users and NH controls equally fast to suppress competitor once it had been activated	• CIS group and NS controls equally fast to suppress competitor once it had been activated
Degree of consideration	*Peak* • Both groups fixated cohort more than unrelated • Both groups fixated rhyme more than unrelated, **but CI users showed more rhyme activation than the NH group** *RT* • **CI users showed late cohort activation, NH group did not** • CI users showed late rhyme activation, NH group did not	*Peak* • Both groups fixated cohort more than unrelated • Both groups fixated rhyme more than unrelated, **but the groups were equal in their degree of rhyme activation** *RT* • **Both groups showed late cohort activation equally** • CIS group showed late rhyme activation, NS group did not

Appendix A

Stimuli

Target	Cohort	Rhyme	Unrelated
batter	baggage	ladder	monkey
bees	beach	peas	cap
Bell	bed	well	can
berry	barrel	fairy	cannon
bowl	bone	pole	nest
Bug	bus	rug	cane
carrot	carriage	parrot	building
Cat	cab	bat	net
chips	chin	lips	boat
coat	comb	goat	badge
dollar	dolphin	collar	hamster
Fish	fist	dish	belt
ghost	goal	toast	bag
horn	horse	corn	box
letter	lettuce	sweater	turkey
money	mother	honey	wagon
mountain	mousetrap	fountain	window
mouse	mouth	house	chain
paddle	package	saddle	waiter
pickle	picture	nickel	donkey
plate	plane	gate	dress
Rake	race	lake	soup
road	roll	toad	cake
rocket	robin	pocket	castle
Rose	robe	hose	band
sandal	sandwich	candle	necklace
snail	snake	pail	web
tower	towel	shower	penguin
wizard	whistle	lizard	bottle