How do you deal with uncertainty? Cochlear Implant users differ in the dynamics of lexical processing of non-canonical inputs.

Bob McMurray,

Dept. of Psychological and Brain Sciences, Dept. of Communication Sciences and Disorders, Dept. of Otolaryngology, University of Iowa

Tyler P. Ellis, and

Dept. of Communication Sciences and Disorders, University of Iowa

Keith S. Apfelbaum

Dept. of Psychological and Brain Sciences, University of Iowa, and Foundations in Learning, Inc.

Abstract

Objectives.—Work in normal hearing (NH) adults suggests that spoken language processing involves coping with ambiguity. Even a clearly spoken word contains brief periods of ambiguity as it unfolds over time, and early portions will not be sufficient to uniquely identify the word. However, beyond this temporary ambiguity NH listeners must also cope with loss of information due to reduced forms, dialect, and other factors. Recent work suggests that NH listeners may adapt to increased ambiguity by changing the dynamics of how they commit to candidates at a lexical level. Cochlear Implant (CI) users must also frequently deal with highly degraded input, in which there is less information available in the input to recover a target word. We asked here whether their frequent experience with this leads to lexical dynamics that are better suited for coping with uncertainty.

Design.—Listeners heard words either correctly pronounced (*dog*), or mispronounced at onset (*gog*) or offset *(dob)*. Listeners selected the corresponding picture from a screen containing pictures of the target and three unrelated items. While they did this, fixations to each object were tracked as a measure of the timecourse of identifying the target. We tested 44 post-lingually deafened adult CI users in two groups (23 used standard electric only configurations, and 21 supplemented the CI with a hearing aid), along with 28 age matched age-typical hearing (ATH) controls.

Results.—All three groups recognized the target word accurately, though each showed a small decrement for mispronounced forms (larger in both types of CI users). Analysis of fixations showed a close time locking to the timing of the mispronunciation. Onset mispronunciations delayed initial fixations to the target, but fixations to the target showed partial recovery by the end of the trial. Offset mispronunciations showed no effect early, but suppressed looking later. This pattern was attested in all three groups, though both types of CI users were slower and did not commit fully to the target. When we quantified the degree of disruption (by the mispronounced

---

Corresponding Author Bob McMurray, W314 SSH, Dept. of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, 319-335-2408 (voice), Bob-mcmurray@uiowa.edu.

forms) we found that both groups of CI users showed less disruption than ATH listeners during the first 900 msec of processing. Finally, an individual differences analysis showed that *within the CI users*, the dynamics of fixations predicted speech perception outcomes over and above accuracy in this task, and that CI users with the more rapid fixation patterns of ATH listeners showed better outcomes.

Conclusions.—Post-lingually deafened CI users process speech incrementally (as do ATH listeners), though they commit more slowly and less strongly to a single item than do ATH listeners. This may allow them to cope more flexibly with mispronunciations.

Keywords

Cochlear Implants; Lexical Access; Eye-tracking; Lexical Competition; Mispronunciations

INTRODUCTION

Accurate speech perception requires listeners to confront variability. Even normal hearing (NH) listeners must deal with the fact that the acoustic form of a word varies with changes in talker (e.g., Ladefoged & Broadbent, 1957), speaking rate (e.g., Summerfield, 1981) and phonetic context (e.g., Daniloff & Moll, 1968). Moreover, words are not always pronounced in full citation form, appearing as variants like *I dunno, p'lice* or *samwich* (Ernestus & Warner, 2011; Gow, 2003; Ranbom & Connine, 2007). Not all of this variation is lawful – the same words pronounced in the same laboratory session take markedly different forms (Newman, Clouse, & Burnham, 2001), and 40% of the variance in phonetic measurements is not explained by lawful causes like talker, and coarticulation (McMurray & Jongman, 2011). Thus, speech perception must be flexible and cope with variation and noise.

People who use cochlear implants (CIs) face an extreme version of this problem. While CIs restore substantial hearing, they poorly approximate the acoustic signal. While Normal Hearing (NH) listeners discriminate thousands of frequencies, a typical CI encodes envelopes within overlapping frequency bands corresponding to about 20 electrodes (and typical CI users see a functional benefit from fewer: Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001; Mehr, Turner, & Parkinson, 2001). Information below about 500 Hz (e.g., fundamental frequency) can be lost, and the periodic vocal signal is replaced with pulsing electrical stimulation. Thus, CIs do not preserve many subtle speech cues, like dynamically changing formant frequencies. Consequently, in addition to the variabilty faced by all lisetners, CI users may also face increased uncertainty due to a loss of information in the input.

Despite this impoverished input, post-lingually deafened adult CI users are generally successful at speech perception, particularly in good listening conditions (Dorman, Hannley, Dankowski, Smith, & McCandless, 1989; Fishman et al., 1997; Stickney, Zeng, Litovsky, & Assmann, 2004; Tyler, Lowder, Parkinson, Woodworth, & Gantz, 1995) (for a review, see Wilson & Dorman, 2008). However, performance is not always robust, breaking down in challenging conditions or tasks (Balkany et al., 2007; Fetterman & Domico, 2002; Friesen et al., 2001; Fu, Shannon, & Wang, 1998; Helms et al., 1997; Stickney et al., 2004). Nonetheless, many CI users report satisfaction and gains in quality of life (Hirschfelder,

Gräbel, & Olze, 2008; Lassaletta, Castro, Bastarrica, Sarriá, & Gavilán, 2005). This suggests they may employ compensatory mechanisms to cope with their impoverished signal.

It is often assumed that the primary route to better CI outcomes is better signal quality (e.g., frequency separation). Speech perception outcomes are predicted by factors like insertion depth, electrode design, the number of electrodes, processing strategy and residual acoustic hearing (Henry, Turner, & Behrens, 2005; Neher, Laugesen, Søgaard Jensen, & Kragelund, 2011; Summers & Leek, 1994), as well as frequency encoding precision (Henry et al., 2005; Scheperle & Abbas, 2015). However, non-audiological factors like cognitive ability (general intelligence: Holden et al., 2013), CI experience (Hamzavi, Baumgartner, Pok, Franz, & Gstoettner, 2003; Oh et al., 2003) and cortical function (Green, Julyan, Hastings, & Ramsden, 2005) also predict outcomes, pointing to central adaptations. However it is unclear *how* these cognitive factors lead to better outcomes.

Some insight may be gained by examining how NH listeners cope with common forms of ambiguity or information loss. As we describe shortly, basic research in NH listeners has documented that even with a perfect signal, speech is always partially or temporarily ambiguous (Bard, Shillcock, & Altmann, 1988; Marslen-Wilson, 1987). This ambiguity requires cognitive processes (such as graded activation and competition among lexical candidates) that manage uncertainty. This stands in contrast to a pure focus on signal quality, as it highlights the fundamental role of central processes. Deeper understanding of how these cognitive mechanisms are deployed (or adapted) in hearing-impaired listeners may improve assessment and outcomes.

Non-Canonical Forms

Any production of a word may not match the idealized (canonical) form stored in the lexicon for many reasons. Many non-canonical productions are lawful. For example, place assimilation causes coronals like /n, t, d/ to adopt the place of articulation of neighboring segments (e.g., *green boat* → *greem boat*). While assimilated forms (*greem*) do not match the intended target, the regularity of this process allows perceptual processes to "undo" the assimilation to identify the word (Gaskell & Marslen-Wilson, 1996; Gow, 2003). Other non-canonical forms, like dialect differences or allophonic differences (*butter* can be pronounced with a full /t/ or a flap), could be learned as part of the lexical templates for words (Bradlow & Bent, 2008; Ranbom & Connine, 2007).

NH listeners are adept at coping with lawful non-canonical forms (Bent & Holt, 2013; Norris, 1994), and changes in surface form due to allophonic variation (Ranbom & Connine, 2007), assimilation (Gaskell & Marslen-Wilson, 1996; Gow, 2003) or dialect (Kraljic, Brennan, & Samuel, 2008). This is accomplished via flexible processing schemes that are sensitive to variation, or that tolerate "slop" in the input. Further, frequent non-canonical productions can lead to long-term adaptation (Bradlow & Bent, 2008).

More challenging are *unpredictable* non-canonical forms, such as speech errors, an unknown dialect, or a new talker with idiosyncratic articulations. This problem is analogous to at least part of the problem faced by CI users for two reasons. First, because the CI input is impoverished, some information is lost and unrecoverable. Second, the auditory form of a

word heard by a CI user is distinct from the form learned before loss of hearing (though this is amenable to adaptation). Thus, at a broad level CI users face a similar problem to that of NH listeners hearing mispronounced forms. They must deal with a loss of speech information, and cope with uncertainty that cannot be "undone". Thus, the way NH listeners deal with unpredictable non-canonical forms may be analogous to the problems faced by CI users.

Cognitive adaptations may help listeners cope with uncertainty more flexibly. Indeed, recent studies suggest that when confronted with non-systematic variation like sporadically deleted phonemes (McQueen & Huettig, 2012) or reduced forms (Clopper & Walker, 2017), NH listeners rapidly adapt to support better processing. Less is known about how CI users deal with non-canonical productions. This is important as CI listeners more frequently encounter non-canonical inputs (due to their CI). Moreover, non-canonical forms offer an opportunity to probe experimentally how CI users may adapt to uncertainty.

Post-lingually deafened CI users reach asymptotic performance in typical speech perception outcome measures after two years of use (Hamzavi et al., 2003; Oh et al., 2003; Tyler et al., 1995). Part of this adaptation occurs at the levels of how sound is encoded and speech cues are used, such as in the mapping between frequency and location on the cochlea (Reiss, Perreau, & Turner, 2012), or weighting of phonetic cues (Moberly et al., 2014; Winn, Chatterjee, & Idsardi, 2012). However, CI users may also adapt at higher levels. A potential locus for such adaptation is the process of mapping speech onto lexical candidates (Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014; McMurray, Farris-Trimble, Seedorff, & Rigler, 2016). Adaptations at this level may help when the perceptual signal cannot be recovered (information is lost). When listeners expect non-canonical inputs, they may avoid strong lexical commitments to keep options available in case later information (e.g., semantic) can help recover from the ambiguity (Clopper & Walker, 2017; McMurray et al., 2016; McMurray, Tanenhaus, & Aslin, 2009).

This study examined the dynamics of lexical access following mispronunciations of common words in post-lingually deafened CI users. Stimuli were clearly spoken forms that did not consistently match the expected form of the target word. Our interest was not in recognition of mispronounced forms, per se, but rather to use these as a representative domain in which to understand how lexical competition in CI users responds to non-canonical forms of a word. Such forms allowed us to experimentally manipulate whether a word was canonical or non-canonical, without reducing the clarity of the phonetic information entering the CI (since all the phonemes could be clearly pronounced). Although this study alone cannot attribute these changes directly to adaptation, it points to differences in cognitive processing that could derive directly from the degraded input or may be the locus for adaptation.

We start with an overview of work on how NH listeners process non-canonical inputs. We then discuss the dynamics of lexical access to identify ways it could differ in CI users. Finally, we discuss research on lexical access in CI users before turning to the experiment.

The dynamics of lexical access

Even for NH listeners, the fact that words unfold incrementally creates brief periods of ambiguity. For example, during the earliest moments of *wizard* (e.g., after *wi-*), one could imagine two types of processing. First, the system could wait until the end of the word before accessing lexical entries (delayed commitment). This delayed commitment approach limits the ambiguity, but the possibility of embedded words (e.g., *car go* or *cargo*) makes it difficult to know when a word ends. Second, the system could partially activate many candidates (e.g., *wizard, wick, whistle*) and winnow this list as information arrives (an immediate activation approach). While immediate activation creates an early period of ambiguity (when multiple words are active), it may allow for more flexible decision-making, as a rejected (but partially active) competitor could be reactivated if later information favors it.

NH listeners engage immediate activation (Dahan & Magnuson, 2006; Weber & Scharenborg, 2012). Recent work has documented this using eye-tracking in the Visual World Paradigm (VWP; Salverda, Brown, & Tanenhaus, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to estimate moment-by-moment lexical activation. In this task, listeners see pictures of candidate words on a monitor, and one is named. The listener clicks on the corresponding picture with a mouse. In order to plan their response, listeners must fixate the correct item. As eye-movements can be made rapidly, listeners typically fixate multiple objects between the onset of the stimulus and the decision. Consequently, the degree to which listeners fixate each referent at any moment offers an estimate of how strongly that word is considered.

Building on prior work with priming and gated stimuli (Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989), VWP studies have shown that multiple lexical candidates are considered (activated) as soon as any acoustic information arrives that potentially matches items in the listener's lexicon (Allopenna, Magnuson, & Tanenhaus, 1998). For example as *wi-*is heard, listeners briefly fixate *wizard* and *whistle.* The degree of commitment is not solely a product of momentary match to the input – it is also sensitive to word frequency (Marslen-Wilson, 1987), and inhibition from other words (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Moreover, the set of candidates also includes words that do not necessarily reflect the serial order of phonemes in the input: rhyme competitors (e.g., gable after hearing table, Allopenna et al., 1998; Connine, Blasko, & Titone, 1993), embedded words (e.g., bone in trombone, Luce & Cluff, 1998), and anadromes (e.g., cat after hearing tack, Toscano, Anderson, & McMurray, 2013). This work as a whole suggests that competition is a cognitive process that reflects consideration of a wide range of potential interpretations of a word, and is not simply timelocked to the unfolding input.

Partial consideration of words that do not fully match the input may be helpful for dealing with non-canonical productions. If a word has been mispronounced (e.g., *dable*), the ability to partially activate a competitor that does not match at onset (*table*) allows more flexibility. In this case, the fact that there is no viable word *dable* could allow *table* to rise to the top (McMurray et al., 2009). Alternatively, subsequent sentence context may favor *table* (Connine, Blasko, & Hall, 1991; Szostak & Pitt, 2013). However, if the listener had

completely ruled out *table* upon hearing /d/, it may be harder to reactivate it from context. Thus, theoretically, the initial partial consideration or activation of a range of competitors may offer listeners flexibility for dealing with non-canonical forms.

Theoretically, flexibility could also be hindered by partial commitment or activation for competitor words. For example, if listeners commit to one word overly much (even if this is just partial) it could inhibit potential alternatives (Dahan et al., 2001; Luce & Pisoni, 1998) making it more difficult to recover if one of those alternatives turns out to be the intended word (McMurray et al., 2009). Onset mispronunciations in particular could lead to such deleterious results, as listeners "dig in" to an incorrect competitor. In such cases, delaying commitment until more information is available could prevent entrenchment to incorrect interpretations. Whether or not a partial commitment is helpful (keeping options open) or deleterious (digging into the wrong item) may be a matter of the degree of activation it receives and may differ across listeners or contexts. Thus, it is an open question as to what circumstances (and for what listeners) such partial activation is beneficial, particularly in listeners with hearing impairments.

Consistent with this nuanced view, deviations from canonical forms at onset lead to immediate decrements in word recognition (Marslen-Wilson & Zwitserlood, 1989; Norris, 1994; Swingley, 2009). However, recognition is not fully blocked: NH listeners can recover as more of the word is heard (McMurray et al., 2009; Swingley, 2009), with recovery reflecting the degree of mismatch (Creel, Aslin, & Tanenhaus, 2006; Milberg, Blumstein, & Dworetzky, 1988; White, Yee, Blumstein, & Morgan, 2013).

The dynamics of competition are also sensitive to the timing of the deviation. Swingley (2009) used the VWP to chart the moment-by-moment activation of lexical candidates after onset and offset mispronunciations (e.g., for *dog*: *tog* and *dob* respectively) in NH adults. Although the recognition decrement was similar for both, the timecourse differed. For onset mispronunciations, fixations to the target were depressed initially, but recovered as the word unfolded. For offset mispronunciations fixations were initially unaffected and decreased following the mispronounced phonemes. This time-locking suggests that when listeners access known words from partially matching non-word input, they employ a similar immediate activation strategy as with familiar words.

The dynamics of lexical access may be adaptable. The dynamics of lexical access improve through adolescence in NH children (Fernald, Perfors, & Marchman, 2006; Rigler et al., 2015; Sekerina & Brooks, 2007), and in brief laboratory training, listeners can adapt the dynamics of lexical activation in response to frequent presentation of lexical neighbors (Kapnoula & McMurray, 2016), reduced forms (Clopper & Walker, 2017) or missing information (McQueen & Huettig, 2012). This suggests lexical activation dynamics could be tuned to cope with listening demands.

A lexical response to unpredictable variation in CI users?

Recent work suggests the dynamics of lexical competition in CI users differ in ways that could help deal with degraded input. This could be either an automatic consequence of a

degraded input (e.g., slowing down could be serendipitously helpful); or it could reflect a longer term more cognitive or central strategy for coping with uncertainty.

Farris-Trimble et al. (2014) used the VWP to examine how listeners activate lexical candidates and suppress competitors. NH listeners show the typical pattern of rapid eye-movements to the target word, and rapid declines in fixations to competitors. However, compared to NH controls, CI users were slower to fixate the target and took longer to disengage from competitors, even though accuracy was high. This heightened consideration of competitors suggested that CI users may strategically keep competitors active (i.e. accessible, in case later occurring information requires a revision). Such a strategy would allow increased flexibility when dealing with the high likelihood that a portion of the input was misheard.

McMurray et al. (2016) asked whether these changes occur at the perceptual level (the way acoustic cues are coded) or at the lexical level. They examined how CI users respond to continuous changes that signal phonemic differences (e.g., voice onset time, distinguishing /b/ from /p/). Overt identification results suggested that CI users coded these fine-grained differences accurately with only small differences from NH listeners. Crucially, fixations were analyzed relative to the response on each trial. CI users exhibited a pattern of fixations that was not consistent with a difference in the mapping between cues like VOT and categories. Instead, they showed heightened activation for competitors across the board, suggesting they may maintain multiple options at the lexical level. Such differences could be ideal for dealing with unpredictable non-canonical inputs.

This work suggests two differences that could be adaptive for CI users: 1) maintaining activation for competitors (not fully committing); and 2) slowing a decision (committing more slowly). Such differences in lexical competition could, in principle, benefit CI users when hearing mispronounced forms, as they may help recover from speech errors by avoiding a costly over-commitment. For example, by not strongly committing to *talk* when *dog* is mispronounced as *tog,* CI users may retain the ability to recover gracefully when later context arrives. In contrast, the canonical view is that CI users' speech perception is fragile and falls off in adverse listening conditions (e.g., in noise). Mispronounced or otherwise non-canonical forms would appear to be just such a case, compounding the challenge of processing already degraded speech. Thus, it is unclear whether the altered dynamics of lexical competition may be helpful.

The Present Study

We used mispronounced forms to ask how CI users cope with non-canonical inputs. At a purely empirical level, this is important for understanding how CI users perceive speech in daily life, where mispronunciations and similar non-canonical productions are common. However, CI users may also process such items differently. This could be a serendipitous consequence of the fact that they are often slower (so they commit less strongly to an erroneous interpretation), or have less access to fine-grained differences (making *tog,* for example, not as bad an exemplar of *dog*). However, their long-term experience with a degraded signal may also lead them to tune lexical competition dynamics to cope with uncertainty.

Thus, a study of the recognition dynamics of mispronounced items may reveal some of the differences in word recognition that CI users exhibit when dealing with the degraded speech they confront because of the CI (even for canonical forms). Our immediate goal is not to disentangle why they show such differences – they could arise out of their lack of auditory sensitivity, slower processing, or an explicit strategy. Moreover, our goal is also not to detect a model of processing specific to mispronunciations (the mechanics of deploying such a strategy would be odd, as listeners may not know a form was mispronounced until the end of the word). Rather, we seek to profile the temporal dynamics of word recognition for mispronounced forms in this population to ask if the general profile for lexical access that has been shown in CI users for canonical inputs has consequences for how they deal with non-canonical forms.

CI users heard non-canonical (mispronounced) words, and we charted their real-time commitment to lexical candidates using the VWP (similar to Swingley, 2009). Stimuli were non-words that differed minimally from a target. After each stimulus, participants clicked on the visual referent of a word that best matched the token they heard. Competing referents were always phonologically and semantically unrelated to the target; so, participants could identify the target picture easily. We compared the timecourse of fixations for mispronounced and normal forms to chart the timecourse of the decrement, and listeners' recovery from it.

As a purely empirical level, our use of the VWP allows a more complete picture of word recognition by charting the entire timecourse of processing. That is we can ask what listeners are thinking at 300 msec after word onset, 400 msec after onset, and so forth. These early moments of processing can be particularly informative as at these times, the listener may know whether the stimulus is a real world or a mispronounced form (a nonword). Consequently, differences in the fixation data reflect not just the speed of processing (which would be expected to be slower for nonwords), but rather reflect the degree to which various candidates are considered—that is, what the listener thinks they are hearing—at any given moment.

The displays on each trial contained only unrelated items. This was done for two reasons. First we wished to match prior work with mispronounced forms by Swingley (2009). Second, by eliminating close competitors we minimize the role of perceptual errors in driving fixations; for example, if a subject heard *dog* and looked at the *dock* it could be because they misheard the /g/ as a /k/. By ensuring that there was no viable alternative on the screen this minimizes the role of such confusions, giving a clearer measure of lexical activation dynamics. Moreover it should be pointed out that there are numerous demonstrations that fixations in the VWP are sensitive to non-displayed competitors (Dahan et al., 2001; Kapnoula, Packard, Gupta, & McMurray, 2015; Magnuson et al., 2007; McMurray et al., 2009), so this choice does not invalidate the VWP as a measure of the dynamics of lexical activation.

As the dynamics of responding to non-canonical forms reflect the timing (Swingley, 2009) and degree (Milberg et al., 1988; White et al., 2013) of mispronunciations, we examined both onset and offset mispronunciations of varying degrees. Onset mispronunciations should

elicit earlier effects than offset due to the sequential nature of auditory processing, and multi-feature mispronunciations should elicit larger effects than single-feature. It is not clear if CI users will differ with respect to these factors.

Several outcomes were possible. First, mispronunciations may compound with CI users' degraded input to cause catastrophic performance decrements. Alternatively, differences in the lexical activation dynamics of CI users may better accommodate mispronunciations, making them more flexible. Finally, the signal degradation faced by CI users could make them insensitive to mispronunciations; if CI listeners coarsely code the input, then small changes in pronunciation may not be detected, leading to reduced effects of mispronunciation.

We limited this study to post-lingually deaf CI users (who lost their hearing in adulthood), as prior work suggests pre-lingually deaf CI users show markedly different dynamics of word recognition (delayed commitment; McMurray, Farris-Trimble, & Rigler, 2017).Within this group, we tested a large number of CI users with varying device configurations and speech perception abilities, to determine 1) if these strategies differed among configurations; and 2) if any particular pattern of lexical access dynamics predicts better outcomes.

With respect to the first issue, some CI configurations provide greater access to acoustic hearing by amplifying whatever residual hearing exists in either the ear contralateral to the CI (bimodal configurations) or the ear ipsilateral to it (hybrid CIs). The availability of residual acoustic hearing generally improves speech perception (e.g., Dorman et al., 2014; Gifford, Dorman, McKarns, & Spahr, 2007; Gifford et al., 2013; Tyler et al., 2002) (Dorman et al., 2009; C. C. Dunn, Perreau, Gantz, & Tyler, 2010; Gantz, Turner, & Gfeller, 2006; Gifford et al., 2013) (though see, McMurray et al., 2016). However, some hybrid and bimodal CI users may lose acoustic hearing after implantation. Moreover, some hybrid CIs offer fewer electrodes than standard CIs (at least for shorter eletrodes), potentially limiting the benefit of electric hearing. Thus, it important to focus on the hearing mode and ability as a whole, not just the devices worn.

The comparison between electric-only (CI_E) and acoustic+electric (CI_{AE}) hearing offers additional granularity for understanding how listeners cope with non-canonical forms. If mispronunciations compound with degraded input to hurt performance in CI users, CI_{AE} listeners' relatively better speech perception may lead to fewer problems with mispronounced forms. Alternatively, if CI users' adaptations are driven by the quality of the peripheral input, CI_{AE} users (particularly those with good residual hearing) may show less adaptation.

With respect to the second issue, differences in the dynamics of word recognition have only been established at the group level in CI users, and it is not clear if these online measures predict outcomes independently of offline measures like accuracy in the task. Here, we use a correlational approach to explore whether differences in lexical processing derive from the degraded input (poor input $\rightarrow$ delayed commitment) or whether they help cope with it.

METHODS

Participants

We targeted three listener groups: 1) Age-Typical Hearing (ATH) listeners who do not have hearing impairment requiring clinical intervention; 2) Standard CI users with profound deafness in both ears (thresholds >85 dB at all frequencies), and use solely electrical stimulation from the CI (CI_E users); and 3) Acoustic+Electric CI users (CI_{AE}) who use either a bimodal or hybrid CI configuration. We tested 72 participants, targeting approximately equal numbers in each group. This led to 23 CI_E users, 21 CI_{AE} users and 28 ATH listeners (see Table 1 for summary; Supplement S1 for a complete description). All participants were monolingual[1], native speakers of English with adequate visual acuity for the task (corrected if necessary). One additional CI_E user was tested and excluded for very low accuracy (<70%).

CI users were recruited from a larger ongoing study at the University of Iowa. CI users had a post-lingual onset of deafness and at least one year of device experience, as speech perception stabilizes by this point (Hamzavi et al., 2003; Oh et al., 2003; Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997). On the day of test, CI users received an audiological exam including device fitting of both the CI(s) and any hearing aids in use, as well as an audiogram of acoustic hearing. CI users were classified as CI_E or CI_{AE} on the basis of these tests of acoustic hearing: users with aided thresholds (<85 dB, hearing aid only) for 125–1000 Hz on either ear were treated as CI_{AE}, otherwise they were CI_E (see Supplement S1).

Participants who did not use CIs were recruited from the University of Iowa community. Their hearing was screened at 250, 500, 1000, 2000 and 4000 Hz. Seventeen had pure tone thresholds at or below 25 dB HL at all tested frequencies. As listeners could adjust the volume during the experiment to a comfortable level, we also included some listeners slightly above this threshold. Eight participants did not pass but had thresholds below 40 dB (the level at which clinical intervention is typically recommended) at one or more frequencies. This was a low (and age-typical) level of hearing loss and not consistent across frequencies and ears (most frequencies/ears were fine); thus these participants were retained. For two participants, hearing screenings were not conducted. However, both reported normal hearing, and performed at 100% correct. This group was referred to as age-typical hearing (ATH)[2].

Testing was completed in a single two hour session. ATH participants received $30, and CI users received either audiological services or a check compensating them for a host of studies across a 1–2 day period. All participants underwent informed consent in accordance with University IRB approved protocols.

Sample Size and Power.—We targeted 20 participants in each group on the basis of prospective power analyses that computed minimum detectable effects (MDEs) given the available sample. We slightly over-recruited to account for attrition or participants not meeting eligibility. While power analyses are not easily performed for mixed models, we assumed ANOVA (widely assumed to be less powerful) with three groups (ATH, CI_E, CI_{AE}), and one within-subject factor with three levels (correct, single- and multi-feature

mispronunciations). We assumed α=.05, β=.8, and σ=.5 (for within subject factors). For the obtained sample, MDE was f=.31 for between subject effects, f=.15 for within subject, and f=.17 for the interaction.

Individual Difference Measures.—Language and non-verbal skills were evaluated using the Peabody Picture Vocabulary Test—Revised (D. M. Dunn & Dunn, 2007), and the Block Design and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence—2nd Edition (Wechsler & Hsiao-Pin, 2011). ANOVAs were conducted comparing WASI and PPVT scores by listener group. We found a significant main effect for PPVT ($F(2, 69)=4.1$, $p=.021$), driven by better vocabulary in ATH relative to CI_E listeners ($t(49)=2.3$, $p=.027$) and CI_{AE} listeners ($t(47)=2.7$, $p=.009$). These differences were small (~3/4 of a SD), and in the supra-normal range (at which oral vocabulary does not exert a large influence on lexical activation in the VWP (McMurray, Samelson, Lee, & Tomblin, 2010). There was no effect for WASI ($F<1$).

Design

Items consisted of 10 sets of four words (Supplement S2). Words were CVC words referring to common objects. Sets were constructed such that all four items were phonologically and semantically unrelated. The same sets were used for all participants. On each trial, one set was selected at random and pictures corresponding to the four items were displayed. One of the objects was named (either correctly or mispronounced), and the participant clicked on it. Across trials, all four items in each set were named in both correct and mispronounced forms. As there was no phonological and semantic relation between the target and competitors, even for mispronounced forms, the target was readily identifiable. This paradigm roughly models a situation in which a degraded word must be identified using discourse context.

Mispronunciations were manipulated along two dimensions. First, mispronounced phonemes occurred either at word onset or offset (not the vowel). Second, mispronunciations altered either one or multiple features of the target phoneme. Single-feature changes always changed either voicing or place of articulation; multi-feature changes included manner of articulation and at least one other feature. Mispronunciations never formed a real word. Target words and mispronounced forms were selected to maximize phonetic diversity to minimize the influence of idiosyncratic differences in perceiving particular acoustic cues.

The statistical design crossed location of the mispronunciation (offset vs. onset) with degree of mispronunciation (correct, single feature or multi-feature). To complete a 2×3 design, correct trials were randomly assigned to either the onset or offset condition. Participants heard four repetitions of each correctly pronounced token, and one repetition of four possible mispronounced variants, totaling 320 trials.

While participants performed this task, fixations were recorded to estimate lexical activation dynamics. We focused solely on target fixations (the foils were unrelated). For target fixations, the VWP has good test/retest reliability of .65 - .70 for timing related parameters (e.g., the slope; Farris-Trimble & McMurray, 2013).

Our goal was to study the dynamics of successful lexical access—stimuli should be as audible as possible to maximize accuracy. Thus, participants conducted the experiment in their everyday listening mode (e.g., bilateral users used both CIs, bimodal users used a CI and their hearing aid), and could adjust the volume in the sound field to a comfortable listening level.

Auditory Stimuli

Stimuli were recorded by a male native speaker of English with a standard Midwestern American dialect in a sound attenuated booth at a sampling rate of 44.1 kHz. To maintain a constant prosody, words were recorded within the carrier phrase "He said ___", resulting in a declarative intonation with a falling tone. Each word was recorded five times. The best exemplar was then selected and excised from the carrier phrase, as words would ultimately be presented in isolation. One hundred msec of silence was appended at stimulus onset. The resulting files were amplitude normalized in Praat (Boersma & Weenink, 2009). Final stimuli were inspected for clarity, and to remove extraneous clicks etc. Auditory stimuli were presented in the sound field via JBL Professional 3 Series amplified speakers at ±26° from center. Stimuli had an average duration of 504 msec (not including the 100 msec of silence).

Visual Stimuli

Visual stimuli were color line drawings of the intended referent, created using a standard lab procedure (McMurray et al., 2010). First, several candidates were selected from a commercial clipart database. Candidates were viewed by a focus group to select the most representative images with a similar style, and to identify any changes to be made. Selected images were edited to ensure uniform size and brightness, obtain more prototypical colors and orientations, and remove backgrounds or distracting elements. Finally the final image was approved by a lab member with experience in the VWP who was not involved in their creation.

Procedures

After informed consent, participants were seated in a sound-attenuated booth in front of a computer monitor. Next, the eye-tracker was calibrated. Participants were then familiarized with the 40 words and their images. During familiarization, each picture was presented at screen center along with its orthographic label. Participants observed each word-image pair and pressed the space bar to advance to the next one. Order was randomized for each participant.

After familiarization, testing began. On each trial, participants saw four images on the screen, with a small blue dot in the center. Images were 300×300 pixels situated 50 pixels from the edge of a 17" LCD monitor at 1280×1024 pixels. After 750 msec, the blue dot turned red. At this point, the participant clicked on it to hear the auditory stimulus. Clicking the dot ensured that the participant's mouse and gaze were at screen center at trial onset. The auditory stimulus was played 100 msec after clicking the dot, and the dot disappeared. After hearing the stimulus, participants clicked on the image that most closely matched the word they heard.

Eye tracking recording and processing

Eye movements were recorded with an SR Research Eyelink 1000. Participants rested their chin on a table-mounted chinrest, and eyes were tracked by a remote camera under the monitor. The Eyelink 1000 samples gaze using both light pupil and corneal reflection at two msec intervals. The eye-tracker was calibrated using the standard nine point display. Drift correction was conducted every 32 trials. Both eyes were tracked (when possible), but only the eye with the best calibration (lowest average error) was used for analysis.

Eye-movements were recorded from the beginning of a trial to the response. For trials ending later than 3000 msec post-stimulus, data were truncated; for trials ending before this time, the final fixation was extended to 3000 msec (though there were few differences after 2500 msec, so figures reflect that shorter window). Consequently, the timecourse of fixations displays an asymptote corresponding to the word the listener "settled" on (Farris-Trimble et al., 2014)

The Eyelink 1000 automatically parses point of gaze into saccades and fixations. We combined saccades and the subsequent fixation into a single unit, a "look", starting at the onset of the saccade, ending at the offset of the fixation, and directed to the average point of gaze during the fixation. Look location was computed in screen coordinates, and the boundaries of the images on the screen were extended by 100 pixels to account for noise in the eye-track. This resulted in at least 124 pixels between adjacent regions of interest.

RESULTS

First, we examined the accuracy of the mouse click. Second, we asked how the patterns of fixating the target object are affected by the degree of mispronunciation and by hearing group. Third, we examined the degree to which listeners are impacted by the mispronunciation during real-time processing. Finally, we conducted an individual differences analysis to relate the real-time measures of lexical processing to outcomes.

Accuracy

The analysis of the mouse click accuracy asked whether listeners recognized the words despite mispronunciations, and whether different types or degree of mispronunciations decreased accuracy (Figure 1). This is a closed set task with unrelated foils, so absolute accuracy is inflated relative to more traditional open set tests of word recognition. Here our focus is on relative performance among groups and conditions. For correctly pronounced words, all three listener groups performed well (ATH: M=99.9% SD=0.3%; CI_E: M=99.2%, SD=1.8%; CI_{AE}=98.6%, SD=2.0%), and no individual was worse than 91.8% correct. All groups showed an effect of mispronunciation, with decrements for single-feature mispronunciations and larger ones for multi-feature. This effect was larger for CI users. While there was little difference within CI users for onset mispronounced forms, for offset mispronunciations, CI_{AE} users showed a larger decrement. However, even for the most severely mispronounced words, CI users' accuracy still exceeded 84%. Thus, there were enough trials on which listeners recovered from the mispronunciation to examine fixations.

For analysis, we used a logistic mixed effects model with accuracy as the dependent measure. Mispronunciation was coded as two orthogonal contrast codes: whether or not the target was mispronounced (−.66 for correct; +.33 for either incorrect form), and Degree (correct: 0, single feature: −.5, multi-feature: .5). Location of mispronunciation was binary (onset: .5; offset: −.5), and listener type was coded as two contrast variables: ATH (+.5) vs. CI (−.5), and CI_E (−.5) vs. CI_{AE} (+.5; ATH=0). Random effects were determined by comparing models of different complexity (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) to identify the model with the random slopes that offered a significantly better fit than reduced models (Supplement S3). The resulting model had random intercepts of subject and base word (the canonical form: e.g., for the set *dog*, *gog*, *mog*, the base word was *dog*), and a random slope of location × degree on subject. Models were fit with the nlme package (ver 3.1–120) in R (ver 3.2).

Results are in Table 2. We found a significantly lower accuracy for CI users than ATH listeners (B=2.08, p<.0001), and significantly lower accuracy for mispronunciations (B=−2.02, p<.0001) and for multi- than single-feature mispronunciations (B=−1.40, p<.0001). There was also an effect of location (B=−0.83, p=.0015) with larger decrements at onset than offset. While there was no main effect of CI type (CI_E vs. CI_{AE} contrast), this did interact with whether there was a mispronunciation (B=−0.90, p=.0021). There were also interactions of Location and Mispronunciation, and CI vs. ATH.

We followed up these interactions by splitting the dataset by location, and running the same model. For onset mispronunciations, there was a main effect of CI use (B=1.49, p<.0001), but no difference between CI_E and CI_{AE} listeners (B=.54, p=.21). Both mispronunciation contrast codes were significant (Mispron: B=−2.56, p<.0001; Degree: B=−1.06, p<.0001), and neither interacted with listener factors. For offset mispronunciations, the pattern was similar with one exception. The CI vs. ATH contrast was significant (B=3.00, p<.0001), but not the CI_E vs. CI_{AE} contrast (B=.03, p=.95). Both mispronunciation contrasts were also significant (Mispron: B=−1.33, p=.001; Degree: B=−1.64, p=.0007). But this time, mispronunciation interacted with CI type (B=−1.16, p=.008) with a bigger mispronunciation decrement in CI_{AE} users (Figure 1A).

Analysis of Fixations

Our primary analyses examined the timecourse of fixations to the target as a function of mispronunciation. We analyzed only trials on which the listener responded correctly, as our goal was to analyze the processes leading up to accurate lexical access (not the trials on which these processes went awry). Eye movements initiated earlier than 300 msec after stimulus onset were ignored. As auditory stimuli began with 100 msec of silence, and it takes 200 msec to plan an eye movements, 300 msec is the earliest time when input-driven eye movements could occur.

Figure 2 shows the proportion of trials on which listeners fixated each object every 4 msecs. Both ATH (Figure 2A) and CI (Figure 2B) listeners made rapid fixations to the target, with little consideration of unrelated items. Similarly to Farris-Trimble et al. (2014), CI users showed delayed and slower target fixations with reduced fixations at asymptote (Figure 2C).

Figure 3 shows the same data by mispronunciation type. There was a close time-locking of fixations to the mispronunciation. In ATH listeners (Panel A), onset mismatch led to an immediate delay followed by recovery, whereas offset mispronunciations had no influence initially (the mispronunciation had not been heard yet), but depressed the asymptote. CI users (Panels B, C) showed similar effects. Both types of CI users showed slower rise times and lower asymptotic fixations across all word types. This confirms the timecourse of processing seen in earlier studies and extends it to non-canonical forms (though it is not clear if CI users perform worse or better on these forms relative to canonical forms—our primary question).

For statistical analysis, we fit a logistic function to the fixation data for each subject / condition (Farris-Trimble & McMurray, 2013; Farris-Trimble et al., 2014). We then used parameters like slope or asymptote to represent the dynamics of processing as a function of mispronunciation condition. Target fixations were modeled with a four parameter logistic (Figure 4). Model fits were excellent (ATH: Mean r=.981; CI: Mean r=.981). Seven fits (of 462) were excluded because the data did not conform to the expected shape. All 7 fits (3 CI, 4 ATH) were in the multi-feature offset mispronunciation condition. Those participants were retained for other conditions.

This function has free parameters for lower and upper asymptotes, the crossover (the time when the curve is halfway between asymptotes), and slope (derivative at crossover). From these we computed two measures. First, Maximum was used as a measure of how strongly the listener committed to the target at asymptote. As it was bounded at 1.0, it was scaled with the empirical logit function prior to analysis. Second, Crossover and Slope both measured timing. These were log-transformed to achieve normality. As these were highly correlated (r=−.795), the log scores were z-scored and averaged (crossover was multiplied by −1), to measure Timing (higher=faster).

We used these two measures in a series of mixed effects models. Separate models were run for the Max and Timing. Fixed effects included the location of the mispronunciation, degree of mispronunciation and listener type (all coded as in the accuracy analysis), as well as two- and three-way interactions. Significance was estimated using the lmerTest package in R. For all models, random slope structure was determined using model comparison. Results of model comparisons and the complete formulas are in Supplement S3. Figure 5 shows both measures as a function of listener group and condition.

Documenting the mispronunciation effect in ATH listeners.—The first analysis examined the ATH data only to document that onset and offset mispronunciations had different effects on the time course of processing (Figure 3A, replicating Swingley, 2009). Thus, our model had only location and degree as fixed effects (Supplement S3B for random effects).

Results are in Table 3. For both measures, there were a significant location × mispronunciation and significant location × degree interactions. Thus, we split the data by location (Supplement S4A for additional visualizations), and conducted additional analyses with only Mispronunciation and Degree (single vs. multi-feature). For *Max* there was a

significant effect of both Mispronunciation and Degree for both onset and offset mispronunciations (mispronunciation led to lower Maxima, and multi-feature was even lower). However, the effects were much smaller (had smaller coefficients) for onsets (Mispron: B=−.12, p<.0001; Degree: B=−.06, p<.0001) than offsets (Mispron: B=−.26, p<.0001; Degree: B=−.23, p<.0001). In contrast, for Timing, there were no effects for offset mispronunciations (Mispron: B=−.08, p=.37; Degree: B=.16, p=.14), while there were significant effects for onset (Mispron: B=−1.74, p<.0001; Degree: B=−.87, p<.0001). Thus, this analysis confirms a different timecourse of processing for onset and offset mispronunciations in ATH listeners. Onset mispronunciations had a large effect on the timing of fixations, with a minimal effect (a smaller coefficient) on the asymptote, while offset mispronunciations had no effect on the timing but a large effect on the asymptote. Supplement S5 reports parallel analyses with the CI users that show similar patterns.

Effect of CI use on word recognition for canonical forms.—We next asked whether the three listener groups differed in the dynamics of lexical access. We examined only correctly pronounced trials to establish a baseline (on which to compare mispronounced trials). Models related each property of the fixation dynamics to two contrast codes for listener group. The only random effect was a random intercept of subject (fixed effects were between-subject). As Table 4 indicates, CI users were slower to activate the target (Timing) and reached a lower asymptote (Max). There was no further difference between CI_E and CI_{AE} users. Thus, even for canonical forms, CI users were slower and less committed at the end of processing.

Interaction of CI use and response to mispronunciation.—Finally, we asked if the location × mispronunciation interaction(s) was moderated by listener group (CI vs. ATH, and CI_E vs. CI_{AE}). This was done by adding listener contrasts to the ATH models described above and running separate models for Max and Timing. The Max included random slopes of all three main effects, but no interactions; the Timing model contained only a random slope of Location (Supplement 3C for details).

Results (Table 5) are complex, and many of the main effects (e.g., of location or mispronunciation) are subsumed under interactions. Moreover, as described in the next section, these properties (Max and Timing) could interact in unexpected ways. Thus, we treat this analysis as an omnibus test, asking whether or the effect of mispronunciations on the timecourse of processing is moderated by listener group before describing an analytic approach that permits a more direct examination of this question[3].

First, consistent with our first analysis, we observed main effects of CI vs. ATH on both properties (Row 1). Across mispronounced and canonical forms, CI users were slower (Timing), and committed less strongly (Max). This can be seen in Figure 5 as higher (faster) Timing (panel A) and higher Max (panel B) roughly across the board for ATH than CI users. There was also a marginal effect of CI_E vs. CI_{AE} (Row 2) with CI_{AE} showing slightly faster dynamics. This appears most clearly in Figure 5A in the onset condition.

Second, there was continued strong evidence of a location × mispronunciation interaction (Row 12). As in the prior analysis, onset mispronunciation showed a large decrement in

Timing (Panel A, left side) and a minimal effect on Maximum (Panel B, left side); whereas offset mispronunciation (right sides of both panels) showed the converse. Both effects were enhanced for multi-feature mismatch (the Location × Degree interaction; Row 15). This is seen in Figure 5, where mispronunciation has a large effect on Timing for onset mispronunciations (Panel A, left side), but almost no effect for offset (right). In contrast, for Max, there was a smaller effect on timing for onset (left) than offset (right).

Third, for Timing, whether the listener was a CI user was a significant moderator of the Location × Mispronunciation interaction by (Row 7), and a marginal moderation of the Location × Degree interaction (Row 16). As Figure 5A indicates, for onset mispronunciation (left side) the effect of mispronunciation was smaller for CI than ATH users. Post-hoc tests examined the onset mispronunciations only separately by group. For both groups, there was a strong effect of mispronunciation (ATH: B=−1.74, p<.0001; CI: B=−1.36, p<.0001), but the magnitude of the effect was smaller for CI users. Similarly, both groups showed an effect of degree (ATH: B=−.87, p<.0001; CI: −.65), but again with a smaller effect for CI users. A final post-hoc model examined only the onset mispronounced trials (both single- and multi-feature), and showed that the small difference in Timing between CI and ATH listeners (relative to the much larger difference between correct and mispronounced forms, Figure 5A) was significant (B=.35, p<.0001), and did not interact with degree of mispronunciation (B=−.22, p=.16).

There were no significant interactions with the CI_E vs. CI_{AE} contrast (though several marginal). Thus, the next analyses collapse across CI users (Supplement S5 compares them).

This analysis suggests post-lingually deafened CI users process speech incrementally (like ATH listeners) and respond similarly to mismatch. However, relative to their slower timecourse of activation, they show less delay from onset mispronunciations than ATH listeners.

Disruption of Lexical Access by Non-Canonical Input

Our goal was to examine the effect of mispronunciation *relative to* the baseline dynamics of activating correctly pronounced words. This required an examination of the difference of conditions. However, the timecourse of activation is not a simple additive result of Max and Timing. These functions are inherently non-linear and these two measures may be correlated among individuals. As a result, it may not be accurate to treat a difference of any property in isolation as indicative of a particular difference in processing.

Thus, for a more global measure of differences in these curves, we utilized the Bootstrapped Difference of TimeSeries (BDOTS: Oleson, Cavanaugh, McMurray, & Brown, 2017; Seedorff, Oleson, & McMurray, in press) which statistically identifies *when* two time-series differ. BDOTS starts by fitting the logistic function to individual listeners' patterns of fixations in each condition. From these fits, confidence intervals are computed using bootstrapping. Next, t-tests are conducted at each time-step with a family-wise error adjustment that takes into account the auto-correlation among t-statistics. BDOTS is implemented for pairwise comparisons (both between and within), so we conduct targeted two-way comparisons.

We started with two BDOTS analyses of ATH listeners comparing fixations after correct forms to fixations after onset or offset mispronounced stimuli (Figure 6A, Table 6, row 6A). We found statistically significant differences between onset-mispronounced and correct stimuli from 467 to 3000 msec, and between offset mispronounced and correct stimuli from 552 to 3000 msec. The earlier effect for onset mispronounced stimuli reflects the rapid disruption of lexical access when the initial phoneme mismatches (e.g., *gog* for *dog*).

To assess how listeners cope with non-canonical forms, we computed the difference in fixations between correct and mispronounced forms for each subject. These curves—which we term the *mispronunciation (MP) disruption*—reflect how much a mispronunciation impedes lexical access over time. A higher MP disruption indicates greater difficulty. Figure 6B shows MP disruption for onset and offset (single feature) mispronunciations (in ATH listeners). BDOTS (Table 6, Row 6B) shows more disruption for onset MPs early in processing (504 – 1220 msec), but greater disruption from offset MPs later (1672 to 3000 msec).

We next turned to our primary question: do CI users and ATH listeners show differential disruption from non-canonical forms. We note that in several cases, poor curve fits led to very high standard errors around the function. These appeared to derive from data that did not conform to the logistic function (but still yielded reasonable R^2 values above .8); for example when participants did not reach a stable asymptote. Thus, for these analyses, we reran the analysis with a more restrictive set of curves. We report results of both in Table 7 (more restricted sets are indicated with *), and focus our description (and Figure 7) on the restricted set.

We started by considering onset mispronunciations (Figure 7A, B; Table 7). Surprisingly, CI users showed *less* disruption for about 300 msec starting at 540 msec. This was reversed by 1212 msec when they showed slightly (but significantly) more disruption. A similar pattern was observed with multi-feature mispronunciations with a longer-lasting effect: CI users again showed *less* disruption, from 532 to 924 msec, before reversing at 1684 msec (the later reversal was not significant in the full sample). For offset mispronunciations (Figure 7C, D) we found no significant differences between CI users and ATH listeners in any analyses. Follow-up analyses (Supplement S6) compared CI_E to CI_{AE} and found no differences.

These analyses suggest surprising accommodation on the part of CI users. For non-canonical forms that differ at onset, at early moments during the processing CI users were less impacted by the mispronunciation (though they showed a cost later). For forms differing at offset there was no difference (despite generally lower accuracy [e.g., Figure 1B]). So *when* precisely is this benefit observed? When we adjust the 540 – 900 msec window for the 100 msec of silence at word onset and the 200 msec oculomotor delay this is 340 – 600 msec, somewhat early in the space of the trial. In fact, this benefit occurs at a point where the CI users' fixations (e.g., in Figure 2B, C) have not yet reached 50% of their maximum value for onset mispronounced trials (roughly 950 msec) and just after ATH users cross this threshold (at roughly 832 msec). Thus this is during the first half of the process of spoken word recognition.

Individual Differences

Finally, we asked whether these measures of real-time lexical access predict outcomes. CNC word recognition scores (collected as part of the larger project) were available for 42 of the 44 CI users (CI_E: N=19, CI_{AE}: N=23). We used these to address two questions: 1) Do the dynamics of lexical activation predict outcomes? 2) Do the dynamics of lexical access for canonical forms predict those of non-canonical forms? As these questions are not primary goals of this study, they are exploratory.

Bivariate Correlations: CNC.—There were a large number of measures of interest, which share variance with each other. However, with only 42 participants, multiple regression can include at most 3–4 predictors to avoid inflated variance estimates. Thus, we started by computing pairwise correlations. We used the Max and Timing measures in each condition as estimates of lexical processing dynamics. To estimate the differential response to non-canonical forms, we computed the difference between correctly pronounced and incorrectly pronounced forms in accuracy and in Max and Timing. This was only done for onset mispronunciations for Timing (which showed the largest differences), and offset for Maximum.

Before conducting this analysis we examined correlations among and within the predictor measures. Measures within a class tended to be highly correlated with each other. For example, accuracy in the five different pronunciation conditions had an average correlation of .735. Within class correlations averaged .576 across all measures. Thus, within a class of measures one should not make strong claims for independence (e.g., Max in the single feature onset mispronunciation is not distinct from the multi feature onset condition). However, across classes, correlations were much lower (Accuracy → Max: r=.197; Accuracy → Timing: r=.320; Max → Timing: r=.294). Thus, the different classes of measures may tap more independent aspects of word recognition.

We next examined correlations of each measure with CNC word recognition (Table 8). *Accuracy* was highly correlated with CNC scores (*r*~.6 across conditions in column 1). This is somewhat surprising given that 1) our task was closed set (vs. open set vocal responding); and 2) this association was robust even in the correctly pronounced condition (where CI users were near ceiling). The large number of trials (80) may have allowed a finer grained assessment of ability even near ceiling. It was notable that these correlations even held in mispronounced conditions – here the ability to derive the right answer was a product of CI users' abilities to use contextual information (since they had to infer *dob* must match *dog,* given the lack of viable competitors on that trial). This suggests that these contextual inference mechanisms may be a part of good CI performance. Accuracy decrements between correct and mispronounced forms were negatively correlated with outcomes. CI users with greater mispronunciation decrements tended to have better speech perception outcomes. This reinforces the notion that these decrements reflect in part the sensitivity to fine-grained acoustic information (White et al., 2013).

For *fixation measures,* we contrast two hypotheses. First, these measures could reflect how similar a CI user is to ATH—more rapid Timing and higher Maxima should predict good performance. Alternatively, there may be benefits to slower lexical commitments. This

predicts slower Timing and lower Maxima will be associated with better outcomes. Results support the former. Timing on correct trials showed a robust correlation of .67 with outcomes – CI users with steeper slopes showed better CNC word recognition. Maximum was positively correlated (r=.44) – higher maxima predicted better speech perception. Thus, more ATH-like dynamics of lexical processing predict better outcomes. Finally, when we examined change scores between Max or Timing (as a function of mispronunciation), we found that for onset mispronunciation, larger changes in Timing predicted better CNC word recognition (single feature: r=.41; multi: r=.33). For offset, this was significant for large changes in maximum (multi: r=.51), and only marginal for smaller changes (r=.28). These findings support the idea that *change* in dynamics may reflect sensitivity to fine-grained detail.

Do the dynamics of lexical processing uniquely predict outcomes?

Given these findings, we next wanted to know if the lexical activation dynamics are redundant predictors. Do they just track overall accuracy in the task or do they reveal something unique? To address this, we conducted a series of hierarchical regressions. On the first step, we predicted CNC word recognition from accuracy. We then added a lexical processing measure to ask whether the dynamics of processing predict CNC word recognition *over and above* accuracy.

We started with Timing on the correct trials (reflecting automaticity of word recognition under ideal circumstances). On the first level, accuracy predicted 52.8% of the variance ($F(1,40)$=44.8, p<.0001). On the next level, Timing accounted for an additional 10.5% unique variance (F $(1,39)$=11.2, p=.002). Of course, with accuracy near ceiling, Timing may have picked up additional variance that could not be extracted from the lower variance accuracy measure. Thus, in a follow up model, we added two additional accuracy measures with lower means: the accuracy on the two multiple feature mispronunciation conditions. In this model, accuracy now accounted for 58.9% of the variance ($F(3,38)$=18.1, p<.0001), but on the second level, slope was still significant (R^2 =.058, F $(1,37)$=6.05, p=.019). This approach was followed for Max as well. *Maximum* was significant when only correct accuracy was in the model (single accuracy: R^2 =.50, F $(1,40)$=4.60, p=.038), but not when all three accuracy measures were used: R^2 =.021, F $(1,37)$=2.04, p=.162)[4]. Thus, measures of lexical activation dynamics (particularly slope) show a relation to outcomes that is not captured by accuracy.

Stability of dynamics.

We next investigated the stability of the parameters of lexical competition across individuals. Within correct trials, there was a robust correlation between Timing and Maximum (r=.747, p<.0001). This correlation is not an artifact of the curvefitting procedure. In this function, the estimate of Maximum is mathematically independent of slope and crossover (Timing). Thus, these correlations suggest that these properties collectively reflect a stable constellation of traits within listeners: CI users with more automatic processing (better Timing) also tend to show increased asymptotic looking (e.g., a more confident decision).

We next asked whether the dynamics of lexical access on correct trials predict those for non-canonical forms (see Table 9). This would imply that individual listeners have a stable pattern of dynamics across different levels of input degradation. We observed a remarkable concordance in some cases. The maximum on correct trials was highly correlated (r~.63) with the maximum on all four types of mispronounced trials. This is particularly notable on offset mispronounced trials, as offset mispronunciation had a large effect on maximum looking (Figure 3). The stability of the Max measure (reflecting the degree of final commitment to a single word) may reflect general language ability, which prior work has linked to Max (McMurray et al., 2010).

Timing was also highly correlated across correct and offset mispronounced trials, but this was to be expected as these parameters largely reflect processes at word onset (and the onsets are the same in an onset mispronounced form). However, Timing was not correlated with itself across correct and *onset* mispronounced trials (r~.11). On these trials, CI users are asked to *recover* from initially misleading information. This suggests that this recovery may not be strongly shaped by the general pattern of lexical activation (e.g., for canonical forms).

DISCUSSION

This study had several key findings. First, both CI users and ATH listeners showed accuracy decrements with mispronounced forms, and these were bigger for CI users (Figure 1). Second, parallel to prior work (Farris-Trimble et al., 2014), relative to ATH listeners CI users were slower to recognize even correctly pronounced forms and committed less strongly (Figure 2C, Figure 3). This also applied to mispronounced forms. Third, during real-time lexical access, both groups showed immediate sensitivity to the location and degree of mispronunciation (similar to Swingley, 2009, with NH listeners), though these differed quantitatively in CI users (Figure 3). Fourth, relative to their slower baselines, CI users showed surprisingly less disruption (than ATH listeners) for onset mispronounced forms early in processing (Figure 7). There was no difference for offset mispronounced forms. Fifth, real-time processing measures are stable predictors of outcomes, over and above accuracy (the hierarchical regressions). We discuss each of these findings and several caveats.

Accuracy Decrements

As this was a closed set task with unrelated foils, absolute accuracy was higher than what would be observed in traditional open set tasks. Nonetheless, the relative accuracy across conditions and participants was informative and predicted outcomes. All groups showed accuracy decrements with mispronounced forms (though this was small for ATH listeners). As these were single feature differences, the fact that CI users showed a decrement at all indicates that they are sensitive to small differences in acoustic detail.

Counterintuitively, CI_{AE} users showed a greater mispronunciation decrement (lower accuracy) than CI_E users for offset mispronunciations (Figure 1). This is unlikely to be due to uneven sampling: CI_{AE} users (M=.750) had slightly better CNC word recognition than CI_E users (M=.682) though this was not significant (t(40)=.96, p=.34). The larger mispronunciation decrement in CI_{AE} users was surprising given the wealth of evidence that

residual acoustic hearing improves speech perception (Dorman et al., 2009; Dorman et al., 2014; C. C. Dunn et al., 2010; Gantz et al., 2006; Gifford et al., 2007; Gifford et al., 2013; Tyler et al., 2002). However, this may parallel McMurray et al. (2016) which found that CI_{AE} users (relative to CI_E users) uniquely show decrements with fricatives (s/ʃ). The authors hypothesized that this may be because CI_{AE} users come to over-rely on acoustic hearing. While this is a good strategy in general, it fails on fricatives which have no low-frequency acoustic information. Here, the comparatively better sound quality experienced by CI_{AE} users may lead them to overestimate their abilities. As a result, non-canonical forms may be more disruptive for CI_{AE} users than for CI_E users, who expect less precise input. However, we cannot rule out that other demographic and/or audiological factors may have contributed to this finding (and it is noteworthy that it only appeared in the offset mispronunciation condition).

Incremental Processing in CI users

At the broadest level, both CI users and ATH listeners showed similar patterns of real-time processing (Figure 3). In both groups, single feature onset mispronunciations (*gog* for *dog*) led to immediate delays in lexical access (for ATH listeners, crossovers were delayed by 218 msec, in CI by 177 msec), and both groups of listeners did recover. Offset mispronunciations (*dob* for *dog*) showed no early effects in both groups, but such forms limited the ultimate asymptote of fixations. The fact that these patterns were observed in both groups suggests that while CI users may be slower to activate words (Figure 2C), they show incremental uptake of information similarly to ATH listeners (see, Farris-Trimble et al., 2014).

This similarity in the dynamics is impressive given the degraded input CI users face. It is not guaranteed. McMurray et al. (2017) examined lexical access dynamics in prelingually deaf CI users. These CI users showed a much larger delay – over 200 msec – and as a result, they showed reduced competition from onset competitors (a form of delayed commitment). This demonstrates that incremental processing is not always attested in word recognition; consequently, the incrementality of adult CI users is impressive.

The pattern of lexical dynamics was not different between CI_E and CI_{AE} users: both accessed the lexicon at similar rates (the *slope* and *crossover* did not differ between CI groups [Table 5]). This conforms with Farris-Trimble et al. (2014). It suggests the gains in word recognition accuracy from residual acoustic hearing may not be because of more direct information for word identity, but rather may derive from mechanisms like segmentation, streaming speech from noise, and voice recognition that are not needed for this experiment.

Nonetheless, while the pattern of fixations in both groups is consistent with an incremental processing model (McClelland & Elman, 1986), CI users were delayed by about 130 msec (for crossover in correct forms, M_{CI}=734 msec, M_{ATH}=605 msec). This is consistent with prior work (Farris-Trimble et al., 2014), and suggests they may commit more slowly to lexical candidates than ATH listeners. They also showed lower maxima (M_{CI}=.86, M_{ATH}=.92), suggesting they commit less fully. Both are important for understanding their relative performance with non-canonical forms.

Accommodation to Non-canonical Inputs

At the macro level, fixation patterns in both CI users and ATH listeners showed similar sensitivity to non-canonical forms. However, closer examination displayed important differences. At early points in processing, ATH listeners showed larger immediate impacts of onset mispronunciations than CI users. This benefit was observed in two places. First, it can be seen in the analysis of Timing (Figure 5A), where the drop from correct to mispronounced forms was much less for CI users than for ATH listeners. That is, timing in CI users showed less of an effect of mispronunciation. Second, and more dramatically, it was observed in MP disruption (Figure 7A, B): ATH listeners showed greater disruption early for onset mispronunciations. Later in processing, this relative benefit disappeared and even reversed. Intriguingly no benefit or cost was observed for offset mispronunciations. This null effect is also important. CI users could have shown a larger, more catastrophic decrement, but for offset mispronunciations, they showed no more relative disruption than ATH listeners.

Crucially, in characterizing these differences CI users show a quantitative, not qualitative, departure from ATH listeners. Our goal, was not to determine if CI users exhibit a specific profile of competitoin for mispronounced forms. Rather, CI users are slower and less committed than ATH listeners more generally (even for canonical forms), and this may have carryover benefits for mispronounced forms. Thus, these analyses rested on dynamics of recognizing mispronounced forms, relative to canonical forms. The early benefit observed in MP disruption may thus be an emergent property of a more general profile of processing shown by CI users.

This slower activation pattern or reduced commitment may derive from a number of sources which are not mutually exclusive. CI users may simply be slower because the input is degraded – each segment of the input offers less evidence for a word in the lexicon because it is less discriminable. However, there may also be long-term adaptations to uncertainty as well. For example, McQueen and Huettig (2012) showed that even when the acoustic form of a target word is unaffected, listeners' implicit expectations of a noisy input (e.g., from a carrier phrase) can influence the dynamics of lexical access. However, whether the effects show here are purely a bottom-up phenomenon or also partially expectation driven, the fact that lexical access unfolds at a different rate has emergent consequences for how CI users handle mispronunciations.

This relative benefit for CI users with onset mispronounced stimuli was brief, lasting 200–400 msec (Figure 7; Table 7). However, it was also early in the timecourse of processing, appearing roughly 300 msec into processing (adjusted for the oculomotor delay). In contrast, the only decrement for mispronounced stimuli in CI users was observed late, after about 900 msec (also adjusted). This earliness is worth noting and may be critical to the real world: in everyday language processing, speech arrives quickly enough that listeners do not have 1200 msec (the point at which CI users show a cost, Figure 7, Table 7) to fully process a word. Rather the early states of processing—at which CI users are less affected by degradation—may be the basis of much downstream language processing. Thus, the fact that CI users showed this benefit early could be helpful for everyday language.

These findings rule out two alternatives for how CI users could have responded to mispronunciations. First, CI users could have been insensitive to mispronunciations. If input coding was coarse, mispronunciations would not have registered as incorrect to CI users. However, CI users exhibited both an overall accuracy decrement (Figure 1), and sensitivity to mispronunciations in the fixation record (Figure 3). This suggests the implant is transmitting sufficient information that even a single phonological feature mismatch is noticeable fairly rapidly. Second, mispronunciation did not lead to catastrophic failure to recognize words. The combination of degraded acoustics and non-canonical forms might have overwhelmed CI users. Instead, CI users could cope with them. Although accuracy was worse than ATH listeners, performance did not fall below 85% in any condition (Figure 1). Equally importantly, when we considered the correct trials only, the fixation record did not reflect heightened competition, and in some conditions suggested less disruption (Figure 7).

Critically, it is not that the CI users are objectively "better" than ATH users at perceiving non-canonical forms. Even in onset mispronunciations, Timing in CI users was slower than ATH listeners (Figure 5A), and of course, accuracy in the mispronounced conditions was lower. Rather, this benefit is seen relative to their own baseline. That is, given that CI users are already slower to activate the target word (Figure 2C), onset mispronunciations cause them less additional trouble than they do ATH listeners (Figure 7A, B).

The reduced MP disruption seen in CI users may derive from the fact that lexical commitment is delayed for them (Farris-Trimble et al., 2014). ATH listeners' word recognition is strongly impaired by mispronunciations immediately as they occur. Most likely, ATH listeners rapidly commit to an interpretation that is consistent with the mispronunciation (e.g., *gong* for *gog*); this may make it harder or more difficult to revise later when the final /g/ is heard. In contrast, CI listeners keep their options open, entertaining the target as a possible referent despite the mispronunciation. This makes them more flexible in integrating the mispronounced speech with the referents on the screen. This delayed commitment strategy is not likely unique to mispronounced forms; rather this appears to be something that they engage even for canonical forms, but it may be particularly helpful for mispronounced forms (and revealed by them).

An account based on the slower growth of lexical commitment can also explain why benefits were observed for onset mispronunciations but not offset mispronunciations. When hearing an onset mispronunciation (e.g., *gog* for *dog*) delaying a commitment helps avoid a garden-path interpretation. However, for offset mispronunciations (*dob* for *dog*), a delay leads them to downweight the initial (correct) information, waiting for information that is partially misleading. The fact that this benefit was only observed for onset but not offset mispronunciations suggests that this is not a simple product of CI users' different "baselines" (with correct forms). Recall that for canonical forms CI users show two distinct effects – a slower Timing, and a lower Max. And onset mispronunciations affected Timing, while offset affected Max. If it was simply the case that CI users' relative lack of disruption for mispronounced forms was driven only by differences in the baseline (canonical forms), they should have been seen in both types of mispronunciations. Thus, the pattern of relative benefits is consistent with a theoretical account in which differences in the general properties of competition in CI users have emergent consequences for non-canonical forms.

Delayed commitment may come with costs, however. Late in the trial, ATH listeners recover well from the mispronunciations, and commit more strongly to a single (correct) referent. In contrast, CI users show reduced commitment to correct referents late in the trial—by delaying a commitment they may never reach the same degree of confidence. Moreover, in running speech, this delay could compound across words in a sentence leading to serious costs. While keeping their options open early is helpful, it may be less so later.

A critical question is the source or locus of this delay. Clearly a part of this is due to the fact that the bottom up input is impaired: each portion of the input contains less distinct information to favor one word over another. This would be an immediate effect—independent of long term experience, and consistent with the fact that NH listeners under CI simulations also show delayed lexical access (Farris-Trimble et al., 2014; McMurray et al., 2017). However, part of this may also be a long-term adaptation to uncertainty – listeners implicitly encounter frequent errors of speech perception and at an implicit level this could led them to slow lexical access (as shown by McQueen & Huettig, 2012, in NH listeners). To be clear, we are not proposing that any such adaptations need be a conscious or controlled process (though they could require general resources: Zhang & Samuel, 2018). More likely, such changes (if present) derive from a long term tuning of the lexical system. Critically, however, they derive from experience and they are not solely based on the level of peripheral degradation.

Whether this is purely a bottom-up consequence of the input maintaining partial activation for competitors is theoretically helpful in NH language processing (Clopper & Walker, 2017; McMurray et al., 2009), and this study suggests potentially increased flexibility in CI users (Farris-Trimble et al., 2014; McMurray et al., 2016). There is not yet evidence that it is adaptive for them. If such a strategy is adaptive, how do we account for CI users' reduced accuracy? Importantly, this lexical strategy cannot augment the amount of information in the signal; rather its goal is to preserve flexibility in case later information arrives to constrain the choice. In this study, words were presented in isolation, making the functional benefits of this strategy less relevant (particularly for offset mispronunciations where no further information is coming).

When faced with uncertainty, CI listeners hedge their bets, avoiding strong commitment to one candidate. However, it is not clear how this is achieved. It could be a product of long-term adaptation to CI input, or to a gradual loss of hearing (for some listeners) prior to receiving a CI. This could account for the gradual performance and satisfaction gains over their first year of device use (Hamzavi et al., 2003; Oh et al., 2003; Tyler et al., 1995). However, NH listeners show similar strategies with noise vocoded (CI simulated) speech (Farris-Trimble et al., 2014). This could derive from rapid adaptation, and work in NH listeners shows that perceptual and lexical processes adapt rapidly in a one or two hour session (Bradlow & Bent, 2008; Maye, Aslin, & Tanenhaus, 2008; McQueen & Huettig, 2012).

Alternatively, it could be that this "softening" of the lexical commitment is just what happens to a lexical competition system confronted with noise. The CI users in this study were confronted with both degraded input (due to the CI) and mispronounced forms, while

the ATH listeners heard only mispronounced forms. Several studies have now documented that when ATH listeners hear CI simulation forms of canonical forms, their dynamics of word recognition closely resemble the slower and less committed dynamics of CI users (Farris-Trimble et al., 2014; McMurray et al., 2017). Thus, reduced interference for onset mispronunciations may not be specific to CI users, but something that emerges whenever the input is noisy. These two hypotheses cannot be distinguished by this study—work with CI simulation would help. However, the benefits of slowing lexical access strategy could be attained under either account.

One important remaining issue is the degree to which this approach would generalize beyond the closed set task with isolated words used here. The closed set task was necessary to characterize the dynamics of processing, and to some extent it was intended to model referential constraints potentially imposed by discourse. However, in open set tasks, this slow down could create extra effort, or it could compound over many words in running speech to be more deleterious. Thus, future work should examine generalizability to more realistic tasks.

Alternative explanations.—One alternative interpretation is that these effects are a statistical artifact of CI users' increased error rates for all mispronunciation conditions. While we excluded trials with incorrect responses, these higher errors may have led to differences in eye-movement dynamics unrelated to processing differences. If CI listeners make errors on some proportion of trials, some subset of *correct* trials should have also resulted from guessing; with a 15% error rate, approximately 5% of correct trials would have arisen from lucky guesses (4 alternatives). Eye movements during trials where participants guess are likely to be more uniformly distributed, and so may appear to indicate delayed commitment.

However, this interpretation is unlikely to lead to our results for several reasons. First, error rates were low for CI listeners, suggesting that the number of guessing trials could lead to only small changes in data. Second, effects were robust across degrees of mispronunciation (Figure 6), despite the smaller error rate for single-feature mispronunciations. If effects arose because of lucky guesses, then these effects would be larger when error rates are higher. It is thus unlikely that our pattern of delayed commitment in CI users arises because of guessing.

Individual Differences

We showed relationships between real-time lexical processing and speech perception outcomes. In our hierarchical regression, these real-time measures (particularly Timing) significantly predicted CNC scores over and above accuracy in this task (which itself was a surprisingly strong predictor). Both processing measures were also correlated with each other. These inter-correlations suggest a stable profile of good and poor lexical access in CI users. Thus, measures of real-time processing (e.g., automaticity of word recognition) may be an important target of assessment and remediation in CI users (this may be particulary imporant in prelingually deaf children who show larger delays, McMurray et al., 2017).

CI users with more ATH-like real-time processing (steeper slopes, earlier crossovers, higher maxima) showed better outcomes. This appears to conflict with our claim that deferring a

commitment or "hedging one's bets" is adaptive. It need not, for two reasons. First, as we described these strategies are primarily beneficial for integrating multiple sources of information (e.g., later context). However, in the open-set CNC words measure there is no contextual information. Thus, these strategies would not be helpful in the CNC words task (even as they are in daily life). A more contextualized outcome measure (or even real world measures like listening satisfaction) may pick up such effects. Second, a large portion of the variance in lexical activation dynamics likely reflects peripheral signal encoding (independent of any differences at the lexical level). CI users with better peripheral input can show more ATH-like dynamics because they receive better input. However, over and above that, they may still delay or hedge to manage uncertainty. This requires simultaneously measuring lexical activation dynamics with peripheral encoding, measures which were not available here.

The possibility of adaptive processes is supported by several correlational findings. First, measures like the maximum or slope on correct trials predict performance on mispronounced trials. This suggests some stability in these parameters across CI users. However, the lack of correlation of slope and crossover between correct and mispronounced forms suggests that CI users may be doing something optional with non-canonical stimuli—something that is not solely a product of the peripheral input (perhaps engaged to cope with poorer input). Whether or not this is adaptive is an important avenue for future work.

Conclusions

Non-canonical forms introduce ambiguity for lexical recognition. However, CI listeners' lexical recognition was robust in the face of non-canonical inputs. CI users commit more slowly or less fully to lexical candidates, and this may be adaptive when coping with degraded or non-canonical input. Consequently, early in processing CI users showed less sensitivity to mispronunciation. This suggests that the dynamics of spoken word recognition may exhibit a different profile after cochlear implantation, exhibiting a profile of processing that may be attuned to the level of uncertainty in the input.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

Allopenna P, Magnuson JS, & Tanenhaus MK (1998). Tracking the time course of spoken word recognition using eye-movements: evidence for continuous mapping models. Journal of Memory and Language, 38(4), 419–439.

Balkany T, Hodges A, Menapace C, Hazard L, Driscoll C, Gantz BJ, . . . Payne S (2007). Nucleus Freedom North American clinical trial. Otolaryngology Head and Neck Surgery, 136, 757–762. [PubMed: 17478211]

Bard EG, Shillcock RC, & Altmann GT (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. Perception & Psychophysics, 44(5), 395–408. [PubMed: 3226889]

Bent T, & Holt RF (2013). The influence of talker and foreign-accent variability on spoken word identification. The Journal of the Acoustical Society of America, 133(3), 1677–1686. [PubMed: 23464037]

Boersma P, & Weenink D (2009). Praat: doing phonetics by computer (Version Version 5.1.05). Retrieved from http://www.praat.org/

Bradlow AR, & Bent T (2008). Perceptual adaptation to non-native speech. Cognition, 106(2), 707–729. [PubMed: 17532315]

Clopper C, & Walker A (2017). Effects of Lexical Competition and Dialect Exposure on Phonological Priming. Language and Speech, 60(1), 85–109. [PubMed: 28326994]

Connine CM, Blasko D, & Hall M (1991). Effects of subsequent sentence context in auditory word recognition: temporal and linguistic constraints. Journal of Memory and Language, 30, 234–250.

Connine CM, Blasko D, & Titone D (1993). Do the beginnings of spoken words have a special status in auditory word recognition? Journal of Memory and Language, 32, 193–210.

Creel SC, Aslin RN, & Tanenhaus MK (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. Journal of Memory and Language, 54(1), 1–19.

Dahan D, & Magnuson JS (2006). Spoken-word recognition In Traxler MJ & Gernsbacher MA (Eds.), Handbook of Psycholinguistics (pp. 249–283). Amsterdam: Academic Press.

Dahan D, Magnuson JS, Tanenhaus MK, & Hogan E (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes, 16(5/6), 507–534.

Daniloff R, & Moll K (1968). Coarticulation of Lip Rounding. Journal of Speech, Language, and Hearing Research, 11(4), 707–721.

Dorman MF, Gifford R, Lewis K, McKarns S, Ratigan J, Spahr A, . . . Thedinger BS (2009). Word recognition following implantation of conventional and 10-mm hybrid electrodes. Audiology and Neurotology, 14(3), 181–189. [PubMed: 19005252]

Dorman MF, Hannley MT, Dankowski K, Smith L, & McCandless G (1989). Word recognition by 50 patients fitted with the symbion multichannel cochlear implant. Ear and Hearing, 10, 44–49. [PubMed: 2721828]

Dorman MF, Loizou P, Wang S, Zhang T, Spahr A, Loiselle L, & Cook S (2014). Bimodal Cochlear Implants: The Role of Acoustic Signal Level in Determining Speech Perception Benefit. Audiology and Neurotology, 19(4), 234–238. [PubMed: 24992987]

Dunn CC, Perreau A, Gantz B, & Tyler RS (2010). Benefits of localization and speech perception with multiple noise sources in listeners with a short-electrode cochlear implant. Journal of the American Academy of Audiology, 21(1), 44. [PubMed: 20085199]

Dunn DM, & Dunn LM (2007). Peabody picture vocabulary test: Manual: Pearson Assessments.

Ernestus M, & Warner N (2011). An introduction to reduced pronunciation variants. Journal of Phonetics, 39(3), 253–260.

Farris-Trimble A, & McMurray B (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. Journal of Speech Language and Hearing Research, 56, 1328–1345.

Farris-Trimble A, McMurray B, Cigrand N, & Tomblin JB (2014). The process of spoken word recognition in the face of signal degradation: Cochlear implant users and normal-hearing listeners. Journal of Experimental Psychology: Human Perception and Performance, 40(1), 308–327. [PubMed: 24041330]

Fernald A, Perfors A, & Marchman VA (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. Developmental Psychology, 42(1), 98–116. [PubMed: 16420121]

Fetterman BL, & Domico EH (2002). Speech Recognition in Background Noise of Cochlear Implant Patients. Otolaryngology -- Head and Neck Surgery, 126(3), 257–263. [PubMed: 11956533]

Fishman KE, Shannon RV, & Slattery WH (1997). Speech Recognition as a Function of the Number of Electrodes Used in the SPEAK Cochlear Implant Speech Processor. Journal of Speech, Language, and *Hearing Research*, 40(5), 1201–1215.

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

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

Gantz BJ, Turner C, & Gfeller KE (2006). Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/nucleus hybrid implant. Audiology and Neurotology, 11(Suppl 1), 63–68. [PubMed: 17063013]

Gaskell MG, & Marslen-Wilson WD (1996). Phonological variation and inference in lexical access. Journal of Experimental Psychology: Human Perception and Performance, 22(1), 144. [PubMed: 8742258]

Gifford RH, Dorman MF, McKarns SA, & Spahr AJ (2007). Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. Journal of Speech, Language, and Hearing Research, 50(4), 835–843.

Gifford RH, Dorman MF, Skarzynski H, Lorens A, Polak M, Driscoll CLW, . . . Buchman CA (2013). Cochlear implantation with hearing preservation yields significant benefit for speech recognition in complex listening environments. Ear and Hearing, 34(4), 413–425. [PubMed: 23446225]

Gow DW (2003). Feature parsing: Feature cue mapping in spoken word recognition. Perception & Psychophysics, 65(4), 575–590. [PubMed: 12812280]

Green KM, Julyan PJ, Hastings DL, & Ramsden RT (2005). Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness. Hearing Research, 205(1), 184–192. [PubMed: 15953527]

Hamzavi J, Baumgartner W. d., Pok SM, Franz P, & Gstoettner W (2003). Variables affecting speech perception in postlingually deaf adults following cochlear implantation. Acta oto-laryngologica, 123(4), 493–498. [PubMed: 12797584]

Helms J, Müller J, Schön F, Moser L, Arnold W, Janssen T, . . . Hochmair-Desoyer I (1997). Evaluation performance with the COMBI 40 Cochlear Implant in adults: A multicentric clinical study. Journal for Oto-Rhino-Laryngology, Head and Neck Surgery, 59, 23–35.

Henry BA, Turner CW, & Behrens A (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. The Journal of the Acoustical Society of America, 118(2), 1111–1121. [PubMed: 16158665]

Hirschfelder A, Gräbel S, & Olze H (2008). The impact of cochlear implantation on quality of life: The role of audiologic performance and variables. Otolaryngology -- Head and Neck Surgery, 138(3), 357–362. [PubMed: 18312885]

Holden LK, Finley CC, Firszt JB, Holden TA, Brenner C, Potts LG, . . . Skinner MW. (2013). Factors Affecting Open-Set Word Recognition in Adults with Cochlear Implants. Ear and Hearing, 34(3), 342–360. [PubMed: 23348845]

Kapnoula E, & McMurray B (2016). Inhibitory processes are plastic: Training alters competition between words. Journal of Experimental Psychology: General, 145(1), 8–30. [PubMed: 26709587]

Kapnoula E, Packard S, Gupta P, & McMurray B (2015). Immediate lexical integration of novel word forms. Cognition, 134(1), 85–99. [PubMed: 25460382]

Kraljic T, Brennan SE, & Samuel AG (2008). Accommodating variation: Dialects, idiolects, and speech processing. Cognition, 107(1), 54–81. [PubMed: 17803986]

Ladefoged P, & Broadbent DE (1957). Information conveyed by vowels. The Journal of the Acoustical Society of America, 29(1), 98–104.

Lassaletta L, Castro A, Bastarrica M, Sarriá MJ, & Gavilán J (2005). Quality of life in postlingually deaf patients following cochlear implantation. European Archives of Oto-Rhino-Laryngology and Head & Neck, 263(3), 267–270.

Luce PA, & Cluff MS (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. Perception & Psychophysics, 60(3), 484–490. [PubMed: 9599997]

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

Magnuson JS, Dixon J, Tanenhaus MK, & Aslin RN (2007). The dynamics of lexical competition during spoken word recognition. Cognitive Science, 31, 1–24. [PubMed: 21635286]

Marslen-Wilson WD (1987). Functional parallelism in spoken word recognition. Cognition, 25(1–2), 71–102. [PubMed: 3581730]

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

Matuschek H, Kliegl R, Vasishth S, Baayen H, & Bates D (2017). Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305–315.

Maye J, Aslin RN, & Tanenhaus MK (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. Cognitive Science: A Multidisciplinary Journal, 32(3), 543–562.

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

McMurray B, Farris-Trimble A, & Rigler H (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. Cognition, 169, 147–164. [PubMed: 28917133]

McMurray B, Farris-Trimble A, Seedorff M, & Rigler H (2016). The effect of residual acoustic hearing and adaptation to uncertainty in Cochlear Implant users. Ear and Hearing, 37(1), 37–51.

McMurray B, & Jongman A (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review, 118(2), 219–246. [PubMed: 21417542]

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

McMurray B, Tanenhaus MK, & Aslin RN (2009). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. Journal of Memory and Language, 60(1), 65–91. [PubMed: 20046217]

McQueen JM, & Huettig F (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. The Journal of the Acoustical Society of America, 131(1), 509–517. [PubMed: 22280612]

Mehr MA, Turner CW, & Parkinson A (2001). Channel weights for speech recognition in cochlear implant users. The Journal of the Acoustical Society of America, 109(1), 359–366. [PubMed: 11206164]

Milberg W, Blumstein S, & Dworetzky B (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. Bulletin of the Psychonomic Society, 26(4), 305–308.

Moberly AC, Lowenstein JH, Tarr E, Caldwell-Tarr A, Welling DB, Shahin AJ, & Nittrouer S (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? Journal of Speech, Language, and Hearing Research, 57(2), 566–582.

Neher T, Laugesen S, Søgaard Jensen N, & Kragelund L (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? a). The Journal of the Acoustical Society of America, 130(3), 1542–1558. [PubMed: 21895093]

Newman R, Clouse S, & Burnham J (2001). The perceptual consequences of within-talker variability in fricative production. Journal of the Acoustical Society of America, 109, 1181–1196. [PubMed: 11303932]

Norris D (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52(3), 189–234.

Oh S. h., Kim C. s., Kang EJ, Lee DS, Lee HJ, Chang SO, . . . Koo JW (2003). Speech perception after cochlear implantation over a 4-year time period. Acta oto-laryngologica, 123(2), 148–153. [PubMed: 12701730]

Oleson JJ, Cavanaugh JE, McMurray B, & Brown G (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. Statistical Methods in Medical Research, 26(6), 2708–2725. [PubMed: 26400088]

Ranbom L, & Connine CM (2007). Lexical representation of phonological variation. Journal of Memory and Language, 57, 273–298.

Reiss LAJ, Perreau AE, & Turner CW (2012). Effects of Lower Frequency-to-Electrode Allocations on Speech and Pitch Perception with the Hybrid Short-Electrode Cochlear Implant. Audiology and Neurotology, 17(6), 357–372. [PubMed: 22907151]

Rigler H, Farris-Trimble A, Greiner L, Walker J, Tomblin JB, & McMurray B (2015). The slow developmental timecourse of real-time spoken word recognition. Developmental Psychology, 51(12), 1690–1703. [PubMed: 26479544]

Salverda AP, Brown M, & Tanenhaus MK (2011). A goal-based perspective on eye movements in visual world studies. Acta Psychologica, 137(2), 172–180. [PubMed: 21067708]

Scheperle RA, & Abbas PJ (2015). Relationships Among Peripheral and Central Electrophysiological Measures of Spatial and Spectral Selectivity and Speech Perception in Cochlear Implant Users. Ear and Hearing, 36(4), 441–453. [PubMed: 25658746]

Seedorff M, Oleson JJ, & McMurray B (in press). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more).

Sekerina IA, & Brooks PJ (2007). Eye movements during spoken word recognition in russian children. Journal of Experimental Child Psychology, 98, 20–45. [PubMed: 17560596]

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

Summerfield Q (1981). Articulatory rate and perceptual constancy in phonetic perception. Journal of the Acoustical Society of America, 7(5), 1074–1095.

Summers V, & Leek MR (1994). The internal representation of spectral contrast in hearing-impaired listeners. The Journal of the Acoustical Society of America, 95(6), 3518–3528. [PubMed: 8046143]

Swingley D (2009). Onsets and codas in 1.5-year-olds' word recognition. Journal of Memory and Language, 60, 252–269. [PubMed: 20126290]

Szostak CM, & Pitt MA (2013). The prolonged influence of subsequent context on spoken word recognition. Attention, Perception, & Psychophysics, 75(7), 1533–1546.

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

Toscano JC, Anderson ND, & McMurray B (2013). Reconsidering the role of temporal order in spoken word recognition. Psychonomic Bulletin & Review, 20, 1–7. [PubMed: 23090749]

Tyler RS, Lowder MW, Parkinson AJ, Woodworth GG, & Gantz BJ (1995). Performance of adult ineraid and nucleus cochlear implant patients after 3.5 years of use. Audiology, 34, 135–144. [PubMed: 8561691]

Tyler RS, Parkinson AJ, Wilson BS, Witt S, Preece JP, & Noble W (2002). Patients utilizing a hearing aid and a cochlear implant: speech perception and localization. Ear and Hearing, 23(2), 98–105. [PubMed: 11951854]

Tyler RS, Parkinson AJ, Woodworth GG, Lowder MW, & Gantz BJ (1997). Performance over time of adult patients using the Ineraid or Nucleus cochlear implant. The Journal of the Acoustical Society of America, 102(1), 508–522. [PubMed: 9228814]

Weber A, & Scharenborg O (2012). Models of spoken-word recognition. Wiley Interdisciplinary Reviews: Cognitive Science, 3(3), 387–401. [PubMed: 26301470]

Wechsler D, & Hsiao-Pin C (2011). WASI-II: Wechsler abbreviated scale of intelligence: Pearson Assessments.

White KS, Yee E, Blumstein SE, & Morgan JL (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. Journal of Memory and Language, 68(4), 362–378. [PubMed: 24065868]

Wilson BS, & Dorman MF (2008). Cochlear implants: A remarkable past and a brilliant future. Hearing Research, 242, 3–21. [PubMed: 18616994]

Winn MB, Chatterjee M, & Idsardi WJ (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. The Journal of the Acoustical Society of America, 131(2), 1465–1479. [PubMed: 22352517]

Zhang X, & Samuel AG (2018). Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. Journal of Memory and Language, 100, 32–50.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
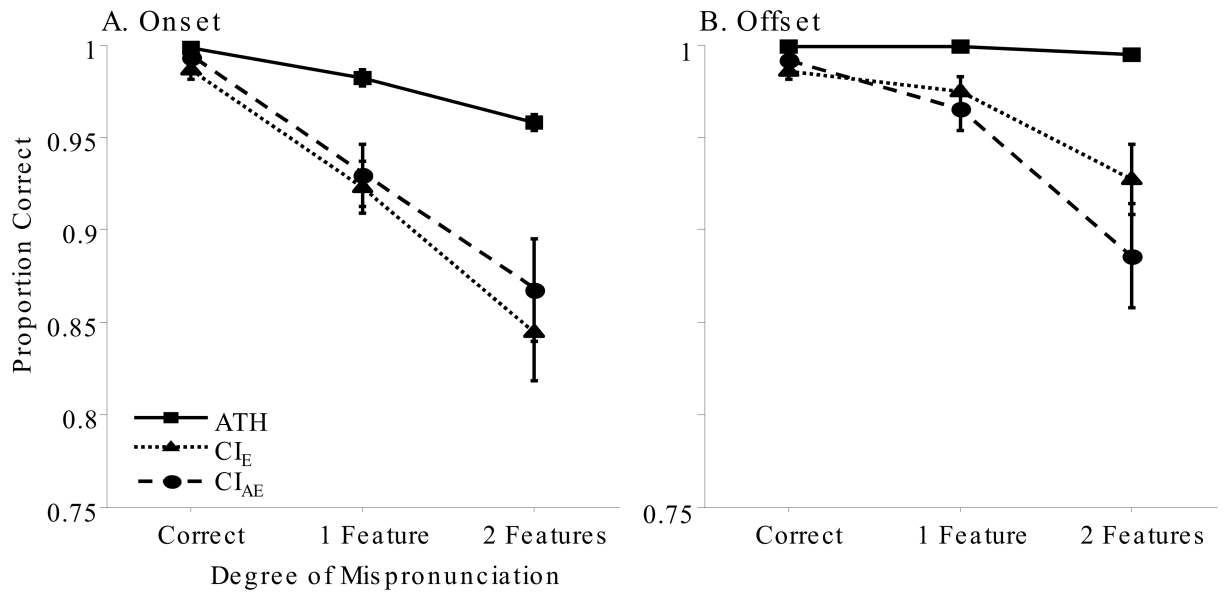
Figure 1:
Average accuracy as a function of listener group and degree of mismatch for onset mispronunciations (panel A) and offset mispronunciations (panel B).
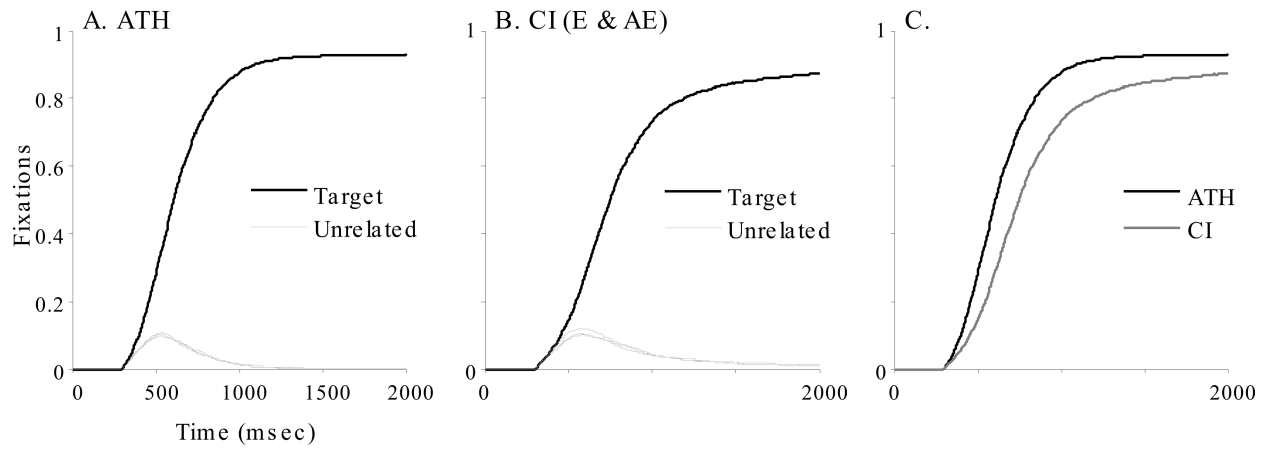
Figure 2:
A) Average fixations to the target and the three unrelated objects as a function of time for NH listeners on correct trials; B) for both groups of CI users; C) Target fixations for ATH and CI listeners.
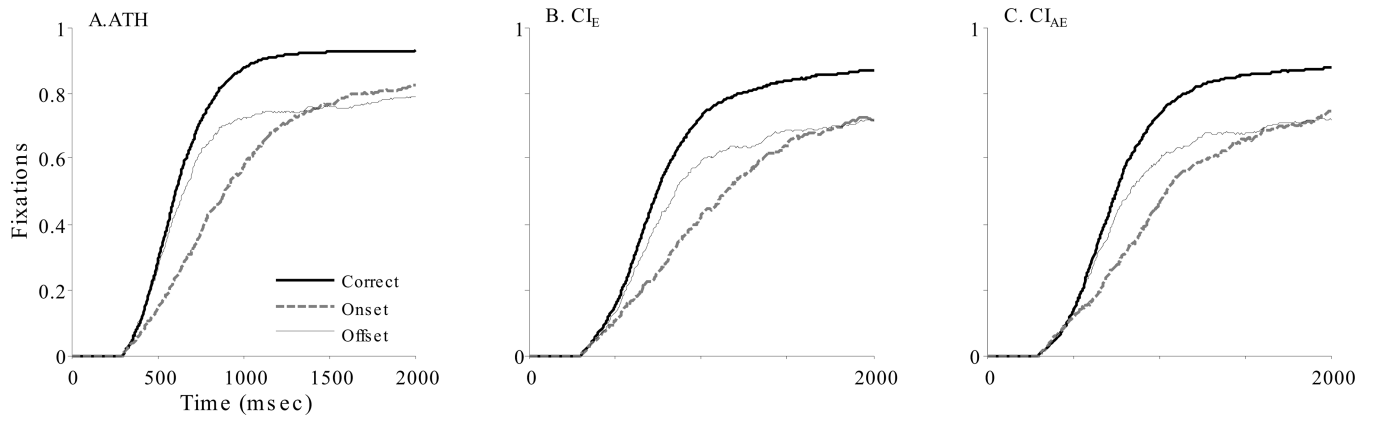
Figure 3:
Effect of mispronunciation type (single feature only) on fixations to the target for A) ATH listeners; B) CI_E users and C) CI_{AE} users.

The logistic function graph shows Fixations (y-axis, 0 to 1) versus Time (msec) (x-axis, 0 to 2500), with labels for Max, Min, Slope, and Crossover. The inset equation reads:

$$\text{Logistic}(T) = \cfrac{\text{max - min}}{1 + e^{\left(\frac{-4 \cdot \text{slope}}{\text{max - min}}(T - \text{crossover})\right)}} + \text{min}$$
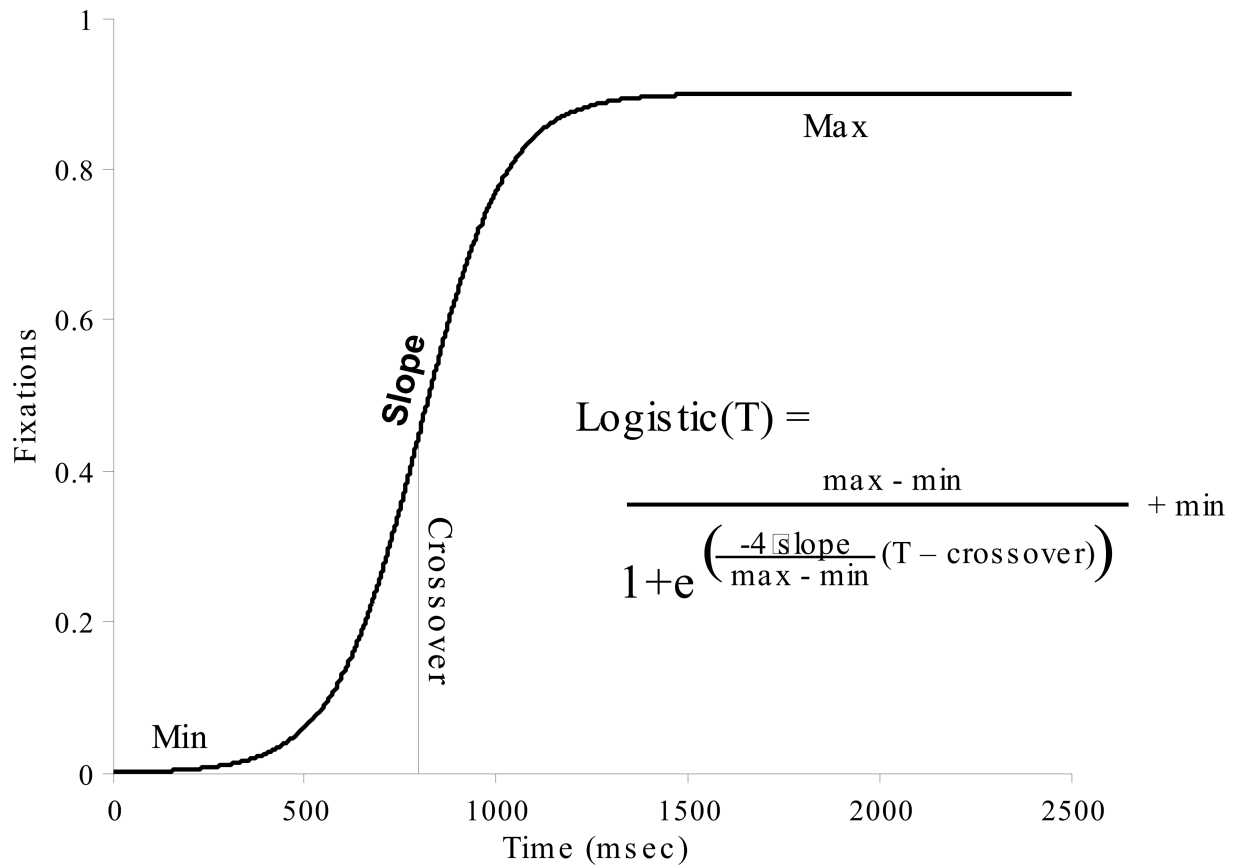
Figure 4:

Logistic function used to model target fixations. Equation is given in inset. This function has four parameters: min (the lower asymptote), max (the upper asymptote), crossover (the point in time where the function is halfway between min and max, and slope (the derivative at the crossover).
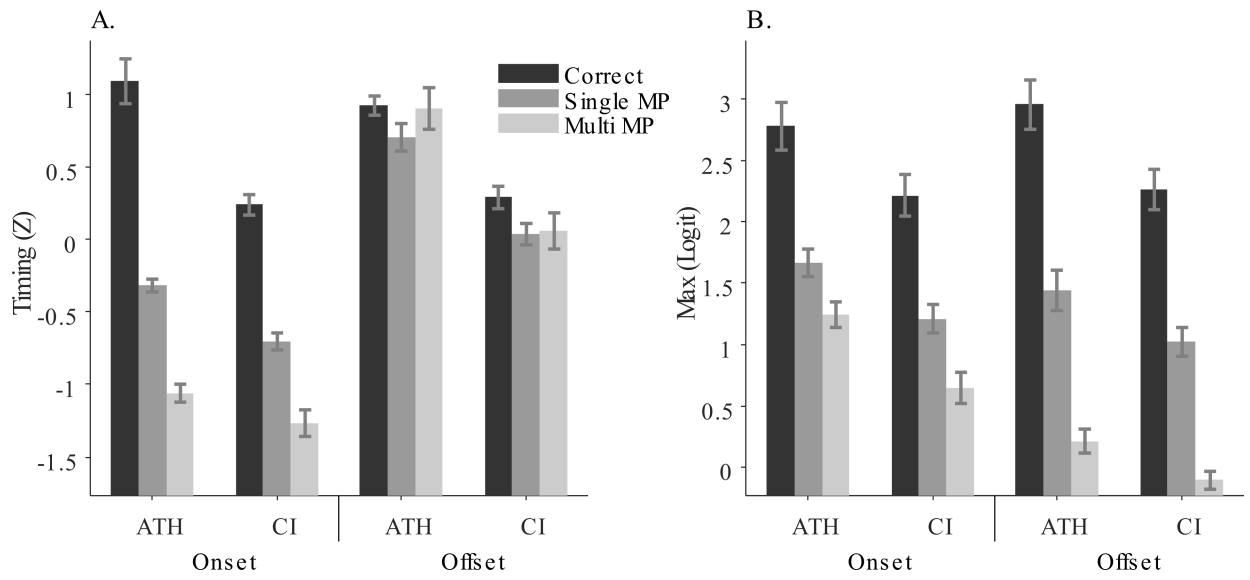
Figure 5:
Timing (A) and Max (B) parameters as a function of Location of mispronunciation and CI use (X axis), and type of mispronunciation (grouped bars)
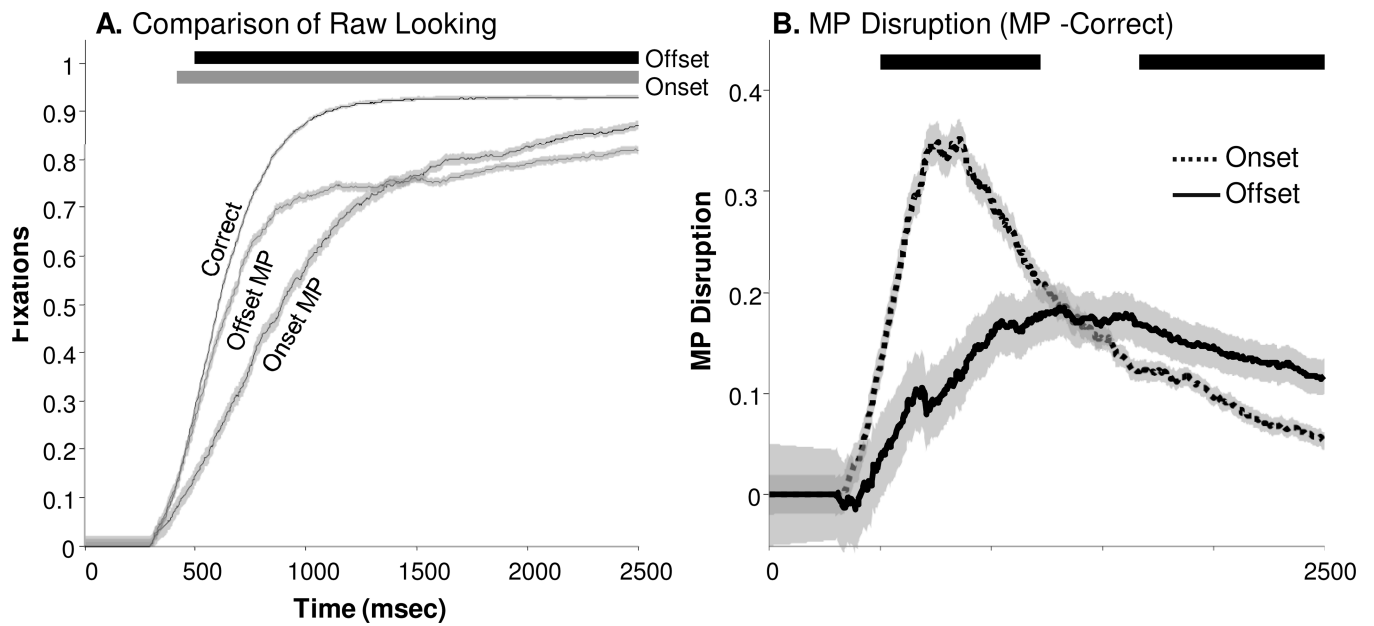
Figure 6:
BDOTS analysis of ATH listeners hearing correct or single-feature mismatch stimuli. Black bar indicates significant region. See Table 5 for statistical results. A) BDOTS directly comparing looks to target between stimulus conditions (Table 5, Row 1); B) BDOTS analysis of difference in fixations (MP disruption measure) comparing onset and offset mispronunciations
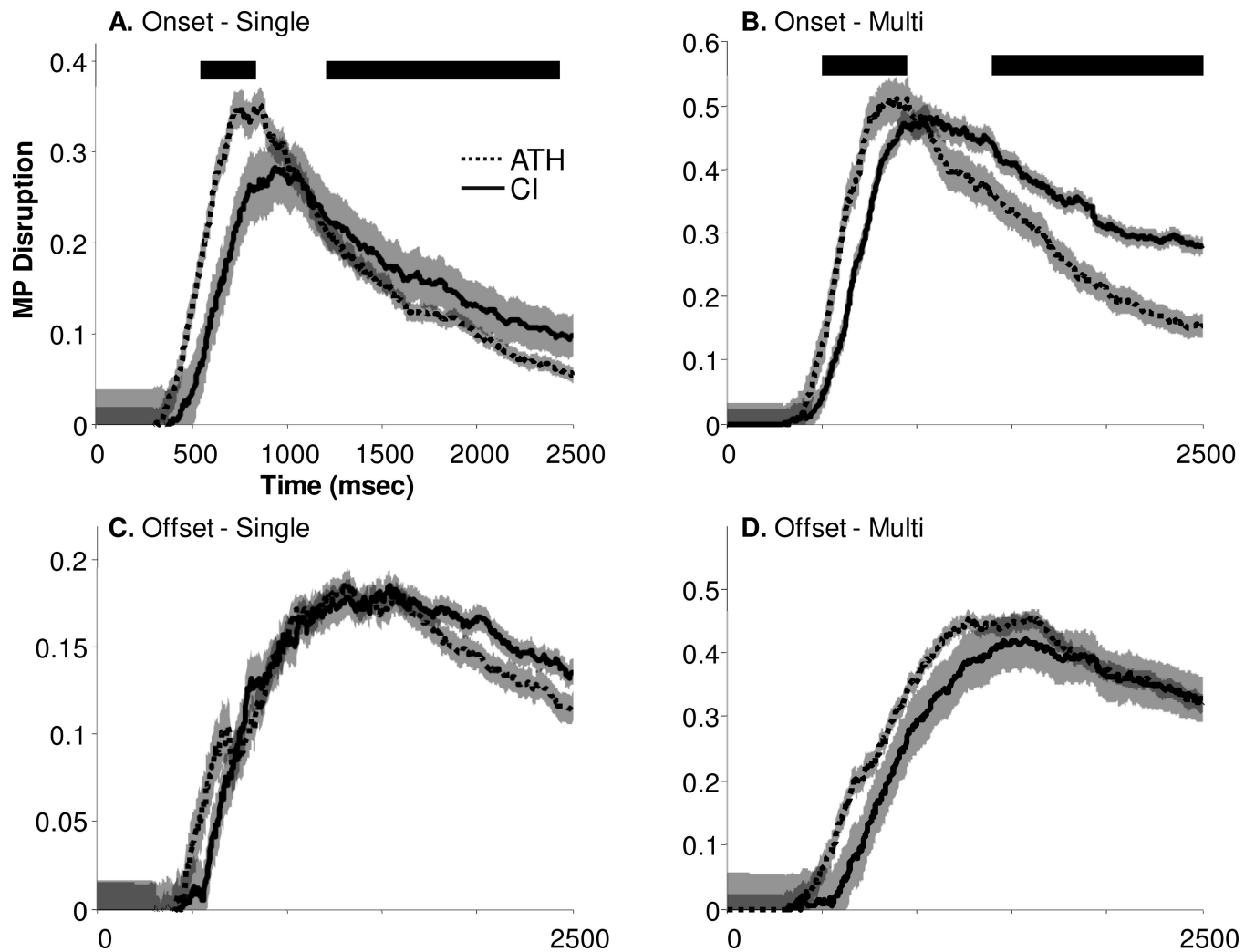
Figure 7:
Comparison of MP disruption effect (difference in target fixations between correct and incorrect forms) between CI users and NH listeners. Significant regions are indicated with black bars (see Table 6 for corresponding statistics). A high MP disruption indicates more difficulty created by non-canonical form. A) Single feature onset mispronunciations; B) Multi-feature onset mispronunciation; C) Single feature offset mispronunciation; D) Multi-feature offset mispronunciation.

Table 1:

Summary of participant characteristics. Numbers in parentheses reflect SD. WASI scores reflect average of two subtests converted to a standard score.

	ATH	CI_E	CI_{AE}
N	28	23	21
Devices		Bilateral: 13 Unilateral: 10	Bimodal: 7 Hybrid: 14
Age (years)	52.1 (9.4)	55.0 (6.9)	52.4 (12.1)
Age of Implant (years)		41.5 (8.4)	47.6 (11.4)
Device Use (years)		13.5 (7.3)	4.9 (3.8)
PPVT (standard)	110.3 (9.9)	103.0 (12.7)	101.8 (11.8)
WASI (standard)	97.3(10.0)	98.0(12.2)	99.5 (10.5)

Table 2:

Results of a binomial mixed effects model examining accuracy as a function of listener group, location of mispronunciation and degree of mispronunciation.

	B	SE	Z	P	
(Intercept)	5.39	0.21	25.4	<.0001	*
CI vs ATH	2.08	0.38	5.5	<.0001	*
CI_E vs CI_{AE}	0.34	0.35	1.0	0.33	
Mispronunciation	−2.02	0.28	−7.3	<.0001	*
Degree	−1.40	0.27	−5.1	<.0001	*
Location	−0.83	0.26	−3.2	0.0015	*
Location × Mispron	−1.16	0.52	−2.2	0.027	*
× Degree	0.67	0.55	1.2	0.23	
Mispron × CI vs ATH	0.55	0.54	1.0	0.31	
× CI_E vs CI_{AE}	−0.90	0.29	−3.1	0.0021	**
Degree × CI vs ATH	−0.15	0.57	−0.3	0.79	
× CI_E vs CI_{AE}	−0.10	0.22	−0.5	0.65	
Location × CI vs ATH	−1.18	0.54	−2.2	0.03	*
× CI_E vs CI_{AE}	0.40	0.30	1.3	0.19	
Location × Mispron × CI v ATH	−0.76	1.10	−0.7	0.49	
× CI_E vs CI_{AE}	0.57	0.65	0.9	0.38	
Location × Degree × CI v ATH	0.54	1.15	0.5	0.64	
× CI_E vs CI_{AE}	0.41	0.48	0.9	0.39	

Table 3:

Results of two mixed models examining estimated parameters of the target fixations as a function of mispronunciation location and degree for the ATH listeners.

DV	Effect	B	SE	df	t	P	
Max	Location	0.07	0.009	134.1	7.9	<.0001	*
	Mispronunciation	−0.14	0.010	134.1	−13.9	<.0001	*
	Single vs. Multi	−0.13	0.012	134.1	−11.0	<.0001	*
	Location × Mispron	0.12	0.020	134.1	6.2	<.0001	*
	Location × Degree	0.16	0.023	134.1	7.2	<.0001	*
Timing	Location	−1.04	0.059	133.8	−17.6	<.0001	*
	Mispronunciation	−0.91	0.063	133.8	−14.5	<.0001	*
	Single vs. Multi	−0.35	0.073	133.9	−4.9	<.0001	*
	Location × Mispron	−1.66	0.125	133.8	−13.2	<.0001	*
	Location × Degree	−1.04	0.145	133.9	−7.2	<.0001	*

*
= p<.05.

Table 4:

Results of mixed models examining estimated parameters of the target fixations as a function of listener group.

Parameter	Effect	B	SE	t	df	p	
CI vs. NH	Max	0.06	0.026	2.2	69	0.031	*
	Timing	0.68	0.102	6.7	69	<.0001	*
CI_E vs. CI_{AE}	Max	0.01	0.032	0.3	69	0.77	
	Timing	0.04	0.128	0.3	69	0.75	

*
: $p<.05$.

Table 5:

Results of omnibus mixed models predicting each logistic parameter as a function of mispronunciation location and degree, as well as the moderation of those factors by listener type. Note that primary factors (e.g., from Table 3) are indicated with a gray background, moderators in white.

	Predictor	DV	B	SE	t	df	P	
1	CI vs ATH	Max	0.070	0.020	3.4	68.9	0.001	*
		Timing	0.609	0.088	6.9	68.7	<.0001	*
2	CI_E v CI_{AE}	Max	−0.014	0.026	−0.5	69.1	0.6	
		Timing	0.195	0.110	1.8	69.1	0.081	+
3	Location	Max	0.067	0.008	8.4	69.9	<.0001	*
		Timing	−0.887	0.056	−15.9	69.9	<.0001	*
4	× CI v ATH	Max	0.012	0.016	0.7	69.5	0.47	
		Timing	−0.254	0.114	−2.2	69.6	0.029	*
5	× CI_E v CI_{AE}	Max	0.010	0.021	0.5	70.8	0.62	
		Timing	0.100	0.143	0.7	70.7	0.48	
6	Mispronounced	Max	−0.167	0.008	−21.9	68.3	<.0001	*
		Timing	−0.848	0.044	−19.1	270.4	<.0001	*
7	× CI v ATH	Max	0.049	0.016	3.1	68.1	0.0026	*
		Timing	−0.109	0.091	−1.2	270.2	0.23	
8	× CI_E v CI_{AE}	Max	−0.027	0.020	−1.4	68.9	0.17	
		Timing	0.234	0.114	2.1	270.9	0.041	*
9	Degree	Max	−0.157	0.011	−14.6	67.2	<.0001	*
		Timing	−0.345	0.052	−6.7	272.3	<.0001	*
10	× CI v ATH	Max	0.053	0.022	2.4	66.7	0.019	*
		Timing	−0.024	0.106	−0.2	271.8	0.82	
11	× CI_E v CI_{AE}	Max	−0.075	0.028	−2.7	68.2	0.0085	*
		Timing	0.237	0.134	1.8	273.7	0.077	+
12	Location × Mispron	Max	0.112	0.013	8.7	130.0	<.0001	*
		Timing	−1.331	0.089	−15.0	270.4	<.0001	*
13	× CI v ATH	Max	0.019	0.026	0.7	129.6	0.48	
		Timing	−0.535	0.182	−2.9	270.2	0.0036	*
14	× CI_E v CI_{AE}	Max	0.002	0.033	0.0	130.9	0.96	
		Timing	0.021	0.228	0.1	270.8	0.93	
15	Location × Degree	Max	0.154	0.015	10.2	133.6	<.0001	*
		Timing	−0.809	0.104	−7.8	272.3	<.0001	*
16	× CI v ATH	Max	0.016	0.031	0.5	132.4	0.61	
		Timing	−0.370	0.212	−1.7	271.7	0.082	+
17	× CI_E v CI_{AE}	Max	0.031	0.039	0.8	136.3	0.43	
		Timing	−0.022	0.267	−0.1	273.5	0.94	

*: p<.05

$^{+}$: p<.1.

Table 6:

Results of BDOTS analysis for Figure 6. For fit quality, number of fits are shown either assuming autoregressive error (AR1) or not (if better fits could be obtained without it).

Panel	Comparison	Fit quality			Subjects	Statistics	Significant Regions
		Criteria	AR1	Non AR1			
6A	Correct vs. Onset MP	$R^2>=0.95$ $R^2>=0.8$ $R^2<0.8$ Dropped	53 0 0 0	3 0 0	NH: 28	$\alpha^* = 0.0014$ $\rho = 0.9982$	Correct > Onset 420 – 3000
	Correct vs. Offset MP	$R^2>=0.95$ $R^2>=0.8$ $R^2<0.8$ Dropped	44 6 0 1	5 0 0	NH: 27	$\alpha^* = 0.0012$ $\rho = 0.9972$	Correct > Offset 496 – 3000
6B	Onset MP vs. Offset MP	$R^2>=0.95$ $R^2>=0.8$ $R^2<0.8$ Dropped	100 6 0 0	5 1 0	NH: 28	$\alpha^* = 0.0018$ $\rho = 0.9990$	Onset > Offset 504 – 1220
							Offset > Onset 1672 – 3000

* α is the α adjusted for family wise error

ρ is the autocorrelation among t-statistics used to compute α^*

Significant regions described in msec.

Table 7:

Results of BDOTS analysis for Figure 6 comparing ATH and CI users within sub conditions.

Panel	Condition	Fit quality			Subjects	Statistics	Significant Regions
		Criteria	AR1	Non AR1			
7A	Onset / single feature	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	125 1 0 3	11 1 0	NH: 28 CI: 41	$\alpha^* = 0.0021$ $\rho = 0.9994$	CI < ATH 540 – 832 ATH < CI 1212 – 2436
7B	Onset / multi feature	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	97 6 0 15	11 2 0	ATH: 25 CI: 32	$\alpha^* = 0.0018$ $\rho = 0.9989$	CI < ATH 532 – 924 ATH < CI 1684 – 3000
7B*	Onset / multi feature (restricted sample)	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	81 5 0 27	8 9 0	ATH: 19 CI: 26	$\alpha^* = 0.0018$ $\rho = 0.9988$	CI < ATH 500 – 944 ATH < CI 1400 – 3000
7C	Offset / single feature	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	112 9 0 1	18 3 0	ATH: 28 CI: 43	$\alpha^* = 0.0022$ $\rho = 0.9995$	$\varnothing$
7C*	Offset / single feature (restricted sample)	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	111 9 0 4	14 2 0	ATH: 27 CI: 41	$\alpha^* = 0.0022$ $\rho = 0.9994$	$\varnothing$
7D	Offset / Multi feature	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	67 30 4 11	6 13 4	ATH: 25 CI: 36	$\alpha^* = 0.0026$ $\rho = 0.9996$	$\varnothing$
7D*	Offset / Multi feature (restricted sample)	$R^2>=0.95$ $R^2>=0.8$ $R^2<.0.8$ Dropped	56 24 4 23	5 9 4	ATH: 22 CI: 27	$\alpha^* = 0.0021$ $\rho = 0.9991$	$\varnothing$

*refers to the adjusted α.

ρ is the observed autocorrelation among test statistics. Significant regions are given in msec.

Table 8:

Correlation of accuracy and eye-movements parameters in each condition with CNC scores. MP refers to difference between correctly and mispronounced (multiple) conditions.

	Accuracy	Max	Timing
Correctly pronounced	0.727*	0.436*	0.666*
Onset – Single	0.594*	0.231	0.272+
Onset – Multiple	0.623*	0.321*	0.049
Offset – Single	0.615*	0.351*	0.540*
Offset – Multiple	0.707*	−0.088	0.437*
MP Onset - Single	−0.359*		0.413*
MP Onset – Multiple	−0.563*		0.325*
MP Offset - Single	−0.510*	0.284+	
MP Offset - Multiple	−0.672*	0.514*	

*: p<.05

+: p<.1.

Non-significant correlations are in gray text.

Table 9:

Correlations between properties of the fixation dynamics on correct trials (e.g., the general dynamics of lexical access) and the corresponding properties on mispronounced trials.

Trial Type	Property	
	Max	**Timing**
Onset – Single	0.758[*]	.261[+]
Onset – Multiple	0.601[*]	−.034
Offset – Single	0.772[*]	0.791[*]
Offset - Multiple	0.383[*]	0.699[*]

[*] : p<.05

[+] : p<.1

non-significant correlations are in gray text.