Cognition 169 (2017) 147–164

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally

Bob McMurray[a,b,c,d,*], Ashley Farris-Trimble[e], Hannah Rigler[a]

[a] Dept. of Psychological and Brain Sciences, University of Iowa, United States
[b] Dept. of Communication Sciences and Disorders, University of Iowa, United States
[c] Dept. of Otolaryngology, University of Iowa, United States
[d] DeLTA Center, University of Iowa, United States
[e] Dept. of Linguistics, Simon Fraser University, Canada

ARTICLE INFO

Keywords:
Speech perception
Spoken word recognition
Cochlear implants
Incremental processing
Vocoded speech
Lexical access

ABSTRACT

Spoken language unfolds over time. Consequently, there are brief periods of ambiguity, when incomplete input can match many possible words. Typical listeners solve this problem by immediately activating multiple candidates which compete for recognition. In two experiments using the visual world paradigm, we examined real-time lexical competition in prelingually deaf cochlear implant (CI) users, and normal hearing (NH) adults listening to severely degraded speech. In Experiment 1, adolescent CI users and NH controls matched spoken words to arrays of pictures including pictures of the target word and phonological competitors. Eye-movements to each referent were monitored as a measure of how strongly that candidate was considered over time. Relative to NH controls, CI users showed a large delay in fixating any object, less competition from onset competitors (e.g., *sandwich* after hearing *sandal*), and increased competition from rhyme competitors (e.g., *candle* after hearing *sandal*). Experiment 2 observed the same pattern with NH listeners hearing highly degraded speech. These studies suggests that in contrast to all prior studies of word recognition in typical listeners, listeners recognizing words in severely degraded conditions can exhibit a substantively different pattern of dynamics, waiting to begin lexical access until substantial information has accumulated.

1. Introduction

Language unfolds over time, and early portions of the signal are often insufficient to recognize a word. For example, a partial auditory input like /wɪ…/ is consistent with *wizard, with, winner,* and *will,* and this ambiguity will not be resolved for several hundred milliseconds. Consequently, even a clearly articulated word has a large (but temporary) form of ambiguity among many lexical candidates. As a result of this, even normal hearing (NH) adults confront and manage a brief period of ambiguity every time they recognize a word. This process is now well understood in typical listeners (Dahan & Magnuson, 2006; Weber & Scharenborg, 2012). And understanding how typical listeners deal with this normal temporary ambiguity, may help understand situations in which listeners confront much greater ambiguity, for example, listeners who face significant loss of acoustic detail because they use a cochlear implant (CI).

There is consensus that NH listeners solve the problem of temporary ambiguity with some version of immediate competition (Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris & McQueen, 2008). As listeners hear a word, multiple candidates are partially activated in parallel (Allopenna, Magnuson, & Tanenhaus, 1998; Marslen-Wilson & Zwitserlood, 1989). As the signal unfolds, some candidates drop out of consideration (Frauenfelder, Scholten, & Content, 2001), and more active words inhibit less active ones (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998) until only a single word remains. The alternative—what we term a *wait-and-see* approach—suggests information accumulates in a memory buffer and listeners wait to initiate lexical access until sufficient information is available to identify the target word. This account is largely hypothetical and has received almost no empirical support, but the contrast between wait-and-see and immediate competition has motivated much work in word recognition (Dahan & Magnuson, 2006; Weber & Scharenborg, 2012).

Immediate competition has a number of advantages over wait-and-see. It does not require a dedicated memory buffer to store auditory information prior to lexical access. It also does not require a dedicated segmentation process – the system can consider multiple segmentations of a string (e.g., *car#go#ship vs. cargo#ship*) and let competition sort it

* Corresponding author at: Dept. of Psychology, University of Iowa, Iowa City, IA 55242, United States.
E-mail address: Bob-mcmurray@uiowa.edu (B. McMurray).

http://dx.doi.org/10.1016/j.cognition.2017.08.013
Received 14 May 2016; Received in revised form 27 August 2017; Accepted 31 August 2017
Available online 14 September 2017
0010-0277/

out (McClelland & Elman, 1986; Norris, 1994). Finally, by maintaining partial activation for multiple alternatives, listeners may have more flexibility in dealing with variable input if an initial commitment turns out to be wrong (Clopper & Walker, 2017; McMurray, Tanenhaus, & Aslin, 2009).

The ubiquity of this conceptualization is underscored by work on individual differences, development and communicative impairment. All of the non-young-adult groups that have been studied to date exhibit some form of immediate competition. This includes toddlers (Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999), adolescents (Rigler et al., 2015) children with SLI (Dollaghan, 1998; McMurray, Samelson, Lee, & Tomblin, 2010), people undergoing cognitive aging (Revill & Spieler, 2012), and postlingually deafened adults who use cochlear implants (CIs) (Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014). While the dynamics of lexical access in these groups differs quantitatively (and in interesting ways) from typical adults, all of these groups also exhibit behavior broadly consistent with immediate competition.

We set out to characterize the dynamics of lexical access in prelingually deaf children who use Cochlear Implants (CIs). Many studies have characterized word recognition *accuracy* in this population, but few have examined processing. While we expected (and found) quantitative differences, lexical processing in this group also in differed in marked ways from immediate competition accounts. This suggested these listeners might be doing something a closer to wait-and-see. We then demonstrated a similar finding with NH adults hearing extremely degraded speech. These studies raise the possibilities that immediate competition is not the only option for dealing with temporally unfolding inputs, and that demands of severely degraded speech can lead to a range of solutions – solutions that may require somewhat different cognitive architectures.

1.1. Word recognition in prelingually deaf CI users

CIs directly electrically stimulate the auditory nerve to provide profoundly deaf people with the ability to perceive speech. In the normal auditory system, frequency is coded topographically along the basilar membrane. CIs work by inserting an electrode along the basilar membrane with multiple channels that directly electrically excite localized portions of the basilar membrane to the degree that each electrode's characteristic frequency is present in the input (see Niparko, 2009). CIs result in some loss of information and systematic distortion from the original acoustic signal. CIs are generally good at transmitting rapid changes in the amplitude envelope of speech. However, because of a limited number of channels (as well as electrical "bleed" between channels), they only transmit a relatively coarse representation of the frequency structure – harmonics are lost as are rapid spectral changes (particularly within a channel). Fundamental frequency is typically not present in CI input (as electrodes are not typically inserted deeply enough for those frequencies), and periodicity in the signal is replaced by rapid electrical pulses. Despite these limitations, adults who use CIs generally show good speech perception (Francis, Chee, Yeagle, Cheng, & Niparko, 2002; Holden et al., 2013).

In prelingually deaf children, CIs generally offer sufficient input to supporting speech perception and functional oral language development (Dunn et al., 2014; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000; Uziel et al., 2007). However, outcomes are highly variable, and speech perception and language may take many year of device use to fully develop (Dunn et al., 2014; Geers, Brenner, & Davidson, 2003; Gstoettner, Hamzavi, Egelierler, & Baumgartner, 2000; Svirsky et al., 2000). Outcomes are related to a variety of audiological, medical and demographic factors. Earlier implantation tends to lead to better outcomes (Dunn et al., 2014; Kirk et al., 2002; Miyamoto, Kirk, Svirsky, & Sehgal, 1999; Nicholas & Geers, 2006; Waltzman, Cohen, Green, & Roland, 2002), though there is not evidence for a sharp cut-off or critical period (Harrison, Gordon, & Mount, 2005). Better pre-implantation hearing and a longer duration of CI use both lead to better outcomes (Dunn et al., 2014; Nicholas & Geers, 2006). At the same time, variability is a persistent problem that cannot always be linked to medical and/or audiological factors.

A standard measure of speech perception outcomes in CI users is open set word recognition, the ability to produce an isolated word that is presented auditorily. Common examples of this are tests like the CNC (consonant nucleus coda), or PBK (phonetically balanced kindergarten) word lists, or the GASP (Glendonald Auditory Screening Procedure) and LNT (Lexical Neighborhood Test). These are commonly seen as measures of speech perception. However, performance is also affected by lexical and cognitive processes like working memory (Cleary, Pisoni, & Kirk, 2000; Geers, Pisoni, & Brenner, 2013; Pisoni & Cleary, 2003; Pisoni & Geers, 2000) and sequence learning (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Importantly, performance is also influenced by lexical factors like frequency and neighborhood density (Davidson, Geers, Blamey, Tobey, & Brenner, 2011; Eisenberg, Martinez, Holowecky, & Pogorelsky, 2002; Kirk, Pisoni, & Osberger, 1995), suggesting that word recognition in children who use CIs reflects competition among words (much like NH adults). As a whole, these studies suggest that a better understanding of the cognitive processes that underlie spoken word recognition in children who use CIs may be crucial for understanding variable outcomes in this population. This highlights the need to understand real-time lexical competition in CI users. However, only two studies have examined this.

Farris-Trimble et al. (2014) examined postlingually deafened adult CI users as well as NH listeners hearing spectrally degraded CI simulations. Participants were tested in a 4AFC version of the Visual World Paradigm (VWP; Allopenna et al., 1998); they heard target words like *wizard* and selected the referent from a screen containing pictures of the target, a word that overlapped with it at onset (a cohort, *whistle*), a rhyme competitor (*lizard*), and an unrelated word (*baggage*). Fixations to each object were monitored as a measure of how strongly it was considered. Both CI users and NH listeners hearing degraded speech showed eye-movements consistent with immediate competition: shortly after word onset they fixated the target and cohort, and later suppressed competitor fixations. However, there were also quantitative differences. Adult CI users and listeners under simulation were slower to fixate the target and fixated competitors for longer than NH listeners. Thus, degraded hearing quantitatively alters the timing of the activation dynamics, though listeners still show immediate competition.

Grieco-Calub, Saffran, and Litovsky (2009) studied prelingually deaf two-year-old CI users using the "looking while listening" paradigm (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). Toddlers heard prompts like "*Where is the shoe?*" while viewing screens containing pictures of the target object (*shoe*) and an unrelated control (*ball*). CI users were slower than NH toddlers to fixate the target, suggesting differences in real-time processing. However, participants were only tested in a two-alternative version of the task with no phonological competitors, making it difficult to map differences in speed of processing onto the broader process of sorting through the lexicon.

Unlike postlingually deafened adults, CI users who were born deaf must acquire phonemes and words from a degraded input; consequently they exhibit delayed language development (Dunn et al., 2014; Svirsky et al., 2000), and their poor input likely leads to substantially different representations for auditory word forms than those of post-lingually deafened adults (who acquired words from clear input). For example, under exemplar type accounts of word recognition (Goldinger, 1998), clusters of exemplars for neighboring words are likely to overlap considerably, as much of the distinctive acoustic information is lost by the CI. Even in under models in which words are represented in terms of phonemes, the distributions of acoustic cues for individual phonemes will overlap creating overlap among phonemes (which will cascade to make lexical templates more similar). Under either account, mental representations of words learned under the degradation imposed by a CI will be more overlapping and harder to discriminate. The