Cognition 166 (2017) 184–200

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Using prosody to infer discourse prominence in cochlear-implant users and normal-hearing listeners

Yi Ting Huang *, Rochelle S. Newman, Allison Catalano, Matthew J. Goupell

University of Maryland, College Park, United States

ARTICLE INFO

Article history:
Received 9 September 2016
Revised 9 May 2017
Accepted 19 May 2017

Keywords:
Cochlear implants
Prosody
Acoustic cues
Discourse prominence
Eye-tracking

ABSTRACT

Cochlear implants (CIs) provide speech perception to adults with severe-to-profound hearing loss, but the acoustic signal remains severely degraded. Limited access to pitch cues is thought to decrease sensitivity to prosody in CI users, but co-occurring changes in intensity and duration may provide redundant cues. The current study investigates how listeners use these cues to infer discourse prominence. CI users and normal-hearing (NH) listeners were presented with sentences varying in prosody (accented vs. unaccented words) while their eye-movements were measured to referents varying in discourse status (given vs. new categories). In Experiment 1, all listeners inferred prominence when prosody on nouns distinguished categories ("*SANDWICH*" → not sandals). In Experiment 2, CI users and NH listeners presented with natural speech inferred prominence when prosody on adjectives implied contrast across both categories and properties ("*PINK horse*" → not the orange horse). In contrast, NH listeners presented with simulated CI (vocoded) speech were sensitive to acoustic differences in prosody, but did not use these cues to infer discourse status. Together, this suggests that exploiting redundant cues for comprehension varies with the demands of language processing and prior experience with the degraded signal.

1. Introduction

Cochlear-implants (CIs) have become the standard of care for individuals with severe-to-profound hearing loss, with over 324,000 users worldwide (NIDCD, 2012). This includes those who are born deaf and learn language through CIs (prelingually deafened) as well as individuals who are born with normal hearing (NH) and receive CIs after losing their hearing later in life (postlingually deafened). However, despite incremental advancements in signal processing, the acoustic input conveyed through CIs remains severely degraded. Consequently, while some CI users comprehend speech exceedingly well, others gain little benefit from their devices (Blamey et al., 2013; Holden et al., 2013). Research exploring this variability has focused on device-related and biological factors that modulate the quality of the signal (Boons et al., 2012; Geers & Sedey, 2011; Holden et al., 2013; Sarant, Harris, Bennet, & Bant, 2014). This assumes that limited comprehension must reflect missing information within the degraded signal. Yet, years of psycholinguistic research has demonstrated that comprehension involves *more* than signal properties and includes the cognitive and linguistic processes that mediate interpretation, e.g., word recognition, syntactic parsing, pragmatic inferencing. These procedures are executed nearly effortlessly in NH listeners, but less is known about how they unfold under conditions of signal degradation, as is the case for CI users. Thus, it remains unclear whether variation in language comprehension among CI users is solely a product of the poor signal itself, or also the result of differences in the higher-level processes that interpret the degraded signal.

The current study investigates this question by focusing on the interpretation of prosody, the sound structures that link elements of meaning within and across sentences. Prosody is often conveyed through pitch changes on accented words (Bolinger, 1986; Cooper, Eady, & Mueller, 1985; Cruttenden, 1997; Ladd & Morton, 1997; Lieberman, 1960; Pierrehumbert & Hirschberg, 1990; Terken, 1991), but these cues are severely limited in CIs. However, co-occurring changes in intensity and duration exist in natural speech (Bard & Aylett, 1999; Breen, Fedorenko, Wagner, & Gibson, 2010; Cole, Mo, & Hasegawa-Johnson, 2010; Fowler & Housum, 1987; Lam & Watson, 2010; Kochanski, Grabe, Coleman, & Rosner, 2005; Wagner & Klassen, 2015; Watson, Arnold, & Tanenhaus, 2008), and these are well preserved in CIs. Yet, little is known about how these redundant cues are recruited by CI users. Beyond its clinical relevance, answers to this question will shed light on

* Corresponding author at: Department of Hearing and Speech Sciences, University of Maryland College Park, 0100 Lefrak Hall, College Park, MD 20742, United States.

E-mail address: ythuang1@umd.edu (Y.T. Huang).

http://dx.doi.org/10.1016/j.cognition.2017.05.029
0010-0277/

current debates on the acoustic correlates of discourse prominence (Arnold & Watson, 2015; Cole et al., 2010; Isaacs & Watson, 2010; Kochanski et al., 2005; Watson, 2010) and inform our understanding of how prior experience impacts the recruitment of novel acoustic cues during comprehension (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Kleinschmidt & Jaeger, 2015; Kraljic & Samuel, 2006; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003).

In the remainder of the Introduction, we will provide an overview of why speech signals are degraded for CI users and briefly summarize how this impacts overall comprehension. Then, we will discuss why prosody may be particularly informative in isolating the role of language processing in understanding degraded speech, and briefly describe previous work on this topic in the clinical and theoretical literature. Finally, we will consider questions left open by prior work and introduce two experiments investigating these issues.

1.1. Signal degradation and speech comprehension

CIs provide partial hearing by circumventing non-functioning inner hair cells in the cochlea (i.e., the transducers of sound to the nervous system), directly exciting the spiral ganglia attached to the auditory nerve. Like a functioning cochlea, CIs divide and analyze an incoming acoustic signal into separate frequency channels that follow the basic organization of the nervous system, where high frequencies are analyzed near the base of the cochlea and low frequencies are analyzed near the apex. However, there are notable differences in how CIs convey sound relative to a normally functioning auditory system. Due to current technological limitations, CIs convey only frequencies that are considered essential for speech understanding (about 200–8000 Hz). Signal encoding also occurs through far fewer distinct frequency-specific channels, only 12–24 electrode contacts compared to the fine-grained resolution achieved by the 3500 inner hair cells of a functioning cochlea. Together, this spectrally degrades the acoustic signal. In addition, while CIs convey changes in the slow-varying temporal envelope of a signal, they omit the rapid fine structure associated with fundamental frequency and harmonic structure. As a consequence, CIs can capture changes in the intensity and duration of the speech signal, but they severely degrade variation in voice pitch.

One common way to investigate how this degraded signal impacts speech comprehension is to present NH listeners with acoustic signals that are created using similar algorithms as CI processors (called "vocoding"), but convey the temporal envelopes using uninformative noise carriers rather than electrical pulse trains. When this is done, similar patterns of performance have been found across NH listeners presented with vocoded speech and high-performing CI users (Friesen, Shannon, Başkent, & Wang, 2001). Despite massive distortions in the acoustic signal, vocoded speech is surprisingly intelligible (Rosen, Faulkner, & Wilkinson, 1999; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Shannon, Zeng, & Wygonski, 1998). NH listeners can often identify isolated consonants/vowels and repeated words in sentences with greater than 90% accuracy (Shannon et al., 1995). At-ceiling recognition of voicing and manner changes for consonants can occur with only three frequency channels. Thus, even with limited training, NH listeners spontaneously recruit non-spectral cues (such as duration and intensity changes) to comprehend phonetic features within a degraded signal.

Overall, these patterns are consistent with prior work demonstrating the robustness of speech perception. NH listeners are strikingly resilient to a wide array of changes to the speech signal. For example, when presented with unfamiliar phonemic categories that are shifted along a continuum, they rapidly recalibrate to novel vowel spaces (Maye et al., 2008) and consonant contrasts (Kraljic & Samuel, 2006; Clayards et al., 2008; Norris et al., 2003). Adaptation to some aspects of vocoded speech is also similarly quick. Improvements in intelligibility are found over the course of 30–40 sentences, e.g., from 0% accuracy to greater than 70% accuracy (Davis et al., 2005). Thus, at the level of phonetic categories, NH listeners readily adapt to a degraded signal, and can remap relationships between the acoustic signal and linguistic categories.

Yet, despite evidence of impressive adaptability among NH listeners, massive variability in speech comprehension exists among CI users. Differences due to biological factors (e.g., age at implantation, onset of hearing loss) and device-related factors (e.g., CI processors, number of electrodes) likely play a role. However, roughly 25–65% of individual variation (Blamey et al., 2013; Holden et al., 2013) is thought to reflect an assorted blend of cognitive abilities such as working memory, processing speed, and linguistic knowledge (Burkholder & Pisoni, 2003; Cleary, Pisoni, & Kirk, 2000; Geers, Pisoni, & Brenner, 2013). Isolating the precise role of these factors can be challenging (Budenz et al., 2011; Collison, Munson, & Carney, 2004; Heydebrand, Hale, Potts, Gotter, & Skinner, 2007; Holden et al., 2013; Leung et al., 2005). For example, improvements in spoken word recognition are predicted by measures of verbal learning in CI users six months after implantation (Heydebrand et al., 2007). However, among CI users with five or more years of experience, no relationship is found between word-recognition abilities and measures of verbal intelligence (Collison et al., 2004).

The equivocality of prior research highlights the limitations of an individual-differences approach to isolating causal factors in heterogeneous populations. When multiple dimensions vary, it is difficult to distinguish which ones facilitate comprehension while also controlling for other correlated properties. Moreover, even when relationships are found, it is unclear whether they are *caused* by the processing of specific signal properties or simply emerge as byproducts whenever comprehension is difficult. Importantly, comprehending speech through CIs requires *more* than just (re) mapping a degraded acoustic signal onto different phonetic categories than those in natural speech. Since informative cues to meaning will vary under conditions of signal degradation, the entire linguistic system must be altered to exploit these relationships in lexical, morphological, syntactic, and pragmatic processing. Presumably these changes depend on a wide array of cognitive skills, past experiences, and compensatory strategies. Thus, rather than relying on highly aggregated measures of comprehension (e.g., accuracy of sentence identification) and cognitive abilities (e.g., verbal working memory), effects of signal degradation on CI users may be better isolated through finer-grained methods that account for underlying linguistic processes. Comparing CI users and NH listeners may yield additional insights into how acoustic signals are interpreted in two very different instantiations of a single language, and distinguish questions of what cues are available in the input (e.g., pitch, intensity, duration) from how listeners use them (e.g., inferring meaning).

1.2. Prosody in cochlear-implant research

The current study focuses on the comprehension of prosody, an area where links between acoustic cues and linguistic meaning are well documented. During communication, speakers recruit prosody to ask questions, convey emotions, and mark prominence. These distinctions are often cued with pitch changes (Bolinger, 1986; Cooper et al., 1985; Cruttenden, 1997; Ladd & Morton, 1997; Lieberman, 1960; Pierrehumbert & Hirschberg, 1990; Terken, 1991), which are severely degraded in CIs. Unsurprisingly,

CI users face difficulties with prosody. Relative to NH listeners, they are less accurate at distinguishing questions from statements (Chatterjee & Peng, 2008; Meister, Landwehr, Pyschny, Walger, & von Wedel, 2009; Peng, Chatterjee, & Lu, 2012; Van Zyl & Hanekom, 2013), assessing speaker emotions (Gilbers et al., 2015; Hopyan-Misakyan, Gordon, Dennis, & Papsin, 2009; Luo, Fu, & Galvin, 2007; Nakata, Trehub, & Kanda, 2012), and isolating word stress (Morris, Magnusson, Faulkner, Jönsson, & Juul, 2013).

Nevertheless, CIs *do* reliably convey alternative prosodic cues like duration and intensity changes, raising questions of how these cues are interpreted during comprehension. Outside of prosody, recent evidence suggests that prolonged experience with a degraded signal leads CI users to adopt unique strategies for exploiting informative cues (Winn, Chatterjee, & Idsardi, 2012). When distinguishing vowels in two minimal pairs (e.g., "*hit*" vs. "*heat*"), NH listeners often relied on less informative spectral (pitch) changes in vocoded speech, suggesting carry-over preferences for a salient cue in natural speech. In comparison, CI users were less sensitive to spectral cues, relying instead on more informative duration changes in a degraded signal. This predicts that, in the case of prosody, a degraded signal may lead CI users to infer meaning on the basis of available duration (and possibly intensity) cues when pitch changes are minimal.

Unfortunately, these issues are difficult to isolate in prior work, which often aggregates diverse prosodic phenomena under a single umbrella. This approach conflates variation due to communicative functions (e.g., question formation, emotional state, contrastive focus) and units of analysis (e.g., acoustic changes over single words or entire phrases). This is also problematic in that some prosodic phenomena appear to be cued entirely by pitch, whereas others include changes in duration and intensity as well. For example, in the well-studied test case of question-statement prosody, recent evidence suggests that acoustic cues are limited to pitch changes (Srinivasan & Massaro, 2003). Thus, with minimal variation in intensity and duration, it is unsurprising that CI users are worse at distinguishing questions from statements. Importantly, this need not imply that CI users would have similar difficulties with other prosodic phenomena.

1.3. Prosody in psycholinguistics research

To test this possibility, the current study focuses on the area of discourse prominence. It is well known that speakers will often distinguish previously unmentioned (new) categories from previously mentioned (given) ones by accenting nouns in utterances. For example, saying "*Give me the PENCIL*" implies contrast with another category that had been under discussion (e.g., if prior attention was on the marker).[1] Moreover, speakers distinguish properties within mentioned categories by accenting prenominal modifiers like adjectives, e.g., "*Give me the PINK horse*" to imply contrast with the red horse. Variation in discourse structure of this kind has been discussed under many names in linguistics (e.g., given vs. new, topic vs. comment, unfocused vs. focused) and psycholinguistics (e.g., predictable vs. unpredictable, important vs. unimportant, accessible vs. less accessible). Importantly, these phenomena are associated with a set of highly correlated acoustic cues (Ladd, 1996; Lieberman, 1960; Wagner & Watson, 2010). Speakers tend to produce higher pitch, greater intensity, and longer duration when referring to new referents, which are also typically less predictable, less accessible, and more important to the discourse (Breen et al., 2010; Wagner & Klassen, 2015; Watson et al., 2008).

Nevertheless, there remains debate about which cues are necessary and/or sufficient for marking prominence. Traditional theories have argued that pitch changes are primary (Bolinger, 1986; Cooper et al., 1985; Ladd & Morton, 1997; Terken, 1991) and map directly onto variation in meaning (Cruttenden, 1997; Pierrehumbert & Hirschberg, 1990). However, evidence from large-scale corpus analyses suggests that prominence judgments (i.e., evaluating which syllables perceptually 'stand out') are more strongly predicted by intensity and duration changes with minimal contributions of pitch (Cole et al., 2010; Kochanski et al., 2005). Curiously, yet another story emerges in experimental studies on the comprehension of prosody. These findings reveal that NH listeners exploit multiple cues in natural speech to infer prominent referents (Arnold, 2008; Dahan, Tanenhaus, & Chambers, 2002; Ito, Bibyk, Wagner, & Speer, 2014; Ito & Speer, 2008; Sekerina & Trueswell, 2012; Terken & Nooteboom, 1987; Weber, Braun, & Crocker, 2006), but they rely on pitch but not duration changes when cues are isolated in (re)synthesized speech (Bartels & Kingston, 1994; Isaacs & Watson, 2010).

However, it is also possible that sensitivity to correlated changes in prosody may depend in part on one's prior experience with the signal. For example, pitch cues may be highly salient and serve as a primary cue to prosody for NH listeners who encounter these changes regularly in their input. In contrast, individuals who have a history of listening to degraded signals (like CI users) might learn to rely on secondary cues like duration and intensity when pitch is severely limited. Thus, rather than construing cue-to-meaning mappings as fixed relationships, it may be useful to consider how comprehension strategies vary with the informativity of contextual cues. This approach is in line with Bayesian models, which highlight ways in which NH listeners exploit novel cues under conditions of certainty but default to canonical representations when uncertainty arises (Gibson, Bergen, & Piantadosi, 2013; Kleinschmidt & Jaeger, 2015; Levy, Bicknell, Slattery, & Rayner, 2009). To understand relationships between acoustic cues and discourse prominence, the current study exploits a natural experiment that exists among CI users and NH listeners. Comparing comprehension of natural speech and vocoder simulation of CI speech will reveal the degree to which pitch changes are a primary cue to prominence for NH listeners. Comparing comprehension across CI users and NH listeners with vocoded speech will reveal how prior experience shapes comprehension strategies.

1.4. Current study

The current study differs from prior CI research in two additional ways. First, rather than examining prosody in the context of isolated words (Chatterjee & Peng, 2008; Gilbers et al., 2015; Hopyan-Misakyan et al., 2009; Peng et al., 2012; Van Zyl & Hanekom, 2013), we employ full sentences to approximate the real-world challenges of spoken-language comprehension. Second, rather than relying on metalinguistic judgments or response-categorization tasks, we examine prosody effects using the visual-world eye-tracking paradigm (Salverda & Tanenhaus, in press). This method of assessing comprehension has been used to study diverse populations (Arnold, 2008; Dahan et al., 2002; Huang & Snedeker, 2011; Isaacs & Watson, 2010; Ito & Speer, 2008; Ito et al., 2014; Sekerina & Trueswell, 2012; Weber et al., 2006), with recent applications to CI users (Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014; Winn, Edwards, & Litovsky, 2015). Since eye movements are often made without conscious reflection, they provide an implicit measure of interpretation that isolates initial speech processing from later-emerging challenges (for more discussion, see Farris-Trimble et al., 2014).

Based on the design developed by Dahan et al. (2002), Experiment 1 presented listeners with sentences which featured

[1] Throughout this paper, capitalization indicates accented stress.

accented (e.g., "*Now put the SANDWICH on the triangle*") or unaccented nouns (e.g., "*Now put the sandwich on the triangle*"). Eye-movements were measured to a visual display (Fig. 1), which contained the corresponding Target (e.g., sandwich) and a phonological cohort Competitor (e.g., sandals). The preceding sentence manipulated the discourse status of the Target. In the given condition, the same object was mentioned in the first and second sentences (e.g., "*Put the sandwich on the star*"). In the new condition, different objects were introduced across sentences (e.g., "*Put the sandals on the star*"). When all acoustic cues are present in natural speech (Arnold, 2008; Dahan et al., 2002), NH listeners are expected to increase Target fixations after unaccented nouns in the given condition and accented nouns in the new condition. For example, after hearing "*Put the sandals on the star. Now put the SAND. . .*," they will be more likely to look to the sandwich (new referent) compared to the sandals (given referent). Importantly, when the signal is degraded, fixations across conditions will reveal the degree to which intensity and/or duration changes can be used to infer discourse prominence.

One possibility is that all listeners will be sensitive to redundant acoustic cues and rapidly recruit them to infer likely referents. Cue-trading of this kind has been documented in syntactic-ambiguity resolution, whereby NH listeners infer boundary tones based on pitch when duration cues are ambiguous, but switch to duration when pitch cues are ambiguous (Beach, 1991). If similar relationships exist for interpreting discourse status, CI users and NH listeners may infer prominence via intensity and/or duration changes when pitch cues are degraded. Yet, another possibility is that pitch cues are privileged for conveying prominence. In a design similar to the current study, Isaacs and Watson (2010) found that NH listeners did *not* recruit duration changes when pitch cues were absent in resynthesized speech. Parallel patterns may emerge with vocoded speech. Finally, a third possibility is that experience with a degraded signal may lead to different comprehension strategies across groups (Gilbers et al., 2015; Winn et al., 2012). Since intensity and duration changes are available in the input to CI users, they may be particularly informative for these listeners. Thus, relative to NH listeners, CI users may be more successful at exploiting these cues in degraded signals to infer discourse prominence.

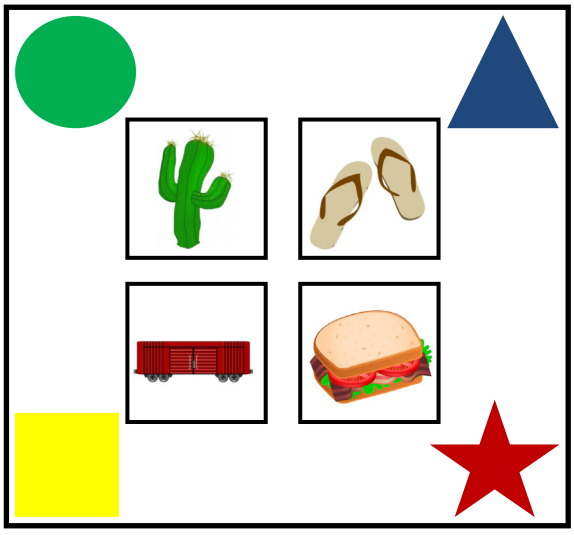

Fig. 1. In Experiment 1, sample display for a critical trial. In the sentence "Now put the sandwich on the triangle," the Target is the sandwich and the Competitor is the sandals.

2. Experiment 1: nouns

2.1. Methods

2.1.1. Participants

Sixty-seven NH and 25 CI users participated in this study. From this group, data from three NH listeners were excluded from the sample because of experimenter error or equipment failure, leading to a final sample of 64 NH listeners. NH listeners self-reported having normal hearing in both ears and were recruited through the student population at the University of Maryland College Park. Half the NH listeners participated in the natural-speech condition, while the other half participated in the vocoded-speech condition.

Data from all CI users were included. Table 1 illustrates that they were an average age of 54 years (SD = 16 years). Seven CI users were unilaterally implanted, and 18 were bilaterally implanted. During the study, CI users' everyday speech processors were used with their typical settings. Since some unilateral users may access low-frequency information like pitch through hearing aids in the alternate ear, their hearing aids were removed and their alternate ear was occluded by earplugs to ensure they were receiving information primarily from their CIs.[2] CI users had used their CIs for an average of 7.4 years (SD = 4.3 years). Five CI users reported to be prelingually deafened and 18 users to be postlingually deafened. All participants (both CI users and NH listeners) were English speakers and had normal or corrected-to-normal eyesight.

2.1.2. Procedure

Listeners were tested individually in a quiet room. They sat in front of a podium divided into four shelves (upper left, upper right, lower left, and lower right), where pictures originally appeared. On the outside corners of the podium were target locations (square, star, triangle, and circle), where pictures could be moved. A camera at the center of the display was focused on listeners' faces and recorded the direction of their gaze while they were performing the task. A second camera, located behind the listeners, recorded both their actions and the location of the items in the display. See Snedeker and Trueswell (2004) for validation of this method of eye-tracking against automated approaches.

At the start of each trial, the experimenter took out four pictures and placed each one on a shelf in a pre-specified order. Trials consisted of a pair of spoken instructions, asking listeners to pick up a target picture and move it to a target location. Instructions were played through an X-Mini II Capsule Speaker, placed about one foot away from the listeners. After each instruction, the listeners produced an action. Pictures remained in their locations in between instructions. The second instruction always asked listeners to move pictures to new locations where they were not already placed. Once listeners produced actions for each of the two instructions, the trial ended, the pictures were removed from the display, and the next trial began. After completing the experiment, participants received a verbal debriefing about the study. Among the CI users, two participants noted awareness that certain words were emphasized in the sentences, and three mentioned the presence of similar-sounding words. No participant (NH listeners or CI users) mentioned noticing relationships between accenting and prior mention.

[2] Because they were not formally evaluated, three CI users may have had some access to pitch cues. However, follow-up analyses confirmed that reported effects remained the same when their data were excluded, suggesting that they alone were not responsible for the overall patterns.

Table 1
In Experiment 1, demographic characteristics of CI users.

Subject #	Age (in years)	Sex	# of CIs	CI use (in years)	Onset of deafness
1	68	F	Unilateral	7	Postlingual
2	51	F	Bilateral	10	Prelingual
3	21	M	Unilateral	9	Postlingual
4	26	M	Bilateral	8	Prelingual
5	50	M	Bilateral	4	Prelingual
6	61	F	Bilateral	8	Postlingual
7	51	M	Bilateral	4	Postlingual
8	73	M	Bilateral	10	Postlingual
9	55	F	Bilateral	3	Postlingual
10	69	F	Bilateral	7	Postlingual
11	62	F	Unilateral	16	Prelingual
12	55	F	Bilateral	9	Postlingual
13	67	M	Bilateral	2	Postlingual
14	19	M	Unilateral	16	Prelingual
15	51	F	Unilateral	Unknown	Postlingual
16	53	F	Bilateral	2	Prelingual
17	63	–	Bilateral	6	Postlingual
18	68	F	Bilateral	16	Postlingual
19	50	F	Unknown	Unknown	Unknown
20	46	F	Bilateral	7	Postlingual
21	61	M	Unknown	1	Postlingual
22	26	M	Unilateral	6	Postlingual
23	69	F	Unilateral	Unknown	Postlingual
24	76	F	Bilateral	7	Postlingual
25	49	F	Bilateral	4	Postlingual

2.1.3. Materials

Instructions for the critical trials represented four cells from a 2 × 2 design (Dahan et al., 2002). The first variable manipulated discourse status. In the given condition, the second sentence referred to the same target picture as the one mentioned in the first sentence (sentences 1a and 1b below). In the new condition, the second sentence referred to a different target picture from the one mentioned in the first sentence (sentences 1c and 1d). The second variable manipulated prosody. In accented trials, the noun corresponding to the target picture in the second sentence had increased prosodic stress relative to the noun corresponding to the location (sentences 1a and 1c). This relationship reversed in unaccented trials (sentences 1b and 1d).

(1) a. Given-accented: Put the sandwich on the star. Now put the SANDWICH on the triangle.
b. Given-unaccented: Put the sandwich on the star. Now put the sandwich on the TRIANGLE.
c. New-accented: Put the sandals on the star. Now put the SANDWICH on the triangle.
d. New-unaccented: Put the sandals on the star. Now put the sandwich on the TRIANGLE.

Instructions were pre-recorded by a female, native English speaker in a noise-reducing sound booth, using a Shure SM51 microphone at a 44.1-kHz sampling rate and 32-bit precision. Vocoded stimuli were created by first passing natural-speech recordings through a 1st-order high-pass Butterworth filter with a 1200-Hz cutoff frequency. This slightly diminishes most vowel information below 1200 Hz and enhances most consonant information above 1200 Hz, similar to what occurs in CIs. To simulate the limited frequency range and small number of channels within CIs, the signal was then divided using eight 4th-order Butterworth bandpass filters with contiguous and logarithmically spaced corner frequencies from 300 to 8500 Hz. Within each channel, the slow variation in the amplitude of the signal (the temporal envelope) is extracted by passing the signal through a 2nd-order low-pass Butterworth filter with a 400-Hz cutoff frequency. This removes the rapid variations in acoustic sound pressure (the temporal fine structure). For CI users, the remaining temporal envelopes are conveyed by modulating high-rate electrical pulse trains to different frequency regions in the inner ear and auditory nerve. For NH listeners, the temporal envelopes are conveyed using narrowband noises that are filtered with the same Butterworth bandpass filters used to perform the sound analysis. The eight channels are summed into a single waveform and normalized to have the same energy as the natural speech. At each step, forward-backward filtering doubled the order of filters to avoid temporal smearing of the signal and preserve critical acoustic cues. All analyses were implemented in Matlab (the Mathworks; Natick, MA).

CI users were presented with natural-speech sentences, which were then altered by the processing of their devices. NH listeners were presented with either natural- or vocoded-speech versions. Analyses revealed no differences in pitch, intensity, or duration during the carrier phrase prior to the critical noun (i.e., "*Now put the...*") in natural and vocoded stimuli (p's > 0.20). In contrast, they confirmed the presence of relevant acoustic properties on the critical noun (i.e., "*...sandwich/SANDWICH*"). Table 2 illustrates average change in pitch (fundamental frequency as measured by the peak of the autocorrelation function), intensity (as measured by the root-mean-square of the waveform), and duration (in ms) across items. For natural speech, pitch and intensity changes were greater in accented compared to unaccented trials. However, duration did not differ significantly, a point that we will return to in Section 4. For vocoded speech, intensity was greater in accented compared to unaccented trials. Since the harmonic structure of speech is replaced with noise in these stimuli, measuring pitch is meaningless. Duration remains unaltered across stimuli type, but greater intensity changes were found in vocoded compared to natural speech ($t(15) = 2.77$, $p < 0.05$). Altogether, these analyses confirm that acoustic cues to prosody were present in natural and vocoded stimuli.

Visual displays featured four pictures placed on a shelf in fixed locations (see Fig. 1). In critical trials, Targets were referents of the second sentence (e.g., sandwich), and Competitors were phonological cohort members of the Targets (e.g., sandals). The temporary ambiguity at word onset allows us to examine how prosody generates expectations of discourse status and whether this changes

Table 2
In Experiment 1, average changes and statistical comparisons of pitch (i.e., change in fundamental frequency), intensity (i.e., loudness level, dB measured relative to full scale), and duration on the critical noun in accented and unaccented trials.

	Accented	Unaccented	Comparison
Natural speech			
Pitch change	292 Hz	212 Hz	$t(15) = 4.95$, $p < 0.001$
Intensity	−29.40 dB	−33.04 dB	$t(15) = 2.97$, $p = 0.01$
Duration	547 ms	526 ms	$t(15) = 1.66$, $p = 0.12$
Vocoded speech			
Pitch change	Not applicable	Not applicable	Not applicable
Intensity	−29.77 dB	−35.83 dB	$t(15) = 4.84$, $p < 0.001$
Duration	547 ms	526 ms	$t(15) = 1.66$, $p = 0.12$

when pitch cues are degraded. Targets and Competitors were paired with two pictures that were phonologically and semantically unrelated (e.g., cactus, boxcar). Across four presentation lists, four versions of 16 picture-set items were created and counterbalanced such that each list contained four different items in each condition and each version of an item appeared just once in every list. Since listeners were randomly assigned to one presentation list, each heard only one version of a picture-set, and saw the same number of picture-sets in the four conditions. Across all lists, critical trials were randomly presented with eight filler trials. Unlike critical trials, Targets in filler trials did not share phonological onsets with other objects in the picture-set. Also, sentences within each instruction pair always referred to different pictures.

2.1.4. Coding

Across all listeners, approximately 0.2% of trials were excluded from subsequent analyses because of experimenter error. Data from all other trials were coded in the following manner. First, trained research assistants watched videotapes of the listeners' actions for the second sentence in the critical trials and noted the picture and location that was selected on each trial. Correct trials were coded as ones where listeners selected the target picture and placed it in the target location. Incorrect trials were coded as ones where listeners made no actions at all or selected a picture or location that was not mentioned in the instructions.

Second, trained research assistants coded eye-movements using Vcode, a frame-by-frame annotation software (Hagedorn, Hailpern, & Karahalios, 2008). Coding began at the onset of the second instruction and ended with the onset of the corresponding action. Changes in the direction of gaze were coded as towards one of the quadrants, at the center, or missing due to looks away from the display or blinking. Missing frames accounted for 11.7% of coded frames for NH listeners with natural speech, 15.8% for NH listeners with vocoded speech, and 20.1% for CI users. These data were excluded from further analysis. Remaining looks were then recoded based on their relation to the instruction: (1) Target looks; (2) Competitor looks; (3) Other looks to unrelated pictures. Twenty-five percent of trials were checked by second coders who confirmed fixation directions for 92.8% coded frames for NH listeners with natural speech, 92.1% for NH listeners with vocoded speech, and 89.4% for CI users. Disagreements between the two coders were resolved by a third independent coder.

2.2. Results

We analyzed actions and fixations using a series of mixed-effects models. Discourse status (given vs. new), prosody (accented vs. unaccented), and listener group (NH listeners with natural speech vs. NH listeners with vocoded speech vs. CI users) were modeled as fixed-effects variables. Subjects and items were simultaneously modeled as random-effects variables, with random intercepts only. We also constructed models that included random slopes for the fixed-effects factors (discourse and prosody) and interaction terms, but none resulted in a significant improvement in model fit ($p > 0.05$) or affected estimates of fixed effects. Thus, we report simpler models without random slopes. Main effects and interactions were first isolated through likelihood ratio tests across models. Within final models, parameter-specific p-values were estimated through z-statistics in logistic regressions (Jaeger, 2008) and normal approximation of *t*-statistics in linear regressions (Barr, Levy, Scheepers, & Tily, 2013). Analyses were implemented through the lme4 software package in R (Bates, Maechler, & Bolker, 2013).

2.2.1. Actions

Action accuracy was high across all listener groups (>90%). Using logistic mixed-effects models, we found greater accuracy in the given compared to new condition (M = 98% vs. 93%; $\chi^2(1) = 27.42$, $p < 0.001$). This suggests that comprehension was easier when second sentences referred to previously mentioned referents. Overall accuracy also differed across listener groups ($\chi^2(2) = 15.54$, $p < 0.001$). Relative to CI users (M = 92%), accuracy was higher for NH listeners with natural speech (M = 98%; $z = 3.71$, $p < 0.001$) and vocoded speech (M = 96%; $z = 2.52$, $p < 0.05$). However, accuracy did not differ across stimulus types for NH listeners ($z = 1.54$, $p > 0.10$). This suggests that while still highly accurate, CI users faced more difficulty with this task. There was no additional effect of prosody or interaction with discourse status (p's > 0.50).

2.2.2. Eye-movements

Initial analyses confirmed that Target looks did not differ across conditions prior to the onset of the critical noun (p's > 0.60). To assess sensitivity to prosody, our analyses focused on fixations that occurred from the onset of the critical noun to sentence offset. The average length of this period was 1433 ms. Figs. 2 and 3 illustrate that all listeners eventually looked to the Target more than other objects, demonstrating correct identification of referents based on the noun. Moreover, changes in Target and Competitor fixations reveal prosody effects across all listener groups. In the given condition, NH listeners exhibit a sharper rise in Target looks for unaccented natural and vocoded speech. Similarly, CI users show fewer looks to Competitors in this context. In the new condition, fixation patterns appropriately reverse. Now, all listeners were less likely to consider Competitors following accented trials compared to unaccented trials.

Yet, closer inspection also reveals differences in how quickly listeners responded to speech cues. While fixation changes were closely time locked to noun onset for NH listeners with natural speech, they were more sluggish in the other groups. This is particularly striking in NH listeners with vocoded speech, who often lingered on previously mentioned objects in the presence of signal degradation (i.e., sustained Competitor looks after noun onset in Fig. 3). These early differences create challenges in defining a single region of analysis since delays in word recognition would have cascading impacts on inferring discourse status. Since our primary question focused on *what* cues listeners used to interpret prosody rather than *when* they did so, we decided to time-lock our analyses to the onset of word recognition. This approach is common in developmental research, where children's eye-movements can often be slower than adults' (Dautriche, Swingley, & Christophe, 2015; Huang, Zheng, Meng, & Snedeker, 2013). It is also logically akin to traditional methods that benchmark to an informative linguistic event (Salverda & Tanenhaus, in press). However, rather

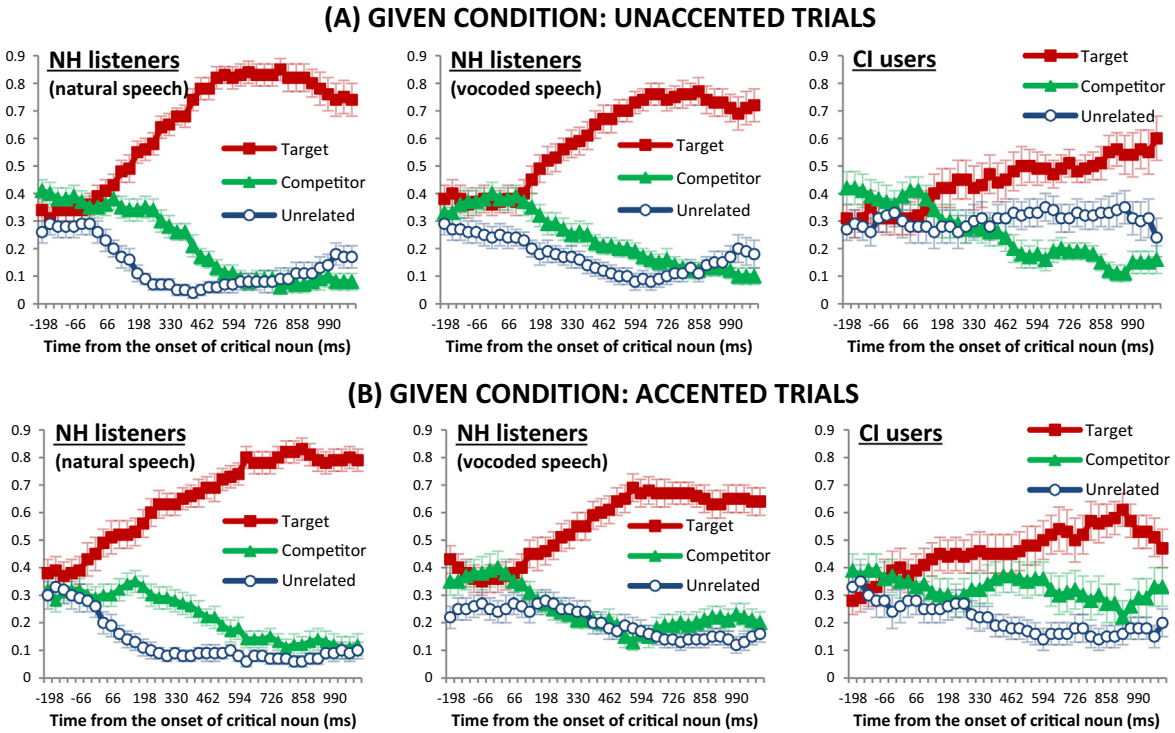

Fig. 2. In Experiment 1, proportion of fixations to Target, Competitor, and Unrelated objects from noun onset in (A) unaccented and (B) accented trials of the given condition.

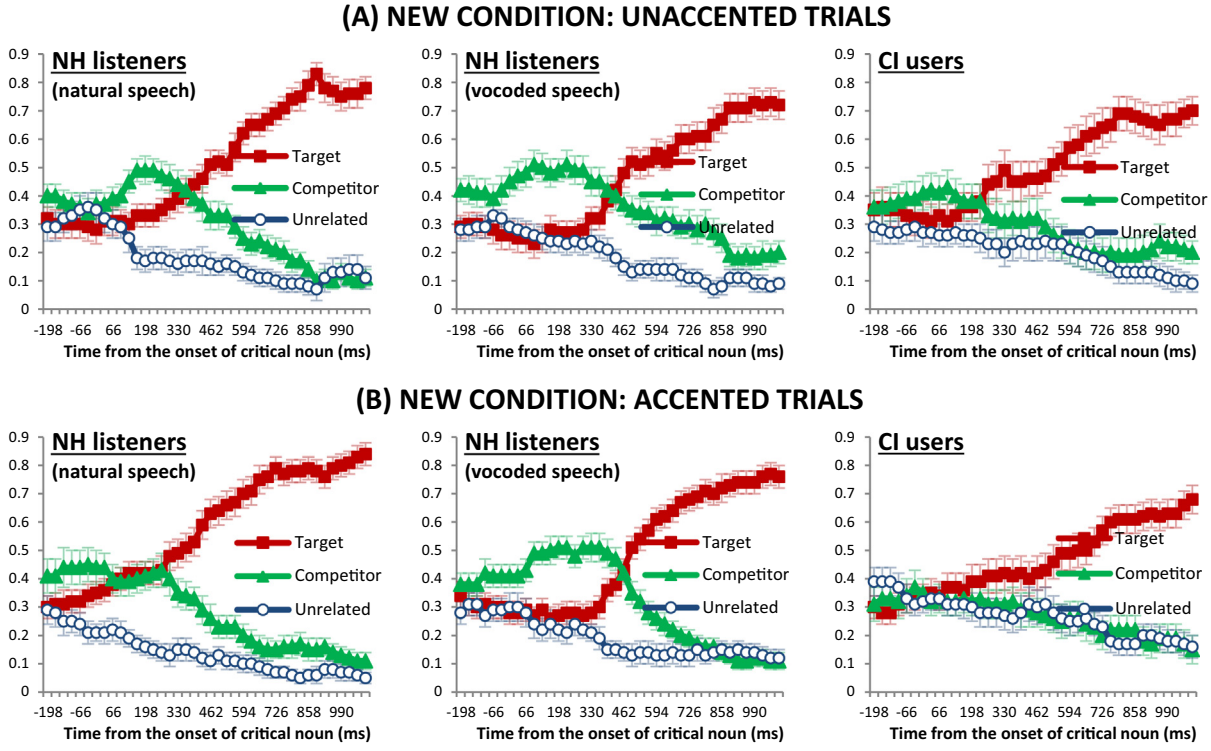

Fig. 3. In Experiment 1, proportion of fixations to Target, Competitor, and Unrelated objects from noun onset in (A) unaccented and (B) accented trials of the new condition.

than defining this point as its onset within the speech stream, we hone in on the earliest point in which it affects fixations.

To implement this, we took advantage of recent extensions of cluster-based permutation analysis to eye-tracking data (Dink & Ferguson, 2015). This approach offers a method for testing significant effects across individual time windows, while simultaneously correcting for multiple comparisons (Maris & Oostenveld, 2007). Within level of prosody, we first tested whether fixations to the unmentioned object were greater for the new compared to given condition in each 33-ms time bin. Statistically significant

($p < 0.05$, two-tailed) and adjacent time bins were clustered into a single window. Then, we compared sum-statistics of this cluster-based window to sum-statistics of iterative comparisons of randomly shuffled data. This later case served as a null hypothesis if critical nouns had no effect on fixations. Cluster-based time windows were considered significant if the probability of observing a cluster of the same size or bigger in the randomized data was less than 5%. Based on this approach, the onset of word recognition in NH listeners with natural speech was 200 ms after the critical noun, which matched what is typically founded in visual-world eye-tracking experiments (Allopenna, Magnuson, & Tanenhaus, 1998). In contrast, NH listeners with vocoded speech were delayed until 400 ms after noun onset. Moreover, CI users showed a dual pattern, whereby word recognition began at the 200 ms window in unaccented trials but was delayed until the 600 ms window in accented trials.

Next, we defined a standard 500 ms window from these time points and calculated Target preference as the proportion of Target looks over Target plus Competitor looks for each trial. We analyzed Target preference using linear mixed-effects models. Overall, cross-model comparisons revealed a significant interaction between discourse status and prosody ($\chi^2(1) = 11.10$, $p < 0.001$). Consistent with prior work, Target preference in the given condition was greater in unaccented compared to accented trials (M = 74% vs. 70%). This pattern appropriately reversed in the new condition (M = 61% vs. 69%). Importantly, our analyses also revealed a 3-way interaction between discourse status, prosody, and listener group ($\chi^2(6) = 24.16$, $p < 0.001$). We conducted follow-up analyses to unpack this pattern within levels of discourse status.

In the given condition, there was a significant effect of listener group ($\chi^2(2) = 14.01$, $p < 0.001$) but no reliable effect of prosody ($\chi^2(1) = 1.96$, $p > 0.15$). Fig. 4a illustrates that Target preference was smaller for CI users compared to NH listeners presented with natural speech ($t = 2.51$, $p < 0.05$) and NH listeners presented with vocoded speech ($t = 3.89$, $p < 0.001$). However, the latter groups did not differ from each other ($t = 1.47$, $p > 0.10$). This suggests that CI users were generally slower than NH listeners at restricting reference to the Target in this task, even when delays in the onset of word recognition were accounted for (i.e., regions offset by 200–600 ms). Moreover, despite their inexperience with degraded signals, NH listeners presented with vocoded speech used disambiguating information on the noun to distinguish Targets from Competitors, and were faster do so relative to CI users after accounting for general delays in word recognition (i.e., regions offset by 400 ms). We will return to this point in the Discussion. Importantly, cross-model comparisons revealed no interaction between group and prosody ($\chi^2(2) = 0.05$, $p > 0.90$). This suggests that prosody differences were comparable across groups.

In the new condition, there was a significant effect of listener group ($\chi^2(2) = 13.22$, $p < 0.01$) as well as a predicted effect of prosody ($\chi^2(1) = 12.91$, $p < 0.001$). Fig. 4b illustrates that Target preference was smaller for NH listeners presented with natural speech compared to NH listeners presented with vocoded speech ($t = 3.58$, $p < 0.001$) and CI users ($t = 2.75$, $p < 0.01$), but the latter groups did not differ from each other ($t = 0.60$, $p > 0.50$). This pattern is likely driven by the fact that regions of analysis were shifted later for listeners who encountered degraded speech, and increased Target looks emerge when disambiguating information on the noun is available. Importantly, all listeners revealed a greater Target preference in accented trials compared to unaccented trials ($t = 3.58$, $p < 0.001$). Moreover, the magnitude of this prosody effect did not interact with listener group ($\chi^2(2) = 0.17$, $p > 0.90$). Together, these findings demonstrate strong prosody effects when pitch cues were present, in the case for NH listeners with natural speech. Moreover, while signal degradation delayed the onset of word recognition for CI users and NH listeners with vocoded speech, it did not reduce the size of their prosody effect.

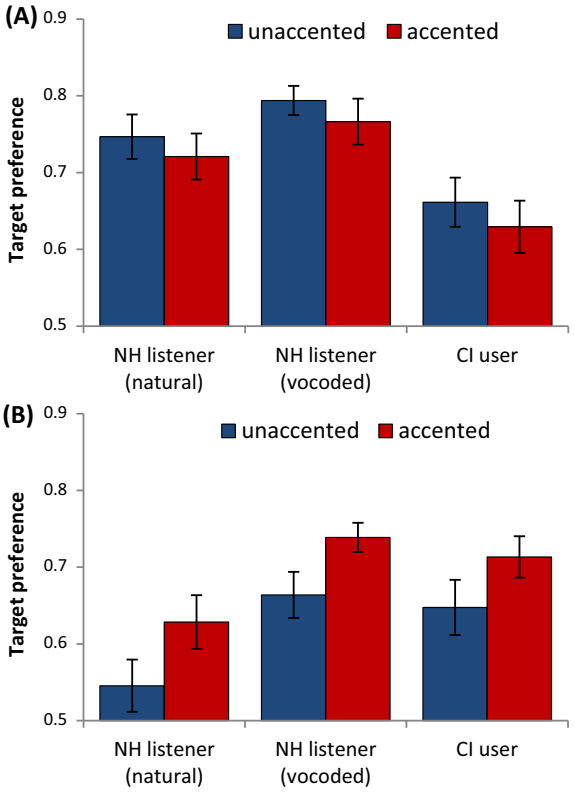

Fig. 4. In Experiment 1, Target preference during regions of analysis in listener groups in the (A) given condition and (B) new condition.

2.3. Discussion

In Experiment 1, we found that listeners inferred the discourse status of accented nouns, both when speech was natural and when it was severely degraded through a vocoder or CI sound processing. When pitch cues were available, NH listeners generated increased fixations to given targets following unaccented compared to accented nouns, and this appropriately reversed for new targets (Arnold, 2008; Dahan et al., 2002). Similar patterns emerged when pitch cues were diminished for CI users and NH listeners presented with vocoded speech. Importantly, the lack of interactions between prosody and listener groups suggests that intensity changes – in the absence of pitch changes – can be exploited with minimal experience under some circumstances. While this pattern may appear at odds with prior work failing to find inferencing of discourse status without pitch cues (Isaac & Watson, 2010), the presence of intensity changes in the current study likely played a role (see Section 4 for more details).

However, Experiment 1 also revealed ways in which CI users differ from NH listeners. While action accuracy was high in CI users, it was lower than in NH listeners. This pattern was somewhat surprising since similarities in the prosody effect suggest that everyone was sensitive to acoustic information on the critical nouns. Moreover, it does not appear to be a general effect of signal degradation since accuracy did not differ for NH listeners presented with natural and vocoded speech. Nevertheless, closer inspection of action errors in CI users suggests that they had particular difficulty resolving cohort competition (e.g., sandals vs. sandwich). Selection of Competitors (M = 9%, SD = 8%) was more than twice as likely compared to all other errors combined, e.g., select-

ing unrelated pictures or placing them in incorrect locations (M = 4%, SD = 8%). Similarly, word recognition was time-locked to the onset of the accented noun for NH listeners presented with natural speech and shortly thereafter with vocoded speech, but it did not emerge until after word offset in CI users. Both patterns are consistent with existing evidence of extended cohort effects in CI users (Farris-Trimble et al., 2014). Interestingly, our results suggest that the procedures for rejecting lexical competitors in word recognition (e.g., hearing "*sandwich*," realizing it is not the cohort competitor "*sandals*") is distinct from inferring discourse status through prosody cues (e.g., hearing "*SANDWICH*," realizing it is not the previously mentioned "*sandals*"). Challenges with the former did not preclude CI users from benefitting from the latter. In fact, it may be that prolonged experience with signal degradation enables CI users to exploit top-down cues to comprehension (e.g., discourse expectations) to overcome the inherent challenges of bottom-up processing (e.g., word recognition).

To explore this possibility, Experiment 2 examined prosody effects when they occur on adjectives. The basic design follows Experiment 1. On each trial, the first sentence introduced the discourse status of a referent using an adjective-noun phrase (e.g., "*Put the orange horse/fish on the star*"). The second sentence manipulated prosody through an accented (e.g., "*Now put the PINK horse on the triangle*") or unaccented adjective (e.g., "*Now put the pink HORSE on the triangle*"). When multiple acoustic cues are available (Ito & Speer, 2008; Ito et al., 2014; Sekerina & Trueswell, 2012; Weber et al., 2006), NH listeners presented with natural speech prefer given categories for accented adjectives and new categories for unaccented adjectives. Thus, if cue-trading relationships reflect general effects of phonetic adaptation to a degraded signal, then all listeners should exhibit prosody effects of similar magnitudes just as they did in Experiment 1. Moreover, to the extent that cohort competition delayed word recognition for CI users in Experiment 1, these effects may be minimal in Experiment 2 when reference restriction relies on phonologically unambiguous words (e.g., "*pink*" shares no overlap with "*orange*").

Importantly, comparing prosody effects on nouns and adjectives may reveal differences in comprehension strategies that listeners employ when faced with acoustic degradation. In particular, while NH listeners readily exploit intensity cues on vocoded nouns to infer discourse prominence in Experiment 1, it remains unclear whether this ability reflects rapid adaptation to the co-occurrence statistics or abstract generalization of structural similarities across natural and vocoded speech (see Kleinschmidt and Jaeger (2015) for a detailed discussion of this distinction). In the former case, NH listeners may have become aware of the fact that intensity changes in vocoded speech are sometimes associated with reference to new categories. This kind of rapid bottom-up learning may be akin to well-documented cases of rapid adaptation to talker-specific phonemic categories (Kraljic & Samuel, 2006; Clayards et al., 2008; Maye et al., 2008; Norris et al., 2003). However, it is also possible that NH listeners' successes with vocoded speech was supported by a broader realization that intensity cues in vocoded speech signal discourse prominence in much the same way that pitch cues function in natural speech.

These two possibilities make different predictions about prosody effects on adjectives. While interpreting accented nouns involves mapping relevant acoustic cues to category contrast only (e.g., hearing "*SANDWICH*," realizing it is not the sandals), doing so for accented adjectives requires mapping cues to both category and property contrasts (e.g., hearing "*ORANGE horse*," realizing it is not the orange fish or pink horse). The presence of multiple options may lead to greater uncertainty about how novel cues map onto discourse prominence. If NH listeners rapidly acquire abstract knowledge of cue-trading relationships for discourse prominence in vocoded speech, then they should apply this understanding to infer that accented adjectives imply contrast across properties within a given category. This should lead to performance that is comparable to natural speech. If, however, NH listeners relied on a more limited awareness of how intensity cues on vocoded nouns predict new categories within a specific context, then they may be less likely to converge on relevant relationships when multiple cue-to-category mappings are available for adjectives. Thus, they may differ from CI users, who may be more likely to recruit structured knowledge of how non-pitch cues map onto discourse prominence by virtue of their extensive experience with the degraded vocoded signal.

3. Experiment 2: adjectives

3.1. Methods

3.1.1. Participants

Sixty-seven NH and 25 CI users participated in this study. From this group, data from three NH listeners were excluded because of experimenter error or equipment failure, leading to a final sample of 64 listeners. Half participated in the natural-speech condition, while the other half participated in the vocoded-speech condition. Data from one CI user was excluded because of substantial difficulties in completing the task, leading to a final sample of 24 listeners. Table 3 illustrates that CI users were an average age of 56 years (SD = 17 years) and used their devices for an average of 6.6 years (SD = 4.6 years). Fourteen were unilaterally implanted, and 10 were bilaterally implanted. Five were prelingually deafened, and 19 were postlingually deafened. NH listeners were all new participants, but 12 CI users in Experiment 2 had also participated in Experiment 1.[3] All listeners were recruited and tested in the same manner as those in Experiment 1 and were English speakers with normal or corrected to normal eyesight.

3.1.2. Procedures

The procedure was identical to Experiment 1.

3.1.3. Materials

The materials were similar to Experiment 1 with key differences. Instructions referred to pictures using adjective-noun phrases. Prosody manipulated accenting of adjectives in the second sentence. In accented trials, adjectives referring to target pictures increased stress relative to nouns (sentences 2a and 2c). This relationship reversed in unaccented trials (sentences 2b and 2d). Vocoded versions were created from natural recordings. Analyses confirmed no differences in pitch, intensity, or duration in the carrier phrase prior to the adjective (i.e., "*Now put the…*") in natural and vocoded stimuli (p's > 0.20). They verified relevant acoustic differences across condition on critical adjectives (i.e., "*…pink/PINK*"). Table 4 illustrates that pitch, intensity, and duration changes were greater in accented compared to unaccented trials in natural speech. For vocoded speech, changes in intensity were greater in accented compared to unaccented trials. Unlike Experiment 1, there were no significant differences in the size of the intensity changes across natural and vocoded speech ($t(15) = 1.58$, $p > 0.10$). Also, unlike Experiment 1, duration differences co-occurred with intensity changes in both natural and vocoded speech.

[3] The average time between experiments was 16 months (range: 10–21 months), thus there was little reason to believe that participating in one would significantly influence performance in the other. Follow-up analyses in Experiment 2 also confirmed no differences in prosody effects for CI users who had participated in Experiment 1 and those who had not (p's > 0.80).

Table 3
In Experiment 2, demographic characteristics of CI users.

Subject #	Age (in years)	Sex	# of CIs	CI use (in years)	Onset of deafness
1	76	F	Bilateral	7	Postlingual
2	79	M	Unilateral	5	Postlingual
3	68	M	Unilateral	11	Postlingual
4	80	M	Unilateral	3	Postlingual
5	61	F	Unilateral	4	Prelingual
6	54	F	Bilateral	2	Prelingual
7	70	F	Unilateral	7	Postlingual
8	70	F	Unilateral	Unknown	Postlingual
9	61	F	Bilateral	Unknown	Unknown
10	57	M	Unilateral	5	Postlingual
11	53	M	Bilateral	4	Postlingual
12	23	M	Unilateral	9	Postlingual
13	27	M	Unilateral	6	Postlingual
14	52	F	Unilateral	Unknown	Postlingual
15	56	F	Unilateral	6	Postlingual
16	50	F	Bilateral	5	Postlingual
17	50	M	Unilateral	11	Prelingual
18	41	M	Bilateral	2	Postlingual
19	57	F	Bilateral	9	Postlingual
20	35	F	Unilateral	15	Unknown
21	21	M	Unilateral	18	Prelingual
22	62	F	Bilateral	6	Prelingual
23	65	F	Bilateral	1	Postlingual
24	73	F	Unilateral	3	Postlingual

Table 4
In Experiment 2, average changes and statistical comparisons of pitch (i.e., change in fundamental frequency), intensity (i.e., loudness level, dB measured relative to full scale), and duration on the critical adjective in accented and unaccented trials.

	Accented	Unaccented	Comparison
Natural speech			
Pitch change	278 Hz	200 Hz	$t(15) = 16.16$, $p < 0.001$
Intensity	−21.54 dB	−26.31 dB	$t(15) = 3.09$, $p = 0.007$
Duration	359 ms	290 ms	$t(15) = 3.48$, $p = 0.003$
Vocoded speech			
Pitch change	–	–	–
Intensity	−29.52 dB	−35.59 dB	$t(15) = 3.68$, $p = 0.002$
Duration	359 ms	290 ms	$t(15) = 3.48$, $p = 0.003$

(2) a. Given-accented: Put the orange horse on the star. Now put the PINK horse on the triangle.
b. Given-unaccented: Put the orange horse on the star. Now put the pink HORSE on the triangle.
c. New-accented: Put the orange fish on the star. Now put the PINK horse on the triangle.
d. New-unaccented: Put the orange fish on the star. Now put the pink HORSE on the triangle.

Visual displays in critical trials featured four object pictures that varied in category and property. Fig. 5 illustrates that Targets were referents of the second sentence (e.g., pink horse), and Competitors were different-category objects of the same color (e.g., pink fish). At adjective onset, referential ambiguity allowed us to examine prosody effects distinct from phonological cohort competition (see Experiment 1). Targets and Competitors were paired with two pictures from the same categories, but differing in color (e.g., orange horse and fish). Four versions of each base item were used to create four counterbalancing lists such that each list contained four items in each condition and each base item appeared just once in every list. Across all lists, 16 critical trials were randomly presented with eight filler trials. Targets in the filler trials did not share category or property with other objects in the set. Sentences within each instruction pair always referred to different pictures.

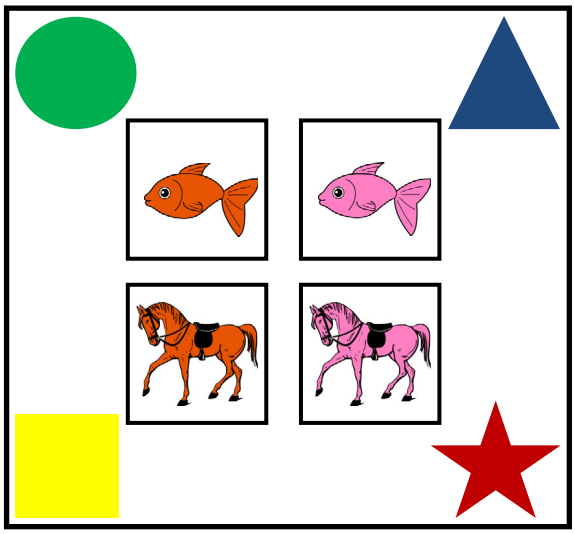

Fig. 5. In Experiment 2, sample display for a critical trial. In the sentence "Now put the pink horse on the triangle," the Target is the pink horse and the Competitor is the pink fish. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1.4. Coding

Across listeners, approximately 0.1% of trials were excluded from subsequent analyses because of experimenter error. The remaining data were coded in the manner described in Experiment 1. Missing frames accounted for 14.7% for NH listeners presented with natural speech, 13.7% for NH listeners presented with vocoded speech, and 18.7% for CI users. Twenty-five percent of the trials were checked by second coders who confirmed the direction of fixations for 92.8% coded frames for NH listeners presented with natural speech, 92.1% for NH listeners presented with vocoded speech, and 89.4% for CI users.

3.2. Results

The data were analyzed in the manner described in Experiment 1.

3.2.1. Actions

Action accuracy was numerically higher than in Experiment 1 and near ceiling across all listener groups (>95%). This is consistent with the hypothesis that prior challenges were due to cohort competition (e.g., "*sandals*" vs. "*sandwich*"). Since target adjectives in Experiment 2 were phonologically distinct (e.g., "*orange*" vs. "*pink*"), they were less likely to be confused. We found no significant effects of discourse status, prosody, or listener group (p's > 0.10).

3.2.2. Eye-movements

Initial analyses confirmed that Target looks did not differ across conditions prior to the onset of the critical adjective (p's > 0.50). To assess sensitivity to prosody, our analyses focused on fixations from adjective onset to sentence offset. The average length of this period was 1930 ms. Figs. 6 and 7 illustrate that all listeners restricted reference to the correct Target. However, changes in Target and Competitor fixations reveal that prosody effects for adjectives emerged for NH listeners presented with natural speech and CI users, but not NH listeners presented with vocoded speech. In the given condition, NH listeners presented with natural speech avoid looks to Competitors in accented trials, and CI users converge on Targets in this context. In contrast, NH listeners presented with vocoded speech generate more incorrect Competitors looks. In the new condition, NH listeners presented with natural speech and CI users are less likely to consider Competitors in unaccented trials compared to accented trials. Relative to these groups, NH listeners presented with vocoded speech show a more diminished pattern.

Closer inspection reveals that fixation latencies again varied with group. However, this pattern did not resemble the pattern in Experiment 1. Within level of prosody, cluster-based permutation analyses compared looks to the unmentioned object for the new compared to given condition. This revealed that word recognition began 200 ms after adjective onset in NH listeners presented with natural speech and CI users. The absence of delays in the latter group suggests that challenges with word recognition may be specific to contexts with salient cohort competitors (e.g., "*sandals*" vs. "*sandwich*"). In contrast, NH listeners presented with vocoded speech showed a dual pattern: 200 ms window in unaccented trials, but 500 ms window in accented trials. This suggests that signal degradation may lead to general delays in word recognition for listeners who have little experience with this input. Across all groups, we defined a standard 500 ms window from the onset of word recognition and calculated average Target preference in this region. Similar to Experiment 1, cross-model comparisons revealed an interaction between discourse status and prosody ($\chi^2(1) = 7.14$, $p < 0.01$). Consistent with prior research, Target preference in the given condition was greater in accented compared to unaccented trials (M = 69% vs. 65%) and reversed in the new condition (M = 56% vs. 63%). Importantly, our analyses also revealed a 3-way interaction between discourse status, prosody, and listener group ($\chi^2(6) = 28.81$, $p < 0.001$). To unpack this, follow-up analyses focused on fixed effects within levels of discourse status.

In the given condition, there was a significant effect of listener group ($\chi^2(2) = 49.86$, $p < 0.001$) but no reliable effect of prosody ($\chi^2(1) = 1.54$, $p > 0.20$). Fig. 8a illustrates that Target preference was greater for NH listeners presented with vocoded speech compared to NH listeners presented with natural speech ($t = 5.15$, $p < 0.001$) and CI users ($t = 3.27$, $p < 0.001$). It was also greater for NH listeners presented with natural speech compared to CI users ($t = 7.93$, $p < 0.001$). Similar to Experiment 1, the advantage found in NH listeners presented with vocoded speech is likely driven by the fact that regions of analysis were shifted later, thus providing disambiguating information on the noun. Differences between NH listeners presented with natural speech and CI users suggest that the latter may experience greater uncertainty during reference restriction, even when the onset of word recognition is similar (both groups time-locked to 200 ms after adjective onset). As in Experiment 1, cross-model comparisons revealed no interaction between listener group and prosody ($\chi^2(2) = 0.06$, $p > 0.50$).

However, a different pattern emerged in the new condition. Unlike prior analyses, main effects of prosody ($\chi^2(1) = 5.88$, $p < 0.05$) and listener group ($\chi^2(2) = 56.37$, $p < 0.001$) co-occurred with an interaction between these factors ($\chi^2(2) = 10.96$, $p < 0.01$). Similar to the given condition, Fig. 8b illustrates that overall Target preference was greater for NH listeners presented with vocoded speech compared to NH listeners presented with natural speech ($t = 8.25$, $p < 0.001$) and CI users ($t = 6.66$, $p < 0.001$). The latter groups did not differ from each other ($t = 0.63$, $p > 0.50$). Comparisons within levels of prosody revealed that Target preference was appropriately greater in unaccented compared to the accented trials in NH listeners presented with natural speech and CI users ($t = 3.30$, $p < 0.001$). Importantly, the size of this prosody effect did not differ across the two groups ($t = 0.94$, $p > 0.90$). In contrast, a *reverse* pattern was found in NH listeners presented with vocoded speech, but this difference did not approach significance ($t = 0.58$, $p > 0.50$). This suggests that unlike nouns, interpreting prosody on adjectives varies with signal quality (leading to larger effects in NH listeners presented with natural speech compared to vocoded speech) and prior experience with the degraded signal (leading to larger effects in CI users compared to NH listeners presented with vocoded speech).

3.3. Discussion

In Experiment 2, we found that when all acoustic cues are present, NH listeners recruit prosody on adjectives to infer discourse prominence. This is most clearly seen in reference to new categories, where Target fixations were greater following unaccented compared to accented adjectives. This pattern replicates prior work with natural speech (Ito & Speer, 2008; Ito et al., 2014; Sekerina & Trueswell, 2012; Weber et al., 2006). However, when pitch cues were degraded in vocoded speech, NH listeners showed limited inferencing via intensity and duration cues. This contrasts with Experiment 1, where NH listeners used intensity changes to infer prominence with nouns. This also contrasts with CI users who inferred likely referent for nouns *and* adjectives in the presence of signal degradation, and exhibited patterns that mirrored NH listeners presented with natural speech. Together, these results suggest that prior experience with degraded signals may be necessary to exploit non-pitch cues under some circumstances. We will explore this point in detail in Section 4.

In the meantime, we first wanted to rule out alternative explanations for why NH listeners failed to recruit intensity and duration changes on vocoded adjectives. One possibility is that they simply failed to perceive these cues, which in turn blocked inferences of discourse prominence. However, acoustic analyses of vocoded speech revealed similar intensity changes for nouns and adjectives ($F(1,30) = 0.01$, $p > 0.80$), and even greater duration changes for adjectives compared to nouns ($F(1,30) = 3.84$, $p < 0.10$). This demonstrates that acoustic cues were clearly present in both contexts. Moreover, eye-tracking data suggest that NH listeners were in fact sensitive to these acoustic differences on vocoded adjectives. Following adjective onset, they generated more Target fixations in accented compared to unaccented trials (M = 81% vs. 77%; $t = 1.71$, $p < 0.05$). Importantly, though, this difference did not vary with discourse status ($t = 0.14$, $p > 0.80$). Thus, they perceived the acoustic distinctions between accented and unaccented trials, but did not use these cues to predict the identity of the subsequent noun. This pattern suggests that perceptual sensitivity alone does not guarantee inferencing of meaning.

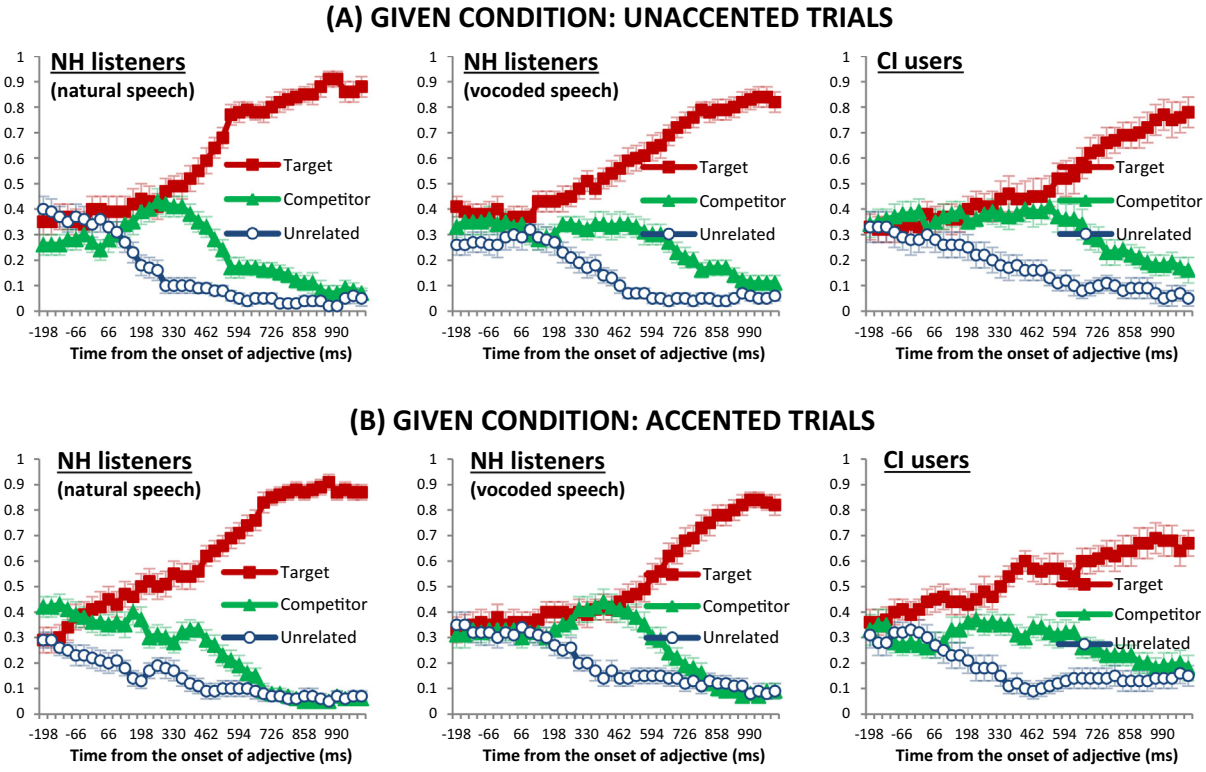

Fig. 6. In Experiment 2, proportion of fixations to Target, Competitor, and Unrelated objects from adjective onset in (A) unaccented and (B) accented trials of the given condition.

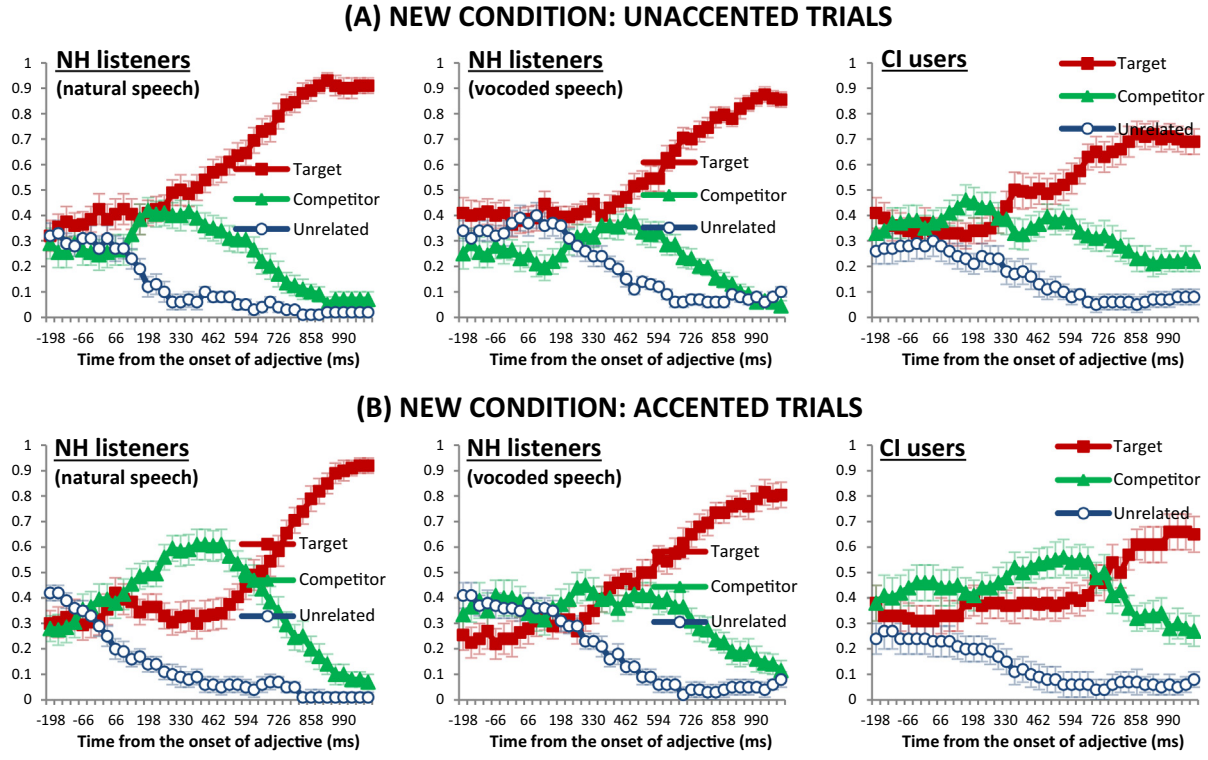

Fig. 7. In Experiment 2, proportion of fixations to Target, Competitor, and Unrelated objects from adjective onset in (A) unaccented and (B) accented trials of the new condition.

A second possibility is that prominence relies on distinct cues for nouns and adjectives. It has been noted that speakers increase word duration at the end of phrases to plan for subsequent content (Bard et al., 2000; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). Effects of phrase-final lengthening may interact with syntactic properties of English. In particular, nouns often occupy phrase-

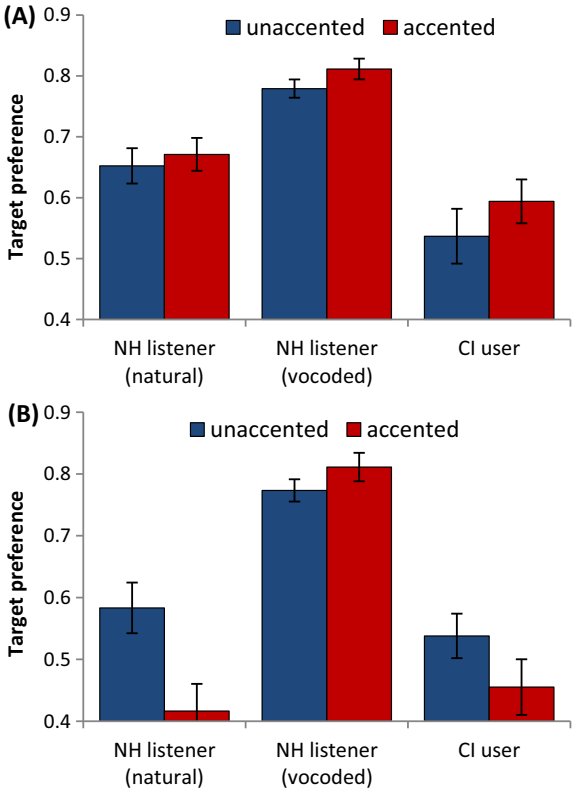

Fig. 8. In Experiment 2, Target preference during regions of analysis in listener groups in the (A) given condition and (B) new condition.

final positions (e.g., "*Now put the sandwich*…"), thus accenting may exaggerate duration changes that can be detected in both natural and vocoded speech. In contrast, adjectives often occur prenominally (e.g., "*Now put the orange horse*…") (Mintz, 2003; Saylor, 2000), thus they may be less likely to benefit from effects of phrase-final lengthening. This suggests that listeners may rely on pitch changes to a greater extent when interpreting accented adjectives. These cues are readily available in natural speech, but not vocoded speech.

However, analyses of natural speech tokens in the current study revealed greater pitch changes across accented and unaccented nouns ($p < 0.01$), but no differences in duration ($p > 0.10$). This suggests that phrase-final lengthening does not always exaggerate duration differences on nouns. In fact, to the extent that word duration often correlates with production difficulty (Bard & Aylett, 1999; Fowler & Housum, 1987; Lam & Watson, 2010), it is argued to be a less reliable cue to discourse prominence (Arnold & Watson, 2015; Watson, 2010). Moreover, a preference for pitch cues in adjectives fails to explain why prosody effects emerged in CI users. Indeed, if inferencing for adjectives depends on prior experience with pitch cues in the speech signal, one might have expected CI users to exhibit weaker effects compared to NH listeners with natural speech in this context. Yet, there was no difference in performance across groups for adjectives ($p > 0.30$). Similarly, within CI users, there was no evidence of weaker effects for adjectives compared to nouns ($p > 0.80$).

4. General discussion

The current study explored how interpretation of prosody is affected by degraded pitch cues. We found that inferences of discourse prominence depend on both the properties of the signal and the demands of comprehension. In the case of nouns, effects of signal degradation were minimal: CI users and NH listeners presented with vocoded speech inferred prominence to the same degree as NH listeners presented with natural speech. This suggests that listeners can recruit changes in intensity when pitch information is minimized. However, while CI users also interpreted prosody on adjectives, NH listeners were unable to do so with vocoded speech. This suggests that comprehension strategies may vary with prior experience. Since CI users encounter limited pitch cues in everyday communication, they may develop a greater sensitivity to how intensity and duration changes impact intended meaning (see also Winn et al., 2012). In contrast, prior experience with natural speech may make pitch changes particularly salient for NH listeners. Thus, their presence in prosody supports inferences of discourse status across noun and adjective contexts, while their absence leads to more restricted inferencing.

In the remainder of this discussion, we will examine three additional issues related to the current findings. First, we will reconcile CI users' sensitivity to prosody in the current study with the evidence of comprehension difficulty in prior work. Next, we will consider how our findings inform debates about the acoustic correlates of prominence in the psycholinguistics literature. Finally, we will consider the implications of our findings in light of recent accounts of how listeners recruit acoustic cues across variable circumstances and the role of prior experience in adapting to the current context (Kleinschmidt & Jaeger, 2015).

4.1. Reconciling prosody effects in CI users with prior studies

The precocity of CI users in this study raises questions of why they faced difficulties in prior work (Chatterjee & Peng, 2008; Gilbers et al., 2015; Hopyan-Misakyan et al., 2009; Luo et al., 2007; Meister et al., 2009; Morris et al., 2013; Nakata et al., 2012; Peng et al., 2012; Van Zyl & Hanekom, 2013). We consider three factors that may contribute to this discrepancy.

First, as noted in the Introduction, studies vary in what phenomena they tackle. Prior work focuses on how prosody distinguishes questions-statements (Chatterjee & Peng, 2008; Meister et al., 2009; Peng et al., 2012; Van Zyl & Hanekom, 2013) and speaker emotions (Hopyan-Misakyan et al., 2009; Luo et al., 2007; Nakata et al., 2012). While pitch cues are primary in these cases, it is unclear whether changes in intensity and duration reliably exist (Srinivasan & Massaro, 2003). By turning to a context where redundant cues are available (Bard & Aylett, 1999; Fowler & Housum, 1987; Lam & Watson, 2010; Breen et al., 2010; Cole et al., 2010; Kochanski et al., 2005; Wagner & Klassen, 2015; Watson et al., 2008), we show that CI users can interpret prosody to infer prominence. Moreover, our findings suggest that even descriptively prosodic phenomena can vary in their comprehension demands, e.g., prominence on nouns vs. adjectives. Thus, it is unsurprising that success is found in some cases but not others.

The current study also differs in how prosody interpretation is assessed. Previous studies rely on metalinguistic judgments that are made after stimulus presentation, which can be subject to different response criteria across populations (Ratcliff, 1978; Ratcliff, Gomez, & McKoon, 2004). Thus, even if CI users perceive prosodic cues to the same degree as NH listeners, they may face greater uncertainty due to challenges in memory retrieval or metacognitive awareness of hearing difficulties. By recruiting eye-tracking as an implicit assessment of interpretation, we can distinguish between factors that influence comprehension (e.g., signal properties, linguistic processing) from those that affect response generation (e.g., memory decay, action planning, metalinguistic awareness). We also distinguished effects of signal degradation on *when* word recognition begins from those that impact *what* cues are used for interpretation. Together, this suggests that eye-tracking methods may allow for more direct comparisons of lan-

guage processing in groups that differ substantially in cognitive abilities and prior experience.

Finally, prior research often examines prosody in contexts involving single words or short sentences (Chatterjee & Peng, 2008; Hopyan-Misakyan et al., 2009; Luo et al., 2007; Meister et al., 2009; Nakata et al., 2012; Van Zyl & Hanekom, 2013). In contrast, the current study measures interpretation across multiple sentences and in the context of referential scenes. These conditions approximate the range of cues that may facilitate or hinder real-world communication, dynamics that were on display when salient cohort competitors co-occurred with discourse cues in Experiment 1. Action accuracy was higher in the given compared to new condition, suggesting that all listeners benefited when the same referent was mentioned across sentences. Yet, cohort competitors led to increased errors in CI users relative to NH listeners, suggesting that they may face disproportionate difficulties resolving competition introduced by the bottom-up signal (e.g., "*sand*…" can refer to either sandals or sandwich). Nevertheless, CI users exploited prosody cues in this context, raising the possibility that extensive experience with signal degradation motivates comprehension strategies at multiple levels: (1) greater sensitivity to intensity and duration cues during bottom-up analysis and (2) increased attention to top-down cues to ease the challenges of interpreting a noisy signal (see also Gibson et al. (2013) and Levy et al. (2009) for related effects on syntactic parsing in NH listeners). While it is difficult to distinguish the relative contributions of these two processes in the current study, future work examining a wider range of phenomena may shed light on how these interactions contribute to language comprehension in CI users.

4.2. Are pitch cues primary for discourse prominence?

While the current study was not designed to test between specific theories of prosody, our data speak to on-going debates about the acoustic correlates of prominence. Consistent with prior accounts (Bolinger, 1986; Cruttenden, 1997; Pierrehumbert & Hirschberg, 1990) and empirical findings (Bartels & Kingston, 1994; Cooper et al., 1985; Isaacs & Watson, 2010; Ladd & Morton, 1997; Terken, 1991), our data suggest that pitch changes are particularly salient for NH listeners. When this cue is present in natural speech, they will recruit it to infer prominence for both nouns and adjectives. However, when it is degraded in vocoded speech, inferencing occurs in the former but not the latter. This difference cannot be due to the sheer number of acoustic cues available in the speech signal since vocoded nouns featured only intensity changes in our items while vocoded adjectives featured duration changes as well.

Yet, our data also suggest that pitch cues may not be necessary to infer prominence. In the noun context, NH listeners with vocoded speech spontaneously switched to intensity cues, revealing prosody effects that were of the same magnitude as in natural speech. Moreover, across both noun and adjective contexts, CI users recruited non-pitch cues to the same extent as NH listeners with natural speech. This suggests that speakers' use of accenting to imply prominence must be reliably correlated with intensity and/or duration changes, such that listeners (depending on the context, with sufficient experience) can recruit these cues when pitch changes are less available. This pattern is also consistent with prior evidence that pitch may not be essential for inferring prominence (Cole et al., 2010; Kochanski et al., 2005).

Finally, our results speak to possible causes of discrepancies across prior research. In eye-tracking studies, NH listeners fail to infer prominence when pitch cues are absent and/or degraded (Isaacs & Watson, 2009; Experiment 2 in this study). Yet, in corpus analyses, pitch cues are surprisingly poor predictors of prominence (Cole et al., 2010; Kochanski et al., 2005). While these approaches vary in their materials (see Watson (2010) for more discussion), they also differ in how sensitivity to prominence is defined. Eye-tracking studies focus on how listeners recruit acoustic cues to infer meaning-based interpretation (e.g., "*PINK horse*" → not the orange horse). In contrast, corpus analyses ask listeners to note which syllables/words "stand out" (Cole et al., 2010; Kochanski et al., 2005). Our findings suggest that these behaviors are clearly not equivalent. In the case of adjectives, NH listeners presented with vocoded speech generated more Target looks in accented trials compared to unaccented ones, demonstrating that they detected intensity and/or duration changes in the speech signal. Importantly, this ability alone was insufficient for inferring contrast between referents in the discourse. Together, this suggests that isolating the acoustic correlates of prominence requires specifying the comprehension processes they trigger (e.g., detecting cues in the signal, inferring meaning on this basis).

4.3. To adapt or generalize: effects of experience on speech perception

On the face of things, NH listeners' failure to infer discourse prominence for vocoded adjectives is somewhat surprising. After all, prior work reveals evidence of recalibration to novel phonemic categories (Kraljic & Samuel, 2006; Clayards et al., 2008; Maye et al., 2008; Norris et al., 2003) and word recognition in vocoded speech (Davis et al., 2005; Rosen et al., 1999; Shannon et al., 1995, 1998). These effects emerge with strikingly minimal experience. Moreover, NH listeners' success with prosody on vocoded nouns suggests that they are able to use non-pitch cues to infer discourse prominence under some circumstances. Yet, this ability appears to be more limited than that of CI users, who recruit intensity and duration changes for both accented nouns and adjectives.

While additional research is needed to isolate the nature of this difference, our findings are consistent with recent Bayesian models that predict systematic interactions between prior experience and current demands (Gibson et al., 2013; Kleinschmidt & Jaeger, 2015; Levy et al., 2009). In our experiments, all listeners can encode dimensions of contrast within the referential scene prior to the onset of the utterance (e.g., noun: new vs. given category; adjectives: new vs. given category *and* property). To recruit this information for interpreting an accented word, they must exploit acoustic cues in the signal to isolate the relevant dimension of contrast (e.g., nouns imply contrast with a new category; adjectives imply property contrast within a given category). Importantly, when pitch changes are present in natural speech, NH listeners can rely on their experiences with this cue when making this inference. Yet, when pitch changes are degraded in vocoded speech, they now must isolate both novel acoustic cues and relevant mappings over the course of 16 trials. Note that the current study does not offer clear statistical evidence in favor of a cue-to-meaning mapping since prosody is used both felicitously (i.e., accented noun implies new category, accented adjective implies a given category) and infelicitously (i.e., accented noun implies given category, accented adjective implies a new category). Nevertheless, NH listeners converge on the relevant mapping when there is only a single dimension of contrast for nouns. Yet, when multiple dimensions of contrast are present for adjectives, they are able to detect the presence of accenting in vocoded speech (i.e., acoustic differences in intensity and duration), but remain agnostic as to how these cues map onto communicative meaning.

In contrast, CI users may weigh duration and intensity changes more heavily by virtue of their extensive experience with the degraded signal, and thus are able to apply these cues across contexts. Consistent with this possibility, follow-up analyses revealed that in the challenging case of adjectives, individual differences in the magnitude of prosody effects was correlated with years of CI use ($r(20) = 0.48$, $p < 0.05$). In contrast, no such relationship

emerged in the easy case of nouns ($r(20) = 0.17$, $p > 0.40$). Similar distinctions are also found when NH listeners interpret prosody in natural speech. Even when pitch cues are available, NH adults sometimes fail to infer prominence for adjectives (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; see discussion by Ito & Speer, 2008). In NH children, age-related delays emerged when interpreting accented adjectives relative to accented nouns (Arnold, 2008; Ito et al., 2014; Sekerina & Trueswell, 2012). These findings provide converging evidence that beyond signal properties, interpreting prosody involves a set of real-time computations that link acoustic cues to meaning. Thus, unlike cases of rapid recalibration to novel phonemic categories, NH listeners in the current study must retune their language systems at both lower (i.e., tracking intensity and duration changes) and higher levels (i.e., tracking their relationships to discourse status). If latter procedures are more complex for adjectives compared to nouns, interpreting prosody may be more difficult in this case.

Taken together, the current findings inform basic questions of how experience influences speech perception over the course of minutes (for NH listeners) versus years (for CI users). In particular, they provide support for a distinction between adaptation and generalization in speech perception (Kleinschmidt & Jaeger, 2015). Due to minimal pitch cues, vocoded speech has a salient "robotic" quality. Since NH listeners in our study had no prior experience with this signal, they had no basis for assuming that it would be systematically related to natural speech, much less inferring what these instantiations might be. Under these circumstances, the rational strategy may be to *adapt* to the properties of novel input through brute force experience with bottom-up statistics. NH listeners' success with prosody on nouns suggests that they rapidly converge on a notion that increased intensity implies novelty. In contrast, CI users have vast experience with this degraded signal, thus they may *generalize* non-pitch cues to structured meaning categories. For postlingual individuals who receive CIs later in life, interactions in a communicative setting may lead to knowledge of how intensity and duration cues in vocoded speech function similarly to pitch cues in natural speech. By formulating relationships over structured categories (e.g., discourse status), they may recruit greater top-down expectations of how non-pitch cues correlate with meaning across contexts (e.g., nouns vs. adjectives).

This interpretation of the current findings is consistent with prior evidence that top-down knowledge facilitates comprehension of unfamiliar speech. For example, NH listeners recruit lexical information when interpreting phonemic variation along a continuum, judging ambiguous sounds as [f] when they appear in words that end with [f] but reinterpreting them as [s] in words that end with [s] (Norris et al., 2003). In contrast, their responses are unbiased when ambiguous sounds occur in nonwords. Sensitivity to lexical cues may also explain why comprehension of vocoded speech improves when NH listeners are trained with nonsense sentences with known words (e.g., "*The effect supposed to the consumer*") but not when these sentences involve novel words (e.g., "*Chotekine garund pid ga sumeun*") (Davis et al., 2005). Both contexts provide bottom-up experiences with how vocoded speech distorts phonemic representations in English, and neither context conveys information about meaning at the sentence level. Yet, evidence of distinct learning suggests that accessing lexical-level representations may be particularly useful for comprehending degraded speech.

Finally, distinguishing between adaptation and generalization may shed light on varying comprehension demands facing different populations of CI users. Unlike postlingual CI users (who have acquired English through experience with natural speech), prelingual CI users must simultaneously isolate the relevant sounds categories within their language and map these forms to meaning during acquisition. Early implantation offers tremendous advantages for language development (Nicholas & Geers, 2007; Niparko et al., 2010). Yet, even among those who receive CIs before 18 months, comprehension delays persist relative to age-matched NH peers. While prior work relies on aggregated measures of ability (e.g., standardized language assessments), the current study suggests that a finer-grained approach may reveal how developmental outcomes arise from distinct strategies for listening and learning with a degraded signal. Notably, the extended cohort competition found in adult CI users in Experiment 1 (e.g., longer looks to the sandals after "*Now put the SANDWICH…*"; see also Farris-Trimble et al., 2014) also emerges in prelingually deafened 3- to 5-year-olds (Edwards, 2017). This suggests that prolonged experience with a degraded signal may systematically alter relationships between phonological and semantic representations within the lexicon.[4]

5. Conclusion

The current study investigated how CI users and NH listeners interpret acoustic correlates of discourse prominence. Our results demonstrate that the comprehension of prosody involves a dynamic interplay between signal properties and linguistic processing. Much like NH listeners presented with natural speech, CI users and NH listeners presented with vocoded speech recruit intensity changes to infer the discourse status of nouns. Similarly, CI users reveal prosody effects for adjectives as well. In contrast, NH listeners presented with vocoded speech are sensitive to intensity and duration changes in this context, but are unable to use these cues to infer discourse status. Together, these findings suggest that the ability to interpret prosody varies with the real-time demands of mapping acoustic cues to meaning and the range of experiences that different listeners have with making these mappings.

Acknowledgments

We are grateful to Alison Arnold, Kerianna Frederick, Sean Anderson, Lauren Evans, Ana Medina Fetterman, Brittany Jaekel, Melissa Stockbridge, and Erin Walter for their assistance in data collection, coding, and analysis. Portions of this work have been presented at the 2014 MASH Conference on Cochlear Implant Research and the 2015 CUNY Conference on Human Sentence Processing. This work was partially supported by a Tier 1 seed grant from the VPR Office at UMCP and a grant from the National Institute On Aging of the National Institutes of Health (R01AG051603). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.05.029.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition, 108*, 69–99.

[4] Follow-up analyses in the current study found no differences in prosody effects for prelingually versus postlingually deafened CI users (p's > 0.20). This likely reflects our limited sample size since there were only five prelingually deafened individuals in both Experiments 1 and 2.

Arnold, J. E., & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: Performance matters. *Language, Cognition, and Neuroscience, 30*, 88–102.
Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.
Bard, E. G., & Aylett, M. P. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 3, pp. 1753–1756). CA: University of California Berkeley.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.
Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. *Journal of the Acoustical Society of America, 95*, 2973.
Bates, D., Maechler, M., & Bolker, B. (2013). *Linear mixed-effects models using S4 classes. R package version 0.999999-0*.
Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644–663.
Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111.
Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., Burke, E., ... Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology, 18*, 36–47.
Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford University Press.
Boons, T., Brokx, J. P., Dhooge, I., Frijns, J. H., Peeraer, L., Vermeulen, A., ... Van Wieringen, A. (2012). Predictors of spoken language development following pediatric cochlear implantation. *Ear and Hearing, 33*, 617–639.
Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*, 1044–1098.
Budenz, C. L., Cosetti, M. K., Coelho, D. H., Birenbaum, B., Babb, J., Waltzman, S. B., & Roehm, P. C. (2011). The effects of cochlear implantation on speech perception in older adults. *Journal of the American Geriatrics Society, 59*, 446–453.
Burkholder, R. A., & Pisoni, D. B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*, 63–88.
Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research, 235*, 143–156.
Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*, 804–809.
Cleary, M., Pisoni, D. B., & Kirk, K. I. (2000). Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. *Volta Review, 102*, 259.
Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1*, 425–452.
Collison, E. A., Munson, B., & Carney, A. E. (2004). Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 47*, 496–508.
Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *Journal of the Acoustical Society of America, 77*, 2142–2156.
Cruttenden, A. (1997). *Intonation*. Cambridge University Press.
Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314.
Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition, 143*, 77–86.
Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*, 222–241.
Dink, J. W., & Ferguson, B. F. (2015). *eyetrackingR: An R library for eye-tracking data analysis*. Retrieved from <http://www.eyetrackingr.com>.
Edwards, J. (2017). Lexical processing in children with cochlear implants. In *Paper presented at the 5th annual Cochlear implant Mid-Atlantic Seminar on Hearing (MASH), College Park, MD, January*. .
Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 308–327.
Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.
Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 11*, 1150–1163.
Geers, A. E., Pisoni, D. B., & Brenner, C. (2013). Complex working memory span in cochlear implanted and normal hearing teenagers. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 34*, 396.
Geers, A. E., & Sedey, A. L. (2011). Language and verbal reasoning skills in adolescents with 10 or more years of cochlear implant experience. *Ear and Hearing, 32*, 39S–48S.
Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, 110*, 8051–8056.
Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., & Başkent, D. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *i-Perception, 6*, 1–19.
Hagedorn, J., Hailpern, J., & Karahalios, K. G. (2008). VCode and VData: Illustrating a new framework for supporting the video annotation workflow. In *Proceedings of the working conference on advanced visual interfaces* (pp. 317–321). ACM. May.
Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiology and Neurotology, 12*, 254–264.
Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., ... Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and Hearing, 34*, 342–360.
Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology, 15*, 136–146.
Huang, Y., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language, 37*, 644–661.
Huang, Y., Zheng, X., Meng, X., & Snedeker, J. (2013). Assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language, 69*, 589–606.
Isaacs, A. M., & Watson, D. G. (2009). Speakers and listeners don't agree: Audience design in the production and comprehension of acoustic prominence. In *Poster presentation at CUNY 2009: Conference on human sentence processing, Davis, CA, March*. .
Isaacs, A. M., & Watson, D. G. (2010). Accent detection is a slippery slope: Direction and rate of f0 change drives listeners' comprehension. *Language and Cognitive Processes, 25*, 1178–1200.
Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language, 41*, 84–110.
Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*, 541–573.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards mixed effects models. *Journal of Memory and Language, 59*, 434–446.
Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*, 148–203.
Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*, 1038–1054.
Kraljic, T., & Samuel, A. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review, 13*, 262–268.
Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25*, 313–342.
Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition, 38*, 1137–1146.
Leung, J., Wang, N. Y., Yeagle, J. D., Chinnici, J., Bowditch, S., Francis, H. W., & Niparko, J. K. (2005). Predictive models for cochlear implantation in elderly candidates. *Archives of Otolaryngology-Head & Neck Surgery, 131*, 1049–1054.
Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106*, 21086–21090.
Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*, 451–454.
Luo, X., Fu, Q. J., & Galvin, J. J. 3rd, (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification, 11*, 301–315.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.
Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32*, 543–562.
Meister, H., Landwehr, M., Pyschny, V., Walger, M., & von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology, 48*, 38–48.
Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.
Morris, D., Magnusson, L., Faulkner, A., Jönsson, R., & Juul, H. (2013). Identification of vowel length, word stress, and compound words and phrases by postlingually deafened cochlear implant listeners. *Journal of the American Academy of Audiology, 24*, 879–890.
Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *Journal of the Acoustical Society of America, 131*, 1307–1314.
Nicholas, J. G., & Geers, A. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to

profound hearing loss. *Journal of Speech, Language, and Hearing Research, 50*, 1048–1062.
NIDCD (2012). Cochlear implants Retrieved from <http://www.nidcd.nih.gov/health/hearing/pages/coch.aspx>.
Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., ... CDaCI Investigative Team (2010). Spoken language development in children following cochlear implantation. *Journal of the American Medical Association, 303*, 1498–1506.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.
Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification, 20*, 1–11.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, Massachusetts: MIT Press.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.
Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.
Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*, 3629–3636.
Salverda, A. P., & Tanenhaus, M. K. (in press). The visual world paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics*. Hoboken, NJ: Wiley (in press).
Sarant, J., Harris, D., Bennet, L., & Bant, S. (2014). Bilateral versus unilateral cochlear implants in children: A study of spoken language outcomes. *Ear and Hearing, 35*, 396–409.
Saylor, M. M. (2000). Time-stability and adjective use by child and adult English speakers. *First Language, 20*, 91–120.
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109–147.
Sekerina, I. A., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language, 32*, 63–87.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303–304.
Shannon, R. V., Zeng, F. G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104*, 2467–2476.
Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238–299.
Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech, 46*, 1–22.
Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America, 89*, 1768–1776.
Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*, 145–163.
Van Zyl, M., & Hanekom, J. J. (2013). Perception of vowels and prosody by cochlear implant recipients in noise. *Journal of Communication Disorders, 46*, 449–464.
Wagner, M., & Klassen, J. (2015). Accessibility is no alternative to alternatives. *Language, Cognition and Neuroscience, 30*, 212–233.
Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*, 905–945.
Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *Psychology of Learning and Motivation, 52*, 163–183.
Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*, 1548–1557.
Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49*, 367–392.
Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *Journal of the Acoustical Society of America, 131*, 1465–1479.
Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing, 20*, 153–165.