JSLHR

Research Article

Gated Word Recognition by Postlingually Deafened Adults With Cochlear Implants: Influence of Semantic Context

Chhayakanta Patro[a] and Lisa Lucks Mendel[b]

Purpose: The main goal of this study was to investigate the minimum amount of sensory information required to recognize spoken words (isolation points [IPs]) in listeners with cochlear implants (CIs) and investigate facilitative effects of semantic contexts on the IPs.
Method: Listeners with CIs as well as those with normal hearing (NH) participated in the study. In Experiment 1, the CI users listened to unprocessed (full-spectrum) stimuli and individuals with NH listened to full-spectrum or vocoder processed speech. IPs were determined for both groups who listened to gated consonant-nucleus-consonant words that were selected based on lexical properties. In Experiment 2, the role of semantic context on IPs was evaluated. Target stimuli were chosen from the Revised Speech Perception in Noise corpus based on the lexical properties of the final words.
Results: The results indicated that spectrotemporal degradations impacted IPs for gated words adversely, and CI users as well as participants with NH listening to vocoded speech had longer IPs than participants with NH who listened to full-spectrum speech. In addition, there was a clear disadvantage due to lack of semantic context in all groups regardless of the spectral composition of the target speech (full spectrum or vocoded). Finally, we showed that CI users (and users with NH with vocoded speech) can overcome such word processing difficulties with the help of semantic context and perform as well as listeners with NH.
Conclusion: Word recognition occurs even before the entire word is heard because listeners with NH associate an acoustic input with its mental representation to understand speech. The results of this study provide insight into the role of spectral degradation on the processing of spoken words in isolation and the potential benefits of semantic context. These results may also explain why CI users rely substantially on semantic context.

Accurate understanding of conversational speech requires rapid and efficient recognition of the words embedded in a spoken message. In day-to-day conversation, listeners often tend to speak very fast, and despite its rapid presentation, listeners with normal hearing (NH) are capable of understanding speech very well, possibly because they take advantage of the mental schemas stored in their long-term memory (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). However, it is less clear how such acoustically degraded spoken words conveyed by current-day cochlear implants (CIs) are processed and gain access to mental representations. Semantic context is known to exploit word knowledge, grammar, situational awareness, and knowledge about the topic of conversation and thus helps listeners to compensate for ambiguous speech information (Kalikow, Stevens, & Elliott, 1977; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). As a result, listeners more easily recognize words in sentences compared to in isolation due to the addition of contextual cues (Boothroyd & Nittrouer, 1988; Grant & Seitz, 2000). Recent reports have also suggested that CI users benefit from semantic context despite poor spectrotemporal resolution (Patro & Mendel, 2016; Shafiro, Sheft, Gygi, & Ho, 2012; Shafiro, Sheft, Kuvadia, & Gygi, 2015). However, it is yet unknown whether acoustically degraded contextual cues (as conveyed by CIs) still assist in processing final words in a sentence. The current study investigated how listeners with CIs process spoken words and how semantic context aids in gaining lexical access using a gated word recognition task.

Spectrotemporal Degradation and Cochlear Implantation

A significant consequence of hearing impairment is the inability to understand spoken messages clearly. CIs

[a]Heuser Hearing Institute, Louisville, KY
[b]School of Communication Sciences & Disorders, University of Memphis, TN

Correspondence to Chhayakanta Patro: chhayakantpatro@gmail.com

Editor-in-Chief: Frederick Gallun
Editor: Lori J. Leibold

Received April 17, 2017
Revision received June 15, 2017
Accepted August 28, 2017
https://doi.org/10.1044/2017_JSLHR-H-17-0141

Disclosure: The authors have declared that no competing interests existed at the time of publication.

are among the great success stories as they have become a standard and viable method of intervention for patients with severe-to-profound sensorineural hearing loss. In adults who are postlingually deaf, postimplantation improvements are common in phoneme, word, and sentence recognition (Wilson & Dorman, 2008), and these listeners who are deaf perform very well in quiet listening environments (Blamey et al., 2013). However, their speech understanding in noise is severely affected (Eskridge, Galvin, Aronoff, Li, & Fu, 2012; Friesen, Shannon, Baskent, & Wang, 2001; Van Deun, Van Wieringen, & Wouters, 2010). This inferior performance in noise has been greatly attributed to a drastic reduction in signal fidelity due to the loss of finer acoustic details (temporal fine structure), and thus, the signals being transferred to the electrode array are severely limited (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Zeng et al., 2004). In addition, many other factors that are inherent to CI devices reduce the spectral resolution of the spoken message due to displaced electrode array (Başkent & Shannon, 2006; Holden et al., 2013), limited interaction between electrodes and auditory neurons (Bierer, Faulkner, & Tremblay, 2011; Khan et al., 2005; Qin & Oxenham, 2003), limited electrode channels (Friesen et al., 2001), and possible channel interactions among the implanted electrodes (Chatterjee & Shannon, 1998). As a result, CI recipients hear speech that is spectrotemporally degraded (Dorman, Loizou, Fitzke, & Tu, 1998; Friesen et al., 2001; Fu, Shannon, & Wang, 1998; Qin & Oxenham, 2003; Rosen, 1992; Stickney, Zeng, Litovsky, & Assmann, 2004).

Gated Word Recognition

It is widely demonstrated through computational modeling and behavioral speech recognition paradigms that word recognition occurs when an input acoustic signal activates its lexical representation (Lively, Pisoni, & Goldinger, 1994; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 2000). Models of spoken word recognition claim that the initial segments of a word activate a set of lexical candidates, and listeners recognize words even before they "hear" the whole word (Dirks, Takayanagi, Moshfegh, Noffsinger, & Fausti, 2001; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Slowiaczek, Nusbaum, & Pisono, 1987). This occurs because a cohort of words is triggered by the initial segments of words and a word candidate drops out from the competition when a mismatch occurs between signal acoustics and the activated cohort. This process continues until a single word candidate is chosen that best matches the sensory input. Hence, determination of a "winning candidate" requires interaction between available sensory information and lexical abilities due to this specific and restricted competition process (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Warren & Marslen-Wilson, 1987).

The gating paradigm, which was introduced by Grosjean (1980), is a direct way to assess the assumptions regarding word recognition being a competitive process. In a gating task, listeners are presented successive fragments of a target word and asked to indicate what they think the target word is (Cotton & Grosjean, 1984; Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Tyler & Wessels, 1983, 1985). Listeners are presented limited acoustic cues and are expected to use their higher-level skills to compensate for the missing information. The extent of top-down compensation is studied in terms of isolation points (IPs). An IP refers to the exact point in the word when a listener can correctly guess it. When a word is not repeated correctly, even after presenting the entire speech stimulus, its total duration plus one gate size will be estimated as the IP (Elliott, Hammer, & Evan, 1987; Hardison, 2005; Metsala, 1997; Walley, Michela, & Wood, 1995). Lexical access does not necessarily proceed in one specific direction in time; therefore, word identification may not always unfold from the beginning of the word to the end. Often, prolonged IPs may adversely impact identification of later portions of spoken words and ultimately affect understanding of continuous speech. Therefore, the current study explores the extent to which spectrotemporal degradation as imposed by current-day CIs can influence word prediction as evidenced by IPs in a forward direction.

Role of Semantic Context

Several studies have examined the effects of semantic context on the intelligibility of sentence-final words (Dubno, Ahlstrom, & Horwitz, 2000; Friesen et al., 2001; Obleser & Kotz, 2011; Patro & Mendel, 2016; Pichora-Fuller, Schneider, & Daneman, 1995; Sheldon, Pichora-Fuller, & Schneider, 2008). Semantic context in a spoken message contributes to top-down processing effects, which are known to influence speech recognition, because each word in a sentence helps to shortlist word candidates for the next word, and listeners more easily recognize words in sentences than in isolation (Boothroyd & Nittrouer, 1988; Grant & Seitz, 2000). It is evident from the literature that contextual knowledge contributes immensely in building a conceptual representation of the unfolding sentence, and thus, listeners rely substantially on such contextual information, especially under adverse listening conditions (Patro & Mendel, 2016; Pichora-Fuller et al., 1995; Pichora-Fuller & Souza, 2003; Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008; Wingfield, Lindfield, & Goodglass, 2000). Presumably, listeners with NH perform 20–35 percentage points higher for sentences that are contextually rich compared to contextually poor sentences (Boothroyd & Nittrouer, 1988; Dubno et al., 2000; Friesen et al., 2001; Patro & Mendel, 2016; Pichora-Fuller et al., 1995). Given the advantage of semantic context on a variety of speech recognition tasks, this study evaluated the role that semantic context may play for the recognition of gated words.

The Current Study

In Experiment 1, the objective was to track the time course of lexical access behaviorally in listeners with NH and CIs as evidenced by IPs in a gated word recognition paradigm. The speech degradations in the spectrotemporal domain experienced by typical CI users have detrimental effects on a variety of top-down skills, including lexical access (Başkent, 2012; Başkent & Chatterjee, 2010; Başkent, Eiler, & Edwards, 2009; Benard & Başkent, 2013; Bhargava & Başkent, 2012; Bhargava, Gaudrain, & Başkent, 2014; Chatterjee, Peredo, Nelson, & Başkent, 2010; Nelson & Jin, 2004; Nelson, Jin, Carney, & Nelson, 2003; Patro & Mendel, 2016). A recent study by Wagner, Toffanin, and Başkent (2016) measured the process of word retrieval and perceived listening effort objectively using pupilometry for natural and spectrotemporally degraded speech. They reported that, for processing natural speech, participants showed increased listening effort during lexical competition. However, for degraded speech, participants showed even more listening effort and a prolonged lexical competition process. The authors claimed that spectrotemporal degradation complicates the process of lexical access, increases cognitive load, and introduces uncertainties in word retrieval. However, it is less clear whether these degradations influence the initial acoustic information required (as measured by IPs) for correct identification of words. In Experiment 1, we hypothesized that spectrotemporal degradations (encountered in CIs) would delay the activation, as well as selection, of lexical candidates. Therefore, it was hypothesized that IPs would be prolonged.

In Experiment 2, the objective was to track the time course of lexical access with and without semantic context (for low-predictability [LP] and high-predictability [HP] sentences) in listeners with NH and CIs using sentences whose final words were gated. Prior research has shown that semantic context can increase the speed of word retrieval, because these contextual cues do not activate (or inhibit) the less probable lexical candidates that are incongruent with the sentential meaning (Marslen-Wilson, 1987; Miller, Heise, & Lichten, 1951; Salasoo & Pisoni, 1985). Second, listeners' reliance on the contextual cues depends on residual bottom-up cues (Bhargava et al., 2014; Oh, Donaldson, & Kong, 2016; Patro & Mendel, 2016; Sivonen, Maess, & Friederici, 2006). Considering these factors, it was hypothesized that the CI users would benefit greatly from semantic context, and their ability to predict the final word would be comparable to that of the listeners with NH.

The motivation for this study was to understand how the gated word recognition of CI users differs from the gated word recognition of listeners with NH. In this regard, IPs of CI groups were compared with IPs of listeners with NH for full-spectrum speech. Second, deficits in recognition of gated words can be explained based on poor spectrotemporal resolution conveyed through CI systems. To understand this, we also tested a group of age-matched young listeners with NH with noise-band vocoder processed stimuli and compared their performance with that of CI users. Thus, we could explore whether possible delays in IPs were primarily due to the spectrotemporal degradation or additional factors, such as auditory deprivation, surgical complications, front-end processing, electrode–neuron interface, and/or channel interaction, which also contribute to such deficiencies in processing spoken words. Note that if these additional factors do not contribute to gated word recognition, mean IPs of NH groups (for full-spectrum and vocoded stimuli) would be similar. Therefore, two groups of listeners with NH and a group of CI users participated in the study.

Experiment 1: Gated Word Recognition

Rationale

The first experiment was designed to characterize the initial auditory information required for word recognition in individuals with typical hearing and CI users. The listeners with NH were presented either vocoder (NH-V) or unprocessed (full-spectrum) stimuli, and the listeners with CIs were presented only unprocessed stimuli (through their implants).

Method

The University of Memphis Institutional Review Board approved the study protocols. All participants signed an informed consent and were compensated for their time. Data collection was carried out at the University of Memphis Speech Perception Assessment Laboratory.

Participants

The CI group consisted of 12 adults (seven women, five men; 29–59 years, M = 42.6). All CI users had postlingual onset of bilateral severe-to-profound sensorineural hearing loss (unaided hearing thresholds ranged from 71 dB HL to "no response"). Twenty-four adults with NH (17 women, seven men) whose ages ranged from 21 to 52 years (M = 38.4) also participated in the study. Half of the participants with NH listened to full-spectrum stimuli (12 subjects; eight women and four men), and the other half (12 subjects; nine women and three men) listened to vocoded stimuli. All participants were native speakers of American English with no second language or sign language experience. In addition, the age of the participants in both groups was limited to below 60 years to minimize the influence of age-related cognitive decline. All subjects of appropriate age who fulfilled the hearing criteria, who could complete the tests needed to collect the data, and who spoke English as their first language were included. All CI participants were more than 2 years postimplantation experience to maximize the chances of acclimatization to their device parameters and prescribed settings. All the participants in the experimental group used their CIs set to their prescribed settings. Table 1 provides detailed information regarding the demographics of the CI group. Prior to research participation, monosyllabic word recognition ability was evaluated, and all CI users' scores for the consonant-nucleus-consonant (CNC) word lists (Peterson & Lehiste, 1962) indicated 80% or better performance in quiet.

The individuals with NH with hearing sensitivity within normal limits reflected by air conduction thresholds better

Table 1. Demographic details of the experimental (CI) group.

Subject no.	Gender	Age	Age of onset of hearing loss (years)	Manufacturer/ internal device (right)	Manufacturer/ internal device (left)	Etiology	Years of CI experience (years)
1	Male	29	7	AB/HiRes90K	AB/HiRes90K	Idiopathic	12.0
2	Female	37	9	Co/CI24RE	Co/Nucleus 5	Progressive SNHL	10.0
3	Female	41	11	Co/Nucleus 5	Co/Nucleus 5	Hereditary + noise exposure	5.0
4	Female	32	10	Co/CI24RE	Co/CI24RE	Progressive SNHL	8.5
5	Female	49	12	AB/HiRes90K	NA	Otitis media	12.0
6	Male	41	12	AB/HiRes90K	HA (off during testing)	Hereditary	11.3
7	Male	37	14	AB/HiRes90K	HA (off during testing)	Idiopathic	10.0
8	Female	48	13	AB/HiRes90K	AB/HiRes90K	Otosclerosis	10.0
9	Male	43	17	Co/CI24RE	Co/ Nucleus 5	Sudden HL	9.0
10	Male	52	10	Co/Nucleus 5	Co/Nucleus 5	Ototoxicity + hereditary	2.5
11	Female	59	10	Co/Nucleus 5	Co/Nucleus 5	Otosclerosis	7.8
12	Female	43	17	Co/Nucleus 5	NA	Progressive SNHL/hereditary	10.0

Note. CI = cochlear implant; AB = Advanced Bionics; Co = Cochlear; HA = hearing aid; NA = not available; SNHL = sensorineural hearing loss; HL = hearing loss.

than 20 dB HL at the octave frequencies from 250 to 8000 Hz. The participants had no known history of otologic, speech and language, or psychological problems. One of the participants who belonged to the CI group could not complete the experimental tasks due to mental health issues and thus was excluded from the study. In the experimental group, the CI users' age of onset of hearing loss was 5 years or greater.

Stimuli

Generation of Gated Word Stimuli

Monosyllabic word stimuli from the CNC test were used. Frequency of occurrence and neighborhood density of CNC words across lists were determined using an online calculator developed by Balota et al. (2007). Words that had an average neighborhood size (i.e., 13–26 similar sounding lexical word candidates) and medium to high frequency (nine to 20 times) of occurrences were included. A word list comprising 20 words was created based on their lexical properties (see Appendix A). The mean duration of the target words was 0.600 s (SD = 0.105). The gate size in this study was set to 33.33 ms, as used by Moradi, Lidestam, Saremi, and Rönnberg (2014), who conducted a series of pilot studies to determine an optimal gate size. The gating started from the second phoneme of each word to shorten the test administration time. The shortlisted words were digitally extracted from the CNC CD and then processed using MATLAB to generate progressive 33.33 ms fragments of stimuli to determine each participant's IP.

Noise-Band Vocoding

The "acoustic simulations of CI processing" were produced using a noise-band vocoder that is widely used to simulate CIs (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). We specifically chose to apply an eight-channel vocoder processing because this has been shown to yield speech recognition performance similar to that of the best-performing listeners with CIs (Friesen et al., 2001; Fu, Chinchilla, & Galvin, 2004; Fu & Nogaki, 2005), indicating that listeners with CIs experienced a presumably similar amount of spectrotemporal degradations in their day-to-day life. It is worth mentioning that, despite the functional similarities, the vocoder processing does not accurately reflect the speech processing in actual CIs per se.

The vocoder processing was applied using Angel sound software (Fu, 2012). The preprocessing input bandwidth was 200–7000 Hz. Then the stimuli were filtered into eight analysis bands using a set of fourth-order Butterworth bandpass filters. The band cutoff frequencies were distributed according to the Greenwood (1990) matching function. The temporal envelope in each band was then extracted to modulate a corresponding bandpass-filtered noise. Finally, outputs from the eight channels were then combined to create eight-channel noise-band vocoder stimuli. Noise-band vocoded and original stimuli were equalized in terms of their root-mean-square amplitude (−23.01 dB). A 1-kHz calibration tone was played prior to each data collection session to calibrate the volume unit (VU) meter deflection to "0." The amplitude level of the calibration tone was matched with the experimental stimuli, so that the stimuli could be presented at the desired presentation level.

Participants in both groups underwent tympanometry to make sure they had normal middle ear function (Type A tympanogram with acoustic reflexes present), and then pure-tone audiometry was carried out to confirm that thresholds were within the normal range for the NH-V as well as NH groups and reflective of a severe-to-profound sensorineural hearing loss (unaided responses) for the experimental group. Participants received written instructions about the task, and two trial runs were performed prior to actual data collection. During the trial sessions, the participants were instructed that they would hear initial segments of the target, and then graduated aspects of the stimuli would be presented in a progressive manner. The stimuli were routed from a Dell Precision M4700 laptop to a calibrated diagnostic audiometer (Grason-Stadler, GSI-61) and then to a sound-field speaker. All testing was completed in a sound-treated room, and listeners were seated facing the speaker at a zero-degree azimuth at a 1-m distance.

The participants were instructed that they would hear initial parts of a target word, and they were asked to recognize the target words after each presentation, regardless of their confidence about the identification of the stimulus. Guessing was encouraged. Their spoken responses were digitally recorded for further analysis. Presentation of progressive gates was continued until the target was correctly identified. After the correct identification, three more gates were presented to make sure that the word was recognized correctly and not a mere guess. If the target word was not correctly recognized, presentation of the stimuli was continued until the entire target word was presented. The IP was calculated as the point in the word where the listener first identified the target word successfully with no change in their response after three successive presentations. When a word was not identified correctly, even after the presentation of the entire word, the total duration of the word plus one gate size was used to estimate the IP (Hardison, 2005; Metsala, 1997; Moradi et al., 2013, 2014). There was no time pressure for their response. The presentation of trials was randomized across the subjects. The entire test took 30–40 min to complete.

All listener responses were digitally recorded using a digital audio recorder (Marantz Model PMD660) for reliability purposes. To ensure accuracy in scoring, interjudge scoring reliability was conducted on 30% of the data (15% NH, 15% CI). Using the following formula: (agreements / [agreements + disagreements]) × 100%), interjudge scoring reliability was 95%. The CI users were tested with their recommended CI signal processing settings.

Results

Figure 1 shows the mean IPs (in percentage) for recognizing gated words in the CI and NH groups. Because the IP is a time-based measure, it is possible that the results could be influenced by the total word duration (e.g., a longer word could have a longer IP). Therefore, IP (in ms) for each word was converted to IP percentage (percent of total word length where a given word was correctly recognized) using the following formula:

$$\text{IP(in ms)} \times 100/\text{Total length of word (in ms)}. \quad (1)$$

The results indicated that there was a statistically significant difference in IP percentage between groups as determined by a one-way analysis of variance (ANOVA), $F(2, 33) = 51.12$, $p < .001$. Post hoc comparisons using the Tukey's honest significant difference test indicated that the mean percent IPs for the NH group ($M = 59.56$, $SD = 4.42$) was significantly different than that for the NH-V and CI groups ($M = 85.69$, $SD = 10.84$ and $M = 94.67$, $SD = 9.85$, respectively). In addition, the mean percent IPs for the NH-V and CI groups differed significantly ($p < .05$), suggesting that the NH-V group performed better on the word prediction task. However, when the stimuli were spectrotemporally degraded, the listeners with NH performed poorly, indicating that speech quality (spectral resolution) plays an important role in the encoding of spoken words.

In addition to the IPs, we also compared the word recognition accuracy across the listening groups. A response was counted as "inaccurate" when the listeners did not recognize the target words even after hearing the entire word. The results from a one-way ANOVA suggested that there was a significant difference in performance between the groups, $F(2, 33) = 55.73$, $p < .001$. Post hoc comparisons for word recognition accuracy revealed that the NH group was more accurate ($M = 100.00$, $SD = 0.00$, $p < .05$) than the NH-V group ($M = 83.91$, $SD = 7.15$, $p < .05$) and the CI group ($M = 73.41$, $SD = 8.04$, $p < .05$). Figure 2 shows the mean accuracy score (in percentage).

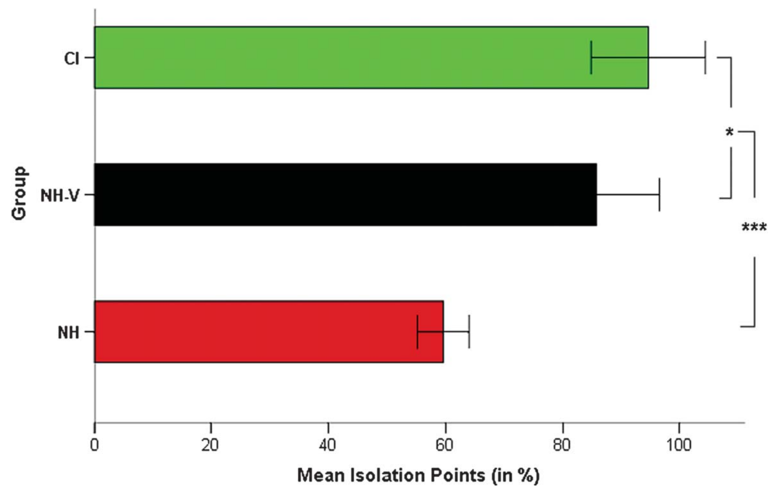
Discussion

In Experiment 1, we hypothesized that spectrotemporal degradations would influence the process of spoken word recognition in general, and listeners would require more speech information to recognize the target words (delayed IPs). Statistical comparison of IPs in listeners with CIs and NH (for vocoder processing as well as full spectrum) showed that the IPs were longer in listeners with CIs, and they required almost 40% more speech information to predict a spoken word. Longer IPs were also observed in listeners with NH when they listened to vocoder processing (see NH-V data in Figure 1). These results in combination suggest that signal fidelity (spectrotemporal resolution) impacts the time course of word recognition, and these results agree with Wagner et al. (2016) that lexical decisions are delayed when spectrotemporal degradation is added to the target stimuli.

In the current investigation, words were extracted from the CNC corpus based on their lexical properties (neighborhood density and frequency of occurrence). The initial phonemes in these words are consonants (e.g., ball), and there is enough evidence to suggest that listeners with CIs displayed many errors in identifying consonants (Dorman, Dankowski, McCandless, Parkin, & Smith, 1991; Dorman & Loizou, 1998; Dorman, Loizou, & Rainey, 1997). The problem persists with listeners with NH when they listened to spectrally degraded speech (Friesen et al., 2001; Shannon et al., 1995). The uncertainty about the initial consonants caused by spectral degradation could complicate the competition process among the lexical items within the mental lexicon and ultimately delay the processing of spoken words.

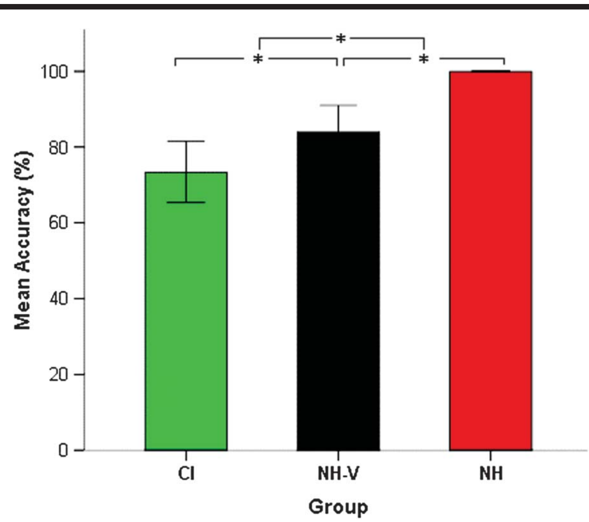
We also observed a minimal but significant difference in performance between NH-V and CI groups in

Figure 1. Results of Experiment 1. Gated word recognition. Bar plots show mean isolation point (IP) percentages in normal hearing (NH), normal hearing with vocoder (NH-V), and cochlear implant (CI) groups. Error bars reflect ±1 *SD* around the means. Listeners with NH performed better (IPs were shorter) for both full-spectrum stimuli and vocoded stimuli than the listeners with CI ($^{*}p < .05$ and $^{***}p < .001$).



IPs. Average IP points appeared to be delayed in CI users because some of them could not recognize a few of the target words accurately despite the presentation of the entire word. Therefore, the total duration of the word plus one gate size was used to estimate the IP as used in most earlier reports (Hardison, 2005; Metsala, 1997; Moradi et al., 2013, 2014). We did observe some whole word misrecognition despite the presentation of the entire target word. In addition, prolonged IPs could also be attributed to additional factors that influence speech recognition in actual CI users that vocoder simulations do not account for (e.g., auditory deprivation, surgical complications, residual neural survival, efficiency of front-end processing devices, and channel interactions).

Figure 2. Mean accuracy data in Experiment 1. Bar plots show percent correct words for normal hearing (NH), normal hearing with vocoder (NH-V), and cochlear implant (CI) groups. Error bars reflect ±1 *SD* around the means. Listeners with NH were more accurate than the other two groups ($^{*}p < .05$).



Experiment 2: Role of Semantic Context on Gated Word Recognition

The second experiment looked at the benefits of having access to semantic context on the processing of spoken words as depicted by IPs. Improvement with semantic context with speech recognition, especially when speech is degraded, has been well established (Bilger, 1984; Pichora-Fuller et al., 1995; Pichora-Fuller & Souza, 2003; Sheldon et al., 2008). As in Experiment 1, the listeners with NH were presented unprocessed as well as vocoded stimuli, and the listeners with CIs were presented unprocessed stimuli.

Method

The same subjects with NH and CIs who participated in Experiment 1 also participated in Experiment 2 (see above for subject demographics and characteristics).

Stimuli

Generation of Gated Word Stimuli

In Experiment 2, the final word in the LP and HP sentences from the Revised Speech Perception in Noise (R-SPIN) test (Bilger, 1984) were used. The HP sentences provide semantic cues for recognition of the final word in the sentence, for example, "*Paul took a bath in the TUB*." Alternatively, the LP sentences provide no contextual cues, making it difficult to understand, for example, "*Miss*

Smith knows about the TUB." As shown in these examples, the target word is the same in both HP and LP sentences, but the content leading to the target word is different, making it easy or difficult to predict. Frequency of occurrence and neighborhood density of final words in HP and LP sentences across lists were determined using the online calculator developed by Balota et al. (2007). The target words had a neighborhood density of 12–30 (similar sounding words) and frequency of occurrences of 9–30. A list containing 30 sentences that had 15 sentences each of LP and HP contexts was finalized (see Appendix B).

The gating parameters were identical to that of Experiment 1 (33 ms and gating started after the first phoneme). The shortlisted sentences were digitally extracted from the R-SPIN CD, and then the final words from the shortlisted R-SPIN sentences were extracted using Adobe Audition Version 3.0. Then the words were processed using MATLAB to generate progressive 33.33-ms fragments of stimuli and then added back to the partial sentences using Adobe Audition to generate the target sentences. While adding the gated words to their corresponding sentences, we made sure that the original duration between the previous word and the target word was maintained. All the experimental stimuli were edited and scaled using Adobe Audition to ensure their root-mean-square amplitude levels were at the same level. Each experimental session began with a 1-kHz calibration tone to calibrate the VU meter deflection to "0." The amplitude level of the calibration tone was matched with the unprocessed R-SPIN sentences.

Noise-Band Vocoding

The R-SPIN sentences were then noise-band vocoded (eight channels) using the Angel sound software. Processing parameters were similar to those used in Experiment 1 (input bandwidth: 200–7000 Hz, eight analysis bands, and Greenwood matching function).

Procedure

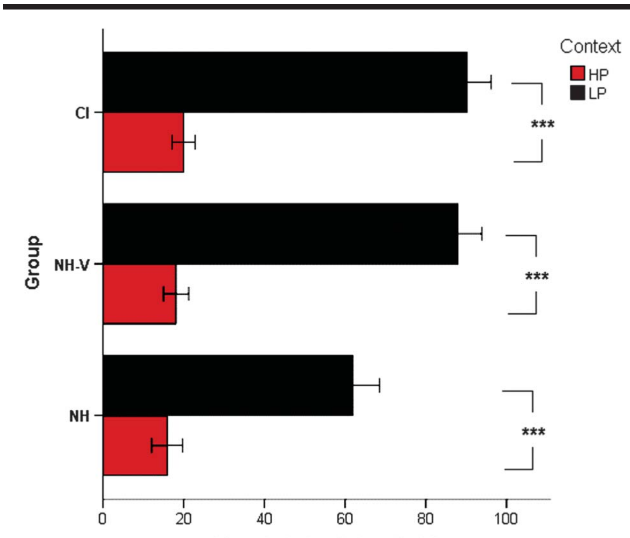
Tympanometry and pure-tone audiometry were carried out to determine hearing thresholds and verify normal middle ear function. Participants were given written instructions about the task, and five sentences that were processed identically as the target stimuli were presented as practice prior to actual data collection to make sure the participants understood the instructions and were responding as desired. The participants were instructed they would hear sentences whose initial segments of the final word and graduated aspects of the stimuli would be presented in a progressive manner. They were instructed to predict the final word in the sentence and respond verbally. The experimental environment, instrumentation, presentation format, and response tasks were identical to that of Experiment 1 (see above for details). Presentation of the stimuli was randomized across the subjects, and each session took 45–60 min to complete. Interjudge scoring reliability was conducted from the recorded responses on 60% of the data (30% NH, 30% CI) to ensure scoring accuracy, and the interjudge scoring reliability was 95%.

Results

Figure 3 shows the mean IPs (in percentage) for HP and LP sentences in the CI and NH groups. A 3 (hearing group: NH, NH-V, and CI) × 2 (context: HP and LP) mixed model ANOVA with repeated measures in the second factor was conducted to examine the effects of hearing status on the IPs for sentences with different contextual conditions. The analyses showed there were significant main effects of hearing status, $F(2, 33) = 54.45$, $p < .001$, as well as context, $F(1, 33) = 4499.94$, $p < .001$. In addition, there was a significant interaction effect, $F(2, 33) = 77.53$, $p < .001$. A post hoc pairwise multiple comparison Tukey's test revealed that context effect (the difference between IPs for HP and LP items) was higher for the NH-V and CI groups compared to the NH group ($p < .001$). However, no significant difference in context effect was observed between the NH-V and CI groups ($p = .574$).

To understand the simple effects of listening groups on each level of semantic context, we evaluated the effects of HP and LP conditions separately on the IPs. A one-way ANOVA was conducted to compare the IPs for HP sentences in NH, NH-V, and CI groups. The analyses showed there was no significant difference in performance for HP sentences, $F(2, 33) = 3.19$, $p = .054$. Post hoc comparisons using the Bonferroni test indicated that mean

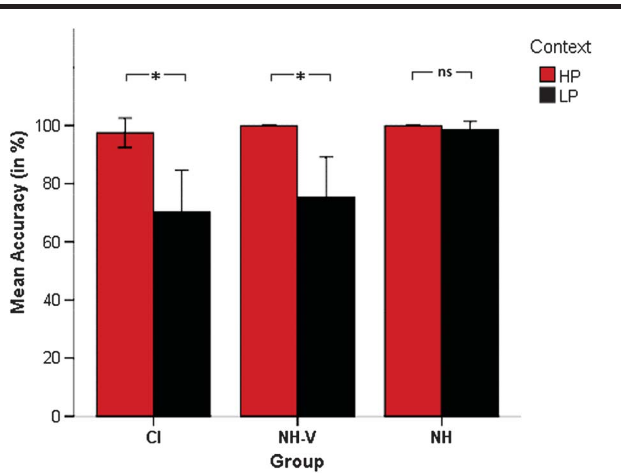
Figure 3. Results of Experiment 2. Gated word recognition with and without semantic context across the groups. Bar plots show mean isolation point percentages for high-predictability (HP) as well as low-predictability (LP) sentences in normal hearing (NH), normal hearing with vocoder (NH-V), and cochlear implant (CI) groups. Error bars reflect +1 *SD* around the means. There was a main effect of semantic context in all the listening groups (***$p < .001$).



IPs did not differ significantly ($p < .05$) across the listening groups. However, the one-way ANOVA analyses for LP sentences across the listening groups revealed that there was a significant difference in performance, $F(2, 33) = 78.17$, $p < .001$. Post hoc comparisons using the Bonferroni test indicated that mean IPs in the NH group were significantly lower ($M = 61.83$, $SD = 6.76$) than the NH-V ($M = 87.91$, $SD = 5.90$) and CI ($M = 90.25$, $SD = 5.83$) groups ($p < .001$). However, no significant difference in IPs was observed between the NH-V and CI groups ($p > .05$) for the LP sentences. Taken together, these results suggested that contextual cues do have an effect on the processing of spoken words. Specifically, our results suggest that, in the absence of semantic context (LP sentences), NH-V and CI groups performed poorly with processing target words. However, it should be noted that, for HP sentences, all of the listening groups performed equally well.

Figure 4 shows the mean accuracy (in percentage) for HP and LP sentences. A 3 (hearing group: NH, NH-V, and CI) × 2 (context: HP and LP) mixed model ANOVA with repeated measures in the second factor was conducted to examine the effects of hearing status on word accuracy for sentences with different contextual conditions. The analyses showed there were significant main effects of hearing status, $F(2, 33) = 22.45$, $p < .001$, as well as context, $F(1, 33) = 75.32$, $p < .001$. In addition, there was a significant interaction effect, $F(2, 33) = 15.92$, $p < .001$. A post hoc pairwise multiple-comparison Tukey's test revealed that predictability contrast (IPs for HP–LP) was higher for the NH-V and CI groups compared to the NH group ($p < .001$). However, no significant difference in context effect in IPs was observed between the NH-V and CI groups ($p = .574$). Thus, the results in combination indicated that when the stimuli were spectrotemporally degraded (CI and NH-V), listeners not only misrecognized the spoken words but also had substantial delay in word recognition as evidenced by IPs. However, in the presence of supportive contextual cues, word recognition became more accurate and faster.

Figure 4. Mean accuracy data in Experiment 2. Bar plots show percent correct words for normal hearing (NH), normal hearing with vocoder (NH-V), and cochlear implant (CI) groups. Error bars reflect ±1 *SD* around the means. All of the groups were more accurate for high-predictability (HP) sentences than for low-predictability (LP) sentences. Listeners with NH were more accurate than the other two groups (*$p < .05$).



Discussion

In Experiment 2, it was hypothesized that spectrotemporal degradations would influence the benefit listeners receive from semantic context, and listeners would require comparatively less speech information to predict the target words successfully when context was present. The results indicated that IPs for HP sentences were significantly shorter than for LP sentences showing a strong contextual advantage, and as mentioned earlier, the listeners with CIs (and listeners with NH listening to vocoder speech) benefitted more from semantic context than the listeners with NH. Such results indicated that the detrimental effects imposed by spectrotemporal degradations do not persist when sentential context is available, and listeners compensate rapidly to achieve similar performance as listeners with NH to encode spoken words (Kaiser, Kirk, Lachs, & Pisoni, 2003; Loizou, 1998; Wang & Humes, 2010). It is likely that listeners with CIs exploit contextual cues routinely to understand speech due to consistent exposure to poor signal quality. Presumably, sentential context available in HP sentences triggers only one or a very restrictive set of target word candidates that match the meaning of the sentence, making it easier to predict target words. Alternatively, for the LP sentences, listeners needed a higher amount of initial speech information to predict the target words, because listeners were unable to form an "informed guess" about the target words, and thus, IPs were longer. Our results also suggested that the NH-V and CI groups benefited equally from contextual cues, suggesting that adults who are postlingually deaf with CIs adapted to degraded signals rapidly and displayed comparable performance to listeners with NH who heard vocoded sentences.

General Discussion

Speech understanding under spectrotemporal degradation has gained a lot of attention in the CI population because many auditory skills are affected by poor signal quality, and listeners with CIs struggle to engage their higher-level compensatory mechanisms to resolve ambiguities in speech (Başkent, 2012; Başkent & Chatterjee, 2010; Başkent et al., 2009; Baskent & Shannon, 2007; Başkent & Shannon, 2006; Benard & Başkent, 2013; Bhargava & Başkent, 2012; Bhargava et al., 2014; Chatterjee et al., 2010; Patro & Mendel, 2016). In Experiment 1, we hypothesized that there would be differences in the amount of speech information required by listeners with NH and CIs due to speech degradation, but the extent to which this process was affected was still unknown. We were also interested in understanding whether the resulting prolongation in IPs was due to poor signal quality resulting from speech degradation alone or if there were any other fundamental differences in speech processing due to long-term experience with CIs. By comparing the performance of listeners with NH to "simulated CI" stimuli with that of actual CI stimulation, we found that signal degradation does play an important role in speech perception (as depicted in the prolongation in IPs for NH-V), but additional factors (see discussion for Experiment 1 for the details) experienced by the listeners with CIs also contributed to lexical processing.

Collison, Munson, and Carney (2004) evaluated the correlation between cognitive–linguistic abilities and gated word recognition in adults with CIs. Interestingly, their results indicated that cognitive abilities do not predict listeners' ability to use initial speech information to recognize spoken words. Collison et al. (2004) claimed that their results may have been due to heterogeneity among their CI group and possible contributing factors, such as age of implantation, duration of device use, and specific causes of hearing loss. Adaptation to implanted devices indeed amplifies one's overall performance in a variety of listening skills (Dorman & Ketten, 2003; Fu, Shannon, & Galvin, 2002; Perkell, Lane, Svirsky, & Webster, 1992; Svirsky et al., 2001). However, the gated paradigm apparently does not seem to completely capture such advantages due to adaptation. Additional objective measures that are sensitive to the time course of lexical processing may need to be investigated further.

It is worth noting that listeners' performance for LP sentences was almost the same as that of gated word recognition performance in Experiment 1. Such results imply that when the contextual cues were misleading (LP), the listeners needed to rely on the available speech information to recognize the target words. In Experiment 2, we observed that clear benefits from semantic context in both the CI and NH-V groups and the computed context effect were higher than that of listeners with NH for full-spectrum speech, because their performance for the LP sentences was poor. On the contrary, these listeners with NH relied less on contextual cues to help them understand speech that was not degraded. Molis et al. (2015) reported that signal degradation (in the presence of background noise) can influence successful prediction of final words in R-SPIN sentences and more so for LP sentences. Lewis et al. (2017) also evaluated the role of semantic context on gated word recognition in children with NH and hearing loss. Their results also indicated that children needed less initial acoustic–phonetic information to achieve accurate word recognition for HP sentences compared to LP sentences regardless of their hearing status. They also reported that children with NH were more accurate in identification of gated words than children with hearing impairment. Our results were consistent with the findings of Molis et al. (2015) and Lewis et al. (2017), showing that CI processing and noise-band vocoding influenced the ability of the listeners to predict the target words when the listeners did not have access to semantic cues. Future studies could more accurately determine the influence of guessing by more directly assessing cloze probability (guessing the last word of the sentence without any stimulation).

When the degraded speech is effortful, listeners deploy additional cognitive resources to maintain high-level speech recognition performance (Mattys, Brooks, & Cooke, 2009; Rönnberg et al., 2013; Van Engen & Peelle, 2014). However, it is assumed that when the embedded words were presented in context, the results suggested it became easier for the listeners to retrieve "most likely" word candidates from memory. We also attribute such findings to the facilitating effect offered by semantic context by CI users as well as listeners with NH for "CI-like" speech. Using similar stimuli (HP and LP sentences with final word gating), Moradi et al. (2014) also demonstrated that semantic context assists in disambiguating target speech, which reduces cognitive load and ultimately expedites word retrieval. Our findings are consistent with Moradi et al. (2014) and can be generalized to their observations for spectrotemporally degraded speech.

A clinical speech recognition test battery does not always capture such time course of word recognition and benefits of context. Typically, clinical testing is focused more on the final product (number of words/key words correct). Few clinical tools tend to explore underlying cognitive mechanisms and the complex temporal dynamics in speech recognition. Thus, even when speech recognition accuracy is high, there are differences in listening effort, speed of processing, and type of processes involved (bottom-up vs. top-down). By understanding different cognitive dimensions involved in speech recognition, device parameters can be optimized to accommodate different listening needs and ultimately improve speech recognition in adverse listening conditions. Furthermore, a wealth of research has been devoted to understanding the adverse effects degraded speech has on speech understanding (Dorman et al., 1997, 1998; Loizou, Dorman, & Tu, 1999; Shannon et al., 1995). Investigating word recognition accuracy indeed provides important insights regarding the effectiveness of different device parameters of interest, but they do not tell us much about how listeners arrive at the final recognition of a word. Likely, this is because listeners resolve several information-processing challenges while perceiving a speaker's intended message. The bottom-up signal contains important acoustic information, and such information from the speech signal is passed onto higher-level top-down processing. An "acoustically not so rich" bottom-up signal might influence the overall process of spoken word recognition in many ways by misrecognition of initial phonemes or activating less probable word candidates delaying the involvement of top-down processing. However, access to semantic context could resolve such lexical ambiguity, facilitate activation of probable candidates, and activate semantic-appropriate lexical items (Gaskell, 2001), thus reducing one's overall listening effort (Winn, 2016) and restricting the set of possible candidates (Brown-Schmidt, Campana, & Tanenhaus, 2005).

Limitations

Some limitations in this study are worth noting. The results of this study were based on the findings from 12 CI users who participated in both experiments. They were high-performing CI users who were postlingually deaf (with good word recognition scores in quiet). Thus, observations from this study may not be applicable to a more general CI population, especially those with lower speech recognition scores. Second, administering a gated task requires continuous priming because listeners are presented successive signal frames of target stimuli until they recognize them successfully with two gates or more. Such a test procedure is not really representative of daily speech communication and is also influenced by some individual communication strategies (Craig & Kim, 1990). It should also be noted that the derived IPs obtained from this gating task do not say much about the nature of lexical competition, the exact time required for lexical access, or shortlisting final candidates. However, it does provide important insights regarding initial speech information required for word recognition and sensory–cognitive interaction. Furthermore, not all of the participants were bilateral users of CIs, whereas all of the listeners with NH received binaural input (when the test stimuli were presented through a loudspeaker). It is possible that the inferior performance displayed by the CI group could also have been due to the lack of binaural stimulation for the unilateral users because bilateral CI users typically display substantial benefits in speech understanding (Ching, Van Wanrooy, & Dillon, 2007; Firszt, Reeder, & Skinner, 2008; Litovsky, Parkinson, Arcaroli, & Sammeth, 2006). Furthermore, recognition of words preceding the target (nontarget) were not evaluated in Experiment 2. Scoring the whole sentence could have helped in understanding whether the target words were too difficult to recognize or words leading to target words were misrecognized.

Conclusion

In summary, in the first experiment, we showed that CI users require a greater amount of initial speech information compared to that of listeners with NH to recognize spoken words. Individuals with NH who listened to vocoded speech also performed poorly on the gated word recognition task, suggesting that the inability in accessing mental representations could be primarily due to poor signal fidelity. Additional factors that are intrinsic to CI systems could also have contributed to the inability to easily process spoken words. Results from the second experiment suggested that CI users could overcome these challenges in processing spoken words and could perform like listeners with NH if sufficient contextual cues were available at least in quiet listening situations.

Acknowledgments

This research was supported in part by the U.S. National Science Foundation under CISE Grant 1231620, awarded to Lisa Lucks Mendel. The authors thank Jerker Rönnberg and Shahram Moradi for their help during stimulus preparation and the subjects for their participation in the study.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. Retrieved from http://elexicon.wustl.edu/default.asp

Başkent, D. (2012). Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation. *Journal of the Association for Research in Otolaryngology, 13*(5), 683–692.

Başkent, D., & Chatterjee, M. (2010). Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hearing Research, 270*(1), 127–133.

Başkent, D., Eiler, C., & Edwards, B. (2009). Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. *The Journal of the Acoustical Society of America, 125,* 3995–4005.

Başkent, D., & Shannon, R. V. (2006). Frequency transposition around dead regions simulated with a noiseband vocoder. *The Journal of the Acoustical Society of America, 119*(2), 1156–1163.

Başkent, D., & Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition. *Ear and Hearing, 28*(3), 277–289.

Benard, M. R., & Başkent, D. (2013). Perceptual learning of interrupted speech. *PLoS One, 8*(3), e58149.

Bhargava, P., & Başkent, D. (2012). Effects of low-pass filtering on intelligibility of periodically interrupted speech. *The Journal of the Acoustical Society of America, 131*(2), EL87–EL92.

Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top–down restoration of speech in cochlear-implant users. *Hearing Research, 309,* 113–123.

Bierer, J. A., Faulkner, K. F., & Tremblay, K. L. (2011). Identifying cochlear implant channels with poor electrode-neuron interface: Electrically-evoked auditory brainstem responses measured with the partial tripolar configuration. *Ear and Hearing, 32*(4), 436–444.

Bilger, R. (1984). Speech recognition test development. *Asha Reports, 14,* 2–15.

Blamey, P., Artieres, F., Başkent, 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*(1), 36–47.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America, 84*(1), 101–114.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution by naïve participants during a task-based unscripted conversation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions* (pp. 153–171). Cambridge, MA: MIT Press.

Chatterjee, M., Peredo, F., Nelson, D., & Başkent, D. (2010). Recognition of interrupted sentences under conditions of spectral degradation. *The Journal of the Acoustical Society of America, 127*(2), EL37–EL41.

Chatterjee, M., & Shannon, R. V. (1998). Forward masked excitation patterns in multielectrode electrical stimulation. *The Journal of the Acoustical Society of America, 103*(5), 2565–2572.

Ching, T., Van Wanrooy, E., & Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification, 11*(3), 161–192.

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*(3), 496–508.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics, 35*(1), 41–48.

Craig, C. H., & Kim, B. W. (1990). Effects of time gating and word length on isolated word-recognition performance. *Journal of Speech and Hearing Research, 33*(4), 808–815.

Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., & Fausti, S. A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing, 22*(1), 1–13.

Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., & Smith, L. (1991). Vowel and consonant recognition with the aid of a multichannel cochlear implant. *The Quarterly Journal of Experimental Psychology, 43*(3), 585–601.

Dorman, M. F., & Ketten, D. (2003). Adaptation by a cochlear-implant patient to upward shifts in the frequency representation of speech. *Ear and Hearing, 24*(5), 457–460.

Dorman, M. F., & Loizou, P. C. (1998). The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. *Ear and Hearing, 19*(2), 162–166.

Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *The Journal of the Acoustical Society of America, 104*(6), 3583–3585.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America, 102*(4), 2403–2411.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America, 107*(1), 538–546.

Elliott, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics, 42*(2), 150–157.

Eskridge, E. N., Galvin, J. J., Aronoff, J. M., Li, T., & Fu, Q.-J. (2012). Speech perception with music maskers by cochlear implant users and normal-hearing listeners. *Journal of Speech, Language, and Hearing Research, 55*(3), 800–810.

Firszt, J. B., Reeder, R. M., & Skinner, M. W. (2008). Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid. *Journal of Rehabilitation Research and Development, 45*(5), 749–767.

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, 110*(2), 1150–1163.

Fu, Q.-J. (2012). Angel Sound (version 5.08.03): Interactive listening rehabilitation and functional hearing test program [Computer software]. Retrieved from http://angelsound.emilyfufoundation.org

Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology, 5*(3), 253–260.

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*(1), 19–27.

Fu, Q.-J., Shannon, R. V., & Galvin, J. J., III. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *The Journal of the Acoustical Society of America, 112*(4), 1664–1674.

Fu, Q.-J., Shannon, R. V., & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of America, 104*(6), 3586–3596.

Gaskell, M. G. (2001). Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes, 16*(5–6), 723–729.

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*(3), 1197–1208.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592–2605.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283.

Hardison, D. M. (2005). Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication, 46*(1), 73–93.

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*(3), 342–360.

Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 46*(2), 390–404.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337–1351.

Khan, A. M., Handzel, O., Burgess, B. J., Damian, D., Eddington, D. K., & Nadol, J. B. (2005). Is word recognition correlated with the number of surviving spiral ganglion cells and electrode insertion depth in human subjects with cochlear implants? *The Laryngoscope, 115*(4), 672–677.

Lewis, D., Kopun, J., McCreery, R., Brennan, M., Nishi, K., Cordrey, E., . . . Moeller, M. P. (2017). Effect of context and hearing loss on time-gated word recognition in children. *Ear and Hearing, 38*(3), e180–e192.

Litovsky, R., Parkinson, A., Arcaroli, J., & Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study. *Ear and Hearing, 27*(6), 714–731.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). San Diego, CA: Academic Press.

Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine, 15*(5), 101–130.

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *The Journal of the Acoustical Society of America, 106*(4), 2097–2103.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences, 103*(49), 18866–18869.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1–2), 71–102.
Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29–63.
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.
Metsala, J. L. (1997). Spoken word recognition in reading disabled children. *Journal of Educational Psychology, 89*(1), 159–169.
Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329–335.
Molis, M. R., Kampel, S. D., McMillan, G. P., Gallun, F. J., Dann, S. M., & Konrad-Martin, D. (2015). Effects of hearing and aging on sentence-level time-gated word recognition. *Journal of Speech, Language, and Hearing Research, 58*(2), 481–496.
Moradi, S., Lidestam, B., & Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Frontiers in Psychology, 4*, 359.
Moradi, S., Lidestam, B., Saremi, A., & Rönnberg, J. (2014). Gated auditory speech perception: Effects of listening conditions and cognitive capacity. *Frontiers in Psychology, 5,* 531.
Nelson, P. B., & Jin, S.-H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America, 115*(5), 2286–2294.
Nelson, P. B., Jin, S.-H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America, 113*(2), 961–968.
Norris, P. (2000). *A virtuous circle: Political communications in post-industrial societies.* Cambridge, United Kingdom: Cambridge University Press.
Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage, 55*(2), 713–723.
Oh, S. H., Donaldson, G. S., & Kong, Y.-Y. (2016). Top-down processes in simulated electric-acoustic hearing: The effect of linguistic context on bimodal benefit for temporally interrupted speech. *Ear and Hearing, 37*(5), 582–592.
Patro, C., & Mendel, L. L. (2016). Role of contextual cues on the perception of spectrally reduced interrupted speecha). *The Journal of the Acoustical Society of America, 140*(2), 1336–1345.
Perkell, J., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *The Journal of the Acoustical Society of America, 91*(5), 2961–2978.
Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders, 27*(1), 62–70.
Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593–608.
Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*(Suppl. 2), 11–16.
Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America, 114*(1), 446–454.
Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*, 31.
Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*(Suppl. 2), S99–S105.
Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 336*(1278), 367–373.
Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*(2), 210–231.
Shafiro, V., Sheft, S., Gygi, B., & Ho, K. T. N. (2012). The influence of environmental sound training on the perception of spectrally degraded speech and environmental sounds. *Trends in Amplification, 16*(2), 83–101.
Shafiro, V., Sheft, S., Kuvadia, S., & Gygi, B. (2015). Environmental sound training in cochlear implant users. *Journal of Speech, Language, and Hearing Research, 58*(2), 509–519.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304.
Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America, 123*(1), 489–499.
Sivonen, P., Maess, B., & Friederici, A. D. (2006). Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience Letters, 408*(3), 220–225.
Slowiaczek, L. M., Nusbaum, H. C., & Pisono, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 64–75.
Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America, 116*(2), 1081–1091.
Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, T. T., & Simmons, P. M. (2001). Auditory learning and adaptation after cochlear implantation: A preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Oto-laryngologica, 121*(2), 262–265.
Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34*(5), 409–420.
Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics, 38*(3), 217–222.
Van Deun, L., Van Wieringen, A., & Wouters, J. (2010). Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear and Hearing, 31*(5), 702–713.
Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience, 8,* 577.
Wagner, A. E., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology, 7,* 398.
Walley, A. C., Michela, V. L., & Wood, D. R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception & Psychophysics, 57*(3), 343–351.

Wang, X., & Humes, L. E. (2010). Factors influencing recognition of interrupted speech. *The Journal of the Acoustical Society of America, 128*(4), 2100–2111.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41*(3), 262–275.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research, 242*(1–2), 3–21.

Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, and Hearing Research, 43*(4), 915–925.

Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing, 20,* 2331216516669723.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). *The Journal of the Acoustical Society of America, 116*(3), 1351–1354.

Appendix A

Stimuli Used in Experiment 1

Number	Word	Frequency of occurrence	Density	Duration (s)
1	Beam	21	16	0.630
2	Boil	12	15	0.692
3	Cape	20	21	0.483
4	Chore	9	20	0.575
5	Ditch	10	15	0.519
6	Dose	12	13	0.716
7	Jar	13	16	0.627
8	Lease	10	24	0.774
9	Loop	17	16	0.319
10	Merge	10	14	0.619
11	Mess	17	17	0.628
12	Pad	8	22	0.647
13	Patch	13	21	0.629
14	Rail	16	26	0.720
15	Ripe	13	16	0.683
16	Thin	17	15	0.464
17	Thumb	12	15	0.632
18	Tooth	17	14	0.546
19	Weep	14	17	0.590
20	Wreck	9	17	0.510

Appendix B

Stimuli Used in Experiment 2

Number	Word	Frequency of occurrence	Density	Duration (s)	Contextual status
1	Coin	10	14	0.721	HP
2	Rag	12	24	0.656	LP
3	Gin	23	20	0.466	HP
4	Hut	13	24	0.568	LP
5	Calf	12	19	0.690	HP
6	Map	13	20	0.554	LP
7	Mate	21	28	0.492	HP
8	Lap	19	26	0.634	LP
9	Deck	23	20	0.568	HP
10	Coach	22	14	0.723	LP
11	Rope	15	23	0.523	HP
12	Bet	20	27	0.394	LP
13	Kick	18	26	0.483	HP
14	Mat	10	28	0.394	LP
15	Fan	18	21	0.339	HP
16	Pill	18	29	0.462	LP
17	Rug	15	30	0.511	HP
18	Shed	12	17	0.623	LP
19	Dive	23	16	0.449	HP
20	Hen	22	24	0.559	LP
21	Cave	9	17	0.621	HP
22	Shell	22	20	0.631	LP
23	Sheep	23	20	0.599	HP
24	Tub	14	18	0.648	LP
25	Gum	15	19	0.667	HP
26	Mouse	10	14	0.723	LP
27	Lid	18	23	0.567	HP
28	Chip	17	22	0.647	LP
29	Juice	12	14	0.769	HP
30	Mice	12	22	0.634	LP

Note. HP = high predictability; LP = low predictability.