The Use of Prosodic Cues in Sentence Processing by Prelingually Deaf Users of Cochlear Implants

Colleen M. Holt,[1,2] Katherine Demuth,[2] and Ivan Yuen[2]

Objectives: The purpose of this study is to assess the use of *prosodic* and *contextual cues to focus* by prelingually deaf adolescent users of cochlear implants (CIs) when identifying target phonemes. We predict that CI users will have slower reaction times to target phonemes compared with a group of normally-hearing (NH) peers. We also predict that reaction times will be faster when both *prosodic* and *contextual* (semantic) cues are provided.

Design: Eight prelingually deaf adolescent users of CIs and 8 adolescents with NH completed 2 phoneme-monitoring experiments. Participants were aged between 13 and 18 years. The mean age at implantation for the CI group was 1.8 years (SD: 1.0). In the *prosodic* condition, reaction times to a target phoneme in a linguistically focused (i.e., stressed) word were compared between the two groups. The *semantic* condition compared reaction time with target phonemes when *contextual cues to focus were provided in addition to prosodic cues.*

Results: Reaction times of the CI group were slower than those of the NH group in both the *prosodic* and *semantic* conditions. A linear mixed model was used to compare reaction times using Group as a fixed factor and Phoneme and Subject as random factors. When only prosodic cues (*prosodic* condition) to focus location were provided, the mean reaction time of the CI group was 512 msec compared with 317 msec for the NH group, and this difference was significant ($p < 0.001$). The provision of contextual cues speeded reaction times for both groups (*semantic* condition), indicating that top–down processing aided both groups in their search for a focused item. However, even with both prosodic and contextual cues, the CI users' processing times remained slower, compared with the NH group, with mean reaction times of 385 msec for the CI users but 232 msec for the NH listeners ($p < 0.001$).

Conclusions: Prelingually deaf CI users' processing of *prosodic* cues is less efficient than that of their NH peers, as evidenced by slower reaction times to targets in phoneme monitoring. The provision of contextual cues speeded reaction times for both NH and CI groups, although the CI users were slower in responding than the NH group. These findings contribute to our understanding of how CI users employ/integrate prosodic and semantic cues in speech processing.

Key words: Children, Cochlear implants, Prosody, Sentence processing.

(Ear & Hearing 2015;XX;00–00)

INTRODUCTION

Speakers use prosody to convey information to their listeners about the structure of the discourse. Prosodic cues, in the form of acoustic changes in fundamental frequency (F0), duration and intensity, provide much information for the listener about the structure of the utterance (Fox 2000; Botinis et al. 2001; Ladd 2008). For example, prosodic cues signal where words and phrases begin and end (Beckman & Pierrehumbert 1986; Turk & Shattuck-Hufnagel 2007; Dilley & McAuley 2008), whether the utterance was intended as a question or a statement (Cruttenden 1997; Fletcher et al. 2002; Ladd 2008) and are used to highlight certain words for the listener (Venditti & Hirschberg 2003; Hirschberg 2004).

Speakers use local F0 movements, or *pitch accents*, to draw listener attention to certain words (Beckman & Pierrehumbert 1986; Welby 2003). Pitch accents form part of the intonational system of English and pitch accents can be distinguished from the utterance-final F0 movements that provide information about whether the utterance is a question or statement (Pierrehumbert 1980; Beckman & Pierrehumbert 1986; Shattuck-Hufnagel & Turk 1996). Although these pitch accents can occur on any lexically stressed syllable of an utterance, speakers use pitch accents to draw listener attention to information that is new, important or in contrast to another item (Rooth 1992; Birch & Clifton 2002; Dahan et al. 2002; Venditti & Hirschberg 2003; Welby 2003; Baumann & Grice 2006). A word is said to have linguistic *focus* when it is emphasized by the speaker using prosodic cues. Speakers change the placement of pitch accents within an utterance to reflect the location of linguistic focus (Venditti & Hirschberg 2003; Welby 2003). For example, in the following exchange, a pitch accent (represented here by the use of CAPS) highlights the new information placed in focus by speaker B:

Speaker A: *So which car did you buy?*
Speaker B: *We bought the Ford STATIONWAGON.*

Consider how the interpretation of speaker B's utterance changes if the location of focus and placement of the pitch accent is changed:

Speaker A: *So which car did you buy?*
Speaker B: *We bought the FORD stationwagon.*

In the second exchange, the word *ford* was assigned a pitch accent and the otherwise accentable *stationwagon* was not. By deaccenting *stationwagon* and accenting *ford*, the speaker draws listener attention to the brand of car bought, rather than the type of vehicle. The location of focus and its associated prosodic cues which include pitch accent placement work together to draw listener attention to the word. The location of focus may be dictated by the preceding context. In the above example, speaker A's question centers on the noun *car*, so a response that focuses and accents the verb would be considered inappropriate:

Speaker A: *So which car did you buy?*
Speaker B: *We BOUGHT the Ford stationwagon.*

While speakers use prosodic cues such as pitch accent placement to appropriately reflect discourse structure, studies have also shown that listeners use these cues to facilitate sentence comprehension (reviewed in Cutler et al. 1997). Although eye-tracking methodology is now widely used to investigate sentence

[1]Audiology and Speech Pathology, The University of Melbourne, Melbourne, Victoria, Australia; and [2]ARC Centre of Excellence in Cognition and its Disorders, Centre for Language Sciences and Department of Linguistics, Macquarie University, Sydney, New South Wales, Australia.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

<zdoi; 10.1097/AUD.0000000000000253>

processing and comprehension (Watson et al. 2006), phoneme monitoring tasks were traditionally used and provide a relatively simple but robust method for obtaining experimental data from a processing perspective. Phoneme monitoring requires participants to press a button as soon as they hear a target phoneme. The reaction time (RT) measurement returned is thought to reflect the ease or difficulty of processing the utterance (Foss 1969; Foss & Lynch 1969). RT in phoneme monitoring tasks is affected by sentence and lexical characteristics of the stimulus sentences as reviewed by Connine and Titone (1996). For example, RTs are faster when the carrier sentences are syntactically simple (Foss & Lynch 1969), when the target-bearing word is a real word (rather than nonword; Frauenfelder et al. 1990; Pitt & Samuel 1995) and when the target-bearing word is a high-frequency word (Eimas et al. 1990; Eimas & Nygaard 1992). The position of the target-bearing word in the utterance also affects RTs, with target-bearing words located early in an utterance responded to more slowly than those located toward the end of utterances (Mehta & Cutler 1988).

Both prosodic and contextual (semantic) cues have been shown to facilitate sentence processing for normally-hearing (NH) listeners. Cutler and Foss (1977) and Shields et al. (1974) reported that phonemes were detected more rapidly when the target-bearing word was pitch accented. This processing advantage remained, even when the prosodic cues were removed from the target-bearing word which was linguistically focused (Cutler 1976). The prosody of the words preceding the target word allowed listeners to anticipate the location of the focused word, with RTs remaining faster to “predicted-accented” words, even in lieu of prosodic cues being available on the target word itself.

A further study by Cutler and Darwin (1981) found that the predicted-accent effect may not be based solely on the use of any particular prosodic cue. Rather, they found that the predicted-accent effect was displayed even when F0 cues were removed and when changes were made to the duration of the stop closure of the target phoneme, so listeners must make use of other prosodic cues, such as duration, in their search for accent location.

The contribution of contextual cues to focus location were also observed by Cutler and Fodor (1979). They manipulated questions which preceded the stimulus sentence to shift the expected location of focus within experimental sentences. They concluded that contextual semantic cues must play an important role in sentence processing, as target phonemes in words focused by the semantic/syntactic context were responded to faster than targets not occurring in anticipated focused positions.

Building on these studies, Akker and Cutler (2003) assessed whether the predicted-accent effect and the use of contextual cues to focus work together, as suggested by Cutler and Fodor (1979). Their results indicated that this is indeed the case: listener RTs to a target phoneme were fastest when the target word is predicted to be accented by *both* the prosody and semantic context. They therefore concluded that these cues work together in the processing of speech to help listeners efficiently locate the most important items of a discourse. Thus, the use of both these cues helps normal NH listeners to effectively and efficiently comprehend sentences in real time.

Cochlear implant (CI) users have difficulty perceiving prosodic cues, with changes in F0 being particularly challenging as the temporal fine structure information exploited by listeners with normal (acoustic) hearing is largely not available to CI users (Moore 2003). However, despite the importance of prosody in speech comprehension, studies of the perception of prosody by CI users are considerably less frequent than those centered on segmental perception abilities. Given the importance of both prosodic and contextual cues for NH listeners during sentence processing, it is therefore important to examine how these cues are processed by listeners with CIs.

CI users have been shown to encounter perceptual difficulty in tracking F0 changes, including tasks involving pitch ranking (Sucher & McDermott 2007; Looi & Radford 2011) and lexical tone discrimination or identification (Ciocca et al. 2002; Lee et al. 2002; Peng et al. 2004). Other studies assessing the perception of prosody as a cue to discourse structure have centered largely on the realisation of questions versus statements (Peng et al. 2008; Peng et al. 2009; Straatman et al. 2010). In these studies, CI users required larger F0 rises than NH listeners to report that they had heard a question. Some studies have targeted the identification of *sentence stress* (i.e., the accented and focused item of an utterance), by CI users and NH individuals (Klieve & Jeanes 2001; Most & Peled 2007; Meister et al. 2011). In the Meister et al. (2011) study, it was found that adult CI users identified an accented and focused word most accurately when all prosodic cues (F0, duration, intensity) were available in combination, as is the case in conversational speech. Despite all cues being available, identification performance was still significantly worse than that of NH listeners. All of these studies assessed the ability of CI users to identify an accented and focused word without the provision of prior context to facilitate that search.

To the authors’ knowledge, phoneme monitoring tasks have not been conducted with CI users. We are not aware of any studies that have directly assessed CI users’ ability to utilize contextual cues to the location of linguistic focus in an utterance. If CI users are less able to identify the word in focus within an utterance, this would affect their participation in complex discourse situations (such as conversation). While effective communication is the goal of CI use, there are few studies which examine spontaneous conversation involving CI users. In complex discourse situations, prosodic cues to the location of important information must be processed and interpreted rapidly, as the listener needs to continuously update their understanding of the conversation in real time. Toe and Paatsch (2013) studied spontaneous conversation in pediatric CI user/NH peer dyads and reported that the CI users dominated the conversation in terms of topic choice and turn length. Dominating the choice of topic relieves the need to continually integrate new information from their conversational partner, and may be related to difficulties with sentence processing (such as rapidly finding the location of the focused word in an utterance).

In this study, we investigated whether difficulty in perceiving prosodic cues affects how CI users integrate prosodic and contextual cues to discourse structure. We used a phoneme monitoring task to examine whether prelingually deaf CI users utilize the cues present in the preceding prosodic contour to predict the location of an accented word (*prosodic* condition). We predicted that CI users would display slower response times to the target phoneme if they were less efficient than the NH listeners at exploiting/processing the prosodic cues present in the utterances. In the *semantic* condition, we assessed whether RTs would be faster when a focus-inducing question is *added* to provide semantic and syntactic cues to the location of the focused word in the stimulus utterance. We expected that both groups would be faster to respond when contextual cues are available as well.

MATERIALS AND METHODS

Participants

A total of 16 adolescents (8 wearing CIs, 8 NH) aged between 13 and 18 years completed the testing. CI users were recruited from patients of the Royal Victorian Eye and Ear Hospital Cochlear Implant Clinic in Melbourne, Australia. These participants were required to be prelingually deaf, with no diagnosed specific language impairment and no less than borderline cognitive ability (i.e., a score of no less than 71 on tests of cognitive ability such as the Wechsler Intelligence Scale for Children) as recorded in medical records. CI users were required to be aural/oral language users. The recruited CI group comprised 6 unilateral CI users (4 females, 2 males) and 2 bilateral CI users (1 female, 1 male). All except subject 8 was congenitally profoundly deaf in both ears (Table 1). The mean age at first CI implant was 1.8 years (SD: 1.0) and the age at second CI for the 2 bilateral users was 14.2 and 6.5 years. All participants were users of Cochlear Ltd devices and CP810 speech processors. Subject 8 used the SPeak coding strategy, but all other subjects used the ACE strategy. The mean age at time of testing was 16.2 years (SD: 1.5). Mean device experience at time of testing was 14.3 years (SD: 1.1). Further demographic information on the CI users is available in Table 1. The NH adolescents (8 females) had no diagnosed language or hearing impairments. The mean age at time of testing for the NH group was 14.5 years (SD: 1.4).

Recording of Stimulus Materials

Stimulus materials were recorded by a 26-year-old adult male speaker of Australian English. The materials were recorded at 44,100 Hz using a Marantz PMD671 solid-state digital recorder coupled with an AKG C520 cardiod condenser head-mounted microphone. Stimuli were RMS normalized to a level of 70 dB.

Phoneme Discrimination Task

Stimuli and Procedure • A two-alternative forced choice discrimination task was conducted to ensure that participants could discriminate between the target phonemes /b p g k/. The speaker recorded the carrier phrase *I said the word X again* along with eight target words (*bat, cat, pat, boat, goat, coat, pet, get*). Each test word was then embedded within the carrier phrase using Praat software (Boersma & Weenink 2007) to create eight test sentences.

Testing was conducted individually in a sound-treated booth. Participants were seated in front of a display monitor and auditory stimuli were presented at 65 dB SPL via a loudspeaker placed 1.5 m directly in front. Participants were instructed to listen using the device or devices that they wore everyday during school classes. The 6 unilateral CI users wore their CI without any device in the ear contralateral to their CI. The 2 bilateral users wore CIs bilaterally. Praat software (Boersma & Weenink 2007) was used to present the stimuli and collect responses. In each trial, participants heard one of the eight test sentences and then used the mouse to select which target word they had heard from the two options available on the screen. The two response options were always minimal pairs featuring two of the target phonemes, for example *goat/coat, pet/bet.* The order of presentation of trials was randomized by the software with a two second interval between trials and a total of 24 trials to complete the testing.

Phoneme Monitoring Task

Stimuli and Procedure • A phoneme-monitoring task was used in two experimental conditions. Phoneme-monitoring requires participants to push a button when they hear a target phoneme. The stimuli manipulated the *prosodic* (PROS) and *semantic* (SEM) contexts of focus. In the PROS condition, the test sentence was presented auditorily, for example, *The girl sewed some buttons onto her school dress* (target phoneme *b*). In the SEM condition, additional semantic context was included by visual and auditory presentation of a question preceding test sentences, for example *What did the girl sew onto her school dress? The girl sewed some buttons onto her school dress* (target phoneme *b*). As pediatric CI users are often reported to have language delays (Spencer et al. 2003; Nicholas & Geers 2007; Geers & Hayes 2011), we facilitated sentence processing using syntactically simpler sentences than those of Akker and Cutler (2003). Sentences used are provided in Appendix 1 (see Supplemental Digital Content 1; http://links.lww.com/EANDH/A237).

Each experimental condition included 16 test sentences and 16 filler sentences. The target phonemes of these sentences were a balanced mix of word initial / p, b, k, g /. All test sentences contained either 10 or 11 syllables and were balanced in terms of the number of syllables preceding and following the target word. Half the target words were monosyllabic and half were trochaic. The position of the target word was varied in the filler sentences, and the filler sentences were randomly interspersed

TABLE 1. Demographic information for the CI group

Subject	Gender	Aetiology	Onset of Profound Deafness (Years)	PTA (dB)	Device Use	Age at Test (Years)	Age at First Implant (Years)
1	Female	Unknown	0	125	Unilateral CI	15.8	1.3
2	Female	Connexin 26	0	98	Unilateral CI	18.0	4.1
3	Female	Genetic	0	103	Bilateral CIs	16.3	2.1
4	Female	Keratitis-ichthyosis-deafness (KID) syndrome	0	125	Unilateral CI	14.7	1.3
5	Female	Cytomegalovirus	0	100	Unilateral CI	16.9	1.3
6	Male	Connexin 26	0	117	Unilateral CI	17.0	1.8
7	Male	Unknown	0	107	Bilateral CIs	13.6	1.1
8	Male	Meningitis	0.8	118	Unilateral CI	17.8	1.8

PTA refers to the pure tone average measured preimplantation for the better ear, averaged across responses measured at 500, 1000, and 2000 Hz.
CI, cochlear implant; PTA, pure-tone average.

with the test sentences in presentation. By varying the position of the target word in filler sentences, we sought to avoid participants pressing the button in anticipation of when the target would occur, rather than on hearing the target phoneme.

To prepare the sentence stimuli for both PROS and SEM conditions, each sentence was recorded twice: once with the test word *focused* (FOC), and once with the corresponding word *nonfocused* (NFOC). The focused word carries the pitch accent. All the test words were extracted from the NFOC carrier sentences and used to replace their counterparts in the FOC sentences. Following Akker and Cutler (2003), this manipulation was made to assess whether listeners could use the cues available in the prosodic contour *preceding* the target word, but not on the word itself. Manipulations were performed using Praat (Boersma & Weenink 2007), with target words spliced at the zero crossing of the initial consonant burst.

Testing was conducted in a sound-treated booth as per the phoneme discrimination task, but with E-Prime software and a serial button box used to present the experiment and collect responses. The sentences of the PROS condition were presented in four blocks (which also contained stimuli for another experiment) and those of the SEM condition in two blocks. Participants completed three training trials before beginning the PROS condition and another three before beginning the SEM condition. Before presentation of the first training trial, the researcher demonstrated the task procedure for the participant. In the training trials, the participant was told to press the button as soon as they heard the target phoneme. Both accuracy and speed were stressed. During the training trials, the researcher asked participants to repeat the target word at the end of the trial as a means of checking that the participant had heard and responded to the target word. The sequence of sentences within blocks was the same for each participant, but the order of presentation of the blocks was counterbalanced. In the PROS condition, participants saw an orthographic representation of the target phoneme on the monitor before hearing the utterance that contained the test phoneme. The task was to press a button when the target phoneme was heard. In the SEM condition, participants first saw a question appears on their monitor, and after a 3-second delay the question was presented auditorily. They then saw the target phoneme followed by an auditory presentation of the utterance stimulus (as in the PROS condition). They responded by pressing a button when they heard the target phoneme. Participants were allowed a break of around 3 min between blocks. RT was calculated as the interval between the onset of the burst release of the target phoneme and the time of the participant's button press.

Predictions

If listeners use sentence prosody preceding the target phoneme to guide the expected location of focus, response time will be fast in the phoneme-monitoring task. In the PROS condition, we expect slower RTs from the CI group if they are less efficient than the NH listeners at processing prosodic cues to the location of focus.

As semantic context can guide the expected location of focus, we expect RTs of both groups to be faster in the SEM condition, where semantic context is available both visually and auditorily to both participant groups.

RESULTS

Phoneme Discrimination Task

Percent correct scores on the two-alternative forced choice phoneme discrimination task were 99.5 for the CI group and 99 for the NH group, indicating that both groups could discriminate between all four target phonemes /p, b, k, g/. Both groups are robust in processing the acoustic cues of the target phonemes.

Phoneme Monitoring Task

Prosodic Condition • As in previous phoneme monitoring experiments (Cutler & Foss 1977; Brunner & Pisoni 1982; Eimas et al. 1990; Akker & Cutler 2003), we treated responses of under 100 msec as likely anticipatory responses, where participants pressed the response button *before* hearing the target. For the PROS condition, we discarded 6 responses in total, all of which were under 100 msec. Three responses were discarded from the CI group and three from the NH group. Each discarded response was from a different participant. Two of these responses were negative, indicating that these 2 participants (1 CI user and 1 NH individual) had, in anticipating the target, pressed the response button *before* presentation of the target phoneme. After discarding these responses, 97.7% of trials from each group were submitted to statistical analyses.

A linear mixed model was created in R (R Core Team 2009) to compare RT to the target phoneme with group (CI or NH) as a fixed factor and subject and phoneme (/b, p, k, g/) as random factors. p values were obtained using Markov chain Monte Carlo sampling. As predicted, the CI user group responded more slowly to the target phonemes and this difference was significant $p < 0.001$. Mean RTs were 512 msec for the CI group and 317 msec for the NH group.

Semantic Condition

In this condition, a contextual question was presented visually and auditorily before the stimulus utterance was presented. One response was discarded from the CI group and 10 were discarded from the NH group. Of the 10 discarded from the NH group, 6 were from subject 10. All discarded responses were less than 100 msec and considered anticipatory responses. On discarding these responses, 99% of trials from the CI group and 92% from the NH group were submitted to statistical analyses.

RTs to the stimuli in the PROS condition were compared with RTs in the SEM condition. A linear mixed model was used to compare RTs with the target phoneme with group (CI or NH) and condition (PROS/SEM) as fixed factors. Subject and phoneme (/b, p, k, g/) were included as random factors. The main effects of group and condition were significant ($p < 0.001$ for both effects), but the interaction of group and condition was not (Fig. 1). The mean RT to the target phoneme when context was provided was 385 msec for the CI users and 232 msec for the NH users. Although the provision of contextual cues speeded response times for both groups (by 25% for the CI users and 27% for NH group), the CI users remained slower than their NH peers.

DISCUSSION

Previous studies with NH listeners have indicated that listeners use prosodic and semantic/contextual cues in their search

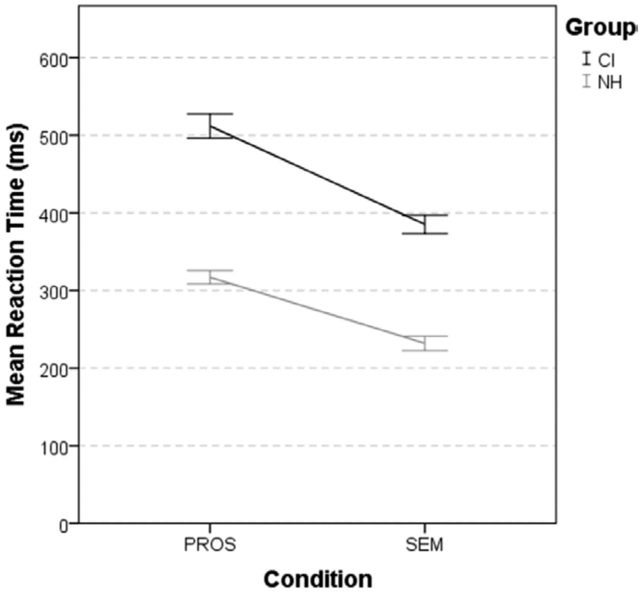

Fig. 1. Mean RT to the target phoneme in msec by the CI user group (*black line*) and NH group (*light grey line*). The CI group RTs were significantly slower than those of the NH group both when contextual cues were available (SEM condition, *right*) and when they were not (PROS condition, *left*). Error bars represent one standard error of the mean. CI indicates cochlear implant; NH, normally-hearing; PROS, prosodic; RT, response time; SEM, semantic.

for a linguistically focused word (Cutler 1976; Akker & Cutler 2003). Poor use of prosodic cues among prelingually deaf CI users has been previously reported in both perception and production (Straatman et al. 2010; Meister et al. 2011). In this article, we asked whether prelingually deaf CI users exploit these prosodic and contextual cues to the location of focus to the same extent as NH listeners. Overall results revealed differences in the processing times of CI users compared with their NH peers. This processing difference cannot be attributed to the difficulty CI users might have encountered in identifying different test consonant phonemes, as both CI users and NH listeners scored over 90% correct in the phoneme discrimination task. This indicates that the CI users could discriminate between stop consonants that differed in place of articulation and/or voicing. This rules out the possibility that RTs in the PROS and SEM conditions reflect perceptual confusion between test consonants.

The *prosodic* (PROS) condition investigated CI users' RT to the target phoneme of a focused word when only prosodic cues preceding the target word were available. Results indicate that the CI group was significantly slower in performing this task than the NH group. We conclude from this that the CI users were not as efficient as the NH participants at exploiting the prosodic cues early in the sentence, and that this slowed their search for the target phoneme. This would be in keeping with Meister et al. (2011), who studied the discrimination of F0, intensity, and duration changes, along with the identification of accented and focused words. For the CI users, they found a moderate correlation between discrimination of individual prosodic cues and identification of a focused word, with better discrimination associated with better identification. Although performance was poorer than that of the NH listeners, the CI users most accurately identified the focused word when all of F0, intensity, and duration cues were available in combination. In our experiment, listeners were presented with all of these cues combined, thus giving the CI users the optimum (and most realistic) environment in which to respond to the target phoneme. As our CI users still reacted more slowly than the NH listeners to the target phoneme, we surmise that this reflects their poorer discrimination of and/or ability to use prosodic cues.

In the semantic (SEM) condition, listeners were provided with additional semantic and syntactic context by means of a focus-inducing question, before presentation of the stimulus utterance. The contextual question was presented both visually and auditorily. We expected that both groups of listeners would benefit from having contextual cues to aid their search for the focused word, and that this would be evidenced by faster RTs. This prediction was supported; both groups were faster to respond when contextual cues were available, illustrating that top–down linguistic expectations were used by both groups to predict which word would be focused in the next utterance. However, despite being able to use this context information to help their search, the CI users remained slower in processing speed compared with their NH counterparts. Moreover, as can be seen by comparing performance by CI users with that of NH listeners (Figs. 1 and 2), the *fastest* speed of the majority of CI users in the SEM condition was *slower* than the speeds of the NH listeners in the PROS condition. In a conversational setting, this difference in processing speed would set our CI users at a distinct disadvantage, as conversation requires participants to rapidly perceive, process, integrate, and respond to incoming information. If CI users are less able to use prosodic cues to help find and incorporate important information in conversation, then their ability to respond in a discourse-appropriate fashion will be hindered.

Individual differences in performance were evidenced for both the NH and CI groups. As this study is preliminary to a larger study of prosody and language processing in CI users, further work will aim to account for these differences seen in

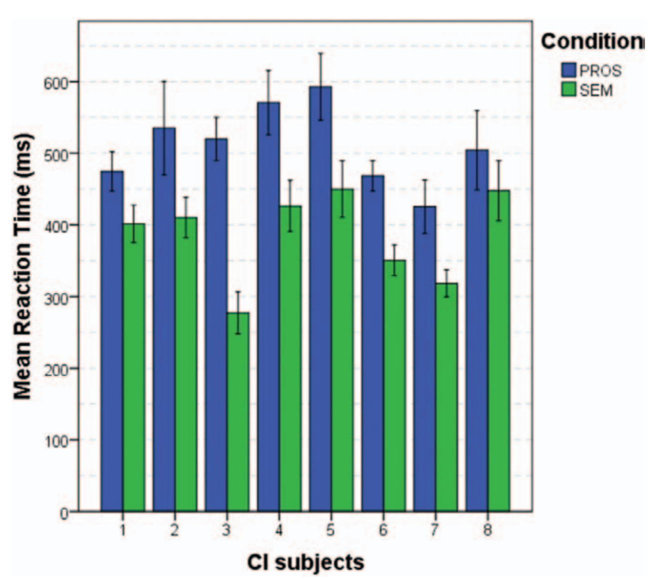

Fig. 2. Mean RT to the target phoneme in msec by the CI users group in the PROS (*blue bars*, prosodic cues only) and SEM (*green bars*, prosodic and contextual cues available) conditions. Error bars represent one standard error of the mean. CI indicates cochlear implant; PROS, prosodic; RT, response time; SEM, semantic.

RTs. The 2 bilateral CI users (subjects 3 and 7) returned the fastest times of the CI group in the SEM condition, when both prosodic and contextual cues were available to aid in the search for the target phoneme. It is possible that availability of binaural redundancy and binaural summation cues led to faster RTs for the 2 bilateral CI users, however only subject 7 returned faster RTs in both conditions. The difference in RT between the PROS and SEM conditions for subjects 8, 14, and 16 suggest that these individuals derived less benefit from the addition of contextual cues than the other participants (Fig. 3). This may suggest a preference for using prosodic cues to focus location over contextual cues. Further investigations are warranted to examine whether individuals show preference in attending to particular cues over others.

Wechsler-Kashi et al. (2014) recently compared RTs on a lexical access (picture naming) task conducted with pediatric CI users and NH counterparts. They found higher nonverbal IQ scores, earlier age at implantation and longer periods of device use to be significantly correlated with faster lexical access speeds. Interestingly, language ability was not correlated with RT. Although neither language ability nor cognitive ability were tested in this study, our CI group was relatively homogenous in terms of age at implant and device experience. We cannot exclude the possibility that differences in nonverbal IQ may account for some of the variation evidenced.

The results also cannot be generalized to performance across multiple speakers and different genders, as the stimuli of this study were recorded by a male speaker only. The fine temporal structure cues used by NH listeners to derive a pitch percept are not available to CI users, and so changes in F0 are difficult for CI users to perceive accurately (Zeng 2002; Moore 2003). CI users can utilize amplitude modulation cues in lieu of fine temporal structure cues up to around 300 Hz (Zeng 2002; McKay 2005; McDermott 2011). As the mean F0 of an average adult female speaker's voice is 225 Hz (de Pinto & Hollien 1982),

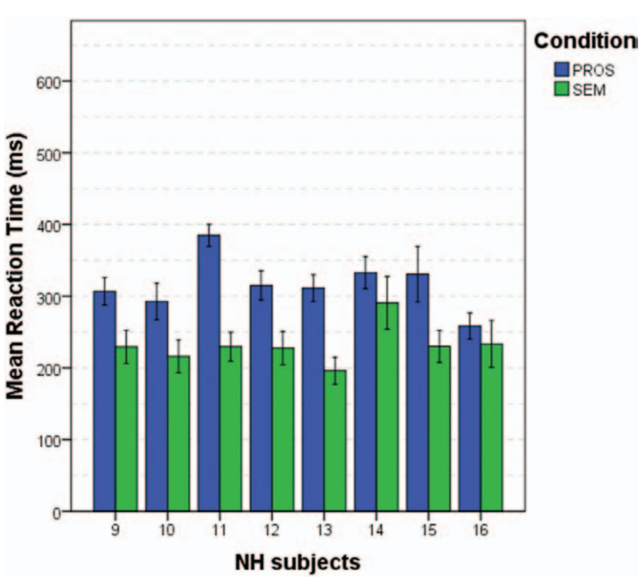

Fig. 3. Mean RT to the target phoneme in msec by the NH group in the PROS (*blue bars*, prosodic cues only) and SEM (*green bars*, prosodic and contextual cues available) conditions. Error bars represent one standard error of the mean. NH indicates normally-hearing; PROS, prosodic; RT, response time; SEM, semantic.

changes in F0 typical of those used to signal intonation may result in F0 values which exceed 300 Hz for female speakers. Consequently, the amplitude modulation cue might not be available for the listeners to utilize when F0 values go above 300 Hz. Straatman et al. (2010) examined the just noticeable difference in F0 perceived by children using CIs. They manipulated F0 contours typical of the rise of a pitch accent occurring over one syllable, from both a male and a female speaker. The just noticeable difference threshold was larger for the stimuli in the female voice (11.3 st) than the male voice (9.4 st). The onset F0 value of the pitch accent recorded by the female speaker was 200 Hz. Given that larger thresholds were seen for changes in F0 produced by female speaker in the study of Straatman et al (2010), it is possible that CI users may perform more poorly in the experimental task of the current study if presented with stimuli in a female voice. As we used stimuli recorded by a male voice only, the results of this study reflect how CI users process the prosodic and semantic cues under optimal conditions when F0 values are below 300 Hz. Future studies are warranted to investigate whether CI users will process these cues differently under nonoptimal conditions, such as using stimuli from female speakers.

The finding of slower processing speeds for prelingually deaf CI users contributes to our understanding of the communicative challenges faced by this group. As noted by Pisoni (2014), to fully understand and explain the variability in outcomes of CI users, we need to study the processes that underlie linguistic performance. The results of this study help us to understand how deficits in the perception of prosodic cues may impact on an individual's language processing speed and efficiency, a process hitherto unexplored with CI users.

ACKNOWLEDGMENTS

We thank Matthew Poole for recording the stimuli and Louise Shum for contributing to data collection. We thank Anne Cutler and Stephanie Shattuck-Hufnagel for discussions in the early stages of this work.

This work was financially supported by Australian Research Council Laureate Fellowship FL130100014 awarded to the second author. This study has approval from the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (Project 08/840H), ratified by the Human Ethics Advisory Committee of The University of Melbourne (ID 1441440).

Parts of this study were presented at *WAPSTI 2014, Workshop on the role of prosody in language learning: Stress, tone and intonation,* Sydney Australia, December 8–9, 2014; and at the *3rd Hong Kong Speech and Hearing Symposium 2014*, Hong Kong, China, October 11–12, 2014. We thank the audience members of these conferences for their valuable feedback.

The authors have no conflicts of interest to disclose.

Address for correspondence: Colleen M. Holt, Audiology and Speech Pathology, The University of Melbourne, 550 Swanston St, Carlton Victoria 3053 Australia. E-mail: cholt@unimelb.edu.au

Received February 11, 2015; accepted October 24, 2015.

REFERENCES

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Biling Lang Cogn*, *6*, 81–96.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *J Pragmat*, *38*, 1636–1657.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonol Yearbook*, *3*, 5–70.

Birch, S., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *J Mem Lang*, *47*, 571–588.
Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.5.16).
Botinis, A., Granström, B., Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Commun*, *33*, 263–296.
Brunner, H., & Pisoni, D. B. (1982). Some effects of perceptual load on spoken text comprehension. *J Verbal Learning Verbal Behav*, *21*, 186–195.
Ciocca, V., Francis, A. L., Aisha, R., et al. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am*, *111*(5 Pt 1), 2250–2256.
Connine, D., & Titone, C. (1996). Phoneme monitoring. *Lang Cognitive Proc*, *11*, 635–646.
Cruttenden, A. (1997). *Intonation*. Cambridge, UK; New York, NY: Cambridge University Press.
Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Percept Psychophys*, *20*, 55–60.
Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Percept Psychophys*, *29*, 217–224.
Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*, 49–59.
Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Lang Speech*, *20*, 1–10.
Cutler, A., Dahan, D., van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Lang Speech*, *40*(Pt 2), 141–201.
Dahan, D., Tanenhaus, M. Chambers, C. (2002). Accent and reference resolution in spoken-language comprehension. *J Mem Lang*, *47*, 292–314.
de Pinto, O., & Hollien, H. (1982). Speaking fundamental frequency characteristics of Australian women: Then & now. *J Phonetics*, *10*, 367–375.
Dilley, L., & McAuley, J. (2008). Distal prosodic context affects word segmentation and lexical processing. *J Mem Lang*, *59*, 294–311.
Eimas, P., & Nygaard, L. (1992). Contextual coherence and attention in phoneme monitoring. *J Mem Lang*, *31*, 375–395.
Eimas, P., Hornstein, S., Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *J Mem Lang*, *29*, 160–180.
Fletcher, J., Stirling, L., Mushin, I., et al. (2002). Intonational rises and dialog acts in the Australian English map task. *Lang Speech*, *45*, 229–253.
Foss, D. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *J Verb Learn Verb Behav*, *8*, 457–462.
Foss, D., & Lynch, R. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. *Percept Psychophys*, *5*, 145–148.
Fox, A. (2000). *Prosodic Features and Prosodic Structure*. New York, NY: Oxford University Press.
Frauenfelder, U. H., Segui, J., Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory. *J Exp Psychol Hum Percept Perform*, *16*, 77–91.
Geers, A., & Hayes, H. (2011). Reading, writing, and phonological processing skills of adolescents with 10 or more years of cochlear implant experience. *Ear Hear*, *32*, 49S–59S.
Hirschberg, J. (2004). Pragmatics and intonation. In D. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 515–537). Oxford, UK: Blackwell Publishing.
Klieve, S., & Jeanes, R. (2001). Perception of prosodic features by children with cochlear implants: Is it sufficient for understanding meaning differences in language? *Deaf Ed Int*, *3*, 15–37.
Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge, UK; New York, NY: Cambridge University Press.
Lee, K., van Hasselt, C., Chiu, S., et al. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int J Pediatr Otorhinolaryngol*, *63*, 137–147.
Looi, V., & Radford, C. J. (2011). A comparison of the speech recognition and pitch ranking abilities of children using a unilateral cochlear implant, bimodal stimulation or bilateral hearing aids. *Int J Pediatr Otorhinolaryngol*, *75*, 472–482.
McDermott, H. (2011). Music perception. In F. G. Zeng, A. Popper, R. Fay (Eds.), *Auditory Prostheses: New Horizons* (pp. 305–339). London: Springer.
McKay, C. (2005). Spectral processing in cochlear implants. *Int Rev Neurobiol, 70,* 473–509.
Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Lang Speech*, *31*(Pt 2), 135–156.
Meister, H., Landwehr, M., Pyschny, V., et al. (2011). The perception of sentence stress in cochlear implant recipients. *Ear Hear*, *32*, 459–467.
Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol Neurotol*, *24*, 243–254.
Most, T., & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing AIDS. *J Deaf Stud Deaf Educ*, *12*, 350–361.
Nicholas, J. G., & Geers, A. E. (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. *J Speech Lang Hear Res*, *50*, 1048–1062.
Peng, S. C., Lu, N., Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiol Neurootol*, *14*, 327–337.
Peng, S. C., Tomblin, J. B., Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear*, *29*, 336–351.
Peng, S. C., Tomblin, J. B., Cheung, H., et al. (2004). Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear*, *25*, 251–264.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, MIT, Department of Linguistics and Philosophy.
Pisoni, D. (2014). Rapid phonological coding and working memory dynamics in children with cochlear implants: Cognitive foundations of spoken language processing. In A. W. Farris-Trimble (Ed.), *Language Acquisition & Language Disorders* (Vol. 56, (pp. 91–112). Philadelphia, PA: John Benjamins.
Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cogn Psychol*, *29*, 149–188.
R Core Team. (2009). R: A language and environment for statistical computing (Version 2.10.1).
Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *J Psycholinguist Res*, *25*, 193–247.
Shields, J., McHugh, A. Martin, J. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *J Exp Psychol*, *102*, 250–255.
Spencer, L. J., Barker, B. A., Tomblin, J. B. (2003). Exploring the language and literacy outcomes of pediatric cochlear implant users. *Ear Hear*, *24*, 236–247.
Straatman, L. V., Rietveld, A. C., Beijen, J., et al. (2010). Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. *J Acoust Soc Am*, *128*, 1884–1895.
Sucher, C. M., & McDermott, H. J. (2007). Pitch ranking of complex tones by normally hearing subjects and cochlear implant users. *Hear Res*, *230*, 80–87.
Toe, D. M., & Paatsch, L. E. (2013). The conversational skills of school-aged children with cochlear implants. *Cochlear Implants Int*, *14*, 67–79.
Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *J Phonetics*, *35*, 445–472.
Venditti, J., & Hirschberg, J. (2003). Intonation and discourse processing. In *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 315–318). Saarbrücken, Germany.
Watson, D., Gunlogson, C. Tanenhaus, M. (2006). Online methods for the investigation of prosody. In S. Sudhoff, D. Lenertov, R. Meyer, et al. (Eds.), *Methods in Empirical Prosody Research* (pp. 259–282). New York: de Grouter.
Wechsler-Kashi, D., Schwartz, R. G., Cleary, M. (2014). Picture naming and verbal fluency in children with cochlear implants. *J Speech Lang Hear Res*, *57*, 1870–1882.
Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Lang Speech*, *46*(Pt 1), 53–81.
Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hear Res, 174,* 101–106.