Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions\textsuperscript{a)}

Cheryl F. Rogers, Eric W. Healy,\textsuperscript{b)} and Allen A. Montgomery

Department of Communication Sciences and Disorders, The Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208

(Received 20 April 2005; revised 22 December 2005; accepted 28 December 2005)

Sensitivity to acoustic cues in cochlear implant (CI) listening under natural conditions is a potentially complex interaction between a number of simultaneous factors, and may be difficult to predict. In the present study, sensitivity was measured under conditions that approximate those of natural listening. Synthesized words having increases in intensity or fundamental frequency (F0) in a middle stressed syllable were presented in soundfield to normal-hearing listeners and to CI listeners using their everyday speech processors and programming. In contrast to the extremely fine sensitivity to electrical current observed when direct stimulation of single electrodes is employed, difference limens (DLs) for intensity were larger for the CI listeners by a factor of 2.4. In accord with previous work, F0 DLs were larger by almost one order of magnitude. In a second experiment, it was found that the presence of concurrent intensity and F0 increments reduced the mean DL to half that of either cue alone for both groups of subjects, indicating that both groups combine concurrent cues with equal success. Although sensitivity to either cue in isolation was not related to word recognition in CI users, the listeners having lower combined-cue thresholds produced better word recognition scores. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2167150]

PACS number(s): 43.66.Fe, 43.64.Me, 43.71.Es [JHG] Pages: 2276–2287

I. INTRODUCTION

It is likely that sensitivity to both intensity and fundamental frequency (F0) is important for the perception of complex spectro-temporal signals such as speech. Variations in intensity over time provide temporal information capable of encoding a number of important speech cues. The difference limen (DL) for detecting changes in the intensity of a broadband noise is approximately 0.5 to 1 dB for normal-hearing (NH) listeners across a wide range of overall levels, but larger near threshold (Miller, 1947). Early experiments employing electrical stimulation of human (Shannon, 1983) and animal subjects (Pfähler, Burnett, and Sutton, 1983) showed both a trend for larger DLs at low levels, and smaller DLs for electrical than for acoustic stimulation. When expressed as Weber fractions, sensitivity to changes in electrical current can be extraordinarily fine. Zeng (2004) describes cases in which a cochlear implant (CI) user can reliably produce an electrical-stimulation DL of 0.08 dB—a value far below that detectable in normal hearing. More generally, the DL measured under direct electrical stimulation is roughly one order of magnitude smaller than the acoustic DL, and this difference has been related to the lack of cochlear compression in electrical stimulation (Zeng, 2004, also see Nelson et al., 1996). However, this extremely fine sensitivity is offset by the small dynamic range of most CI users. Whereas the dynamic range of NH is roughly 120 dB (cf. Bacon, 2004), the difference between threshold and maximum loudness in electrical stimulation has been shown to be approximately 10 to 30 dB (Zeng and Shannon, 1994, 1999). This combination of fine sensitivity and reduced dynamic range combines to produce discriminable steps that number roughly 5 to 45 in CI users (Nelson et al., 1996; Zeng and Shannon, 1999; Kreft, Donaldson, and Nelson, 2004), compared to 50 to 200 in normal hearing (Viemeister and Bacon, 1988).

A number of factors have been shown to affect intensity discrimination for stimulation of single electrodes with electrical pulse trains. Discrimination may be affected by pulse rate (Kreft, Donaldson, and Nelson, 2004), electrode configuration (Drennen and Pfänigst, 2005), proximity to the modiolus (Cohen, Saunders, and Clark, 2001), and electrode position (Nelson et al., 1996). Nelson et al. (1996) reported large variability in the number of discriminable loudness steps within a group of eight CI users. Although some relationship between the size of the discriminable step and threshold or dynamic range was observed, this relation did not hold for different electrode positions within the same subject. More importantly, the number of steps was not simply related to the absolute size of the electrical dynamic range. It has also been found that characteristics of the stimulus can affect discrimination. Wojtczak, Donaldson, and Viemeister (2003) observed smaller DLs for increments applied to continuous carriers relative to those of gated carriers for some subjects and level conditions. Because the meaningful


\textsuperscript{b)} Electronic mail: ewh@sc.edu
intensity variations that occur in speech are often relative to a baseline speech level, these results may be more relevant to everyday signals than those from the gated-carrier conditions usually employed.

The F0 of an utterance carries important information regarding speaker gender, identity, and intent. It is the means by which intonation is encoded, and it is one of the cues signifying lexical stress (cf. Cutler, 2005; Vaissiere, 2005). Normal-hearing listeners generally produce very small F0 DLs when all harmonics exist and both place and temporal cues are present. Classic work demonstrated that the F0 DL for steady-state synthesized vowels is on the order of 0.3 to 0.5 Hz (Flanagan and Saslow, 1958). However, when listeners are forced to rely on temporal cues alone, either by removing the lower resolved harmonics (Cullen and Long, 1986; Houtsma and Smurzynski, 1990) or by employing sinusoidally amplitude modulated (SAM) noise (Formby, 1985), DLs increase by roughly one order of magnitude.

Because the spectral representation provided by the CI electrode array is quite poor, pitch discrimination in these users may depend more strongly upon temporal cues than in normal hearing. Early experiments showed that CI recipients could discriminate the pitch of pulse trains having different repetition rates up to roughly 300 Hz (Shannon, 1983; Tong and Clark, 1985; Townshend et al., 1987). Recent estimates indicate that the DL for rate discrimination (Zeng, 2002) or sinusoidal frequency modulation (FM) detection (Chen and Zeng, 2004—approximately 20 Hz at a standard frequency of 100 Hz) is roughly one order of magnitude larger for CI listeners under direct stimulation relative to NH listeners. Although modern CI devices do not employ pulse trains of varying repetition rate, these values are meaningful because some similarity has been observed between the pitch elicited by repetition rate and that elicited by fixed-rate pulse trains having different SAM rates (McKay, McDermott, and Clark, 1994). Thus, NH listeners have available both spectral and temporal cues to pitch, whereas CI users rely primarily on spectral cues. This is evidenced by the similarity in performance across NH and CI listeners under conditions in which spectral cues are removed and only temporal cues are available.

As with intensity discrimination, a number of factors can influence sensitivity to changes in F0. Although the DL for NH listeners hearing steady-state vowels is quite small, this value increases by roughly a factor of ten when F0 rises or falls in linear fashion over time (Klatt, 1973). Geurts and Wouters (2001) examined F0 discrimination in conditions that more closely resembled those of normal CI listening. The stimuli were synthesized vowels having F0 coded by modulation rate of amplitude-modulated pulse trains. Although sound-field presentation was not employed, the study differed from others by employing direct stimulation of all electrodes in interleaved fashion. The authors found discrimination of a 400-ms steady-state vowel at 150 Hz to average approximately 12 Hz (8%). It was also found that multiple-electrode stimulation, which is more representative of natural CI listening, allowed detection of modulation at smaller depths when compared to stimulation of individual electrodes.

Studies designed to measure discrimination sensitivity in CI users have generally employed direct stimulation of single electrodes using simple pulsatile stimuli. Although essential for isolating particular aspects or stages of processing, direct stimulation bypasses much of the processing that would normally take place in the CI reception of sound. Further, results obtained using simple pulsatile stimuli must be extrapolated to make assumptions about sensitivity to complex spectro-temporal signals such as speech. In sharp contrast, sensitivity to acoustic speech cues by CI listeners under natural listening conditions is a potentially complex interaction between a number of factors. One set of influences is attributable to the CI device and may include pulse rate, electrode configuration, electrode position, proximity to the modiolus, single-versus multi-electrode stimulation, etc. Another set of factors is related to the complexity of the stimulus. Discrimination of various cues may be influenced by nonstationary pitch and/or intensity, and duration of the target segment. Further, despite the large number of studies that have examined sensitivity to individual cues in isolation, little work has examined sensitivity under conditions that simulate the simultaneous frequency and intensity changes that occur in speech.

In the present study, sensitivity to increments in intensity and F0 was examined under pseudonatural listening conditions. The stimuli were synthesized words in which highly controlled manipulations of the cues existed, but these changes had the spectral and temporal characteristics of those that normally occur in speech. In addition, subjects were tested in the free field of a calibrated studio monitor using their everyday clinical speech processors to examine sensitivity when the CI device is performing in its normal mode from microphone to electrode array.

The goals of the present study are (1) to compare sensitivity to increments in intensity and F0 across NH and CI listeners under conditions that more closely approximate those of natural listening, (2) to determine if sensitivity is enhanced when concurrent intensity and F0 increments are present, (3) to determine if this potential enhancement is similar in magnitude across NH and CI listener groups, (4) to examine the relation between sensitivity to intensity increments, F0 increments, and concurrent F0 and intensity increments, to word recognition in CI users.

II. EXPERIMENT 1. SENSITIVITY TO INDIVIDUAL CUES

A. Method

1. Subjects

Two groups of subjects participated. The CI group was composed of six individuals ranging in age from 30–70 years. Five were fitted with the Nucleus 24—Esprit 3G device and one (CI2) was fitted with the Clarion II—PSP device. The subjects with the Nucleus device all used the ACE strategy, monopolar (MP1+2) stimulation, a Q value of 20, and a stimulation rate of 900 Hz/channel. The subject with the Clarion device used the HiRes-P strategy with a stimulation rate of 5156 Hz/channel. Table I provides additional information. As the speech recognition scores reflect,
the CI users were all relatively good performers. The second group was composed of five individuals with NH [audiometric thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz (ANSI, 1996)]. All were native speakers of American English and were paid for their participation. Only one subject (NH4) had previous experience in psychoacoustic tasks.

2. Stimuli and apparatus

The stimuli consisted of the synthesized words, “potato,” “sufficient,” and “allowance” (Words 1–3). These words have a middle stressed syllable that is embedded in different phonetic environments. Productions of each word were made by a male speaker having a standard American dialect, and acoustic measurements of these productions were entered into a Klatt-based speech synthesizer operating at 20 kHz sampling. Each of the resulting synthesized words was clearly identifiable and even allowed NH listeners familiar with the speaker to recognize his voice. Figure 1 displays the amplitude envelopes of the three synthesized words. The standard durations of the middle vowels for Words 1–3 were 140, 70, and 150 ms, respectively. Additional versions of each stimulus word were also created in which the middle vowel had systematic increases in either intensity or F0.

a. Intensity increments. In this series, the intensity of the middle vowel was increased in 0.25-dB steps. Because the speech synthesis application provided an intensity resolution of only 1 dB, stimuli having 1-dB steps were synthesized, and these files were scaled using a waveform editor (Adobe Audition) to obtain the desired 0.25-dB steps. These modifications were performed at zero crossings and included a 5 ms linear rise/fall to eliminate clicks or other artifacts that could potentially aid judgments. Considerable care was taken to ensure that desired specifications were met and that the stimuli varied only in the manner intended. Level measurements of the transduced middle syllables, played in loop mode, indicated that all intensities were within 0.1 dB of the desired value.

b. Fundamental-frequency increments. This series was created by increasing the F0 of the middle syllables in 1 Hz steps, up to 40 Hz, using the speech synthesizer. Prior to modification, the F0 of the middle syllables for Words 1–3 was 132 (±8), 125 (±1), and 91 (±5) Hz, respectively. To assess the accuracy of the F0 modifications, the frequency of each increment step was measured using the waveform editor. Although the speech synthesizer allowed F0 to be specified at high resolution, measurements of the output waveforms indicated that the actual frequencies varied in some cases considerably from the specified values. These actual F0 values were determined using the waveform editor and measurements across the syllable at sample-point resolution, which allowed measurement within 0.03 to 0.2 Hz across conditions. By manipulating the synthesis parameters and re-measuring the resultant sound files, F0-increment stimuli were created for each word that were within 0.3 Hz of the target value for each step in the series up to 15 Hz and within 0.7 Hz for steps above 15 Hz. There was no cumulative drift across the F0 increment values. Fundamental frequency DLs were calculated based upon both the nominal and the measured F0 increment values to examine the influence of these small deviations.

Previous research involving synthesized speech has indicated that a change in F0 may be accompanied by a change in intensity (e.g., Kewley-Port and Watson, 1994), and similar artifacts were observed here. They were relatively small.
at F0 increases of 1–3 Hz, but could be as great as 1.5 dB at the larger frequency increments for all three stimulus words. To correct for these artifacts, each frequency increment stimulus was processed to ensure that its amplitude envelope matched exactly that of the corresponding standard stimulus. Using MATLAB, the envelopes for both the standard and the increment were extracted via full-wave rectification and low-pass filtering at 50 Hz. The waveform of each frequency increment stimulus was then scaled by multiplying it with a correction vector consisting of the ratio between the standard stimulus for F0 increments up to 10 Hz and within 0.1 dB of that of the standard for increments up to 40 Hz.

It was also observed that the manipulation of F0 in the middle syllable affected the onset timing of the subsequent syllable for Words 1 and 2. Therefore, the final syllable for each of the F0 increment stimuli for these words was replaced with that from the standard version. This process was completed using the waveform editor by replacing the final syllable at a particular fixed sample value in the center of the closure of the /t/ for Word 1 (“potato”) or the low-intensity aperiodic phoneme /r/ for Word 2 (“sufficient”) using a 5-ms crossfade to ensure a smooth transition.

Waveforms for stimulus Word 1 (“potato”) are shown in Fig. 2. The upper panel shows the standard synthesized word, and the two panels below it show the stimulus having an increment in either intensity or F0. Throughout the course of stimulus preparation, a pair of individuals having experience preparing psychoacoustic stimuli examined each stimulus carefully and found them to be free of clicks or other artifacts that could possibly aid performance. The digital files were played back from a PC (Edirol UA-5 D/A converter) and presented using an amplified loudspeaker (Mackie HR824) having a flat frequency response (±1.5 dB, 39 Hz–20 kHz). Each standard stimulus was set to play back at a slow-response rms peak level of 70 dB SPL measured 1 m directly in front of the speaker.

3. Procedure

Subjects were administered the Consonant Nucleus Consonant test (CNC, Peterson and Lehiste, 1962) to assess identification of monosyllabic words and their constituent phonemes. The 50-item test was administered in soundfield at 65 dB SPL within a double-walled sound booth. Results are presented in Table I. Also provided are scores for the Hearing in Noise Test (HINT, Nilsson, Soli, and Sullivan, 1994) sentences presented both in quiet and with the standard noise background at +10 dB S/N (CI2 was unavailable for HINT testing).

Difference limens (DLs) were measured for each isolated cue for each of the three stimulus words. Presentation of conditions and words was random, with the exception that listeners heard all three words in either the intensity or F0 condition before advancing to hear all three words in the other condition. Thresholds were obtained using an adaptive three-interval, two alternative forced-choice (3I-2AFC) paradigm with a two-down, one-up decision rule to track the 71% correct point on the psychometric function (Levitt, 1971). The standard always appeared in the first interval and again in either the second or third. The increment step appeared randomly in either the second or third interval and 500 ms of silence separated each interval. A track began with the step stimulus clearly distinguishable from the standard. Subjects selected the different stimulus from the final two intervals by responding on a computer keyboard. A block of trials was terminated after 12 reversals, and the mean of the last eight was taken as the threshold estimate. The step sizes were 2 dB or 3 Hz prior to the fourth reversal, and 0.25 dB or 1 Hz thereafter.

A minimum of three threshold estimates were collected in each condition. Estimates comprising a mean were collected across different days, with each session lasting one to two hours. Testing took place with the subject seated 1 m directly in front of the calibrated loudspeaker within a double-walled sound booth. Trial-by-trial feedback was provided using a computer monitor located adjacent to the speaker. Presentation of the sound files in adaptive fashion and collection of responses were performed using custom software written in MATLAB. All subjects received a minimum of five hours of practice prior to data collection.

Although a criterion for inclusion of intensity threshold estimates based on the standard deviation (SD) of the reversals was set at 1.5 dB, no estimate from the highly-practiced subjects exceeded this value. Further, 89% of the adaptive
tracks had a SD below 0.75 dB. For the NH listeners, the frequency estimate criterion was set to 3 Hz, but no estimate exceeded this value. Further, 96% of the estimates had a SD below 2 Hz. For the CI listeners, the SD criterion defining an acceptable estimate was increased to 6 Hz and estimates not conforming to this (9% of the total) were discarded and replaced. The majority (84%) of the estimates following replacement had a SD below 3 Hz.

The CI subjects completed the experiment using their usual clinical processor, map, and strategy. It was important to ensure that the stimuli fell within the linear input range of the device and were not affected by the limiting of input signals to a maximum output current level. To ensure this operation and still provide stimuli having levels that approximated those of normal speech, autosensitivity was disabled and subjects were instructed to adjust sensitivity within a range that ensured linear operation. It was found that subjects generally chose a similar setting, and that all reported stimulus levels approximating those of normal conversation at that setting. A sensitivity setting of 1 was therefore standardized for all subjects using the Nucleus device. This setting placed the standard 70 dB (SPL) stimulus near the center of the 60–90 dB input dynamic range of the implant. The subject with the Clarion device was programmed with sensitivity set to zero and input dynamic range set to 60 dB (spanning input between 25 and 85 dB SPL).

**B. Results**

1. **Sensitivity**

   a. **Intensity increments.** Intensity DLs and corresponding SDs are shown in Fig. 3. The mean intensity DL (increment size) across words and listeners was 1.3 dB for the NH listeners and 3.1 dB for the CI listeners. This planned comparison between listener groups was significant \((t(9)=-2.65, p =0.03)\), indicating that the NH listeners were on average more sensitive to increments in intensity within the speech stimuli than were the CI listeners. Although DLs were larger on average for the CI listeners, substantial variability was observed within this group, and two of the listeners (CI4 and CI5) produced DLs within the range of the NH means.

   The three stimulus words were chosen to have different phonetic characteristics, which produced substantially different amplitude envelope shapes (see Fig. 1). To examine possible effects related to these differences, one-way repeated-measures analyses of variance (ANOVAs) were performed. The NH listeners \([F(2, 8)=11.77, p =0.004]\) produced intensity DLs that were larger \((p<0.05)\) by Tukey post-hoc for Word 3 (“allowance”), which was continuously voiced and had a smooth amplitude envelope lacking the steplike increment characteristic of the other two words. However, no significant difference was observed across words for the CI listeners \([F(2, 10)=2.05, p=0.18]\).

   b. **Fundamental frequency increments.** Fundamental frequency DLs and corresponding SDs are presented in Fig. 4. The mean DL across words and listeners was 3.2 Hz for the NH listeners and 25.0 Hz for the CI listeners. This planned comparison between groups was significant, indicating that the CI users were less sensitive to increments in F0 in the speech stimuli \((t(9)=-7.44, p=0.001)\). The values are expressed as Weber fractions in Table II.

   An examination of Fig. 4 shows that, although the NH listeners produced generally similar DLs across the three words, several of the CI listeners produced substantially
higher thresholds for stimulus Word 2 ("sufficient"). Indeed, four of the CI listeners produced threshold estimates near the ceiling frequency value of 40 Hz for this word, and so it is possible that even these large DLs slightly underestimate their true thresholds for this word. This observed difference between words is supported by statistical analyses, in which no effect of word was found in a one-way repeated-measures ANOVA for the NH listeners \([F(2, 8) = 2.80, p = 0.12]\). However, a second ANOVA showed that the CI subjects \([F(2, 10) = 11.35, p = 0.003]\) produced larger DLs for Word 2 relative to the other words \((p < 0.05)\) by Tukey post-hoc. These results did not change when the tests were performed on Weber fractions.

To examine the influence of the small deviations from nominal F0 values observed for some of the stimuli, the DLs were recalculated using the measured F0 value of each increment stimulus. Mean thresholds were on average within 0.1 Hz of the thresholds based on nominal frequency and shown in Fig. 4, and only three estimates exceeded a difference of 0.2 Hz from the nominal value.

### 2. Relation to speech recognition

Further analyses were completed to determine if there was a relationship between sensitivity to intensity increments within the speech stimuli and performance on a standard speech-recognition test for the CI users. First it was important to establish that the recognition scores covered an appropriate range. By examining the binomial probabilities (Thornton and Raffin, 1978), it was found that significant differences existed across the word recognition scores produced by the CI users. The lowest of the six scores (54%) was significantly different \((p < 0.05)\) from the top three scores (76%, 80%, 82%); the top score (82%) was significantly different from the bottom three scores (54%, 58%, 64%); and the two middle scores were significantly different from the top and bottom scores, respectively.

Figure 5 shows the relationship between the mean inten-

![FIG. 4. Difference limens for individual subjects hearing synthesized words having a fundamental-frequency increment in the middle syllable.](image)

![FIG. 5. Mean intensity DLs produced by the CI subjects as a function of their performance on the CNC word recognition test.](image)

<table>
<thead>
<tr>
<th></th>
<th>Word 1</th>
<th>Word 2</th>
<th>Word 3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>2.34</td>
<td>3.35</td>
<td>2.68</td>
<td>2.79</td>
</tr>
<tr>
<td>CI</td>
<td>17.88</td>
<td>29.08</td>
<td>18.77</td>
<td>21.91</td>
</tr>
</tbody>
</table>

**Table II. Mean fundamental frequency DLs as Weber fractions (%).**
sity DLs for each word as a function of CNC word recognition for each CI listener. No relation is apparent from the figure, and the correlation between mean DL across the three words for each listener and word recognition was nonsignificant \((r=-0.35, p=0.50)\). Figure 6 shows the relationship between the mean F0 DLs for each word as a function of the word recognition score for each CI listener. Again, no relation is apparent from the figure \((r=-0.67, p=0.14, \text{nonsignificant})\).

III. EXPERIMENT 2. SENSITIVITY TO CONCURRENT CUES

In this experiment, intensity and F0 increases were presented concurrently within a single stimulus word and potential sensitivity enhancement resulting from the ability to combine cues was assessed.

A. Method

The same six CI subjects and five NH subjects who completed Experiment 1 participated in this experiment. The DLs obtained in Experiment 1 were used in this experiment to equate the perceptual equivalence of intensity and F0 increments for each of the three stimulus words. For the NH subjects, the median DL for each word was employed. Because of the increased variability of the CI subjects, different combined-cue stimuli were created for each subject, tailored to their individual DLs for each word.

Psychometric functions relating \(d'\) to increment size are generally linear and pass through the origins (e.g., Turner and Nelson, 1982; Nelson and Freyman, 1986; Nelson et al., 1996; Lentz and Richards, 1997; Moore et al., 1999; Jesteadt, Nizami, and Schairer, 2003). Indeed, these functions are often fit using lines through the origins (e.g., Nelson and Freyman, 1986; Nelson et al., 1996; Moore et al., 1999). These studies provide some assurance that the methodologically simple procedure of dividing the increment at DL into physically equal steps produces increments that are also perceptually equivalent.

For the NH subjects, stimulus Words 1 and 3 had concurrent-cue steps in which the intensity and F0 increases matched 1/8 of the corresponding median DLs. For Word 2, careful measurement of F0 indicated that 1/4-DL steps were the smallest that could be reliably created using the maximum 0.1-Hz nominal resolution of the synthesizer. Frequency increment stimuli (using steps of 0.32, 0.78, and 0.27 Hz for Words 1–3, respectively) were created using the procedures employed in Experiment 1. Corresponding 1/8- or 1/4-DL intensity increases (steps of 0.13, 0.24, and 0.24 dB) were then applied to each F0-increment file. Incrementing series up to two or three DLs were created.

To create the combined-cue stimuli for the CI subjects, the integer F0 increment (one to four hertz) that was closest to 1/8 of the F0 DL for each word and listener was chosen, yielding step sizes that ranged across listeners from 1/10 to 1/7 DL. Corresponding intensity step values were then calculated for each word and listener and applied to the F0 increments. Combined-cue stimuli were not created for the few conditions in which F0 DLs were near the ceiling value in Experiment 1. Each intensity and F0 value was confirmed using the procedures employed in Experiment 1, and found to be within 0.1 dB and 0.1 Hz of the nominal values. The same pair of experienced listeners confirmed that all stimuli were free from artifacts. The bottom panel of Fig. 2 shows the waveform for Word 1 (“potato”) having concurrent intensity and F0 increments.

Difference limens were obtained using the procedures employed in Experiment 1. The step size was increased by a factor of two until the first four reversals were obtained. Words were presented in random order. Individual estimates were discarded and replaced on the rare occasion that the SD of the reversals exceeded 0.35 DL. Three or more estimates were averaged to produce each threshold.

B. Results

1. Sensitivity

Thresholds are expressed as proportions of individual-cue DLs in Table III. Therefore, a DL of 1.0 indicates no advantage of concurrent cues over individual cues. When averaged across words and listeners, the NH subjects produced a DL of 0.50 and the CI subjects produced a DL of 0.51 [planned comparison \(t(9)=-0.10, p=0.92\)]. Thus, these data indicate no overall difference across listener groups in the sensitivity enhancement that results from the presence of concurrent cues. Further, the value of 0.5 indicates that the two cues combined to produce a threshold that was half that of either cue alone.

2. Relation to speech recognition

Figure 7 shows the relationship between the mean combined-cue DLs for each word as a function of CNC word recognition for each of the CI users. These data show a tendency for individuals who were better able to combine the concurrent cues to perform better on the speech recognition tests. In contrast to the results of Experiment 1, the correlation between mean DL across the three words and word recognition was significant \((r=-0.87, p=0.02)\).
TABLE III. Mean combined-cue DLs (relative to individual-cue DLs) and standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>Word 1 “Potato”</th>
<th>Word 2 “Sufficient”</th>
<th>Word 3 “Allowance”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>NH 1</td>
<td>0.54</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>NH 2</td>
<td>0.87</td>
<td>0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>NH 3</td>
<td>0.67</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>NH 4</td>
<td>0.45</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>NH 5</td>
<td>0.63</td>
<td>0.20</td>
<td>0.36</td>
</tr>
<tr>
<td>Mean</td>
<td><strong>0.63</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.35</strong></td>
</tr>
<tr>
<td>CI 1</td>
<td>—</td>
<td>—</td>
<td>0.54</td>
</tr>
<tr>
<td>CI 2</td>
<td>0.52</td>
<td>0.05</td>
<td>—</td>
</tr>
<tr>
<td>CI 3</td>
<td>0.24</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>CI 4</td>
<td>0.38</td>
<td>0.05</td>
<td>—</td>
</tr>
<tr>
<td>CI 5</td>
<td>0.52</td>
<td>0.19</td>
<td>—</td>
</tr>
<tr>
<td>CI 6</td>
<td>0.24</td>
<td>0.04</td>
<td>0.47</td>
</tr>
<tr>
<td>Mean</td>
<td><strong>0.38</strong></td>
<td><strong>0.13</strong></td>
<td><strong>0.43</strong></td>
</tr>
</tbody>
</table>

IV. DISCUSSION

Sensitivity under natural conditions is a potentially complex interaction between a number of factors. As a result, it would be difficult or impossible to predict “real-world” sensitivity based upon the results of studies employing direct stimulation of single electrodes using simple stimuli. In the present study, this question was addressed under conditions in which these various simultaneous factors were allowed to interact in a more natural way. It is important to note that, although they produced a modest range of word recognition scores, the CI users employed in the present study were all relatively good users, and so the current results may be restricted to representing this population.

In Experiment 1 it was found that intensity DLs for spectro-temporally complex stimuli presented in soundfield were larger on average for the CI users than for the NH listeners. This sensitivity for CI users is a result of tradeoffs between fine sensitivity to increments in electrical current and limited dynamic range, combined with differential sensitivity across electrodes within individuals and differences in the number of discriminable steps across users. These tradeoffs together yielded a real-world intensity sensitivity that was poorer than that of NH by a factor of 2.4 in decibel units.

Further, while the NH subjects produced larger intensity DLs for the stimulus word that had an amplitude envelope lacking the abrupt steplike changes characteristic of the other two words, the CI subjects produced DLs that were more uniform. These data therefore suggest that CI users may be less able than NH listeners to exploit abruptness of change in envelope to detect intensity increments. Although gated stimulus presentations have been found to produce larger intensity DLs than conditions in which increments were applied to continuous or semi-continuous pedestals (Wojtczak et al., 2003), this difference was restricted to Clarion users and was not observed in users of the Nucleus device. However, because intensity discrimination in the present study required comparison across a silent interstimulus interval, and because the preceding and following “fringe” syllables provided no cue, the current conditions can be considered variations of gated conditions. It may be worthwhile to examine systematically the influence of abruptness of change on intensity discrimination by different listener groups, by employing simple stimuli and gated- or continuous-stimulus conditions in which increments have systematically varying rise/fall durations.

Discrimination of F0 by the NH listeners in the present study is in accord with classic work employing isolated vowels having nonstationary F0 (Klatt, 1973). Values produced by the CI subjects are roughly one order of magnitude larger and are in accord with recent estimates of rate or FM discrimination (e.g., Zeng, 2002; Chen and Zeng, 2004). Under normal listening conditions, F0 may be encoded not only by temporal cues, but also by changes in the spectral balance of activity across the electrodes as the signal increases in F0 and harmonics move up in frequency. Geurts and Wouters (2001) showed that this “average amplitude cue” could indeed be used by some CI subjects to discriminate F0 in conditions where only spectral balance cues existed. However,
the difference between NH and CI listeners in the present study is approximately the same as that seen when place cues are removed from NH listeners. This result serves to highlight the heavy reliance on temporal cues and the especially poor spectral resolution of modern CIs by suggesting a relative lack of contribution by spectral balance (place) cues to F0 discrimination even when current balance across electrodes is allowed to vary in normal fashion.

Under conditions that were relatively similar to the F0 conditions of the present study, Geurts and Wouters (2001) obtained a mean F0 DL of approximately 8%, compared to the value of roughly 22% in the present study. Although there could be influences of CI processor coding not accounted for in the direct-stimulation technique of the previous study, it is also possible that the values observed here were affected by the natural durations of the speech stimuli and more accurately reflect the real-world abilities of CI users, than do the long-duration stimuli employed in the previous study.

There were differences, sometimes quite large, between the F0 DLs produced by the CI listeners across the three stimulus words. The presence of a higher DL for Word 2 ("sufficient") is especially clear in the three CI subjects who tended to produce the lowest DLs (CI2, CI3, and CI4). The three center syllables differ (by design) along a number of dimensions, including overall bandwidth, position of formants, and amount of spectral change within the syllable. Although Word 2 does have the smallest amount of spectral change within the center syllable, the changes that occur within the other two words are small compared to the F0 DL of the CI listeners. There are also differences in the intensities of the middle syllables relative to the preceding and following syllables, and differences in the depth of modulations when extracted using identical procedures (see Fig. 1).

The word that produced the highest F0 DL was also that having the briefest middle-syllable duration at 70 ms. One possibility is that by equating the slow-response rms peak levels of the three words, the instantaneous peak of the brief middle syllable in Word 2 was nearer the top of the input dynamic range of the CI device, and so the temporal fluctuations were compressed to some degree not present in the other two words. However, intensity DLs were not larger for this word, indicating that any limiting of the brief middle syllable that did occur was not severe enough to disrupt intensity discrimination.

Another possibility is that accurate F0 information was simply less available to the CI users because of the brief duration of the target segment. The results of other studies have not shown this dependence of F0 discrimination on duration. Discrimination by NH subjects is relatively consistent across the range of 70 to 150 ms, even when the stimulus is a complex tone composed of unresolved harmonics (White and Plack, 2003). There is one relevant early study involving two CI subjects (Tong et al., 1982). These authors evaluated effects of duration on rate pitch for stimuli that varied linearly in pulse rate during the course of the stimulus pulse train. Results were mixed, with some tendency for performance to improve with increasing duration, but only for higher initial pulse rates. Thus, although substantial differences exist between the present investigation and the study by Tong et al., those early data would not allow the prediction of the large F0 DLs observed here for Word 2. The cause of these differences in F0 DLs across words may warrant further investigation.

While the present study provides some indication of the sources of the observed sensitivity differences between listener groups, its main goals involved an examination of the magnitude of these differences. However, to address the general question involving influences of the stimuli versus influences of the CI device, control conditions were prepared to compare sensitivity to increments in speech to those in simple stimuli. A complex tone having a F0 of 125 Hz, a bandwidth up to 6000 Hz, and a duration of 150 ms was synthesized and incrementing intensity and F0 series were prepared as before. The F0, spectral range, and duration were selected to resemble those of the middle syllable of Word 1, and the complex tone was presented at an intensity matching that of the middle syllable of Word 1. Therefore, differences between the stimuli primarily involved the presence in Word 1 of surrounding syllables, spectral shaping, a F0 that changed (±8 Hz) over time, and a nonconstant amplitude envelope.

Difference limens for Word 1 and the complex tone were collected from two additional NH listeners and three CI users using the procedures employed earlier. These data are presented in Table IV, along with the proportion of the speech DL produced by the nonspeech stimuli. It was found that the NH listeners experienced a reduction in F0 DLs when simple stimuli replaced the speech stimuli (mean DL for complex tone was on average 0.51 that for speech), but the CI users

<table>
<thead>
<tr>
<th>S#</th>
<th>Word 1</th>
<th>Complex Tone</th>
<th>Proportion</th>
<th>Word 1</th>
<th>Complex Tone</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH 1</td>
<td>2.0 (0.2)</td>
<td>1.2 (0.2)</td>
<td><strong>0.60</strong></td>
<td>0.9 (0.2)</td>
<td>0.8 (0.1)</td>
<td><strong>0.89</strong></td>
</tr>
<tr>
<td>NH 2</td>
<td>2.4 (0.7)</td>
<td>1.0 (0.1)</td>
<td><strong>0.42</strong></td>
<td>0.8 (0.1)</td>
<td>0.5 (0.1)</td>
<td><strong>0.63</strong></td>
</tr>
<tr>
<td>CI 1</td>
<td>22.0 (4.5)</td>
<td>22.4 (0.9)</td>
<td><strong>1.02</strong></td>
<td>3.4 (1.1)</td>
<td>3.1 (1.5)</td>
<td><strong>0.91</strong></td>
</tr>
<tr>
<td>CI 2</td>
<td>11.9 (2.6)</td>
<td>12.0 (3.5)</td>
<td><strong>1.01</strong></td>
<td>2.9 (1.3)</td>
<td>3.1 (0.9)</td>
<td><strong>1.07</strong></td>
</tr>
<tr>
<td>CI 3</td>
<td>24.7 (1.0)</td>
<td>23.0 (2.1)</td>
<td><strong>0.93</strong></td>
<td>3.3 (0.8)</td>
<td>3.5 (1.3)</td>
<td><strong>1.06</strong></td>
</tr>
</tbody>
</table>

The word that produced the highest F0 DL was also that having the briefest middle-syllable duration at 70 ms. One possibility is that by equating the slow-response rms peak levels of the three words, the instantaneous peak of the brief middle syllable in Word 2 was nearer the top of the input dynamic range of the CI device, and so the temporal fluctuations were compressed to some degree not present in the other two words. However, intensity DLs were not larger for this word, indicating that any limiting of the brief middle syllable that did occur was not severe enough to disrupt intensity discrimination.
experienced no such improvement (mean DL for complex tone was on average 0.99 that for speech). Similarly, intensity DLs were reduced, especially for one of the NH listeners, with the change to simple stimuli (mean DL for complex tone was on average 0.76 that for speech), but the CI users experienced no improvement (mean DL for complex tone 1.01 that for speech). These results, although preliminary, indicate that the sensitivity differences observed among listener groups in the present study may be related more to factors associated with the implant, rather than dynamics of the stimuli. However, it should be noted that identifying in detail the relative contributions of the various influences that combine to produce real-world sensitivity may not be a simple matter.

In Experiment 2, concurrent intensity and F0 increments were presented within single words. It was found that, on average, CI users and NH listeners were equally able to combine concurrent cues to improve sensitivity. This combination of two cues served to improve sensitivity to half that of either cue in isolation for both groups. This result indicates that CI users should not be especially hindered by the presence of concurrent cues. Further, these results may potentially serve to simplify our understanding of CI sensitivity to cues that change in more complex fashion. It is clear that sensitivity to some cues in isolation is poorer than that of NH. Because it was found that CI users in the current study were able to exploit concurrent cues in similar fashion to the NH listeners, it is suggested that overall sensitivity to stimuli having concurrent changes in more than one cue may potentially be estimated based simply upon sensitivity to the individual cues.

The CI listeners demonstrate larger variability than their NH counterparts including, as Table III shows, in the combined-cue conditions of Experiment 2 where the stimuli were customized relative to individual sensitivity. It may be potentially interesting to examine psychometric functions relating d' to increment size for individual cues and listeners in an attempt to examine in more detail the cue-combining mechanism and to predict combined-cue sensitivity enhancement for individual subjects. Particularly large or small combined-cue DLs may potentially be explained by differences in the slopes of the psychometric functions for the two contributing cues, together with differences in the relative contributions of the two cues.

It is not uncommon for basic psychoacoustic abilities to relate only weakly to speech reception in hearing-impaired listeners, and a lack of correspondence has been seen previously for CI users (Cazals et al., 1990; Busby and Clark, 1999). However, the lack of a significant relationship between intensity or F0 DLs and word recognition in the present study may be considered somewhat surprising given the nature of the stimuli employed and the fact that some correspondence for CI users has been observed previously (Cazals et al., 1991, 1994; Muchnik et al., 1994; Dorman et al., 1996; Hanekom and Shannon, 1998; Fu, 2002). In contrast to the lack of relationship in Experiment 1, a significant relationship was observed between individual differences in the ability to combine intensity and F0 cues, and word recognition. Although, on average, CI users combine cues to improve sensitivity with the same success as NH listeners, individual differences within the CI group in this ability predicted performance on word recognition. Because the combined-cue stimuli in Experiment 2 were tailored for each CI listener to match their sensitivity to the individual cues in Experiment 1, these data indicate that it is the differential ability to combine cues, separate from the sensitivity to the cues themselves, that correlates to speech reception. Because this pattern of absence versus presence of correlation was observed within identical groups of subjects, using similar stimuli, identical word recognition scores, and identical procedures, the usual difficulty encountered when attempting to interpret results of different studies is avoided. These data therefore suggest, at least for relatively good CI users, that the complex processing required to perform the multidimensional task of combining concurrent cues may more closely predict performance on speech recognition than does simple psychoacoustic sensitivity.

The cues manipulated in the present study are those that signify lexical stress in English. Thus, the present study may have implications for the perception of prosodic cues by CI users. Although the relative weights of the acoustic cues are a matter of some debate (cf. Howell, 1993; Cutler, 2005), sensitivity to intensity and F0 was found to be poorer by factors of 2.4 and 8, respectively. Because the relative sensitivity enhancement resulting from concurrent cues was found to be equal on average across listener groups, these sensitivities may possibly be used to estimate the relative availability of cues to lexical stress in modern CI users in natural settings.

V. CONCLUSIONS

(1) Sensitivity to basic acoustic cues in NH and CI listeners under natural listening conditions may be estimated using stimuli having characteristics that approximate those of normal speech, and presentation methods that allow the interaction of multiple simultaneous factors.

(2) The factors governing intensity discrimination in modern CI devices combine to produce sensitivity in natural settings that is poorer than that of NH by a factor of 2.4 in decibels. This mean value represents relatively good users.

(3) Unlike NH listeners, the CI users did not produce smaller intensity DLs for words having abrupt changes in amplitude envelope.

(4) Fundamental frequency DLs for these relatively good CI users were poorer than those of NH listeners by nearly one order of magnitude, which indicates the relative lack of availability of spectral balance cues and the heavy reliance on temporal cues in CI listening.

(5) The CI subjects produced significantly larger F0 DLs for one of the three stimulus words (Word 2), which happens to have the briefest syllable duration. However, this effect was not observed in intensity DLs.

(6) The sensitivity enhancement resulting from combining concurrent cues was similar across NH and CI listeners. Thresholds were reduced, on average, to half those of either cue alone.
The relation between sensitivity to the individual cues and performance on a standard word recognition test was weak for the CI users. However, individual differences in the ability to combine concurrent cues to improve sensitivity were significantly related to word recognition.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC05795 and by a grant from the University of South Carolina Office of Research. Portions were drawn from a dissertation submitted by the first author to the Graduate School at the University of South Carolina. The authors thank Wendy Potts for audiovisual assistance and Sid Bacon for helpful comments. We are also grateful for the constructive comments of two anonymous reviewers and the Associate Editor on a previous version of this manuscript.

1One of the NH listeners (NH 4) had slightly to moderately elevated thresholds at the highest audiometric frequencies in one ear. The decision to include this individual was based upon the unilateral nature of the threshold elevation and the fact that the elevation was restricted to frequencies above that of the stimulus. His DLs were found to be the lowest of the group in almost every condition.

2All differences observed here using planned comparison are robust enough to also be significant when assessed using ANOVA.

3To investigate whether the high F0 DLs associated with Word 2 may have obscured the relationship between word recognition and F0 sensitivity, correlations were performed between word recognition and F0 DLs for each of the words individually. However, none of these relationships reached statistical significance (p > 0.05).


