The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners

Belinda A. Henry a) and Christopher W. Turner
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

(Received 5 April 2002; accepted for publication 26 January 2003)

The differences in spectral shape resolution abilities among cochlear implant (CI) listeners, and between CI and normal-hearing (NH) listeners, when listening with the same number of channels (12), was investigated. In addition, the effect of the number of channels on spectral shape resolution was examined. The stimuli were rippled noise signals with various ripple frequency-spacings. An adaptive 4IFC procedure was used to determine the threshold for resolvable ripple spacing, which was the spacing at which an interchange in peak and valley positions could be discriminated. The results showed poorer spectral shape resolution in CI compared to NH listeners (average thresholds of approximately 3000 and 400 Hz, respectively), and wide variability among CI listeners (range of approximately 800 to 8000 Hz). There was a significant relationship between spectral shape resolution and vowel recognition. The spectral shape resolution thresholds of NH listeners increased as the number of channels increased from 1 to 16, while the CI listeners showed a performance plateau at 4–6 channels, which is consistent with previous results using speech recognition measures. These results indicate that this test may provide a measure of CI performance which is time efficient and non-linguistic, and therefore, if verified, may provide a useful contribution to the prediction of speech perception in adults and children who use CIs. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1561900]

PACS numbers: 43.71.Ky, 43.71.Es, 43.66.Fe, 43.66.Ts [KRK]

I. INTRODUCTION

While many cochlear implant (CI) users achieve high levels of open-set speech recognition with current speech processors, there is a wide range in performance, with some CI listeners relying on the use of lipreading cues in order to communicate (e.g., Skinner et al., 1994). This significant variability in speech understanding is thought to be a product of variations among implantees in factors such as the number and function of surviving spiral ganglion cells, the placement of the electrodes within the scala tympani, patterns of current distribution within the cochlea, and the status of the central auditory system. Identifying variables in postlinguistically deafened adults that are presumably related to these factors, including the duration of deafness, the duration of CI use, etiology of deafness, and preoperative sentence scores, allows some level of prediction of postoperative speech perception ability (Battmer et al., 1995; Blamey et al., 1992, 1996; Dorman et al., 1990; Gantz et al., 1993; Rubinstein et al., 1999). However, in order to design improved speech processing strategies, and to optimize these strategies for individual CI listeners, it is necessary to develop an understanding of the perceptual mechanisms employed by CI listeners of different performance levels in recognizing speech.

The resolution of complex spectral patterns is one important aspect in perceiving speech. The ability to resolve the spectral peaks associated with the first two or three formants of speech is of primary importance in the identification of vowels, as well as other phonemes. In multichannel CIs, the speech processor typically resolves the frequency components in the signal using bandpass filtering, and the spectral shape information in the speech signal is then represented in the pattern of stimulation across the electrode array. The spectral resolution abilities of CI listeners depend first on the ability of the CI system to provide the spectral detail in the signal, which will be limited by the finite number of discrete channels provided in the particular CI system, and also on the way in which the acoustic signal is processed and represented in the electrical stimulation. Second, spectral resolution will depend on the ability of the individual CI listener to perceive this electrical representation of the spectral information.

The number of stimulating channels in current CI systems is generally limited to between 6 and 22, depending on the device and speech processing strategy, and therefore these speech processors do not preserve the fine spectral detail in the speech signal. It may seem logical to assume that optimizing the transmission of fine spectral detail by maximizing the number of channels may lead to improved speech perception. However, studies of speech recognition in normal-hearing (NH) listeners through acoustic simulations of CI processing indicate that high levels of speech recognition do not require a finely detailed spectral representation of the speech signal (Dorman et al., 1997; Friesen et al., 2001; Shannon et al., 1995; Turner et al., 1995). Rather, these studies have shown that high levels of speech understanding can be achieved in quiet listening conditions with 4 to 12 spectral bands, depending on the speech materials used (a larger number of channels is generally required for high levels of performance when more difficult speech materials are used.

aCurrent affiliation: Department of Communicative Disorders, University of Wisconsin–Madison, Madison, WI 53706. Electronic mail: bahenry@wisc.edu.

such as those comprised of multiple talkers and/or those with low levels of linguistic context). However, when listening in background noise, NH listeners utilize more channels to recognize speech, with performance continuing to increase to at least 16 to 20 channels, depending on the level of the noise (Dorman et al., 1998; Friesen et al., 2001).

While most current CI speech processors therefore provide a sufficient number of channels to represent the spectral detail in the speech signal, at least in quiet listening conditions, most CI listeners are unable to fully utilize this spectral information. The results of speech recognition testing have indicated that spectral resolution is one major factor limiting CI speech perception. Vowel recognition, for example, which as noted above relies primarily on the perception of spectral cues, has been shown in many studies to be highly variable among CI listeners. Ranges in vowel recognition performance of 14% to 79%, reported by van Wieringen and Wouters (1999) for 25 users of the Laura CI system with the Continuous Interleaved Sampling (CIS) strategy, and 36% to 100%, reported by Skinner et al. (1994) for 59 users of the Nucleus CI22M implant and SPEAK strategy, are typical. In addition, while speech perception performance in CI listeners improves as the number of channels is increased, there is an asymptote in performance at four to seven channels, regardless of whether testing is performed in quiet or noisy listening conditions (Dorman and Loizou, 1997, 1998; Dorman et al., 1997; Fishman et al., 1997; Friesen et al., 2001), which is in contrast to NH listeners (see above). These results provide further evidence that CI listeners generally cannot utilize all of the spectral information that is provided by CI speech processors and multiple electrode arrays. Furthermore, while the better-performing CI users can achieve performance levels similar to those of NH listeners when listening with the same number of channels, there is a wide range of abilities across CI listeners, and there is some evidence that poorer-performing CI listeners may not be able to utilize as many channels as better-performing CI listeners (Friesen et al., 2001).

One factor which may limit spectral resolution with a CI is the degree of spectral contrast (the difference in amplitude between the peak and valley). For CI listeners, the wide electrical dynamic range must be compressed into the small electrical dynamic range (typically 3–30 dB), resulting in a reduction in spectral contrast. While NH listeners only require a spectral contrast of between 1 and 3 dB for reasonably accurate vowel identification (Leek et al., 1987; Alcantara and Moore, 1995; Turner and Van Tassell, 1984), Loizou and Poroy (2001) found that the minimal spectral contrast required for high levels of vowel identification accuracy was 4–6 dB for CI listeners.

Psychophysical studies which have investigated the ability to discriminate between stimulation on different electrodes also indicate that the perception of spectral cues is a performance-limiting factor in CI speech perception (Collins et al., 1997; Donaldson and Nelson, 2000; Henry et al., 2000; Nelson et al., 1995; Throckmorton and Collins, 1999). Henry et al. (2000) showed a correlation between speech perception and electrode discrimination in the frequency regions below approximately 2.7 kHz when the stimuli in the electrode discrimination task were presented with randomly varying levels, but no correlation between the two measures in the frequency regions above 2.7 kHz, indicating that fine spectral discrimination may be relatively more important in the vowel-formant regions than in the high-frequency regions, where the cues are mostly broadband. The relationship between the two measures indicates the usefulness of electrode discrimination measures as a method of predicting speech perception in CI users. However, these measures are time consuming to conduct since they are multi-point measures, which are made at multiple positions along the electrode array.

Electrode discrimination measures can only provide an indirect assessment of frequency discrimination ability for acoustic signals. The ability of CI listeners to resolve peaks in the acoustic signal, when listening through their speech processors, is not currently well understood. The present study was therefore designed to address the following specific questions:

1. How closely can peaks in the acoustic spectrum be spaced and still be resolved by CI users of varying performance levels, and does this differ between CI and NH listeners, when listening with the same number of channels?
2. Is the ability to resolve spectral peaks related to speech recognition with a CI?
3. How does the ability to resolve spectral peaks vary with the number of processing channels in CI and NH listeners?

Spectral shape resolution was investigated in this study using a method based on the “ripple phase reversal test” (e.g., Supin et al., 1994, 1999), which was originally developed to measure the resolution of spectral patterns in NH listeners. Using this method, the threshold for resolvable spectral peak spacing is determined using rippled noise stimuli, which are broadband noise signals with spectral ripples. The listeners’ task is to discriminate between two rippled noise stimuli in which the frequency positions of the peaks and valleys are reversed. Ripple density is varied and the threshold is taken as the highest or maximum ripple density at which an interchange in the peak and valley positions in the rippled spectrum (i.e., a ripple phase reversal) can be detected. It is hypothesized that a listener can only discriminate the spectral ripple phase reversal if the rippled spectrum pattern can be resolved. If the spectral pattern is not resolvable due to the ripples being spaced too closely, the reversal cannot be detected. This test is hypothesized to provide a one-point measure of the listeners’ ability to perceive the frequency locations of the peaks in a generic speech-like acoustic signal.

In experiment 1, a modified version of the ripple phase reversal test was used to investigate differences in spectral ripple resolution thresholds among CI listeners having a range of speech perception abilities, when listening to signals through a 12-channel CIS speech processor. In order to estimate the upper limit of spectral shape resolution provided by CI speech processors, spectral ripple resolution was also measured in NH subjects, who listened to noise-band acous-
tic simulations of 12-channel CI processing. It is hypothesized that since the identification of vowels relies primarily on the ability to resolve and identify the frequencies of spectral peaks, measuring the ability to perceive the frequency locations of the peaks in a complex acoustic spectrum using the spectral ripple resolution test may show a relationship with vowel recognition. Therefore, the relationship between spectral peak resolution and vowel identification ability was also examined. In experiment 2, the effect on spectral ripple resolution of limiting the spectral detail that is presented in the acoustic signal, by varying the number of processing channels, was investigated both in CI and NH listeners. The purpose of the second experiment was to investigate the number of effective channels utilized by NH and CI listeners in resolving spectral peaks, and to further investigate the degree of similarity between results obtained using the spectral peak resolution test and those obtained using speech recognition tasks.

II. EXPERIMENT 1: SPECTRAL RIPPLE RESOLUTION AND VOWEL RECOGNITION

A. Methods

1. Subjects

Eight NH subjects (six female, two male), ranging in age from 19 to 39 years, participated in this experiment. Normal hearing was defined as having pure-tone air conduction thresholds ≤15 dB HL at octave frequencies from 125 to 8000 Hz in the tested ear. Twenty-one postlinguistically deafened adults (14 female, 7 male) with CIs also participated. All were users of the Cochlear Corp. CI24M implant and the SPrint speech processor, and had a wide range of speech perception abilities. The individual subject details are shown in Table I. All participants were native American English speakers.

2. Stimuli

a. Rippled noise stimuli. Rippled noise stimuli of 500-ms duration and with peak-to-valley ratios of approximately 30 dB were created from white noise. The white noise was recorded in 16-bit digital form at a 44.1 kHz sampling rate from a white noise generator onto a Macintosh G4 computer. The technique of adding a white noise to a copy of itself with a delay of $T$ seconds, where the reciprocal of the delay gives the frequency of the rippled noise, was employed using Sound Designer II (DigiDesign, Audio Media III) software. The resulting spectrum has peaks linearly spaced at 1/T Hz. The first peak in the spectrum of the noise occurs at 0 Hz when the delayed version of the noise is added to the original in phase (standard stimulus). When the polarity of the delayed noise is reversed, the first peak occurs at 0.5/T Hz (inverted stimulus).

The frequency spacings of the ripples were 8000, 6000, 4000, 3000, 2000, 1500, 1000, 500, 250, 125, 64, 32, and 16 Hz. Examples of rippled noise spectra, both standard and inverted, with ripple frequency spacings of 8000, 2000, and 500 Hz are shown in Fig. 1. The rippled noise stimuli were then processed in two different ways, as described in the signal processing section below, depending on whether they were to be presented to the CI or the NH listeners.

b. Vowel stimuli. Vowel recognition was measured using a closed-set 12-alternative identification procedure. Medial vowel tokens were presented in a /h/-vowel-/d/ context (had, hod, hawed, head, hayed, heard, hid, heed, hoed, hood, hud, who’d). The tokens were digitized natural productions from

TABLE I. Individual subject details. The number of maxima is 8 for the ACE strategy and 6 for the SPEAK strategy unless noted otherwise.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (years)</th>
<th>Duration of profound deafness (years)</th>
<th>CI experience (years)</th>
<th>Etiology</th>
<th>CI Etiology</th>
<th>Age (years)</th>
<th>Duration of profound deafness (years)</th>
<th>CI experience (years)</th>
<th>Etiology</th>
<th>CI Etiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI1</td>
<td>73</td>
<td>12</td>
<td>5</td>
<td>Infection</td>
<td>ACE</td>
<td>72</td>
<td>22</td>
<td>83.3</td>
<td>9.5</td>
<td>54.2</td>
</tr>
<tr>
<td>CI2</td>
<td>46</td>
<td>13</td>
<td>2.5</td>
<td>Congenital, prog</td>
<td>ACE</td>
<td>1800</td>
<td>22</td>
<td>47.3</td>
<td>13.0</td>
<td>37.9</td>
</tr>
<tr>
<td>CI3</td>
<td>62</td>
<td>4</td>
<td>4</td>
<td>Unknown</td>
<td>ACE</td>
<td>700</td>
<td>19</td>
<td>71</td>
<td>9.7</td>
<td>62.1</td>
</tr>
<tr>
<td>CI4</td>
<td>37</td>
<td>2</td>
<td>2.5</td>
<td>Autoimmune</td>
<td>SPEAK</td>
<td>250</td>
<td>19</td>
<td>73.3</td>
<td>12.0</td>
<td>52.5</td>
</tr>
<tr>
<td>CI5</td>
<td>72</td>
<td>8</td>
<td>3</td>
<td>Congenital, prog</td>
<td>SPEAK</td>
<td>250</td>
<td>20</td>
<td>85.3</td>
<td>11.3</td>
<td>43.3</td>
</tr>
<tr>
<td>CI6</td>
<td>72</td>
<td>1</td>
<td>1</td>
<td>Meniere’s disease</td>
<td>ACE</td>
<td>1200</td>
<td>22</td>
<td>71.3</td>
<td>8.6</td>
<td>26.3</td>
</tr>
<tr>
<td>CI7</td>
<td>51</td>
<td>2</td>
<td>2.5</td>
<td>Autoimmune</td>
<td>SPEAK</td>
<td>250</td>
<td>20</td>
<td>48.7</td>
<td>5.5</td>
<td>28.8</td>
</tr>
<tr>
<td>CI8</td>
<td>41</td>
<td>5</td>
<td>1</td>
<td>Congenital, prog</td>
<td>ACE</td>
<td>1800</td>
<td>22</td>
<td>76</td>
<td>17.9</td>
<td>32.1</td>
</tr>
<tr>
<td>CI9</td>
<td>74</td>
<td>25</td>
<td>4</td>
<td>Congenital, prog</td>
<td>ACE, 12</td>
<td>72</td>
<td>22</td>
<td>81.3</td>
<td>10.6</td>
<td>30.4</td>
</tr>
<tr>
<td>CI10</td>
<td>76</td>
<td>40</td>
<td>2</td>
<td>Unknown</td>
<td>CIS</td>
<td>900</td>
<td>6</td>
<td>NA</td>
<td>5.5</td>
<td>23.3</td>
</tr>
<tr>
<td>CI11</td>
<td>40</td>
<td>18</td>
<td>3</td>
<td>Autoimmune</td>
<td>SPEAK</td>
<td>250</td>
<td>20</td>
<td>83.3</td>
<td>8.6</td>
<td>55.8</td>
</tr>
<tr>
<td>CI12</td>
<td>48</td>
<td>2</td>
<td>1</td>
<td>Unknown</td>
<td>CIS</td>
<td>2400</td>
<td>6</td>
<td>69.3</td>
<td>5.6</td>
<td>29.2</td>
</tr>
<tr>
<td>CI13</td>
<td>54</td>
<td>5</td>
<td>3</td>
<td>Unknown</td>
<td>ACE, 10</td>
<td>1200</td>
<td>22</td>
<td>30.7</td>
<td>6.3</td>
<td>9.2</td>
</tr>
<tr>
<td>CI14</td>
<td>80</td>
<td>3</td>
<td>3</td>
<td>Unknown</td>
<td>SPEAK</td>
<td>250</td>
<td>19</td>
<td>62</td>
<td>3.2</td>
<td>20.4</td>
</tr>
<tr>
<td>CI15</td>
<td>37</td>
<td>2</td>
<td>1</td>
<td>Unknown</td>
<td>ACE</td>
<td>720</td>
<td>22</td>
<td>49.3</td>
<td>10.9</td>
<td>45.4</td>
</tr>
<tr>
<td>CI16</td>
<td>48</td>
<td>2</td>
<td>3</td>
<td>Ototoxicity</td>
<td>ACE</td>
<td>1800</td>
<td>22</td>
<td>82.7</td>
<td>5.3</td>
<td>45.8</td>
</tr>
<tr>
<td>CI17</td>
<td>78</td>
<td>56</td>
<td>5</td>
<td>Viral infection</td>
<td>SPEAK</td>
<td>1800</td>
<td>22</td>
<td>71.3</td>
<td>7.4</td>
<td>45.4</td>
</tr>
<tr>
<td>CI18</td>
<td>46</td>
<td>7</td>
<td>3.5</td>
<td>Unknown</td>
<td>ACE</td>
<td>20</td>
<td>900</td>
<td>54.7</td>
<td>15.7</td>
<td>30.8</td>
</tr>
<tr>
<td>CI19</td>
<td>78</td>
<td>4</td>
<td>5</td>
<td>Unknown</td>
<td>ACE</td>
<td>22</td>
<td>900</td>
<td>71.3</td>
<td>7.2</td>
<td>44.2</td>
</tr>
</tbody>
</table>


B. A. Henry and C. W. Turner: Resolution of complex spectral patterns

2863
ten male and ten female talkers, taken from the materials recorded by Hillenbrand et al. (1995). The vowel stimuli were stored in a digital form on a Macintosh G4 computer, after being transferred from CD.

3. Signal processing

a. Cochlear implant subjects. The final processing of the rippled noise stimuli to be presented to the CI listeners was accomplished using the Sound Designer II program. The rippled noise signals were first filtered with a filter that approximated the long-term speech spectrum (Byrne et al., 1994). This low-pass filtering resulted in a distribution of energy across the entire frequency range for the processed signals (since the width of the filters in the CI processor increase with frequency, the energy in the processed signals would be concentrated in the high frequencies if the signals were not low-pass filtered). The overall levels of the rippled noise sound files were then approximately equalized, and linear rise/fall times of 150-ms applied to each stimulus to avoid click effects and other idiosyncratic cues which were shown to be perceptible by NH listeners in a pilot study.

The stimuli were presented to the CI subjects using the Nucleus CI24M implant system and the SPrint speech processor. The CI24M electrode array has 22 intracochlear and 2 extracochlear electrodes. All CI subjects used a pulse width of 25 µs, and stimulus pulses were presented using the monopolar MP1+2 stimulation mode, where current flows between the active intracochlear electrode and both extracochlear electrodes. The CIS speech processing strategy was used to present both the rippled noise and vowel stimuli, with 12 channels, a Q-value of 20, and a pulse rate of 900 pulses per second (pps)/channel, and therefore a total stimulation rate across electrodes of 10800 pps. In the 12-channel CIS strategy implemented in the SPrint processor, the amplitude envelope is estimated within each of the 12 bands during each stimulation period. These amplitudes are converted to electrical stimulus levels, and stimulus pulses representing each band are presented sequentially on the associated electrodes, in a basal to apical order in this experiment. The default frequency to channel allocation for 12-channel CIS [as set by the Cochlear Diagnostic and Programming System (DPS) software] was used, and this provided a frequency range of 187–7937 Hz, and the filter cutoff frequencies shown in Table II. The post-processing spectra for the rippled noise stimuli with peak spacings of 8000, 2000, and 500 Hz (from Fig. 1) are shown in Fig. 2. These channel output levels were measured using the SCILAB (Swiss Cochlear Implant Laboratory software) program, which records the RF transmissions from the speech processor and provides either a graphical display or a list of the electrodes activated and the current levels. The speech processor settings were the same as used in the experiment (i.e., 12-channel CIS, 900 Hz pulse rate), with the electrical stimulation levels set to the average across CI subjects [threshold level (T)=150, comfortably loud level (C)=200]. Note the reduction in spectral contrast from 30 dB in the acoustic signal to between 2 and 5 dB for the processed spectra.

b. Normal-hearing subjects. The NH subjects were presented with both rippled noise and vowel stimuli processed using a 12-channel simulation of CI processing. The acoustic simulation of electrical hearing was accomplished using multiple channels of narrow-band noise modulated by the signal, using MATLAB software. The original rippled noise and vowel signals were band-pass filtered into 12 frequency bands using filters whose cutoff frequencies (see Table II) and frequency response slopes approximated those used in the CI speech processor. Within each band, the signal was transformed into broadband signal correlated noise by randomly assigning negative or positive signs, with a probability of 0.5, to the sampled waveform points (Schroeder, 1968; Shannon et al., 1995; Turner et al., 1995). This resulted in a white noise carrier signal whose amplitude was modulated by the original speech waveform envelope. This broadband noise was then refiltered to spectrally limit the signal to the bandwidth corresponding to the originally filtered signal bands, and the overall amplitudes of each band were corrected to correspond to those of the original stimulus bands. The output of all of the bands was then recombined. The processed rippled noise stimuli underwent a final filtering with a filter that approximated the long-term speech spectrum, an equalizing of the overall levels of the processed signals, and an application of linear rise/fall times of 150 ms, which was identical to that of the rippled noise stimuli to be presented to the CI subjects.

4. Procedures

a. Spectral ripple resolution. Each CI subject was programmed with an experimental speech processor program (map), which differed in all cases from the normal maps used by the subjects (Table I). The 12 active electrodes to be in-

FIG. 1. Rippled noise spectra. Ripple frequency spacings of 8000, 2000, and 500 Hz, with both standard and inverted peak positions are shown.
cluded in each subject’s speech processor program (see Table II) were determined by the default setting in the DPS software. In the cases where a map contained deactivated electrodes which corresponded to the electrodes selected for the experimental map, the next closest available electrode was included instead. T and C levels were measured for each electrode, using standard clinical procedures. Apart from these subject-dependent settings, it was important that all other speech processing parameters be constant across subjects. Therefore, the same speech processor, with the same sensitivity setting (a value of 8), was used for all subjects. Since the audibility of conversational-level signals is controlled mostly by the sensitivity setting of the speech processor and does not vary significantly among CI users, providing the T and C levels are set optimally, a constant sensitivity setting minimizes individual variation in the audibility of the signals. An optimal combination of signal level and sensitivity setting was selected, by measuring the output of the signals. An optimal combination of signal level and sensitivity setting is controlled mostly by the sensitivity setting of the speech processor and does not vary significantly among CI users, providing the T and C levels are set optimally, a constant sensitivity setting minimizes individual variation in the audibility of the signals. An optimal combination of signal level and sensitivity setting was selected, by measuring the output of the speech processor using the SCILAB program, so that the signal peaks resulted in stimulation at approximately 90% of the dynamic range.

The rippled noise stimuli were output via custom software routines using a 16-bit digital-to-analog converter (DigiDesign, Audio Media III), which has a built-in anti-aliasing filter at 20 kHz. For the CI listeners, the stimuli were presented via a direct input into the speech processor, using a 3.5-mm mono phone plug, which defeats the headset microphone so that no external input occurs, while at the same time preserving the microphone pre-emphasis. The stimuli were presented to the NH subjects monaurally using Sennheiser HD 25-SP1 circumaural headphones at an average level of 65 dB SPL. Participants were tested using the same ear throughout the experiment.

A four-interval forced-choice (4IFC) procedure, in which three intervals contained the standard (reference) rippled noise stimulus, while the other interval, chosen at random, contained the inverted (test) stimulus, was used to measure the discrimination abilities of the listeners. The stimuli were separated by silent intervals of 500 ms. The level of the stimuli was varied randomly for each interval within an 8-dB range in 1-dB steps using a Tucker-Davis programmable attenuator (model PATT), in order to minimize the ability of subjects to use loudness cues for identification. Subjects were asked to select one of four buttons labeled numerically on a touch-screen, corresponding to the “different” interval, ignoring any loudness differences between intervals. Feedback was provided. Practice testing was provided at a range of ripple frequency spacings. Following this, psychometric functions were obtained for 18 of the CI listeners (CI1 to CI18) and for 4 of the NH listeners. Percent...
correct scores for the psychometric functions were based on the average of between one and four runs (or sets) of 20 trials of the 4IFC task. In general, scores of 95% or 100% are based on one run (i.e., only one run of the task was administered), scores of less than 40% are based on two runs, and scores between these values (i.e., those near or on the steepest portion of the function) are based on three to four runs of the 4IFC task. Detailed psychometric functions were obtained for all subjects except CI19, C20, and C121, who were recruited to participate in experiment 2. A threshold for ripple resolution was estimated from the psychometric functions at the 70.7% correct level for each subject. To determine this threshold, the data points in the 20% to 100% range were examined, and a straight line fitted to those four data points distributed around the 70.7% level. Each run of 20 trials of the 4IFC task took approximately 5 min to complete.

An adaptive procedure was also used to estimate ripple resolution thresholds for all CI and NH subjects. Thresholds were obtained using a two-down, one-up 4IFC procedure, converging on the 70.7% correct point (Levitt, 1971). The 4IFC task was the same as that detailed above. An experimental run proceeded until 13 reversals were obtained, and the threshold for the run was taken as the mean of the final 8 reversals. All 13 ripple frequency spacings were included in the adaptive procedure, a run of which commenced at the 6000-Hz ripple frequency spacing. Several practice runs (between four and six) were completed initially for each subject.

The final threshold values for each subject were taken as the average of the final five to six adaptive runs.

Using this adaptive procedure, spectral ripple resolution was assessed in two conditions; a broadband condition (0–8000 Hz) and a 3000-Hz low-pass filtered condition. A Kemo filter (VBF8.04) was used, with filter slopes of 30 dB/oct. The reason for this was to compare spectral shape resolution ability over the entire speech frequency range with that in the frequency region which is most relevant to vowel perception, and had previously been shown to be the frequency region in which electrode discrimination and speech recognition were correlated (Henry et al., 2000). All subjects were tested in a double-walled sound treated room.

b. Vowel recognition. The vowel stimuli were output via custom software routines using a 16-bit digital-to-analog converter (see above). These stimuli were presented to the CI subjects in the free-field, positioned approximately 1 m from a speaker, at an average level of 65 dB SPL, and to NH subjects monaurally through Sennheiser HD 25-SP1 circumaural headphones at an average level of 65 dB SPL. All CI and NH subjects participated in the vowel recognition test.

Initial training was provided for the vowel identification task. Each of the 12 words containing the medial vowels was displayed as buttons on a touch-screen. The subject was asked to select the word they wanted to hear, and five examples of that word (i.e., the word was spoken by five different talkers) were then presented. Fifty trials of the training task were conducted (or more as desired by the individual subject) in order to allow the subject to familiarize themselves with the tokens and their labels.

Following the vowel training, two runs (a practice and a test run) of the vowel identification test were administered. A single run of the test consisted of 240 trials (12 vowels × 20 talkers). On each trial, a stimulus token was chosen randomly, without replacement, and following presentation of each token the subject responded by pressing 1 of the 12 buttons on the touch-screen. Feedback was provided during the practice run, but not during the test run. Each run of the vowel test took approximately 25 min to complete.

B. Results

1. Spectral ripple resolution

The psychometric functions representing ripple frequency spacing versus percent correct score are shown for four of the NH listeners in Fig. 3, and for the CI listeners (CI1 to CI18) in Fig. 4. Also shown are the estimates of the 70.7% correct thresholds derived from the psychometric functions. As the spacing between spectral peaks decreases, there is a reasonably monotonic decrease in the ability to discriminate between the standard and inverted stimuli. The consistently monotonic nature of these functions supports the use of an adaptive procedure as a more time-efficient method of estimating ripple resolution thresholds.

The thresholds determined using the adaptive procedure for the NH and CI listeners, and for both the broad-band and low-pass conditions, are shown in Fig. 5. The mean spectral ripple resolution threshold for the NH subjects was 391 Hz, in the low-pass condition. The range was 309 to 555 Hz. The
performance of CI14, CI15, and CI20 on the ripple resolution task was too poor to perform the adaptive procedure, since their thresholds (around 8000 Hz) were above the starting point for the adaptive procedure. Thus, the ripple resolution thresholds shown in Fig. 5 for these listeners were derived from their psychometric functions. The mean spectral ripple resolution threshold for the CI listeners as a group was 2977 Hz in the low-pass condition. Thus, spectral ripple resolution ability was substantially poorer in the CI than the NH subjects. In addition, spectral ripple resolution abilities varied substantially among the CI group, with a range of 802 to approximately 8000 Hz. Thresholds at or near 8000 Hz (CI14, CI15, and CI20) imply very limited spectral shape resolution abilities.

Paired t-tests showed no significant differences between the spectral ripple resolution thresholds obtained in the broadband condition compared to the low-pass condition for the CI listeners (broadband mean = 3093 Hz, low-pass mean = 2977 Hz; p = 0.078). While the difference between the two conditions was mildly significant for the NH listeners (broadband mean = 418 Hz, low-pass mean = 391 Hz; p = 0.038), the mean difference is small and is therefore not of practical importance.

The accuracy of the thresholds determined using the adaptive procedure compared to those derived from the psychometric functions was investigated using regression analysis. Figure 6 shows the adaptively determined threshold plotted against the threshold derived from the psychometric function for each subject. The slope of the regression line was 0.97, which, being very close to 1, and along with the significant correlation between the two measures (r = 0.98, p < 0.0001), indicates a high degree of similarity between the results from the two threshold-estimating methods.

The possible effect of learning through the course of the experiment on the results was assessed by comparing the threshold obtained in the broadband condition at the start compared to the end of the experiment. This analysis is based on data for all CI subjects except CI14, CI15, and CI20, whose thresholds were too high to perform the adaptive procedure, as mentioned above. The results of a paired t-test indicate no significant difference between the first and last runs for either the CI (mean start threshold = 2587 Hz, mean end threshold = 2488 Hz; p = 0.316) or the NH (mean start threshold = 490 Hz, mean end threshold = 508 Hz; p = 0.361) subjects. This indicates that the adaptive ripple discrimination task does not suffer from a lengthy training period to reach stable performance.

The post-processing rippled noise spectra in Fig. 2 show a significant reduction in spectral contrast at the output of the speech processor, which is due to the compression of the acoustic dynamic range into the small electrical dynamic range of CI listeners. It may be hypothesized that dynamic range may contribute to the variability in ripple resolution thresholds among CI listeners. However, the results of a regression analysis between the average dynamic range across electrodes (shown in Table I) and ripple resolution threshold did not indicate a significant relationship between these two measures (r = −0.17, p = 0.47).

III. EXPERIMENT 2: THE EFFECT OF THE NUMBER OF CHANNELS ON SPECTRAL RIPPLE RESOLUTION

In the second experiment, the effect of varying the number of processing channels on spectral ripple resolution in both CI listeners and NH listeners was investigated. Previous studies (see the Introduction) employing speech recognition measures have shown an asymptote in performance at between around 4 to 7 channels for CI listeners, compared to around 4 to 12 channels for NH listeners (in quiet), depending on the speech material used. The purpose of this experiment was to further investigate the degree of similarity between results obtained using the spectral ripple resolution test and those obtained using speech recognition tasks.
A. Methods

1. Subjects

A different group of eight NH adult subjects (seven female, one male, age range 21–22 years) participated in experiment 2. Normal hearing was defined as in experiment 1, and all NH participants were native American English speakers. Eleven of the CI subjects from experiment 1 (CI1, CI2, CI3, CI5, CI6, CI9, CI13, CI17, CI18, CI20, and CI21) also participated in experiment 2.

2. Stimuli

The effect of the number of channels on spectral ripple resolution was assessed using the same rippled noise stimuli as in experiment 1. The signal processing is described below.
3. Signal processing

a. Cochlear implant subjects. The SPrint processor and CIS strategy was again used in experiment 2, and the number of channels was varied using the Cochlear DPS software. Spectral ripple resolution was assessed in seven different conditions, which were with 1, 2, 4, 6, 8, 12, and 16 channels. The filter cutoff frequencies, which are shown in Table II, were determined by the default settings used in the CIS strategy for each of these conditions. As the number of frequency bands was varied, the total bandwidth of 187–7937 Hz remained the same, and the frequency range associated with each electrode increased. The pulse rate per channel was held constant at 900 pps as the number of channels was varied. The electrodes activated for each of the channels are shown in Table II. If a selected electrode was unavailable for activation in an individual’s map, the next closest available electrode was used instead. If testing was conducted on the same day as experiment 1, each subject used the same T and C levels as in experiment 1 for all conditions (and the extra 4 channels were added to the map for the 16-channel condition). If the testing for experiment 2 was conducted during a separate test session, the T and C levels were checked prior to commencing the experiment.

b. Normal-hearing subjects. Spectral shape resolution was assessed in the NH subjects in the same conditions as for the CI subjects. The rippled noise stimuli were processed into the various number-of-channels acoustic CI simulations using the same general methods as those described previously for experiment 1. The cutoff frequencies of the filters used in the simulations approximated those used in the CI speech processor for each condition (see Table II).

4. Procedures

The stimuli were presented in the broadband condition (0 to 8000 Hz). The 12-channel condition was administered...
B. Results

The average of four runs of the adaptive procedure remaining conditions were presented in a randomized order. to assess any learning effects over the course of the experiment. The remaining conditions were presented in a randomized order. The average of four runs of the adaptive procedure (see experiment 1) was taken as the ripple resolution threshold for each condition for each listener.

Figure 8 shows the average spectral ripple resolution thresholds as a function of the number of processing channels for NH and CI listeners. CI subjects: circles; NH subjects: triangles. Error bars represent ± one standard deviation. The downward-pointing arrow indicates that no threshold could be obtained in the one-channel condition.

as the first and last test conditions, in order to assess any learning effects over the course of the experiment. The remaining conditions were presented in a randomized order. The average of four runs of the adaptive procedure (see experiment 1) was taken as the ripple resolution threshold for each condition for each listener.

Both NH and CI subjects were unable to resolve any peaks using a one-channel processor, as indicated by the downward-pointing arrow in Fig. 8. Therefore, the results for this trivial condition could not be included in the statistical analyses. In addition, the data of CI20 were excluded from the statistical analyses, since this subject performed too poorly in all conditions to complete the adaptive task. A mixed model ANOVA, with number of channels (2 to 16) and subject group (NH versus CI) as factors, showed main effects of both the number of channels (p < 0.0001) and of the subject group (p = 0.004), as well as a channel by group interaction (p = 0.004). To further analyze the channel by group interaction, separate repeated measures ANOVAs were conducted on the two subject groups. For the NH subject group, a repeated measures ANOVA showed a significant effect of number of channels (p < 0.0001). Post hoc tests (Sidak) showed statistically significant increases in performance as the number of channels was increased to 16. In Fig. 8, a significant difference between one condition and the next highest number-of-channels condition is represented by an asterisk symbol above the NH curve. All differences were shown to be significant in post hoc tests, except the difference in ripple resolution thresholds between the 8- and 12-channel conditions. For the CI subject group, a repeated measures ANOVA also showed a significant effect of number of channels (p < 0.0001). Post hoc tests (Sidak) indicated a statistically significant difference between the two-channel condition and all other conditions (as shown by the asterisk symbol below the CI curve in Fig. 8), but no other statistically significant differences between conditions. However, the difference between the four- and six-channel condition was close to significance (p = 0.06). These results indicate a plateau in performance at between four and six channels for the CI subject group.

Figure 9 shows the variation in thresholds with number of channels across the individual CI subjects (as well as the average thresholds for the NH group, as shown in Fig. 8). There was substantial variability among CI subjects in the effect of the number of channels on spectral ripple resolution ability. For example, while more closely spaced peaks could be resolved by CI21 as the number of channels increased to 8, increasing the number of channels did not result in an increased ability to resolve spectral peaks for CI20.

The possible effects of learning over the course of the experiment were assessed by comparing the average threshold values from the 12-channel condition at the start to those at the end of the experiment. The results of paired t-tests for both the NH (mean start threshold = 614 Hz, mean end threshold = 666 Hz; p = 0.142) and the CI (mean start threshold = 2106 Hz, mean end threshold = 2105 Hz; p = 0.495) subjects do not suggest any influence of learning on the experimental results.

---

**FIG. 8.** Average spectral ripple resolution thresholds as a function of the number of processing channels for NH and CI listeners. CI subjects: circles; NH subjects: triangles. Error bars represent ± one standard deviation. The downward-pointing arrow indicates that no threshold could be obtained in the one-channel condition.

**FIG. 9.** Spectral ripple resolution thresholds as a function of the number of processing channels for the NH group and for individual CI listeners. Vowel scores are shown in the legend for individual CI subjects. Error bars represent ± one standard deviation. The downward-pointing arrow indicates that no threshold could be obtained in the one-channel condition.
IV. DISCUSSION

The average spectral ripple resolution threshold for the CI group compared to the NH group differed by nearly an order of magnitude when listening with 12 spectral channels (ripple frequency-spacing of approximately 3000 Hz compared to approximately 400 Hz, experiment 1). While acoustic simulations of CI processing cannot exactly replicate CI signals, since, for example, they cannot simulate the effects of current distribution patterns in the cochlea, they do allow some understanding of perceptual abilities with CI-like signal processing. Thus, to the extent that CI simulations provide an estimate of optimal performance, these results indicate that the spectral shape resolution abilities of most CI listeners are considerably poorer than those of NH listeners when listening with the same number of channels. Performance on the spectral ripple resolution task varied widely among CI listeners, with the best-performing CI subject able to achieve a level of performance (threshold of 800 Hz) close to the range of the NH subjects (300 to 600 Hz), while at the other end of the performance range three subjects showed thresholds of approximately 8000 Hz.

It is interesting that as the spectral detail provided in the signal was increased, by increasing the number of processing channels from 1 to 16, the NH listeners were able to use the additional spectral detail to resolve more closely spaced peaks, showing significant increases in performance to the maximum number of channels tested (16 channels). In contrast, while the CI listeners as a group showed significant increases in performance as the number of channels was increased to between four and six channels, as the number of channels was increased further, this increased spectral detail was not able to be used by these listeners to resolve more closely spaced peaks. Furthermore, the number of effective channels utilized by CI subjects to resolve spectral peaks varied widely, from one to eight channels (Fig. 9). These findings are significant in two respects. First, they provide additional evidence that CI listeners are not able to take advantage of the spectral detail provided in the processed signal. Second, these results are generally consistent with the results of previous studies which have investigated the effect of the number of channels on speech recognition (e.g., Fishman et al., 1997; Friesen et al., 2001). As such, they indicate that patterns of performance obtained using the spectral ripple resolution test are similar to those obtained using speech recognition tests, the potential implications of which are discussed below.

In CIs and CI simulations, acoustic spectral shapes are represented by the relative amplitude across channels. Therefore, in order to discriminate between spectral shapes, such as between different formants, or in the specific case of the present study, between the standard and inverted stimuli, listeners must be able to detect and discriminate between relative changes in amplitude across channels. The generally poorer ability of CI listeners to utilize the spectral detail provided in the processed signal is not well understood, but may be due to a lower specificity of neural populations activated in electrical stimulation, due to factors related to the patterns of neural survival and function, and patterns of current distribution in the cochlea. An inability of CI users to discriminate electrodes, which is an ability which has previously been shown to be related to speech recognition ability (see the Introduction), and to detect and discriminate changes in level between electrodes, may therefore be associated with a “blurring” of the spectral peaks in acoustic signals. In addition, the reduced spectral contrast at the output of CI processors, due to the compression of the acoustic dynamic range into the narrow electrical dynamic range, may also affect the ability to resolve the spectral peaks in acoustic signals. Indeed, Loizou and Poray (2001) found that CI listeners required 4–6 dB of spectral contrast for maximal vowel perception, in comparison to 1–3 dB for NH listeners (see the Introduction). While it might be hypothesized therefore that spectral shape resolution may be related to the dynamic range of an individual, studies on the effect of dynamic range on vowel recognition in quiet listening conditions have shown mixed results (Zeng and Galvin, 1999; Loizou et al., 2000). In this study, there was no apparent relation between spectral peak perception and dynamic range (at least when expressed as an average value across frequency). Clearly, further research regarding the effects of the dynamic range of electrical stimulation on the ability to resolve the spectral peaks in acoustic signals is required.

We hypothesized that a listeners’ ability to perceive the frequency locations of the peaks in a generic speech-like acoustic signal (as measured in the spectral ripple resolution task in this study) may be related to vowel recognition. Indeed, there was a significant relationship between ripple resolution thresholds and vowel recognition for CI listeners, with the $r^2$ value of 0.42 indicating that 42% of the variability in vowel scores is accounted for by the regression of vowel score and spectral ripple resolution threshold. These results indicate that those listeners who are better able to determine the positions of the spectral peaks in the acoustic signal, as shown by lower ripple resolution thresholds, are, on average, more readily able to extract vowel information from the signal. While further research is required to investigate the potential clinical applications of these results, it seems possible that the adaptive spectral ripple resolution test may provide a time-efficient and non-linguistic measure which may contribute to the prediction of performance in both adult and child CI users. Should this test be verified, and extended to other speech perception measures, it may be potentially applicable in a clinical setting in guiding the selection of re/habilitation strategies, and selecting and optimizing speech processing strategies for individual CI users, and also in improving the predictive power of models which use variables such as duration of deafness and preoperative sentence scores (e.g., Rubinstein et al., 1999) to describe the variance in CI speech recognition, via the inclusion spectral ripple resolution thresholds as an additional factor.

Finally, while some of the variance in vowel scores is accounted for by the regression of vowel score on spectral ripple resolution threshold, more than half of the variance remains unaccounted for. As discussed in the Introduction, there are many factors, both peripheral and central, which may contribute to the variance in speech perception among CI users, and some of these may be unrelated to the resolution of spectral ripples. For example, one factor which may
have contributed more to the variance in speech perception scores than the variance in ripple resolution scores is the age of the participants, which varied widely from 37 to 80 years in this study. While it does seem highly likely that other factors account for the remaining variance in speech recognition scores, it may also be possible to optimize the rippled noise stimuli to improve the predictive power of this test. For instance, the effects measured in this study might depend to some extent on whether a linear or logarithmic scale is used for the ripple frequency spacing. While ripple spacing on a logarithmic scale would reflect the properties of the normal auditory system, the filters in the CIS strategy implemented in this study use a combination of linear and logarithmic division of the frequency scale (see Table II), and therefore linear spacing of ripples was chosen to provide an initial estimate of ripple resolution abilities in the present study.

V. CONCLUSIONS

The main findings in this study were the following.

(1) There was wide variation in the ability of CI listeners to resolve spectral shapes in the acoustic signal, with spectral ripple resolution thresholds ranging from 800 to 8000 Hz.

(2) Spectral shape resolution was poorer in CI users than NH listeners when listening with the same number of channels (12). While the average ripple resolution threshold for NH listeners was 400 Hz, the average for CI listeners was 3000 Hz.

(3) NH listeners were able to utilize more channels to resolve spectral peaks than CI listeners. Spectral shape resolution increased with the number of channels to 16 for the NH group, while the CI listeners showed a performance plateau at 4–6 channels, which is consistent with previous studies of the effect of the number of channels on speech recognition.

(4) There was a significant correlation between spectral shape resolution and vowel recognition for the CI listeners.

ACKNOWLEDGMENTS

We are grateful to the subjects for their considerable time and effort in participating in this research. We wish to thank Keith Kluender, Fan-Gang Zeng, and two anonymous reviewers for their comments on an earlier version of this manuscript. Thanks also to James Hillenbrand for use of the multtalker vowel test materials, to Norbert Diller and Waikong Lai for use of the SCILAB program, and also to Arik Wald, Bomjun Kwon, Kristine Melis, Anne Torkelson, Courtney Burke, and Jenny Klein for their assistance. Funding for this research was provided in part by NIDCD Grant No. 1R01DC00377.


