

Holes in Hearing

ROBERT V. SHANNON,^{1,2} JOHN J. GALVIN III,¹ AND DENIZ BASKENT²

¹House Ear Institute, Los Angeles, CA 90057, USA

²Department of Biomedical Engineering, University of Southern Californiam, Los Angeles, CA 90089, USA

Received: 20 April 2001; Accepted: 17 September 2001; Online publication: 10 December 2001

ABSTRACT

Previous experiments have demonstrated that the correct tonotopic representation of spectral information is important for speech recognition. However, in prosthetic devices, such as hearing aids and cochlear implants, there may be a frequency/place mismatch due in part to the signal processing of the device and in part to the pathology that caused the hearing loss. Local regions of damaged neurons may create a “hole” in the tonotopic representation of spectral information, further distorting the frequency-to-place mapping. The present experiment was performed to quantitatively assess the impact of spectral holes on speech recognition. Speech was processed by a 20-band processor: SPEAK for cochlear implant (CI) listeners, and a 20-band noise processor for normal-hearing (NH) listeners. Holes in the tonotopic representation (from 1.5 to 6 mm in extent) were created by eliminating electrodes or noise carrier bands in the basal, middle, or apical regions of the cochlea. Vowel, consonant, and sentence recognition were measured as a function of the location and size of the hole. In addition, the spectral information that would normally be represented in the hole region was either: (1) dropped, (2) assigned to the apical side of the hole, (3) assigned to the basal side of the hole, or (4) split evenly to both sides of the hole. In general, speech features that are highly dependent on spectral cues (consonant place, vowel identity) were more affected by the presence of tonotopic holes than temporal features (consonant voicing and manner). Holes in the apical region were more damaging than holes in the basal or middle regions. A similar pattern of

performance was observed for NH and CI listeners, suggesting that the loss of spectral information was the primary cause of the effects. The Speech Intelligibility Index was able to account for both NH and CI listeners' results. No significant differences were observed among the four conditions that redistributed the spectral information around the hole, suggesting that rerouting spectral information around a hole was no better than simply dropping it.

Keywords: cochlear implant, deafness, hearing aids, hearing impairment, hearing loss, speech recognition

INTRODUCTION

In the normal-hearing ear there is a well-defined mapping of spectral information onto specific locations in the cochlea, with high frequencies processed near the base and low frequencies processed near the apex. Acoustic spectral information is parsed into its frequency components mechanically by the passive and active frequency-selective vibration of the basilar membrane. These tonotopically mapped vibrations are converted into neural activation by the hair cells of the cochlea. However, in the hearing-impaired ear there may be a loss of hair cells and/or auditory nerve fibers as a result of the hearing loss pathology, which can alter this normal frequency-to-place mapping.

One type of alteration in the spectral–tonotopic mapping that commonly occurs in cochlear implants is an absolute frequency–place shift, in which spectral information is presented to an electrode that is more basally located than the normal acoustic place for that spectral information. The effect of tonotopic shifting on speech recognition was recently measured in NH

Correspondence to: Robert V. Shannon, Ph.D. • House Ear Institute • 2100 W. Third St. • Los Angeles, CA 90057. Telephone: (213) 353-7020; fax: (213) 413-0950; email: shannon@hei.org

and CI listeners; both NH and CI listeners could tolerate a shift of ± 3 mm, equivalent to a $\pm 60\%$ frequency shift relative to the matched cochlear location with little effect on performance (Dorman et al. 1997; Fu and Shannon 1999). Shifts larger than 3 mm resulted in a sharp decrease in speech recognition.

Another type of alteration in the tonotopic pattern could result from localized loss of hair cells and/or auditory neurons, producing a "hole" in the tonotopic representation. This localized loss would produce an elevated sound threshold at a particular frequency in a hearing-impaired ear or an elevated electrical threshold for one or more electrodes in an implanted ear. The normal clinical solution to such localized elevated thresholds is to increase the gain of the device for that spectral region. However, if there is a complete loss of hair cells and/or neurons in that tonotopic region, increasing the signal level will simply cause the activation to spread to neighboring regions where hair cells and/or neurons are still intact. In this case the spectral information from that band will be received but will activate neurons in a different tonotopic region than intended, producing a local warping in the tonotopic representation of spectral information.

Another potential distortion in cochlear implants could produce a warping of the spectral representation of speech in the cochlea relative to the acoustic representation. For example, the frequency-place mapping would be warped if evenly spaced electrodes activate unequal regions of neurons due to uneven nerve survival or nonuniform current flow. Shannon et al. (1998) measured speech recognition with four-band noise processors under conditions that caused a warping of the entire spectral-to-tonotopic mapping. Spectral information was analyzed into four logarithmically spaced frequency bands; the envelope information from those bands was presented as modulations of four linearly spaced noise bands, causing a logarithmic warping of the normal acoustic frequency-place mapping. Results showed a dramatic decrease in performance to a level equivalent to that produced by a single-channel processor, i.e., the loss of all spectral information. Although four bands of spectral information were still extracted and presented, listeners were unable to make use of any of the spectral information. While the results of that study demonstrated the devastating effects of warping the entire spectral-to-tonotopic mapping, the present study investigates the effects of more localized warping in the tonotopic representation.

As an example of a localized warping in the tonotopic pattern, consider a listener who has no remaining hair cells or neurons in the cochlear region that is normally tuned to 2–4 kHz but who has normal hearing at higher and lower frequencies. This listener's audiogram would show elevated thresholds in the

2–4-kHz region, with maximum elevation toward the lower-frequency end. If a hearing aid is fitted to this patient with 50 dB of gain in the 2–4-kHz range, the patient will be able to perceive sound at those frequencies. However, such amplification will cause information in the 2–4-kHz region to spread and activate hair cells and neurons in the tonotopic region above 4 kHz. Depending on the location and size of the hole, this kind of amplification may improve speech intelligibility. However, Turner et al. (1999) recently demonstrated that amplification of high frequencies could actually reduce speech intelligibility in some hearing-impaired listeners with 60 dB of hearing loss above 2 kHz. They speculated that even though amplification of this frequency region resulted in audibility of spectral information above 2 kHz, the resulting distortion in the frequency-place representation might have been detrimental. Thus, it is possible that amplification, in some cases, could actually reduce speech intelligibility.

Previous work on spectral holes

One of the earliest attempts to quantify the effect of lost spectral information was the development of the Articulation Index (AI). Harvey Fletcher and colleagues varied the cutoff frequencies of high-pass and low-pass filtered speech until each produced equal speech recognition (Fletcher and Steinberg 1929; French and Steinberg 1947). Equal speech intelligibility was found for speech either high-pass or low-pass filtered at about 1900 Hz, though lower cutoff frequency values have been reported more recently for more linguistically simple stimuli. The results suggested that half of the speech intelligibility was found above 1900 Hz and half below 1900 Hz. More detailed studies have since derived frequency-importance functions for various speech materials (e.g., Studebaker and Sherbecoe 1993) in which materials that are more linguistically simple have a crossover frequency lower than 1900 Hz. This basic methodology was updated recently (ANSI S3.5, 1997) and termed the Speech Intelligibility Index (SII).

The AI model integrates recognition probability across frequency bands, an operation that assumes that the recognition probability in each frequency band is independent of the recognition probability in all other bands. However, it is well known that much speech information is redundant and correlated across the frequency spectrum. Several recent studies have demonstrated the highly synergistic additivity of speech information from different spectral regions. Breeuwer and Plomp (1984, 1985, 1986) showed that reasonable speech recognition could be achieved with two octave

bands of frequency information, even when eliminating the midfrequency region of 1000–3000 Hz. Lippmann (1996) demonstrated good speech recognition even when the entire midfrequency region from 800 to 3000 Hz was removed. Warren et al. (1998) found that excellent speech recognition was possible when listening through a few narrow spectral bands (slits). Clearly, large regions of spectral information can be dropped without disrupting speech recognition because of the spectral redundancy of speech. It is not clear if the AI can account for the high levels of speech recognition observed in these studies when large portions of the midfrequency region of spectral information are missing.

The distribution of neural activity in the cochlea may be significantly altered when a region of the cochlea is missing hair cells and neurons. Moore and Glasberg (1997) proposed an excitation pattern model of loudness that could account for the loudness growth functions of hearing-impaired listeners who had a localized loss of inner hair cells, which they termed “dead regions.” Moore and colleagues proposed psychophysical methods for diagnosing these dead regions (Moore et al. 2000; Moore 2001; Moore and Alcantara 2001). Their results suggest that, for a signal located within a dead region, masking does not really occur inside the dead region but rather at the edges of that region where the surviving hair cells respond to both masker and signal.

Vickers et al. (2001) demonstrated that amplification of frequencies up to one octave above the estimated edge frequency of a high-frequency dead region could be beneficial for most subjects. Amplification of frequencies more than one octave higher than the edge of the dead region did not help and actually reduced performance in some listeners. Turner and colleagues demonstrated the potentially detrimental effects of amplification within dead regions for speech recognition (Hogan and Turner 1998; Turner and Cummings 1999). Standard amplification formulas actually led to reduced speech intelligibility in some hearing-impaired patients with more than 60 dB of high-frequency hearing loss. They speculated that the amplification in a dead region produced a spread of activation to healthy, unimpaired regions, resulting in a distortion in the frequency–place pattern. For some patients, this frequency–place distortion produced a significant drop in speech recognition compared with no amplification in the dead region. In this case, amplification and the resulting distortion apparently produced poorer speech recognition than if the speech information was lost altogether.

The hypothesis of the present study is that spectral warping of the tonotopic representation may result in a larger detriment to speech recognition than if the spectral information is simply lost. To most effectively

remediate a hearing loss, it is important to understand the tradeoff between the loss of spectral information that occurs with no amplification and the possible warping of the tonotopic representation caused by amplification. The present study attempts to quantify this tradeoff between the loss of information and distortion in the tonotopic representation.

METHOD

Subjects

Six listeners fitted with the Nucleus-22 cochlear implant and six normal-hearing listeners participated in the present experiment. All subjects were native speakers of American English. All implant users had 20 active electrodes available and had used the SPEAK speech processor for at least 4 years (see Table 1 for more information). Normal-hearing listeners ranged in age from 25 to 52 years and had thresholds better than 15 dB HL at audiometric frequencies between 125 and 8000 Hz.

Speech materials and testing procedure

Speech perception tests used to evaluate the experimental settings were all presented without lip-reading (sound only). The subjects’ tasks were medial vowel identification, medial consonant identification, and word recognition in sentences. Each listener was tested with the experimental speech processors immediately after receiving them (no practice). The listeners were seated 1 m in front of the loudspeaker in a sound-treated room (IAC) and the sound level was set by adjusting a 1000-Hz sinusoid to 70 dB rms on an A-weighted scale. Each speech stimulus was digitally scaled to have an rms level computed over the entire stimulus duration that was equal in level to the calibration sinusoid.

The Nucleus-22 implant device includes a microphone, an adjustable sensitivity setting, and an automatic gain control that determines the acoustic range of sound that is mapped to electrical current (Cochlear Corp. 1995). Because the wide acoustic range of sound must be mapped to the much smaller electrical dynamic range of the implant listener, some amount of signal compression is necessary. Depending on the interactions of these three components, portions of the acoustic signal may be peak- and/or center-clipped. Because the present experiment was concerned with the relative effects of spectrally limited speech, it was important to keep the acoustic input range to the speech processor constant. In this way, any information loss would be due to local holes in the spectral representation and could not be compensated

TABLE 1

Pertinent information about cochlear implant users who participated in the experiment, including baseline vowel, consonant, and sentence recognition scores, as well as average scores for the normal-hearing listeners

Subject	Age (yr)	Gender	CI Ear	Etiology	Insertion depth (rings out)	Duration of implant use (yr)	Frequency allocation table	CUNY (sound only) sentence (%C)	NU-6 words (%C)	Vowel baseline phoneme (%C)	Consonant baseline phoneme (%C)	TIMIT baseline sentence (% word correct)
N3	56	M	R	Trauma/ unknown	3	7	7	79.4	24	59.6	56.6	46.3
N4	40	M	R	Trauma	4	5	9	99.0	70	80.8	85.4	74.5
N7	55	M	R	Unknown	0	2	9	99.0	48	66.7	75.0	67.8
N9	55	F	L	Hereditary	4	7	9	100.0	60	72.1	80.4	64.9
N14	63	M	R	Unknown	0	1	9	47.1	18	34.6	35.4	11.5
N19	74	M	L	Noise- induced	6	8	7			70.8	72.6	63.0
Mean CI baseline performance with 20-electrode SPEAK processor										64.1	67.6	54.7
Mean NH baseline performance with 20-channel noise-band processor										90.6	95.1	97.3

The Nucleus-22 cochlear implant has 22 active stimulating electrodes and 10 passive electrodes that are present for mechanical stiffening, with spacing of 0.75 mm between each electrode ring. The surgeon estimates the insertion depth of the electrode array by counting the number of electrode rings outside the cochleostomy. Frequency tables are standard filter sets that determine the overall frequency range used and which frequencies are assigned to each electrode. Frequency Allocation Table 9 refers to the default table, which assigns the frequency range 150 Hz–10.8 kHz to the 20 electrode pairs; Frequency Allocation Table 7 uses the frequency range 120 Hz–8.7 kHz

for by applying more gain to the acoustic input. Thus, implant listeners were instructed to use the same volume sensitivity setting as their normal processor for all conditions and not to change that setting for the duration of the experiment. Similarly, for normal-hearing listeners, no normalization was applied to the speech signal after the multichannel noise-band processing.

Vowel stimuli were taken from materials recorded by Hillenbrand et al. (1994). Stimuli were presented to the listeners and responses collected by custom software (Robert 1997). Listeners heard 10 presentations (5 male and 5 female talkers) each of 12 medial vowels including 10 monophthongs (/i ɪ ε æ u ʊ α ɔ ɜ ʒ/) and 2 diphthongs (/o e/) presented in a /h/-vowel-/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). Chance level on this test was 8.33% correct with an upper 95% confidence level of 13.3%.

Consonant stimuli (3 male and 3 female talkers) were taken from materials recorded by Turner et al. (1999) and Fu et al. (1998). Consonant confusion matrices were compiled from 12 presentations (6 talkers times 2 repetitions of each talker) of each of 14 medial consonants /b d g p t k l m n f s j v z θ/, presented in an /a/-consonant-/a/ context. Tokens were presented in random order by custom software (Robert 1997) and the confusion matrices were analyzed for information received on the production-based categories of voicing, manner, and place of articulation (Miller and Nicely 1955). Chance performance level for this test was 7.14% correct with an upper 95% confidence level of 11.1%.

Recognition of words in sentences was measured using the Texas Instruments/Massachusetts Institute of Technology (DARPA/TIMIT) corpus of sentence materials (National Institute of Standards and Technology 1990). For each condition, data were collected from each listener for 20 sentences of varying lengths (from 3 to 13 words per sentence, with an average of 6.8 words per sentence). The sentences were of moderate-to-high difficulty, spoken by multiple talkers, and presented with no context or feedback on performance; no sentences were repeated to an individual listener. Each word in a sentence was scored as correct or incorrect and performance under a given condition was expressed as the percentage of correct words.

Signal processing: CI listeners

The Nucleus-22 SPEAK processing strategy divides speech into 20 contiguous frequency bands and assigns the output of each frequency analysis band to one electrode pair (Fig. 1). This frequency-to-electrode assignment is determined by selecting 1 of 15 frequency allocation tables that map various acoustic frequency ranges to electrode locations (Cochlear Corp. 1995). The clinically assigned frequency allocation table was used for the baseline experimental speech processor; four of the CI listeners used the default Table 9 (0.15–10.8-kHz frequency range) and two used Table 7 (0.2–8.6-kHz frequency range). The normal frequency-to-electrode assignments also produce a spectral shift between the lowest frequency analysis band and the most apical electrode location. In a fully inserted Nucleus-22 electrode, the most apical active

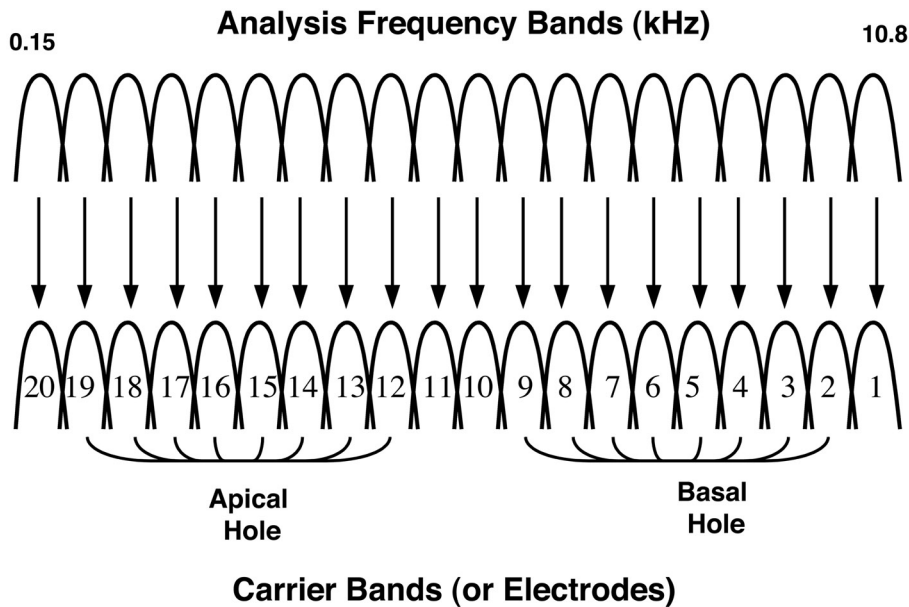


FIG. 1. Schematic representation of signal processing for all listeners. The spectrum from 150 Hz to 10.8 kHz was divided into 20 bands corresponding to equal mm distances along the basilar membrane. The output of each band after processing was used to modulate a biphasic pulse train on a single electrode for CI listeners or to modulate a band of noise for NH listeners. Holes in the spectral representation were created in the apical, middle, or basal regions (apical and basal shown) by turning off 2, 4, 6, or 8 output noise carrier bands or electrodes.

electrode is located approximately 25 mm from the round window. Using Greenwood's (1990) frequency-place formula and assuming a 35 mm cochlear length, this location should correspond to an acoustic frequency of 513 Hz. Thus, the cochlear location of the most apical electrode is more than 4.6 mm basal to the normal cochlear location corresponding to the lowest cutoff frequency of Table 9 (150 Hz) or Table 7 (200 Hz).

Experimental processors were created by eliminating stimulation to 2, 4, 6, or 8 electrodes along the array, resulting in tonotopic holes of 1.5, 3, 4.5, and 6 mm in extent; holes were created in the apical, middle, and basal regions of the cochlea. Four conditions were created in which the spectral information that would normally have been presented to those electrodes was (1) dropped, (2) assigned to the electrode on the apical edge of the hole ("apical"), (3) assigned to the electrode on the basal edge of the hole ("basal"), or (4) evenly split between the apical and basal edge of the hole ("split") (Fig. 2). For the dropped condition, the threshold (T-level) and comfortable listening level (C-level) for selected electrodes were set to the device minimum, making the electrical stimulation inaudible. For the three reassignment conditions, the output of selected frequency analysis channels was routed to nearby electrodes; the target electrode would then receive the electrical stimulation of both its own frequency analysis channel and the channels that would have normally been assigned to the electrodes in the hole.

The SPEAK processing strategy picks the six to ten frequency bands with the highest energy for stimulation approximately every 4 ms (McDermott 1989; Seligman and McDermott 1995). The electrodes assigned

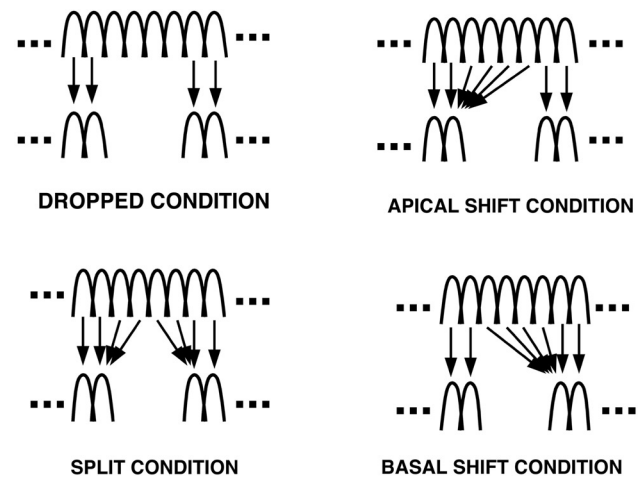


FIG. 2. Schematic representation of the four conditions for remapping spectral information from the hole. The envelope information from the filter bands that would have been assigned to the hole region was either dropped (upper left), presented to the apical edge of the hole (upper right), presented to the basal edge of the hole (lower right), or evenly split to the two edges of the hole.

to the selected bands receive stimulation pulses on that cycle. Because some electrodes were not stimulated in the present experiment, the frequency reassignment conditions sometimes resulted in changes in the per-electrode rate of stimulation. For the electrodes at the edges of the holes, the average pulse rate per electrode was sometimes higher for experimental processors that reassigned the analysis bands from within the hole to an edge electrode. In the "dropped" condition, the effective overall pulse rate was reduced because electrodes in the hole region were stimulated at an inaudible level. These changes in stimulation rate were all dependent on speech materials and the experimental

processor configurations, however, and it is difficult to assess the effects (if any) of variable stimulation rates for each experimental condition. [Fu and Shannon (2000) showed no effect on speech recognition of changes in stimulation rate for rates above 150 pps/electrode.]

Signal processing: NH listeners

For the acoustic processors, the speech signal was divided into 20 contiguous frequency bands by 6th-order Butterworth filters; the cutoff frequencies of the bands were selected to coincide with the filter divisions in frequency allocation Table 9 (0.15–10.8-kHz frequency range) for the SPEAK processor (Cochlear Corp. 1995). The speech envelope was extracted from each filter by half-wave rectification and low-pass filtering with a 3rd-order Butterworth filter (220 Hz cutoff frequency). In the “no hole” condition, the envelope from each band was used to modulate a broadband white noise, which was then filtered by the same 6th-order Butterworth filter used in the frequency analysis. To create holes in the noise-band speech, several output noise bands were turned off. In the dropped condition, the output noise bands were simply omitted from the processed signal. In the reassignment conditions, the envelope waveforms that would have normally been used to modulate the bands in the hole region were summed and added to the envelope waveform for the noise carrier band at the edge of the hole (Fig. 2). This summed modulation envelope was then divided by the number of contributing bands so that its amplitude would not be higher than, and potentially mask, surrounding bands. Finally, all modulated noise bands were summed and presented to the listeners through Tannoy Reveal loudspeakers (see Shannon et al. 1995 for more details on signal processing).

RESULTS

Figures 3, 4, and 5 present the results from the consonant, vowel, and sentence recognition tests, respectively. Results are presented as a difference in percent correct between the score for the experimental condition and the score with no hole, after correcting both scores for chance performance. While there was relatively little variability in scores across NH listeners, CI listeners had a wide range of scores in the no-hole condition (Table 1). Normalizing the results to the no-hole performance reduced the cross-listener variability but preserved the relative decrement in performance due to the hole. In Figures 3, 4, and 5, data from normal-hearing listeners are presented in the top three panels and implant results are presented in the

lower three panels. From left to right, each row presents the results from holes placed in the basal, middle, and apical regions of the cochlea, respectively. Within each panel, results from the four conditions that varied the way spectral information from the hole was distributed are represented by the four curves. In general, speech recognition decreased as the hole size increased and the decrease was larger for apical holes than for basal holes. Overall, there were few significant differences between the four conditions that distributed the spectral information around the hole. Average performance for normal-hearing listeners was generally higher than average implant performance (Table 1).

Consonant recognition (Fig. 3) generally decreased significantly as the size of the hole increased [$F_{(3,240)} = 197.22$, $p < 0.01$ for NH, $F_{(3,240)} = 36.09$, $p < 0.01$ for CI]. For NH listeners, *post hoc* Scheffe tests revealed that 4.5 mm holes in the apical, middle, and basal regions all caused performance to decrease significantly. For NH listeners there was no significant difference in recognition between the four conditions which distributed the spectral information from the hole in different ways (apical shift, basal shift, split, dropped) [$F_{(3,240)} = 1.26$, $p = 0.29$]. For CI listeners, *post hoc* Scheffe tests revealed that 4.5 mm holes in the apical and middle regions significantly worsened consonant recognition, while performance was not significantly affected in the basal region until the hole was 6 mm in extent. For CI listeners, reassignment was significant [$F_{(3,240)} = 4.63$, $p < 0.01$], but the *post hoc* Scheffe tests showed no significant differences between any pair of reassignment conditions. This is probably due to the fact that with large apical holes, one CI listener performed much more poorly than the others when the information was dropped (see Fig. 3, lower-right panel); when this listener was removed from the statistical analysis, there was no significant difference between reassignment conditions for the remaining CI listeners. For this CI listener all reassignment conditions did improve performance compared to when the information was dropped. However, the pattern of performance for this listener in the reassignment conditions was similar to that of other CI and NH listeners; only the performance in the dropped condition was deviant for this listener. Thus, it is probable that the data for this listener in the dropped condition are aberrant. NH listeners as a group had significantly higher overall performance on consonant recognition than CI listeners (Table 1), but, overall, there were no significant differences between NH and CI listeners once scores were normalized to the baseline performance. NH and CI listeners showed similar patterns of consonant recognition for basal and midcochlea holes, but CI listeners showed a smaller decrease in

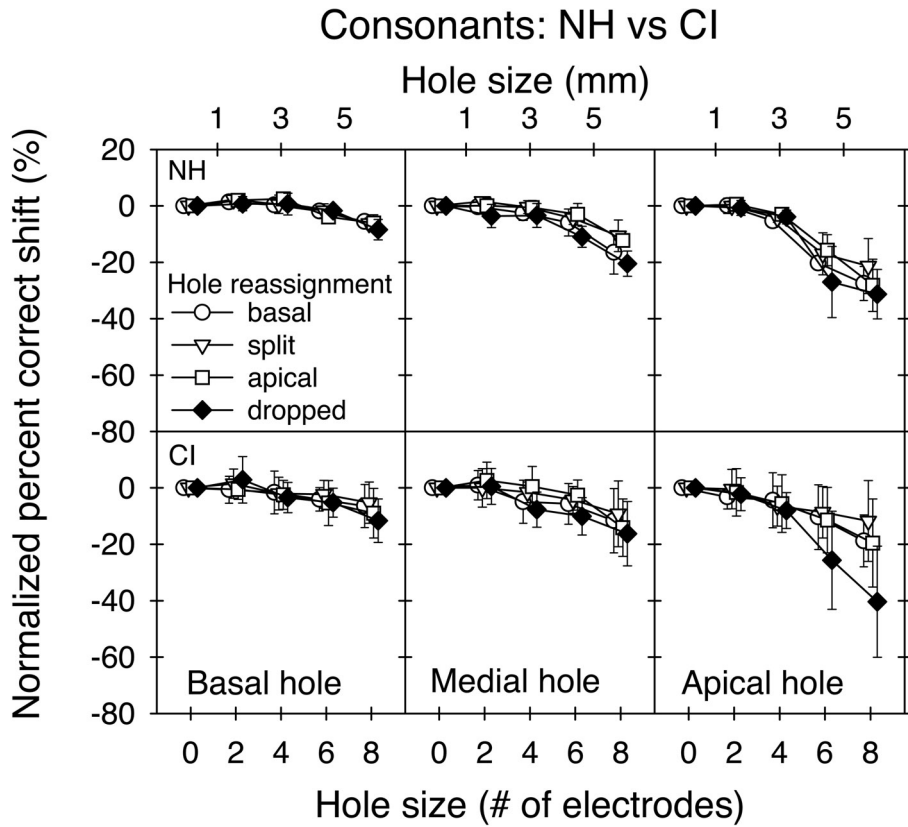


FIG. 3. Recognition of 14 consonants as a function of hole location and size for both NH and CI listeners. The four curves in each panel represent data from the four conditions that remapped the spectral information to the edges of the hole. The error bars show the standard deviation. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm. The y-axis of all panels shows the percent correct normalized to the no-hole condition and adjusted for chance.

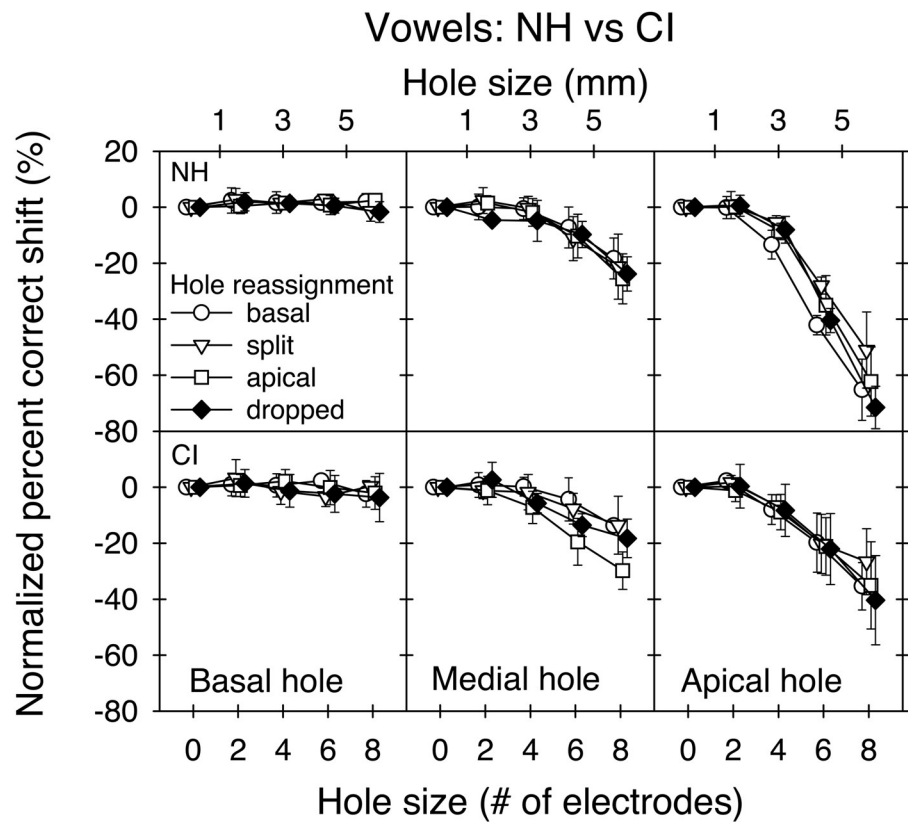


FIG. 4. Recognition of 12 vowels as a function of hole location and size for both NH and CI listeners. The four curves in each panel represent data from the four conditions that remapped the spectral information to the edges of the hole. The error bars show the standard deviation. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm. The y-axis of all panels shows the percent correct normalized to the no-hole condition and adjusted for chance.

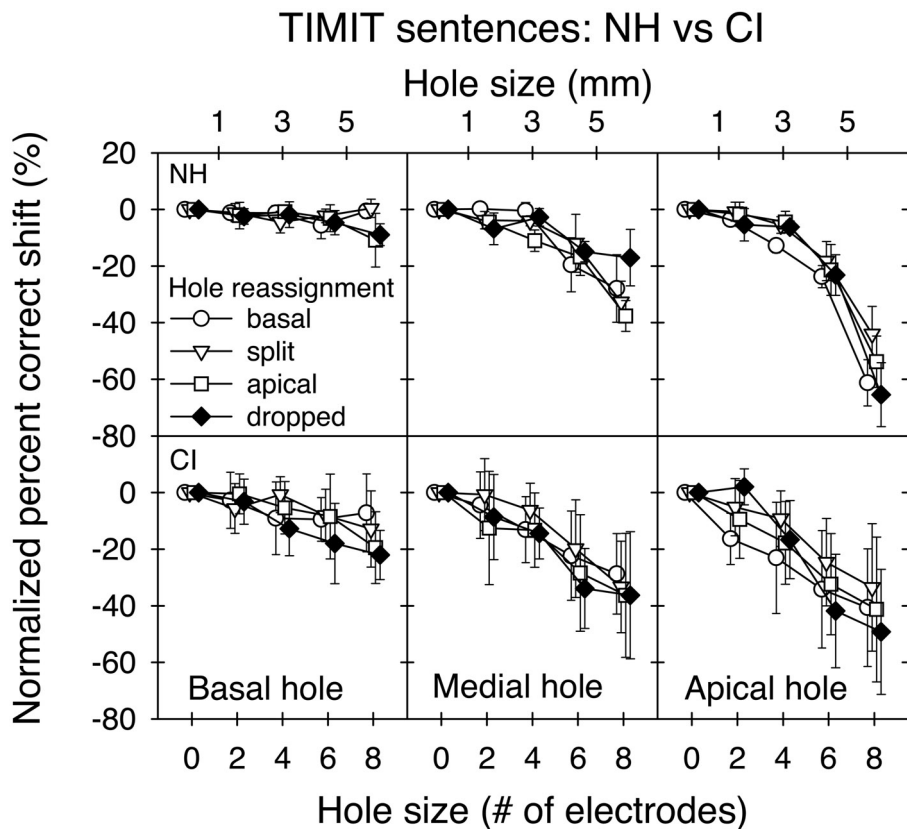


FIG. 5. Recognition of words in TIMIT sentences as a function of hole location and size for both NH and CI listeners. The four curves in each panel represent data from the four conditions that remapped the spectral information to the edges of the hole. The error bars show the standard deviation. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm. The y-axis of all panels shows the percent correct normalized to the no-hole condition and adjusted for chance.

consonant recognition than NH listeners for large apical holes.

Vowel recognition (Fig. 4) fell as the hole became wider [$F_{(3,240)} = 343.69$, $p < 0.01$ for NH, $F_{(3,240)} = 100.36$, $p < 0.01$ for CI], and there was a significant interaction between hole size and hole location [$F_{(3,240)} = 122.55$, $p < 0.01$ for NH, $F_{(3,240)} = 22.84$, $p < 0.01$ for CI]. For both NH and CI listeners, *post hoc* Scheffe tests revealed that there was no significant decrease in vowel recognition for basal holes even up to 6 mm, but that vowel recognition was significantly lower for holes 4.5 mm or larger in the midcochlear region and for holes larger than 3 mm in the apical region. There was no significant difference in recognition between the four conditions that redistributed the spectral information from the hole in different ways (apical shift, basal shift, split, dropped) [$F_{(3,240)} = 0.40$, $p = 0.76$ for NH, $F_{(3,240)} = 0.37$, $p = 0.78$ for CI]. NH listeners as a group had significantly higher overall performance on vowel recognition than did CI listeners (Table 1), but there were no significant differences between NH and CI listeners when scores obtained with holes in the middle and basal regions were normalized to the baseline performance. Only in the apical region did NH and CI listeners perform differently, where CI listeners showed a 35 percentage point drop in performance with a 6 mm hole, while NH listeners

showed a 60 percentage point drop for the same condition. In general, when there are no spectral cues, vowel recognition remains about 30% correct because of nonspectral cues such as duration and overall amplitude (Shannon et al. 1995). In this experiment, NH listeners had a baseline (no hole) vowel recognition score of 90.4%, while CI listeners had a baseline score of 64.1% correct. Both groups of listeners dropped to about 30% correct with a 6 mm apical hole.

Sentence recognition (Fig. 5) generally decreased significantly as the size of the hole increased [$F_{(3,240)} = 109.11$, $p < 0.01$ for NH, $F_{(3,240)} = 39.71$, $p < 0.01$ for CI], and there was a significant interaction between hole size and hole location. For NH listeners, *post hoc* Scheffe tests revealed that there was no significant decrease in performance for basal holes even up to 6 mm, but that sentence recognition was significantly lower for holes 4.5 mm or larger in the middle and apical cochlear regions. It is difficult to gauge the effect of the various experimental conditions for CI listeners, because of the large inter- and intraindividual variability. Nonetheless, for CI listeners, *post hoc* Scheffe tests revealed that performance significantly decreased only when basal holes were 6 mm in extent and that sentence recognition was significantly lower for holes 4.5 mm and larger in the middle and apical cochlear regions. For both CI and NH listeners there were no

significant differences in recognition between the four conditions which distributed the spectral information from the hole in different ways (apical shift, basal shift, split, dropped) [$F_{(3,240)} = 0.38$, $p = 0.77$ for NH, $F_{(3,240)} = 1.06$, $p = 0.37$ for CI]. There were slight differences between the pattern of results for CI and NH listeners. NH listeners showed no decrease in sentence recognition for hole sizes of 3 mm or less in the apical and middle regions of the cochlea, and no decrease in performance for holes in the basal region. CI listeners' performance was not significantly decreased by holes 3 mm or less in the apical and middle regions; however, it was significantly reduced by a 6 mm hole in the basal region, while NH listeners' performance was not. For holes larger than 3 mm, NH listeners' performance decreased sharply with hole size, while CI listeners' performance decreased more gradually. These differences in the shape of the curves between NH and CI listeners are probably due to ceiling effects. NH listeners obtained word recognition scores close to 100% correct for holes up to 3 mm at all cochlear locations; CI listeners, on average, achieved only 65% correct for the baseline condition. However, the better CI listeners had a pattern of performance more similar to that of the NH listeners.

Consonant recognition was less affected by tonotopic holes than was vowel or sentence recognition, probably because consonant recognition relies more on temporal cues than spectral cues. Figures 6 and 7 present the results of an analysis of consonant information received for NH and CI listeners, respectively. Voicing and manner information were both virtually unaffected by the size of the hole at all cochlear locations. These cues are primarily temporal cues (Van Tasell et al. 1987) and require only minimal spectral information for nearly perfect reception (Shannon et al. 1995). [The only exception was for CI listeners, which again was caused by a single listener who performed poorly with large apical holes (note the large error bars in the right column of panels in Fig. 7).] However, for both NH and CI listeners, information received on place of articulation decreased considerably as the hole size was increased, particularly for the apical hole location. This pattern of results is similar to that observed for vowel recognition. The similarity between vowel recognition and information received on consonant place of articulation demonstrates that holes in the tonotopic representation primarily affect spectral cues.

Figure 8 compares NH and CI results averaged across reassignment conditions; in this figure scores were not normalized or adjusted for chance-level performance. The hatched area shows the total range of scores obtained by the 6 CI listeners. Note that the upper edge of the hatched region is similar in shape to NH listeners' pattern of results (but lower in terms

of absolute performance). This suggests that even the best CI listener's performance was not as good as that of NH listeners restricted to the same number of spectral channels. This observation is consistent with the results of previous studies that showed CI listeners were using only the equivalent of 4–8 channels of spectral information (Fishman et al. 1997; Dorman and Loizou 1998; Eddington et al. 1997; Friesen et al. 2001). In those studies, NH performance increased substantially as the number of channels was increased from 8 to 20, but CI performance was unchanged as the number of electrodes was increased from 8 to 20.

DISCUSSION

Holes in acoustic vs. electric hearing

NH listeners were able to achieve higher overall scores than CI listeners. This could be because the NH listeners were using a 20-band processor that used a Continuous Interleaved Sampling (CIS) strategy while CI listeners were using a 20-band SPEAK processing strategy. (CIS processors stimulate each electrode at a high rate, like 800 pulses/s, modulated by the speech envelope from each band. SPEAK performs a 20-band FFT every 4 ms and selects electrodes associated with the six bands with the most energy for stimulation.) However, previous studies have not observed any significant differences between CIS and SPEAK processing with the same number of electrodes in implant listeners (Fishman et al. 1997; Friesen et al. 2001). In those studies, performance of CI listeners was similar for 20-band SPEAK and 8-band CIS processing, and both were similar to the performance of NH listeners with 4–8-channel noise-band processors. While NH listeners' performance continued to improve as the number of bands was increased, implant performance did not improve as the number of electrodes was increased beyond 7 (10 for some test materials). This implies that the difference in overall performance between NH and CI listeners was due to NH listeners' ability to utilize all the spectral information provided by the 20 channels. It is not clear why CI listeners appear to be unable to utilize all the spectral information available to them.

In spite of the difference in overall performance level, the pattern of performance reduction was similar for NH and CI listeners. Some exceptions to this generalization were due to ceiling effects (NH sentence recognition) or floor effects (CI vowel recognition). NH and CI listeners were affected similarly as the spectral information was removed from the signal. The difference in overall performance level between NH and CI listeners may have been a result of CI listeners' inability to use more than 8 spectral channels, but the effect

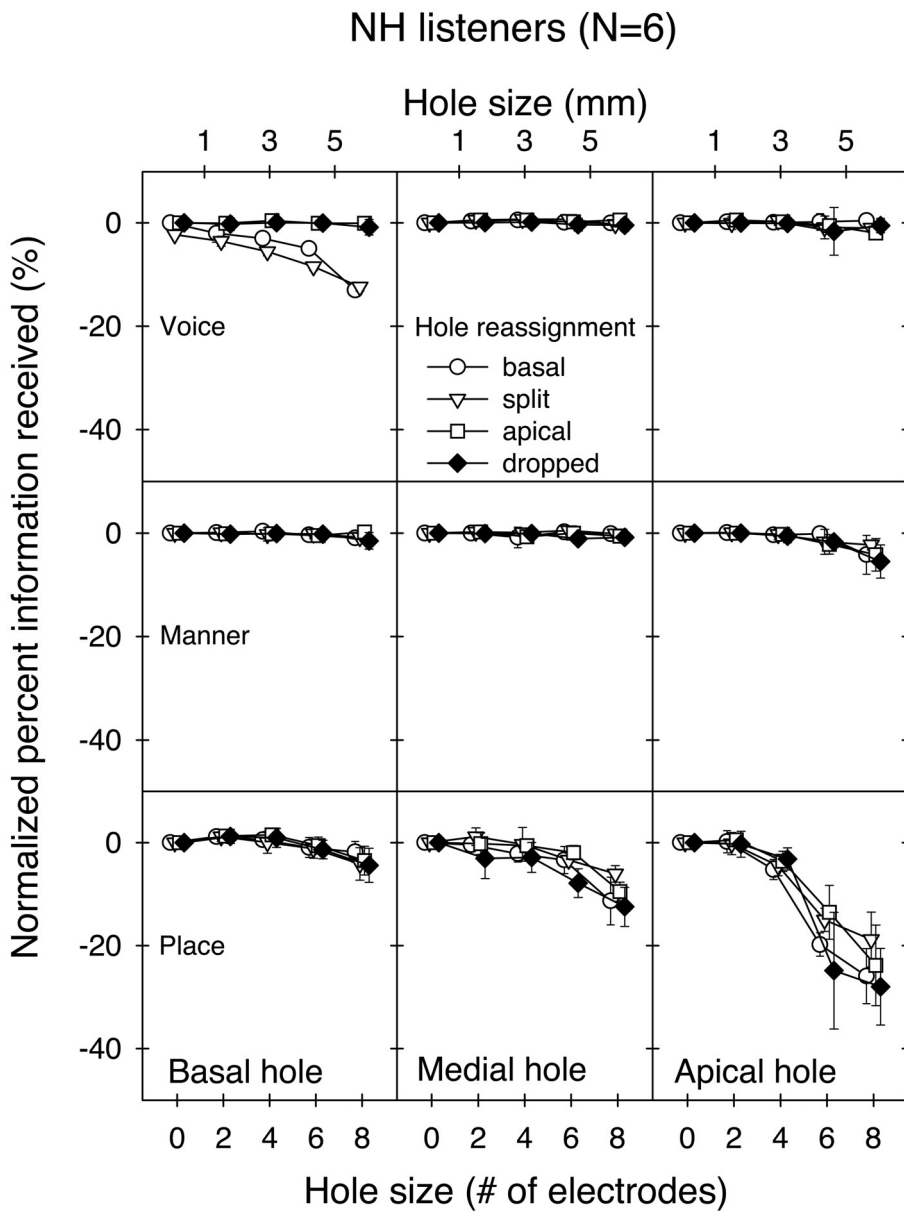


FIG. 6. Consonant information received on the production-based categories of voicing, manner, and place of articulation for NH listeners as a function of hole location and size, for four conditions of remapping. The error bars show the standard deviation. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm.

of a tonotopic hole was similar for both sets of listeners because the same information was removed. This observation suggests that the relative effect of a hole in the spectral information is similar for different levels of spectral resolution.

Predictions of articulation theory

Predictions of the SII for the conditions of the present experiment were computed. Since there were no significant differences between reassignment conditions, the results from all reassignment conditions were averaged for each speech test. SII values were computed using the ANSI 1/3-octave band-importance weighting function [see Table 3, ANSI S3.5 (1997)]. These weights reflect the contribution of specific frequency

bands to speech intelligibility for a standard set of monosyllabic words. Importance weighting functions for other materials can deviate considerably from the ANSI standard function (Studebaker and Sherbecoe 1993). Although the vowel and consonant materials used in the present experiment were different from the ANSI standard word materials, the standard weighting function was used as an approximation for the present calculations.

In the present experiment the frequency range from 150 Hz to 10 kHz was divided into 20 frequency bands. The ANSI specifications apply to the frequency range from 160 Hz to 8 kHz and values are provided only at 1/3-octave intervals. To calculate SII predictions for the present experiment, the ANSI weighting function was converted from frequency to cochlear

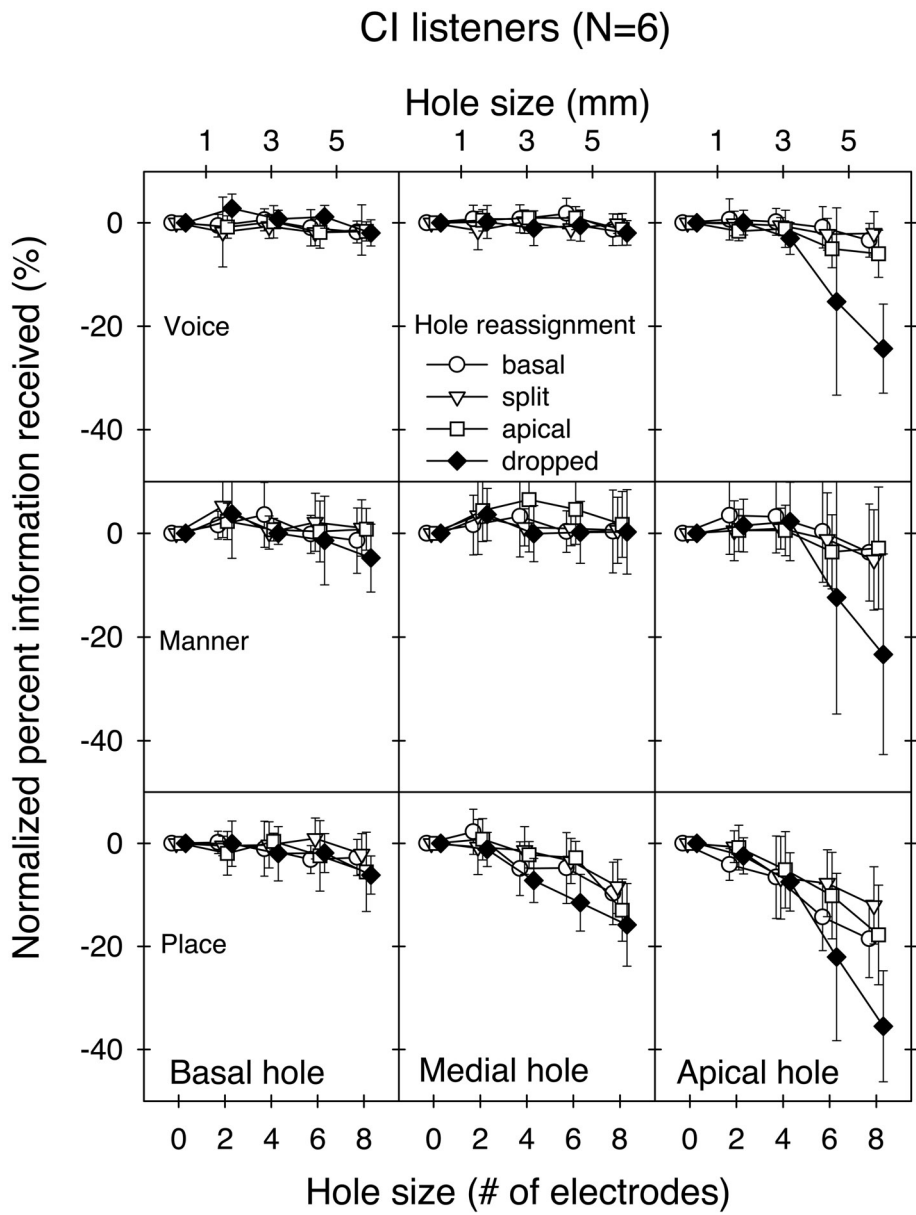


FIG. 7. Consonant information received on the production-based categories of voicing, manner, and place of articulation for CI listeners as a function of hole location and size, for four conditions of remapping. The error bars show the standard deviation. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm.

location according to Greenwood (1990). Weights of zero were assigned to frequencies above or below the ANSI frequency table. The weighting values for each of the 20 filters in the present experiment were interpolated linearly in terms of cochlear location from the ANSI standard values. For the present calculations the weights were increased slightly so that the total integrated SII value for the 20-band condition equaled 1. SII values were then computed by integrating the area under the interpolated weighting function. The SII for the hole conditions was computed by assigning weights of zero to the frequency range of the hole.

NH listeners' normalized speech recognition results for all hole conditions are plotted as a function of SII values in Figure 9. Because there was no significant difference between any of the remapping conditions,

the mean value across remapping conditions for each hole size and location was used. The data were normalized to those for the baseline (no hole) condition so that the curves peak at 0 shift in percent correct where the SII was defined to have a value of 1. These plots were fitted with a sigmoidal function of the form:

$$\%correct = 100 * (1 - 10^{-SII/Q})^n \quad (1)$$

where n and Q are fitting parameters. The optimum parameter values were obtained by minimizing rms errors. Table 2 presents the parameter values of these curves and the corresponding rms error (in percentage points). The resulting best-fit curves are plotted as the solid lines in Figure 9. Note that the value of Q was relatively similar for the three sets of materials, but the value of n was quite different for consonants.

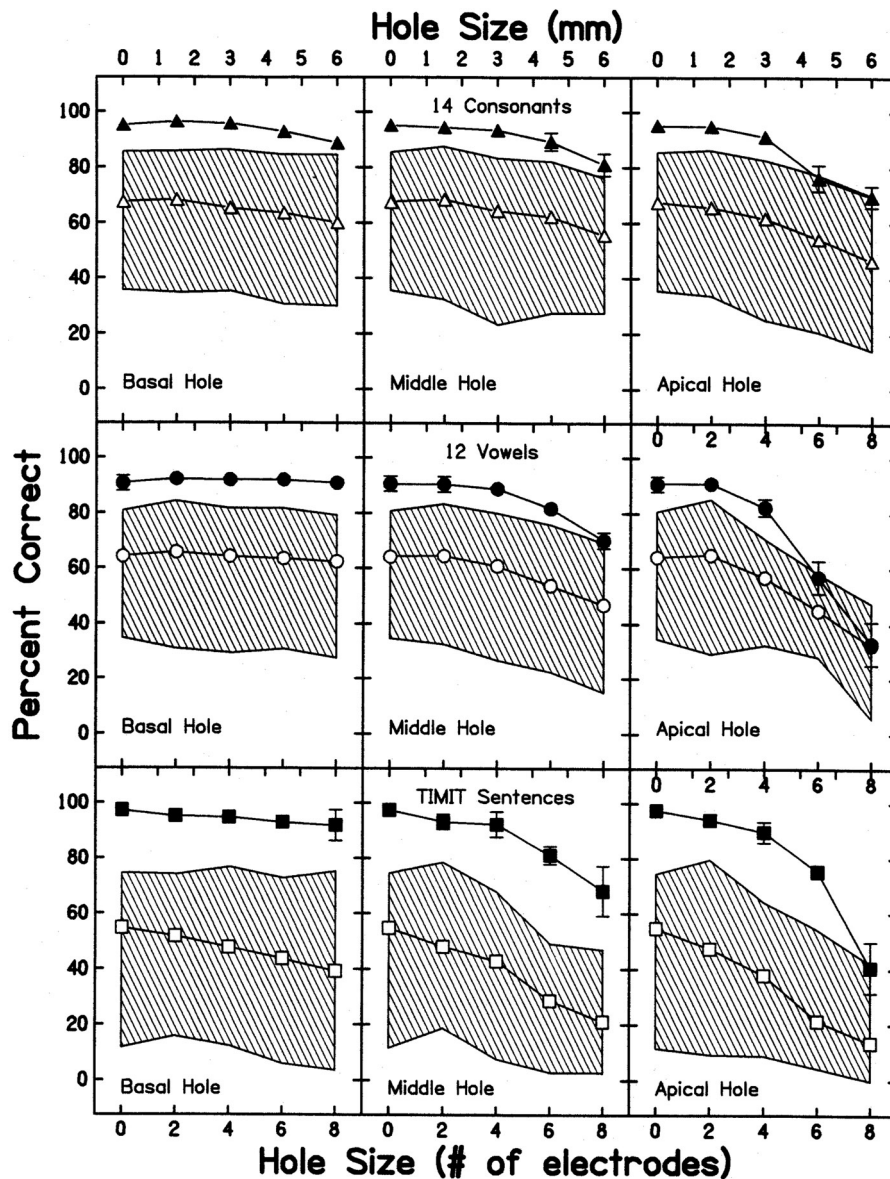


FIG. 8. Comparison of the absolute levels of speech recognition for NH and CI listeners. Results have been averaged across remapping conditions. The filled symbols with the solid lines present mean results for NH listeners and the error bars show the standard deviation; the open symbols with the solid lines present mean results for CI listeners. The bottom x-axis of the panels shows the hole size in terms of the number of electrodes, while the top x-axis shows the hole size in terms of cochlear distance in mm. The y-axis of all panels shows the absolute percent correct. The hatched area represents the total range of scores obtained from CI listeners.

Vowel and sentence recognition required a value of n that was 6–7 times higher than for consonant recognition. This may reflect the fact that consonantal distinctions are less dependent on spectral resolution than vowel or sentence recognition.

The parameters estimated from the aggregate data were then used to model performance for the individual experimental hole conditions. The model results are presented in Figure 10 and compared with the average results from NH listeners. Note that the predictions of SII are relatively close to the actual data, with the rms error of the predictions ranging from 0.64% to 4.44%. This error value is similar to the standard deviation of the measurements and so constitutes a good fit. Although previous results have suggested that separated spectral bands can interact synergistically (Lippmann 1996; Breeuwer and Plomp 1984, 1985,

1986; Warren et al. 1995), the present results suggest that the loss of spectral information due to a hole in the tonotopic representation can be well represented by the traditional SII model. Application of the SII model to the average CI results produced an rms error ranging from 0.8% to 5.7% across conditions; slightly larger rms error values than for NH listeners, but still comparable to the standard deviation of the measurements. Thus, the loss of spectral information due to a hole in the tonotopic representation can be relatively well modeled by the SII for both CI and NH listeners.

Implications for preserving/restoring the lost information

The present results are not encouraging for the prosthetic restoration of lost information due to a hole in

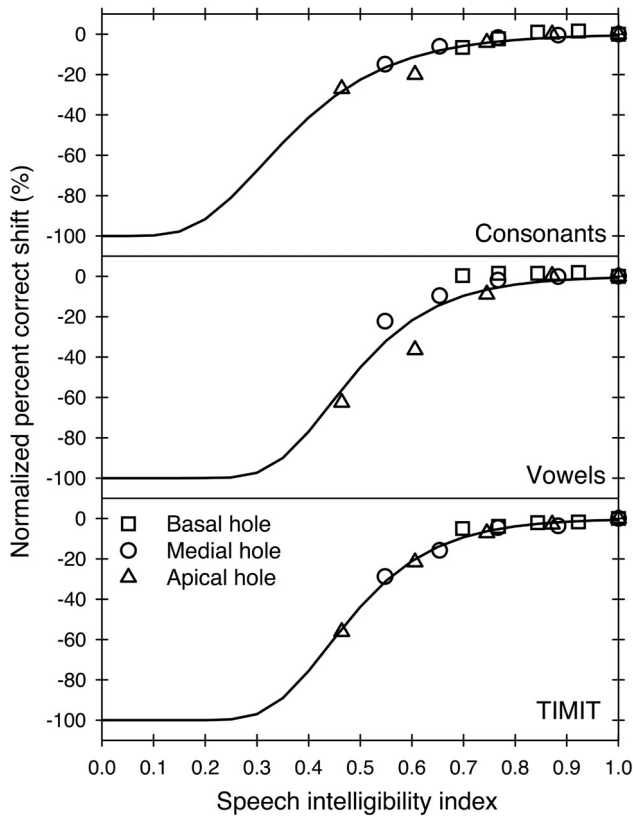


FIG. 9. Speech recognition by NH listeners as a function of the Speech Intelligibility Index (SII) for the three sets of speech materials. The solid lines represent sigmoidal fits to the normalized data by eq. (1).

TABLE 2

Parameter values for SII analysis and the rms error in percentage points					
Speech material	Q	n	Basal hole rms	Middle hole rms	Apical hole rms
Consonants	0.32	9.2	1.16	0.86	2.36
Vowels	0.26	50.0	3.48	3.06	4.44
TIMIT sentences	0.26	48.0	1.22	1.00	0.64

Values were calculated for the average of all reassignment conditions for NH listeners.

the receptor array. The information from the hole region was remapped to the edges of the hole in three different ways and none resulted in improved performance (although none degraded performance either). Although the information in the speech signal was preserved, neither NH nor CI listeners showed any benefit of reassigning the spectral information to the edges of the hole as compared with simply dropping the information. This suggests that the pattern of spectral information becomes unusable if it is not

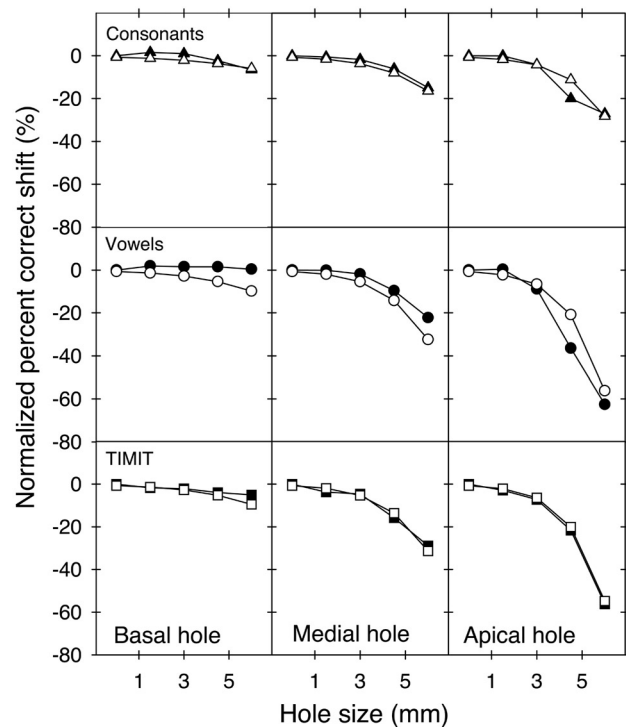


FIG. 10. SII predictions for NH listeners as a function of hole size and location. Filled symbols represent data from NH listeners averaged across remapping conditions and open symbols represent SII predictions.

presented to the correct tonotopic location. This pattern was evident in the detailed analysis of consonant information received. Reassigning the spectral information to the edges of the hole did not change the amount of information received on voicing, manner, or place of articulation. It is still possible that other reassignments could result in a rescue of the information from the hole region but the simple reassignments used in this experiment did not.

This observation is consistent with recent results obtained by Turner and colleagues (Hogan and Turner 1998; Turner and Cummings 1999; Turner et al. 1999). Finding that amplification in patients with more than 60 dB of high-frequency hearing loss could actually degrade speech performance, they suggested that such amplification resulted in the reassignment of spectral information to a lower tonotopic place (due to the spread of excitation in the cochlea near the amplified region). In Turner's studies, such reassignment resulted in a decrease in performance for some listeners and no change in performance for others; no listeners showed an increase in performance as a result of amplification. However, Vickers et al. (2001) found some benefit of amplifying frequencies up to 1 octave above the estimated edge frequency in listeners diagnosed with a high-frequency dead region. This

result implies that some listeners may be able to make use of “reassigned” information.

One issue that complicates the interpretation of the present results is the potential effect of long-term learning. The present results show no benefit of remapping spectral information around a hole when the listeners had no time to practice or adapt to the new mapping. It is possible that listeners might be able to make use of remapped information if given sufficient time to adapt to the new pattern. Van Tasell et al. (1992) found significant improvement on consonant recognition in NH listeners after training on single-channel noise processors. Rosen et al. (1999) found that NH listeners improved significantly after limited training with 4-channel speech processors that had been frequency-shifted by 6.5 mm in terms of cochlear place. However, Fu et al. (submitted) saw incomplete adaptation in CI listeners with frequency-shifted processors even after three months of daily use. It appears that some adaptation is possible in a short time period, but it is still not clear how much of an alteration in the frequency–place mapping can be relearned. Eisenberg et al. (2000) demonstrated that 5-year-old children are not able to process spectrally degraded speech as well as 12 year olds or adults. This result suggests that speech pattern recognition develops over a long time course in normal auditory development. It seems reasonable to speculate that the degree of relearning and the time course of adaptation depend on the severity of the frequency–place mismatch. Frequency–place alterations that require learning only a slight modification of the original pattern recognition may only require minimal adaptation over a limited time. However, alterations that essentially require the acquisition of an entirely new system of pattern recognition may take many years, as happens in normally developing hearing children’s speech acquisition. It is not clear whether the frequency–place remapping around tonotopic holes in the present study could be learned quickly or not. Further research is needed to delineate the relationship between the parameters of frequency–place mapping and the time course of adaptation to the new mapping.

CONCLUSIONS

Both normal-hearing and cochlear implant listeners could tolerate a hole in the tonotopic representation as large as 6 mm (about 1.5 octaves) in the basal cochlear region with only modest decreases in speech recognition. However, when the hole was moved to the middle cochlear region, the effect was more significant, and a hole in the apical region produced a large drop in speech recognition. These results are broadly consistent with predictions based on the articulation index.

Normal-hearing and cochlear implant listeners showed similar patterns of results, suggesting that the loss of information caused by tonotopic holes was the primary factor in performance and that NH and CI listeners were similarly affected by that loss of information. However, the present results are discouraging for the prospect of restoring this lost information: None of the conditions in this experiment that attempted to “rescue” the lost information by remapping it to the edges of the hole produced any improvement.

ACKNOWLEDGMENTS

We are grateful to both the normal-hearing and implant subjects who participated in this experiment for their patience and persistence, and to Brian C.J. Moore and Arthur Boothroyd for helpful comments on the earlier drafts of this manuscript. Funding for this research was provided by NIH grant R01-DC-01526 and NIH contract N01-DC-92100.

REFERENCES

- American National Standards Institute (1997) Methods for Calculation of the Speech Intelligibility Index, S3.5.
- BREEUWER M, PLOMP R. Speechreading supplemented with frequency-selective sound-pressure information. *J. Acoust. Soc. Am.* 76:686–691, 1984.
- BREEUWER M, PLOMP R. Speechreading supplemented with formant frequency information from voiced speech. *J. Acoust. Soc. Am.* 77:314–317, 1985.
- BREEUWER M, PLOMP R. Speechreading supplemented with auditorily-presented speech parameters. *J. Acoust. Soc. Am.* 79:481–499, 1986.
- Cochlear Corporation (1995) Technical Reference Manual. Englewood, CO.
- DORMAN MF, LOIZOU PC. Identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations processors with two to nine channels. *Ear Hear.* 19:162–166, 1998.
- DORMAN MF, LOIZOU PC, RAINEY D. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J. Acoust. Soc. Am.* 102:2993–2996, 1997, DOI: 10.1121/1.420354.
- EDDINGTON DK, RABINOWITZ WR, TIERNEY J, NOEL V, WHEARTY M (1997) Speech Processors for Auditory Protheses, 8th Quarterly Progress Report, NIH Contract N01-DC-6-2100.
- EISENBERG L, SHANNON RV, MARTINEZ AS, WYGONSKI J, BOOTHROYD A. Speech recognition with reduced spectral cues as a function of age. *J. Acoust. Soc. Am.* 107:2704–2710, 2000, DOI: 10.1121/1.428656.
- FISHMAN K, SHANNON RV, SLATTERY WH. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J. Speech Hear Res.* 40:1201–1215, 1997.
- FLETCHER H, STEINBERG JC. Articulation testing methods. *Bell Systems Tech. J.* 8:806–854, 1929.
- FRENCH NR, STEINBERG JC. Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19:90–119, 1947.

- FRIESEN L, SHANNON RV, BASKENT D, WANG X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110:1150–1163, 2001.
- FU Q-J, SHANNON RV, WANG X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *J. Acoust. Soc. Am.* 104:3586–3596, 1998.
- FU Q-J, SHANNON RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J. Acoust. Soc. Am.* 105:1889–1900, 1999, DOI: 10.1121/1.426725.
- FU Q-J, SHANNON RV. Effect of stimulation rate on phoneme recognition in cochlear implants. *J. Acoust. Soc. Am.* 107:589–597, 2000, DOI: 10.1121/1.428325.
- FU Q-J, SHANNON RV, GALVIN J. Performance over time of Nucleus-22 cochlear implant listeners wearing speech processor with a shifted frequency-to-electrode assignment. *J. Acoust. Soc. Am.* (submitted), 2002.
- GREENWOOD DD. A cochlear frequency–position function for several species—29 years later. *J. Acoust. Soc. Am.* 87:2592–2605, 1990.
- HILLENBRAND J, GETTY L, CLARK M, WHEELER K. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97:3099–3111, 1994.
- HOGAN CA, TURNER CW. High-frequency audibility: Benefits for hearing-impaired listeners. *J. Acoust. Soc. Am.* 104:432–441, 1998, DOI: 10.1121/1.423247.
- LIPPMANN RP. Accurate consonant perception without mid-frequency energy. *IEEE Trans. Speech Audio. Proc.* 4:66–69, 1996.
- MCDERMOTT HJ. An advanced multiple channel cochlear implant. *IEEE Trans. Biomed. Eng.* 36:789–797, 1989, DOI: 10.1109/10.32112.
- MILLER G, NICELY P. An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27:338–352, 1955.
- MOORE BCJ. Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids. *Trends Amplific.* 5:1–34, 2001.
- MOORE BCJ, ALCÁNTARA JI. The use of psychophysical tuning curves to explore dead regions in the cochlea. *Ear Hear.* 22:268–278, 2001.
- MOORE BCJ, GLASBERG BR. A model of loudness perception applied to cochlear hearing loss. *Aud. Neurosci.* 3:289–311, 1997.
- MOORE BCJ, HUSS M, VICKERS DA, GLASBERG BR, ALCÁNTARA JI. A test for the diagnosis of dead regions in the cochlea. *Br. J. Audiol.* 34:205–224, 2000.
- National Institute of Standards and Technology (1990) DARPA TIMIT Acoustic–Phonetic Continuous Speech Corpus CD-ROM. Gaithersburg, MD.
- ROBERT ME. AIPSS-ID: Phoneme Identification Software. House Ear Institute Los Angeles, CA, 1997.
- ROSEN S, FALKNER A, WILKINSON L. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J. Acoust. Soc. Am.* 106:3629–3636, 1999, DOI: 10.1121/1.428215.
- SELIGMAN PM, MCDERMOTT HJ. Architecture of the Spectra-22 speech processor. *Ann. Otol. Rhinol. Laryngol.* 104(Suppl 166):139–141, 1995.
- SHANNON RV, ZENG F-G, KAMATH V, WYGONSKI J, EKELID M. Speech recognition with primarily temporal cues. *Science* 270:303–304, 1995.
- SHANNON RV, ZENG F-G, WYGONSKI J. Speech recognition with altered spectral distribution of envelope cues. *J. Acoust. Soc. Am.* 104(4):2467–2476, 1998.
- STUDEBAKER GA, SHERBECOE RL. Frequency–importance functions for speech recognition. In: Studebaker GA, Hochberg I, (eds) *Acoustical Factors Affecting Hearing and Performance*. Allyn and Bacon Boston, 1993. p. 185–204.
- TURNER CW, CUMMINGS KJ. Speech audibility for listeners with high-frequency hearing loss. *Am. J. Audiol.* 8:47–56, 1999.
- TURNER CW, CHI S-L, FLOCK S. Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *J. Speech Hear. Res.* 42:773–784, 1999.
- VAN TASELL DJ, GREENFIELD DG, LOGEMANN JJ, NELSON DA. Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants. *J. Acoust. Soc. Am.* 92:1247–1257, 1992.
- VAN TASELL DJ, SOLI SD, KIRBY VM, WIDIN GP. Speech waveform envelope cues for consonant recognition. *J. Acoust. Soc. Am.* 82:1152–1161, 1987.
- VICKERS DA, MOORE BCJ, BAER T. Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *J. Acoust. Soc. Amer.* 110:1164–1175, 2001.
- WARREN RM, REINER KR, BASHFORD JR JA, BRUBAKER BS. Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Percept. Psychophys.* 57:175–182, 1995.