Speech recognition under conditions of frequency-place compression and expansion

Deniz Baskent[a)]
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089

Robert V. Shannon
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089
and House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

In normal acoustic hearing the mapping of acoustic frequency information onto the appropriate cochlear place is a natural biological function, but in cochlear implants it is controlled by the speech processor. The cochlear tonotopic range of the implant is determined by the length and insertion depth of the electrode array. Conventional cochlear implant electrode arrays are designed for an insertion of 25 mm inside the round window and the active electrodes occupy 16 mm, which would place the electrodes in a cochlear region corresponding to an acoustic frequency range of 500–6000 Hz. However, some implant speech processors map an acoustic frequency range from 150 to 10 000 Hz onto these electrodes. While this mapping preserves the entire range of acoustic frequency information, it also results in a compression of the tonotopic pattern of speech information delivered to the brain. The present study measured the effects of such a compression of frequency-to-place mapping on speech recognition using acoustic simulations. Also measured were the effects of an expansion of the frequency-to-place mapping, which produces an expanded representation of speech in the cochlea. Such an expanded representation might improve speech recognition by improving the relative spatial (tonotopic) resolution, like an ''acoustic fovea.'' Phoneme and sentence recognition was measured as a function of linear (in terms of cochlear distance) frequency-place compression and expansion. These conditions were presented to normal-hearing listeners using a noise-band vocoder, simulating cochlear implant electrodes with different insertion depths and different number of electrode channels. The cochlear tonotopic range was held constant by employing the same noise carrier bands for each condition, while the analysis frequency range was either compressed or expanded relative to the carrier frequency range. For each condition, the result was compared to that of the perfect frequency-place match, where the carrier and the analysis bands were perfectly matched. Speech recognition in the matched conditions was generally better than any condition of frequency-place expansion and compression, even when the matched condition eliminated a considerable amount of acoustic information. This result suggests that speech recognition, at least without training, is dependent on the mapping of acoustic frequency information onto the appropriate cochlear place. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1558357]

I. INTRODUCTION

Speech recognition is adversely affected if the spectral information of speech is presented to an inappropriate cochlear location. For example, several studies have shown a reduction in speech recognition when the speech spectrum is shifted up to higher frequencies (e.g., Daniloff *et al.*, 1968; Fu and Shannon, 1999), or if the frequency-to-cochlear place mapping is distorted nonlinearly (Shannon *et al.*, 1998). Changes in speech recognition as a result of such distortions in the frequency-to-cochlear place mapping are of theoretical interest as an indication of the mechanisms by which speech patterns are stored and retrieved in the central nervous system. In addition, understanding the potential effect of an ap-propriate frequency-place mapping is of critical importance for the design and programming of cochlear implants and hearing aids. These prosthetic devices can stimulate the peripheral auditory system with a tonotopic pattern of information that is distorted relative to the normal acoustic pattern. Such stimulation raises several questions: In the case of hearing loss, can the patient's speech recognition be improved by adjusting the spectral range of speech to match the frequency region of her residual hearing (Braida *et al.*, 1979; Reed *et al.*, 1983)? Or can the resulting frequency-place distortion actually interfere with speech understanding, as shown in some listeners by Turner *et al.* (1999)? Which of these spectral manipulations are more detrimental and so should be avoided?

Several previous studies have addressed the question of how spectral manipulations affect speech understanding. Fu and Shannon (1999) measured vowel recognition in normal hearing (NH) and cochlear implant (CI) listeners when the

---
[a)]Author to whom correspondence should be addressed. Department of Auditory Implants and Perception, House Ear Institute, 2100 W. Third St., Los Angeles, CA 90057. Electronic mail: baskent@usc.edu

acoustic spectral information was mapped to cochlear locations that were shifted apically or basally relative to the acoustic location for that information (for NH listeners) or relative to each listener's clinical frequency-to-electrode map (for CI listeners). They found that vowel recognition was robust to tonotopic shifts up to 3 mm, but dropped significantly for larger shifts. This result matches well with classical studies on frequency shifting (Daniloff *et al.*, 1968; Nagafuchi, 1976; Tiffany and Bennett, 1961).

In a similar study, Dorman *et al.* (1997) measured the effect of a shift in mm between the acoustic frequency range presented and the cochlear range to which it was presented. In the acoustic simulations the analysis filter bands were fixed, and sine wave carriers were shifted in mm along the cochlea relative to the normal acoustic place for that information. Speech recognition performance dropped as the stimulated electrode locations were shifted basally from the normal tonotopic location.

Shannon *et al.* (1998) measured speech recognition under conditions that produced a nonlinear warping of the frequency-place mapping. They used a noise-band vocoder to implement a logarithmic or exponential transformation between acoustic frequency and the normal cochlear place for that frequency. Although four spectral channels of information were presented, listeners' performance with the warped mapping dropped to the same level as that seen with a single-channel noise vocoder. This result suggests that nonlinear frequency-place warping can eliminate listeners' ability to utilize spectral cues.

As an extension to previous studies dealing with frequency-place distortions, the present study explored the effects on speech recognition when the acoustic frequency range delivered is larger or smaller than the normal cochlear range.

Note that neither the present study nor the previous studies discussed above address the potential effects of learning. Research by Rosen *et al.* (1999) showed that, following a short training process, listeners could partially adapt to basalward spectral shifts of as much as 6.5 mm. Another study (Fu *et al.*, 2002) showed significant improvement over the first few days by cochlear implant patients using a 2–4-mm apically shifted frequency-place map, but only little change was observed over the following 3 months. At the end of the 3-month training period consonant and HINT sentence recognition scores were comparable to the baseline scores, while vowel and TIMIT sentence recognition scores were still significantly lower than the baseline scores obtained with the patient's own clinical map before the beginning of the test. In the present experiments the emphasis is on speech pattern recognition without any training. The intention was to test the ability of central pattern recognition mechanisms to accommodate alterations in the peripheral pattern of information with no time to adapt.

Frequency compression has historically been used in an attempt to increase the performance of hearing aids. Most hearing aid users have hearing loss at high frequencies with residual hearing at lower frequencies. To make better use of this residual hearing the spectrum of the speech was lowered and compressed so that the entire speech information was delivered to the audible range of the patient. The main techniques used for this purpose were slow playback, frequency shifting, vocoding, and zero-crossing-rate division. In terms of frequency-to-place mapping most of these manipulations consisted of a compressed apical cochlear shift. Braida *et al.* (1979) reviewed frequency compression/shifting studies and concluded that frequency lowering did not result in any substantial improvement in speech recognition, and often decreased the performance compared to simple amplification. Reed *et al.* (1983) evaluated the effect of frequency lowering on consonant recognition in a more systematic way, parametrically varying the frequency compression scheme from linear compression to nonlinear frequency-place warping. The results from this study confirmed that frequency lowering did not improve consonant recognition. Linear frequency compression, where the whole frequency range was compressed, resulted in worse consonant recognition scores than a frequency warping compression in which only higher frequencies were spectrally compressed and lowered.

These studies provide insight into the mechanisms used by the central nervous system for storing and retrieving tonotopic patterns of speech. If speech patterns were stored in a "positionally relocatable" fashion, then a tonotopic shift that maintained the overall spatial distribution should still be intelligible. This is clearly not the case, because frequency shifting usually reduces speech recognition. If only the relative order of spectral features were important, then monotonic alterations in the tonotopic pattern would still be intelligible. This is also not the case, because frequency shifting usually reduces speech recognition. The present experiment was designed to further quantify the importance of linear compression or expansion of the tonotopic pattern of information (in cochlear mm). If the central pattern recognition system stores and retrieves information in terms of the relative tonotopic pattern, then it might be able to tolerate a substantial amount of linear compression or expansion.

These issues are not only noteworthy in terms of understanding the relative importance of peripheral vs central pattern recognition for speech, but are of critical importance for the design and fitting of cochlear implants and hearing aids. In a cochlear implant, the electrode array is typically inserted into the scala tympani, reaching a depth of 20–30 mm inside the round window. The average insertion depth from 20 Nucleus implant patients was estimated to be 20 mm by Ketten *et al.* (1998). However, newer electrode designs are intended to achieve array insertions as deep as 30 mm (Gstoettner *et al.*, 1999). The active stimulation range is typically 16 mm in length for Clarion I and Clarion II, and 16.5 mm for Nucleus 22 and Nucleus 24 devices. According to Greenwood's (1990) frequency-to-place equation, and assuming a 35-mm cochlear length in humans, this stimulated cochlear region corresponds to an acoustic frequency range of 500–6000 Hz in humans for a 25-mm insertion depth, and an acoustic frequency range of 1–12 kHz for a 20-mm insertion depth of the electrode array. Present cochlear implants

offer only a limited choice of analysis filters, which cannot be changed individually to match a given patient's electrode location. Most commercial implant speech processors assign a wider fixed acoustic frequency range to this limited cochlear region regardless of the length or the insertion depth of the electrode array. For example, Clarion II assigns an acoustic range of 350–8000 Hz (Advanced Bionics Corporation, 2001) and the default frequency allocation of the Nucleus-22 implant (SPEAK strategy Table 9) assigns a frequency range of 150 Hz–10 kHz to the electrodes (Cochlear Corporation, 1995). The latter acoustic range would normally cover a 25-mm range in the cochlea, specifically from 5 to 30 mm from the round window, rather than the 16.5 mm covered by the electrode array. Thus, mapping the larger acoustic frequency range onto the electrode array results in a compression of the frequency-to-place mapping. In some cochlear implant patients there may also be a tonotopic shift due to the discrepancy between the actual electrode location and the acoustic information assigned. The present experiment evaluated the effect of frequency-place compression on speech recognition in normal-hearing listeners in conditions that simulated two implant electrode insertion depths.

In addition to compression, frequency-place expansion was also evaluated. In this condition, the midfrequency region was expanded in terms of its representation in the cochlea, effectively increasing the sensory resolution within this frequency range. This frequency-place expansion is analogous to the ''acoustic fovea'' in bats or cetaceans, where a large portion of the cochlea is devoted to the small frequency region used for echolocation (e.g., Echteler et al., 1994). While this type of expansion results in the loss of some acoustic information, the most critical spectral information is presented to a larger cochlear region, resulting in better neural resolution (increased mm/Hz) within that range.

II. METHOD

A. Subjects

Six normal-hearing listeners, aged 26 to 34, participated in the study. All subjects were native speakers of American English and had thresholds better than 20 dB HL at audiometric frequencies between 125 and 8000 Hz. One subject was excluded from the sentence recognition test because she was already familiar with the sentences in the database. Another 32-year-old subject was added to the sentence recognition test to maintain six subjects for each test.

B. Stimuli

The speech recognition tasks consisted of medial vowel and consonant discrimination, and sentence recognition. All stimuli were presented via a loudspeaker in a sound field at 70 dB on an A-weighted scale.

Consonant stimuli were taken from materials recorded by Turner et al. (1992, 1999) and Fu et al. (1998) at a 44.1-kHz sampling rate. Six presentations (three male and three female talkers) were made of 14 medial consonants /b d f g k m n p s ʃ t θ v z/, presented in an /a/-consonant-/a/ context. Tokens were presented in random order by custom software (Robert, 1998), and subjects were instructed to select the

consonant they heard from the set of 14 possible consonants displayed on the screen. The resulting consonant 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, and the single-tailed 95% confidence level was 11.77% correct based on a binomial distribution.

Vowel stimuli were taken from the phoneme set recorded by Hillenbrand et al. (1995) at a 32-kHz sampling rate. The tokens were presented to the listeners in random order via custom software (Robert, 1998), and subjects were instructed to select the vowel they heard from the set of 12 possible vowels displayed on the screen. Ten presentations (five male and five female talkers) were made of 12 medial vowels, including 10 monophthongs and 2 diphthongs presented in an /h/-vowel-/d/ context (heed, hid, head, had, hod, hawed, hood, who'd, hud, heard, hayed, hoed). Chance level on this test was 8.33% correct, and the single-tailed 95% confidence level was 12.48% correct based on a binomial distribution.

Recognition of words in sentences was measured using the custom software (TIGER SPEECH RECOGNITION SYSTEM developed by Qian-Jie Fu) with Texas Instruments/Massachusetts Institute of Technology (DARPA/TIMIT) corpus of sentence materials (National Institute of Standards and Technology, 1990). The sentences were of moderate-to-hard difficulty, such that individual words were difficult to predict from the context of the sentence, and sentences were spoken by multiple talkers. For each condition, the percent-correct score was acquired for 20 sentences of varying length from each listener. The length of the sentences varied from 3 words to 12 words. The groups of 20 sentences were prepared such that the average word length per sentence was 6–8 words. They were presented without any context information and no sentences were repeated to an individual listener. Sentences were not balanced for difficulty, so 20 sentences were used for each condition to obtain a sample that included varying levels of difficulty. In addition, the order of the presentation of sentences was completely randomized using a random number generator for each subject so that all subjects heard different sentences for different conditions. Consequently differences arising from varying difficulty of the sentences were randomly distributed across different conditions and different subjects. Subjects were asked to repeat what they had heard. The percent-correct score was obtained by counting the percentage of words repeated correctly by the subject.

This study concentrates on effects of frequency-place compression and expansion on speech recognition without any learning effects. We typically observe a short-term adaptation to the test procedure by inexperienced subjects where all scores (regardless of the condition) increased slightly over the first 3 days of the testing. However, the scores remained more or less stable after this initial adaptation period. This was not observed with subjects who already had experience in similar experiments. To minimize any learning effects for the specific experimental conditions no practice was provided on any conditions prior to data collection, even for

TABLE I. Frequency-place mismatch conditions for the 4-channel processor at the simulated 20-mm electrode insertion depth. The name of each condition represents the change in frequency range expressed in mm between the analysis and carrier bands. For each condition the table lists the following information for the analysis bands: cochlear location in mm from the round window, cutoff frequencies for a 4-band processor, and total frequency range. Because the simulated electrode location was fixed, the noise carrier bands covered the frequency range from 1168 to 11 837 Hz in all conditions, and the frequency partition of carrier bands was as shown in the center entry.

Frequency-place mismatch condition	Cochlear location of analysis bands (mm)	Bandpass filter cutoff frequencies for 4 channels (Hz)				Frequency range of analysis bands (Hz)
−5 mm (expansion)	15–9	2476	3080	3822	4736	5 860	2476–5 860
−3 mm (expansion)	17–7	1843	2663	3822	5459	7 771	1843–7 771
−1 mm (expansion)	19–5	1363	2301	3822	6289	10 290	1363–10 290
0 mm (matching)	20–4	1168	2138	3822	6749	11 837	1168–11 837
+1 mm (compression)	21–3	999	1985	3822	7243	13 612	999–13 612
+3 mm (compression)	23–1	722	1710	3822	8337	17 990	722–17 990
+5 mm (compression)	25–0	513	1471	3822	9594	23 762	513–23 762

new subjects, and no feedback was given in any part of the testing. In addition, to reduce a possible adaptation to a particular condition, all conditions with all stimuli were presented to subjects in a completely random order. Therefore, any effects of learning on scores would be distributed across different conditions with different subjects.

C. Signal processing

The noise-band vocoder technique described by Shannon *et al.* (1995) was implemented in MATLAB to generate the stimuli. First, speech materials were bandpass filtered into a number of contiguous frequency bands by sixth-order Butterworth filters. The −3-dB cutoff frequencies were determined according to Greenwood's (1990) frequency-to-place mapping equation. The exact frequency ranges and cutoff frequencies were determined depending on the specific experimental conditions (see Tables I and II). The speech envelope was extracted from each band by half-wave rectification and low-pass filtering using a third-order Butterworth filter whose output was 3 dB down at a cutoff frequency of 160 Hz. The noise carrier bands representing the

stimulation region in the cochlea were obtained from white noise by sixth-order Butterworth filters where the cutoff frequencies (−3-dB points) were determined by the condition. The extracted speech envelopes were used to modulate the noise carrier bands, and all modulated noise bands were combined to form the processed speech. The amplitude level of the processed speech was adjusted such that the original and processed tokens had the same overall rms energy.

D. Mapping conditions

Speech materials were processed using 4, 8, or 16 frequency bands. Two different electrode array locations were simulated, representing insertion depths of 20 and 25 mm from the round window. For all expansion and compression conditions the stimulation region covered by the simulated electrode array was fixed at 16 mm (comparable to the typical length of the electrode array for many implant devices). The 20-mm insertion depth condition simulated an electrode array located between 4 and 20 mm from the round window, and the 25-mm insertion depth condition simulated a location between 9 and 25 mm from the round window. Because

TABLE II. Frequency-place mismatch conditions for a 4-channel processor at the simulated 25-mm electrode insertion depth. For each condition the table lists the cochlear locations of the analysis bands, cutoff frequencies of the bandpass filters, and the total analysis frequency range. The noise carrier bands were fixed between 513 and 5860 Hz with the partition shown as in the center of the table. The +5-mm condition is the one most similar to frequency-to-electrode assignment used in a cochlear implant with a full electrode insertion.

| Frequency-place mismatch condition	Cochlear location of analysis bands (mm)	Bandpass filter cutoff frequencies for 4 channels (Hz)				Frequency range of analysis bands (Hz)
−5 mm (expansion)	20–14	1168	1471	1843	2300	2 864	1168–2 864
−3 mm (expansion)	22–12	851	1262	1843	2663	3 822	851–3 822
−1 mm (expansion)	24–10	611	1081	1843	3080	5 085	611–5 085
0 mm (matching)	25–9	513	999	1843	3310	5 860	513–5 860
+1 mm (compression)	26–8	428	922	1843	3557	6 750	428–6 750
+3 mm (compression)	28–6	290	785	1843	4106	8 944	290–8 944
+5 mm (compression)	30–4	184	665	1843	4736	11 837	184–11 837
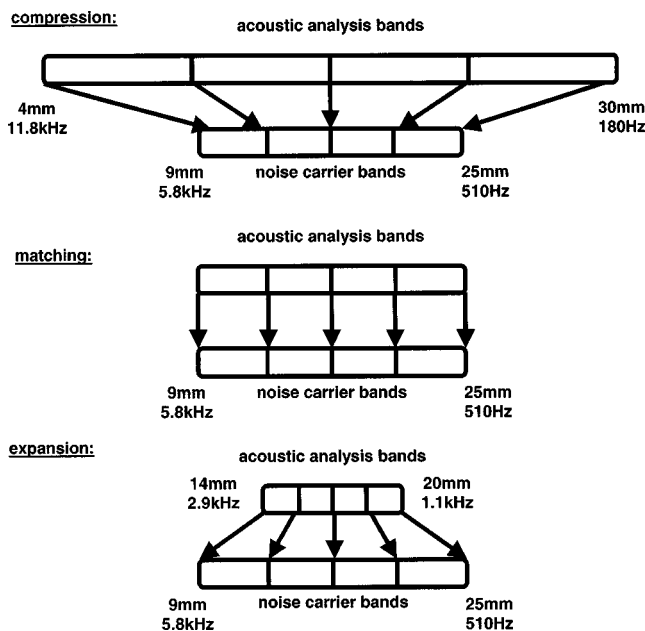
FIG. 1. Frequency-place mapping conditions for 4-channel processor at the simulated 25-mm electrode insertion depth. For this condition the noise carrier bands were fixed (9–25 mm: 510–5800 Hz). The speech envelope was extracted from the analysis bands and used to modulate the noise carrier bands. The top panel shows the +5-mm compression condition schematically: the analysis bands are mapped onto narrower carrier bands. The middle panel shows the 0-mm condition schematically, in which the analysis and carrier bands are matched. The lower panel shows the −5-mm expansion condition schematically, in which analysis bands are mapped onto wider carrier bands.

the stimulation region was fixed at 16 mm, electrode locations were represented by noise bands that were 4 mm wide in terms of cochlear location for the 4-band condition, 2 mm wide for the 8-band condition, and 1 mm wide for the 16-band condition.

In the simulation, the noise carrier bands determine the cochlear location stimulated. The "acoustic analysis bands" are the filters used to process and extract the acoustic envelope information used to modulate the carrier bands (Fig. 1). The distribution of carrier bands was kept fixed for each simulated insertion depth while the analysis bands were systematically altered to create the conditions of frequency-place expansion or compression.

The cutoff frequencies of each analysis band and the related cochlear locations for the 4-band processors are summarized in Table I for the 20-mm simulated insertion depth, and Table II for the 25-mm simulated insertion depth. Filters for the 8- and 16-band conditions were determined by dividing the four bands into two or four equal parts (in mm) and using Greenwood's (1990) formula to determine the acoustic frequencies for the band edges. The 0-mm condition refers to a perfect match between analysis and carrier bands. Thus, the cochlear location and frequency cutoffs listed for the 0-mm condition specify the fixed locations and frequencies of both analysis and carrier bands. Cochlear locations are all specified in terms of mm from the round window, using Greenwood's (1990) formula, assuming a 35-mm cochlear length for the human. In the +5-mm condition, the analysis band range was 5 mm wider than the carrier band range on both

apical and basal ends, causing a frequency-place compression of approximately 2 octaves. Similarly, in the −5-mm condition, the analysis band range was 5 mm shorter on each end, causing a frequency-place expansion of about 2 octaves. The +5-mm compression condition in Table I most closely simulates the typical frequency-place compression observed in the standard clinical map of the Nucleus speech processors. Also, +5-mm compression and −5-mm expansion conditions for the simulated 25-mm insertion depth are shown schematically in Fig. 1, as well as the matching case. Here, one can see that the simulated electrode locations are from 9 to 16 mm from the round window and the frequency range used to modulate each noise carrier band is different for every condition.

As the analysis filters were changed from −5-mm to +5-mm conditions, the amount of acoustic information delivered was changed. An important control condition was included to evaluate the effect of the varying amount of acoustic information. In these control conditions the analysis bands and noise carrier bands were always matched in frequency place. These baseline conditions were not intended to simulate any electrode insertion depth or spacing, because in a cochlear implant the electrode position and length are fixed after the implant surgery. Rather, the baseline conditions only assess the effect of changing the overall acoustic bandwidth. Performance in the baseline condition indicates the effect of the gain or loss of acoustic information resulting from the expansion or truncation of the analysis frequency range. The difference in performance between the baseline condition and the compression–expansion condition is due to the frequency-place distortion only.

III. RESULTS

Percent-correct scores for consonants, vowels, and sentences were obtained with 4-, 8-, and 16-channel processors at simulated insertion depths of 20 and 25 mm. In Figs. 2–9, the number of channels increases from 4 to 8 to 16 in the left, middle, and right panels, respectively, of each figure. The average percent-correct scores of six subjects, corrected for chance $[p = 100*(\text{score-chance})/(100\text{-chance})]$, are plotted for consonant and vowel recognition. The average score of six subjects is plotted for sentence recognition. Within each panel the filled symbols present results from the baseline conditions in which analysis and carrier bands were always matched, and the open symbols present results from the experimental conditions in which the frequency-place mapping was expanded or compressed.

A. Consonants

The consonant recognition scores are presented in Fig. 2 for the 20-mm simulated insertion depth and Fig. 3 for the 25-mm simulated insertion depth.

First, consider the data from the baseline conditions (filled symbols), where the analysis and carrier bands were always matched. Consonant recognition increased only slightly as the analysis (and matched carrier) bands were widened (+5-mm baseline condition) from the simulated electrode range. However, as the analysis and carrier bands were narrowed (−5-mm baseline condition) there was a loss
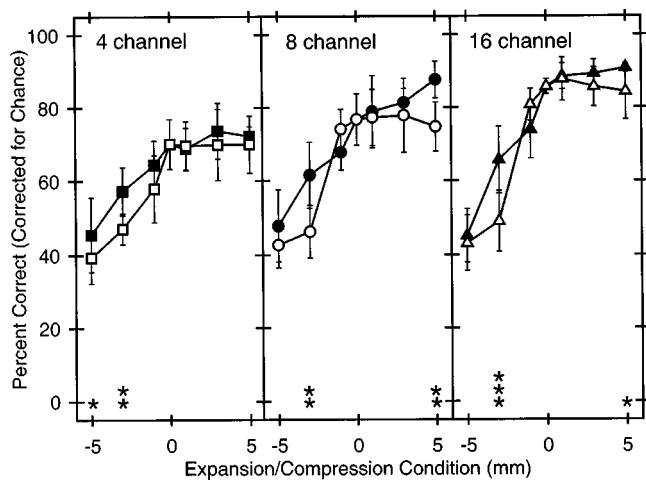
FIG. 2. Consonant recognition percent scores for the simulated 20-mm electrode insertion depth, as a function of compression or expansion in the frequency-place mapping. The number of spectral bands increases from 4 to 8 to 16 in the left, middle, and right panels, respectively. The percent-correct scores represent the average performance of 6 normal-hearing subjects, corrected for chance, and the error bars represent the standard deviation. Filled symbols denote the baseline condition where the carrier bands were always matched to the analysis band frequency range. Open symbols denote the compression–expansion conditions where the carrier bands were fixed and the analysis bandwidth was varied. The stars at the bottom of the figure denote significant differences between the baseline (filled symbols) and mismatch (open symbols) conditions: one star indicates $p < 0.05$, two stars for $p < 0.01$, and three stars for $p < 0.001$.

FIG. 3. Consonant recognition, similar to Fig. 2, but for carrier bands simulating a 25-mm insertion depth.

of acoustic information (due to the reduced acoustical bandwidth) that resulted in lower consonant recognition. This reduction in consonant recognition was more severe in the simulated 20-mm insertion depth condition (Fig. 2) because more low-frequency information was eliminated.

Next, consider the data from the experimental mismatch conditions (open symbols), where the analysis filter frequency range was smaller, equal to, or larger than the simulated electrode length, resulting in frequency-place expansion, matching, or compression, respectively. Note that performance in the mismatched conditions was always equal to or poorer than the baseline conditions. Thus, there are two contributing factors to the reduced performance: (1) the reduction in the amount of information delivered, and (2) the distortion in frequency-place mapping.

A one-way repeated-measures ANOVA test was used to assess the significance of the drop in the performance with
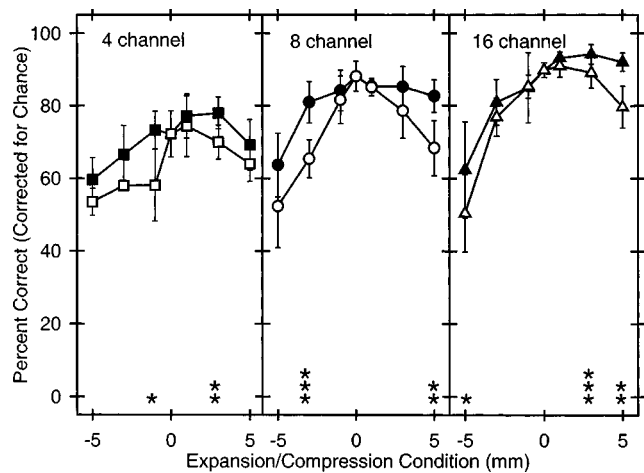
expansion/compression mismatch conditions from the matched condition. Each run included only the expansion or compression percent scores in addition to the 0-mm matched percent score at a particular insertion depth with a specific number of channels. The baseline scores were not included in the ANOVA to isolate the effect of expansion or compression only on speech recognition. The baseline condition was compared to the corresponding mismatch condition with paired t-tests.

The analysis revealed that all frequency expansion conditions reduced performance significantly from the 0-mm matched condition. Corresponding F and p values are shown in Table III. A substantial amount of this drop was due to the loss of acoustic information, as indicated by the filled symbols. An additional drop was observed for some conditions when the frequency-place mapping was expanded (open symbols).

A paired t-test analysis compared the baseline (filled symbols) and expansion (open symbols) performance. Conditions that are significantly different are indicated with stars on the bottom of the figure. One star denotes significant difference with a p value of $p < 0.05$, two stars denote $p < 0.01$, and three stars denote $p < 0.001$. The analysis shows a significant difference for −3-mm expansion for most processors at both insertion depths. The difference was generally not significant for −5-mm expansion, possibly because the

TABLE III. F and p values calculated with one-way repeated-measures ANOVA for expansion and compression mismatch conditions for consonant recognition at 20-mm and 25-mm simulated insertion depths.

Expansion at 20-mm insertion	$F(3,15)$	p	Compression at 20-mm insertion	$F(3,15)$	p
4 channel	54.60	<0.001	4 channel	0.01	1
8 channel	89.05	<0.001	8 channel	0.23	0.87
16 channel	117.79	<0.001	16 channel	0.66	0.59
Expansion at 25-mm insertion	$F(3,15)$	p	Compression at 25-mm insertion	$F(3,15)$	p
4 channel	15.10	<0.001	4 channel	7.40	<0.01
8 channel	62.60	<0.001	8 channel	26.99	<0.001
16 channel	90.26	<0.001	16 channel	15.38	<0.001

performance was limited by a floor effect (previous studies have found 30%–40%-correct consonant recognition even for single-channel noise processors, indicating that this level of performance is possible using only temporal cues: Van Tasell *et al.*, 1987; Shannon *et al.*, 1995). The difference was also not significant for −1-mm expansion, which produced performance similar to the 0-mm matched condition.

Thus, even though the cochlear tonotopic representation of the spectral information was expanded, resulting in improved spectral resolution within the pattern, performance was poorer than the matched condition. This result suggests that improved resolution in the spectral domain does not necessarily improve speech recognition, probably because the information is not in the appropriate cochlear place. In the present experiment, which did not provide any practice or time to accommodate to the new mapping, expansion in frequency-place mapping always resulted in poorer consonant recognition. It is possible that additional practice with the experimental processors would have resulted in improved performance.

Frequency-place compression did not have a significant effect on consonant recognition for the 20-mm simulated insertion depth, yet there was significant decrease from the matched 0-mm condition (10%–20% drop at +5-mm compression) in performance for the 25-mm simulated insertion depth (as shown in Table III). For the extreme condition of frequency-place compression (+5 mm) there was a significant reduction in performance of 10%–15% relative to the baseline condition for 8- and 16-channel processors.

There was no clear difference between the pattern of results for 4, 8, and 16 channels other than the overall improvement in performance with more channels.

B. Consonant feature analysis

Information transmitted on the consonant features of place, manner, and voicing is plotted in Figs. 4 and 5 for 20-mm and 25-mm simulated insertion depths, respectively. Within each figure the top, middle, and lower panels show the percent of information transmitted on place, manner, and voicing, respectively.

Information transmission percent scores are calculated from the confusion matrices where the diagonal entries are the numbers of correct answers and the off-diagonals are the confusions. A measure for the transmission of information is the covariance between the input and the output, as given by

$$\text{Cov}(x,y) = -\sum_{i,j} p_{ij} \log \frac{p_i p_j}{p_{ij}},$$

where p_i, p_j, p_{ij} are directly related to the frequencies of occurrences of stimulus i, response j, and joint occurrence of stimulus i and response j, respectively. Next, the covariance is converted to the information transmission percent score by normalization such that it yields 100% when the subject identifies all phonemes accurately (Miller and Nicely, 1955).

Note that information received on manner at 20-mm simulated insertion and on voicing at both insertion depths was not affected by frequency-place compression, and man-
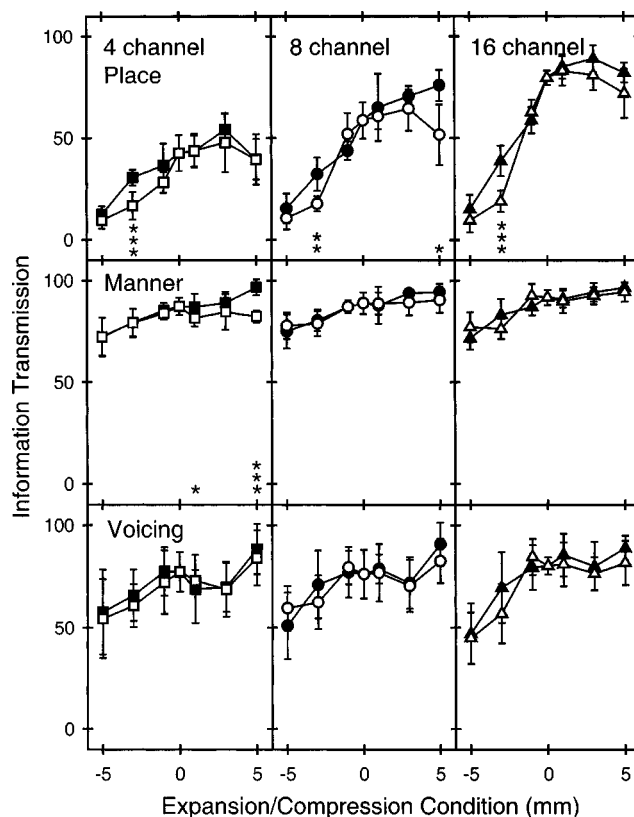

FIG. 4. Information transmission percent scores for consonant features at 20-mm simulated insertion depth as a function of frequency-place mismatch conditions. The features are grouped into production-based categories of place, manner, and voicing, in the top, middle, and bottom rows, respectively. The number of spectral bands increases from 4 to 8 to 16 in the left, middle, and right panels, respectively.

ner information received at 25-mm simulated insertion depth was only slightly affected. Yet, both manner and voicing information transmission scores dropped significantly with expansion (see Table IV for corresponding F and p values). Also, the compression/expansion mismatch scores were similar to baseline scores for manner and voicing, implying that these features are mostly affected by the bandwidth of acoustic information. The performance for both features was similar for different number of channels and for the two simulated insertion depths.

In contrast, place information was strongly affected by expansion at both simulated insertion depths, and by compression at 25-mm simulated insertion depth (Table IV). The overall pattern of the performance changed from 20-mm insertion to 25-mm insertion, and the information received increased with increasing number of channels. This pattern is very similar to the consonant recognition results of Figs. 2 and 3, and therefore it appears that the overall shape of consonant recognition performance was primarily determined by the loss of place information. This observation agrees well with previous studies, which found that manner and voicing cues are more robust to spectral manipulations (Shannon *et al.*, 1998).

C. Vowels

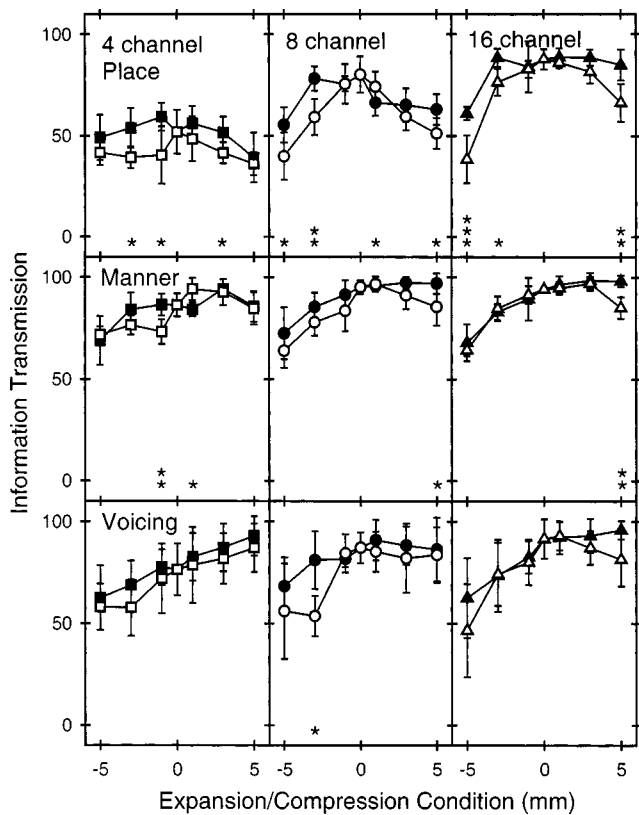Vowel recognition scores for simulated 20-mm and 25-mm insertion depths are presented in Figs. 6 and 7, re-

FIG. 5. Consonant feature information transmission, similar to Fig. 4, but for carrier bands simulating a 25-mm insertion depth.
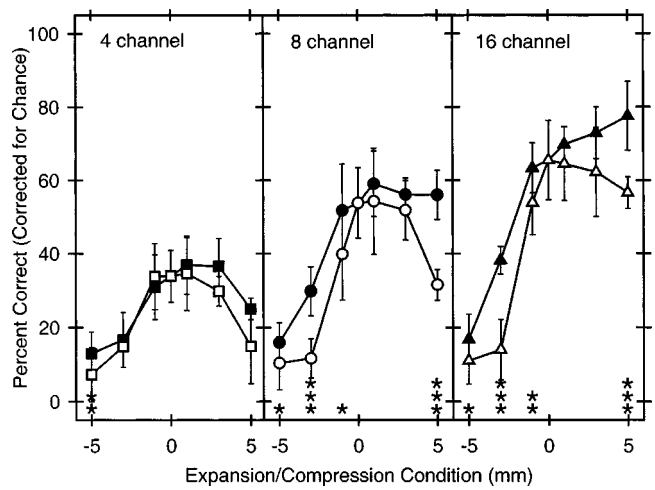

FIG. 6. Vowel recognition percent scores for carrier bands simulating a 20-mm insertion depth, as a function of compression or expansion in the frequency-place mapping. The number of spectral bands increases from 4 to 8 to 16 in the left, middle, and right panels, respectively. The percent-correct scores represent the average performance of 6 normal-hearing subjects, corrected for chance, and the error bars represent the standard deviation. Filled symbols denote the baseline condition where the carrier bands were always matched to the analysis band range. Open symbols denote the compression–expansion conditions where the carrier bands were fixed and the analysis bandwidth was varied. The stars at the bottom of the figure denote significant differences between the baseline (filled symbols) and mismatch (open symbols) conditions: one star indicates $p < 0.05$, two stars for $p < 0.01$, and three stars for $p < 0.001$.

spectively. Vowel recognition was much more strongly affected by frequency-place mismatch than consonant recognition; performance decreased significantly from the matched 0-mm condition with both expansion and compression mis-

match conditions (F and p values obtained by one-way repeated-measures ANOVA are given in Table V). Overall performance improved as more spectral bands were used but the pattern of results was similar for 4, 8, and 16 bands. For 8 and 16 channels, vowel recognition decreased as the analy-

TABLE IV. F and p values calculated with one-way repeated-measures ANOVA for expansion and compression mismatch conditions for consonant feature recognition at 20-mm and 25-mm simulated insertion depths.

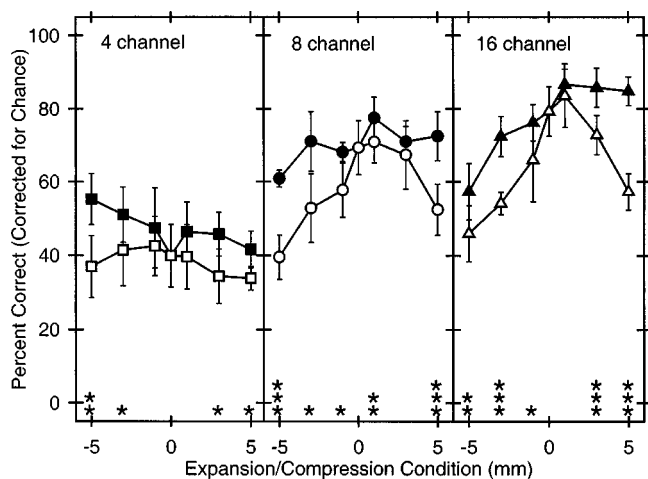| Expansion at 20-mm insertion	Place		Manner		Voicing	
	$F(3,15)$	p	$F(3,15)$	p	$F(3,15)$	p
4 channel	63.42	<0.001	9.09	<0.01	6.90	<0.01
8 channel	63.01	<0.001	10.18	<0.001	12.97	<0.001
16 channel	347.90	<0.001	21.15	<0.001	25.77	<0.001
Compression at 20-mm insertion						
4 channel	1.55	0.24	1.51	0.25	4.47	<0.05
8 channel	1.48	0.26	0.16	0.92	3.52	<0.05
16 channel	2.49	0.10	2.43	0.11	0.62	0.62
Expansion at 25-mm insertion						
4 channel	2.19	0.13	8.98	<0.01	6.10	<0.01
8 channel	38.92	<0.001	24.90	<0.001	16.32	<0.001
16 channel	55.35	<0.001	72.88	<0.001	26.62	<0.001
Compression at 25-mm insertion						
4 channel	7.46	<0.01	3.77	<0.05	1.19	0.35
8 channel	25.38	<0.001	5.06	<0.05	0.75	0.54
16 channel	15.59	<0.001	15.43	<0.001	6.04	<0.01

FIG. 7. Vowel recognition percent scores for noise carrier bands simulating a 25-mm electrode insertion depth, as a function of compression or expansion in the frequency-place mapping.
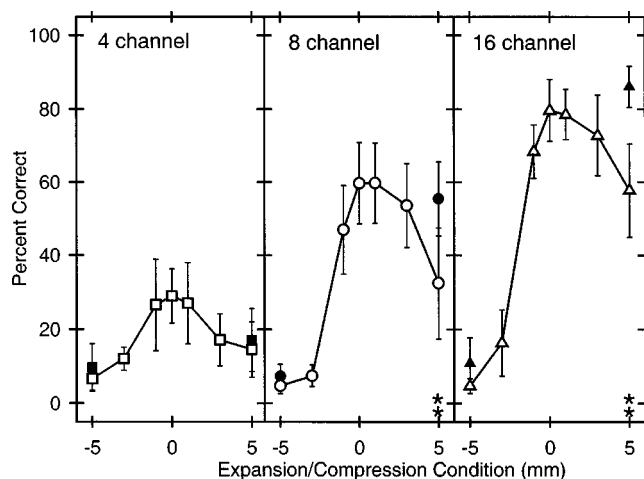

FIG. 8. Sentence recognition percent scores for noise carrier bands simulating a 20-mm electrode insertion depth, as a function of compression or expansion in the frequency-place mapping. The number of channels increases from 4 to 8 to 16 in the left, middle, and right panels, respectively. The percent-correct scores represent the average performance of 6 normal-hearing subjects, and the error bars represent the standard deviation. Filled symbols denote the baseline condition where the carrier bands were always matched to the analysis band frequency range. Open symbols denote the compression–expansion conditions where the carrier bands were fixed and the analysis bandwidth was varied.

sis frequency range was reduced (baseline condition going from +5 mm to −5 mm), and a further drop in recognition was seen for both frequency-place compression and expansion as a result of the mismatch.

As in consonant recognition the stars at the bottom of each panel indicate significant difference between baseline (filled symbols) and mismatch (open symbols) conditions determined by paired t-test.

For 8- and 16-channel processors both frequency-place expansion of −5 mm and compression of +5 mm resulted in a 20%–30% drop in recognition compared to the matched condition (the flat performance with the 4-channel processor at 25-mm insertion depth may have been limited by the overall poor level of performance). Note that the Nucleus cochlear implant processor mentioned above typically uses a frequency-place assignment that is similar to the +5-mm compression condition, which produced a significant reduction in vowel recognition.

Although a reduction in the acoustic frequency range normally causes a drop in performance, Fig. 7 shows an improvement in vowel recognition for the baseline expansion condition for 4-channel processor at 25-mm insertion [$F(6,30)=4.13, p<0.01$]. This improvement could be due either to an increase in resolution or to a better frequency partition of the analysis bands.

Note that there was probably a floor effect for the expansion conditions with the simulated 20-mm insertion, where increasing the number of channels did not increase the intelligibility (Fig. 6, leftmost data point of each panel). Both −5-mm and −3-mm expansion results were close to chance level, which might be due to the loss of all low-frequency information below 1850 Hz (−3-mm condition) or below 2476 Hz (−5-mm condition).

D. Sentences

The percentage of words recognized in TIMIT sentences is presented in Figs. 8 and 9 for 20-mm and 25-mm simulated insertion depths, respectively. Due to the limited number of sentence sets available, the matched baseline performance was measured only for extreme mismatch conditions (−5-mm expansion and +5-mm compression). For all numbers of channels and both simulated insertion depths the best performance was obtained when the analysis and carrier bands were matched.

TABLE V. F and p values calculated with one-way repeated-measures ANOVA for expansion and compression mismatch conditions for vowel recognition at 20-mm and 25-mm simulated insertion depths.

| Expansion at 20-mm insertion	$F(3,15)$	p	Compression at 20-mm insertion	$F(3,15)$	p
4 channel	40.26	<0.001	4 channel	14.28	<0.001
8 channel	78.55	<0.001	8 channel	14.86	<0.001
16 channel	235.08	<0.001	16 channel	2.05	0.15
Expansion at 25-mm insertion	$F(3,15)$	p	Compression at 25-mm insertion	$F(3,15)$	p
4 channel	1.29	0.31	4 channel	1.95	0.17
8 channel	32.67	<0.001	8 channel	22.31	<0.001
16 channel	39.54	<0.001	16 channel	32.94	<0.001
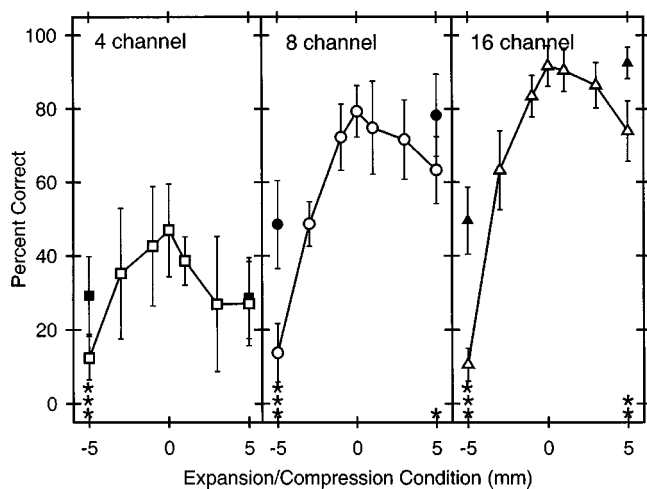
FIG. 9. Sentence recognition percent scores for noise carrier bands simulating a 25-mm electrode insertion depth, as a function of compression or expansion in the frequency-place mapping.

Note that the drop in performance for frequency-place expansion (-5 mm) was dramatic compared to the matched condition (Table VI). Performance at the -5-mm condition drops to 5% correct for the 20-mm simulated insertion depth, and 15% for 25-mm insertion depth for all three spectral resolutions. For the 16-channel processor this drop was 75–80 percentage points. Although much of this loss was due to the information loss (filled symbols), there was an additional 20–40-point drop in performance due to the frequency-place expansion at the 25-mm simulated insertion depth.

A smaller drop in performance was observed for frequency-place compression. There was a drop of 15–25 percentage points from the 0-mm baseline condition to the $+5$-mm baseline condition, even though the overall frequency range increased by about 2 octaves. There was an additional drop of 20–30 percentage points from the $+5$-mm baseline to the $+5$-mm compression condition. This $+5$-mm compression condition is similar to the typical frequency-place assignment used in Nucleus cochlear implants.

IV. DISCUSSION

The best speech recognition performance was always observed in conditions where frequency information was matched to its normal acoustic cochlear place. In most conditions, altering the frequency-place mapping by either compression or expansion resulted in poorer speech recognition.

A. Implications for cochlear implants

The present study observed a reduction of approximately 20 percentage points in vowel and sentence recognition when a frequency range was compressively mapped onto a cochlear range that was smaller by 2 octaves ($+5$-mm compression condition). Even though a broader frequency range of acoustic information is presented in this condition, performance was reduced due to the distortion in the frequency-place assignment. This compressive frequency-to-place mapping is similar to the mapping used in Nucleus cochlear implant systems, in which the acoustic frequency range of 150 Hz to 10 kHz is typically mapped onto electrodes that occupy the cochlear locations that normally respond to an acoustic range of only 500–6000 Hz. This result implies that speech recognition performance in cochlear implants might be improved by as much as 20% if the frequency range for each electrode could be mapped according to the normal acoustic characteristic frequency of that cochlear location.

How can a cochlear implant speech processor be adjusted to achieve the best mapping of frequency information onto the most appropriate cochlear place, given the variability in cochlear length and electrode insertion depth across patients? In implant listeners there is uncertainty in the exact location of the electrodes and further uncertainty as to the location of the stimulated neurons. Recent advances in imaging technology allow sufficient resolution to evaluate the depth of electrode insertion and to detect the presence of any kinks or abnormalities in the electrode carrier (Ketten *et al.*, 1998). However, these imaging procedures are costly, time consuming, deliver large doses of radiation, and still may not provide all of the necessary information. For example, even knowledge of the exact cochlear location of an electrode is no guarantee that the stimulation of neurons is actually occurring at that location. The actual stimulation location can be affected by the pattern of local nerve survival or by unusual current pathways due to bone growth and fibrous blockage. In addition, the actual site of stimulation may be in the spiral ganglion, whereas Greenwood's formula holds for stimulation at the basilar membrane. These factors produce additional uncertainties regarding the appropriate frequency-place mapping in implant patients. The data presented above

TABLE VI. F and p values calculated with one-way repeated-measures ANOVA for expansion and compression mismatch conditions for sentence recognition at 20-mm and 25-mm simulated insertion depths.

Expansion at 20-mm insertion	$F(3,15)$	p	Compression at 20-mm insertion	$F(3,15)$	p
4 channel	18.53	<0.001	4 channel	6.87	<0.01
8 channel	66.88	<0.001	8 channel	8.76	<0.01
16 channel	256.22	<0.001	16 channel	13.91	<0.001
Expansion at 25-mm insertion	$F(3,15)$	p	Compression at 25-mm insertion	$F(3,15)$	p
4 channel	11.64	<0.001	4 channel	4.29	<0.05
8 channel	147.89	<0.001	8 channel	3.59	<0.05
16 channel	233.12	<0.001	16 channel	22.17	<0.001

control for these factors by making the measurements in normal-hearing listeners, in whom the actual stimulation locations can be controlled, at least within the constraints of the normal acoustic spread of excitation. Due to these uncertainties, the optimal frequency-place alignment in an implant patient might best be determined functionally, by adjusting the frequency range and distribution across the implanted electrode array to achieve the best performance. A simple optimizing algorithm, paired with a sensitive phonetic contrast test, could provide an efficient method for converging on an optimal frequency-place mapping for an individual patient, without the costs and risks of x rays and CT scans. The present results may help to define the inherent trade-offs between electrode depth, number of electrodes, and frequency range.

Another uncertainty comes from anatomical and geometrical issues regarding the stimulation of deeper turns of the cochlea. Electrodes in cochlear implants do not always reside between 9 and 25 mm or between 4 and 20 mm inside the round window, the two conditions simulated in this study. The latest generation of CI electrodes, such as Clarion HiFocus, Nucleus Contour, or Med-El Combi40+ offers deeper insertion, possibly up to 30 mm. Even though these specially designed electrodes make it possible to reach more apical locations inside the cochlea, it is unknown if it is possible to stimulate the spiral ganglia corresponding to low frequencies. Cell bodies of the spiral ganglia from the apical turn of the cochlea are located in the modiolus of the cochlear middle turn, and so are physically (and presumably electrically) closer to electrodes in the middle turn than to the medial wall of the cochlea in the apical turn. Studies of pitch have shown little change in pitch with electrode location for electrodes that were deeply inserted into the apical turn, suggesting that there may be a point of diminishing returns in terms of electrode insertion depth (Cohen *et al.*, 1996). And, even with the new electrode designs, the array cannot always be fully inserted due to cochlear ossification or otosclerosis. Thus, the actual location of the implanted electrode is difficult to determine accurately, and the location of the neurons actually stimulated by each electrode adds a further layer of uncertainty.

New studies are presently being conducted in patients with substantial residual hearing, where the electrode array is only shallowly inserted in an attempt to preserve any residual acoustic hearing. In these special cases it may be particularly important to assign the appropriate frequency-place mapping, because the electrically stimulated hearing must combine with residual acoustic hearing. Indeed, preliminary results (Turner and Gantz, 2001; Brill *et al.*, 2001) suggest that combined electric and acoustic hearing is best when the electrodes in the basal turn receive high-frequency information that is matched to their tonotopic location.

B. Trade-off between spectral resolution and overall bandwidth

Some of the present results indicate a trade-off between spectral resolution (number of bands) and overall bandwidth. For a given number of bands there may be an optimal bandwidth—too small a bandwidth would discard too much important speech information, and too wide a bandwidth would increase the frequency range of each band, reducing the relative resolution. Consider the baseline conditions (filled symbols in Figs. 2–9). In these conditions the analysis bands were always matched in frequency to the carrier bands, while the number of bands was held constant. For the 16-band conditions, a larger bandwidth generally produced better performance whereas for 8 bands the performance was generally unchanged as the overall bandwidth was increased relative to the standard matched condition (0 mm). However, when there were only 4 bands of spectral resolution available, there was a complex interaction between the bandwidth and spectral resolution. In many cases, performance dropped both when the bandwidth was increased, and decreased relative to the standard matched condition. As bandwidth decreased the relative spectral resolution increased, but this was not sufficient to offset the loss of information. As bandwidth increased the additional spectral information was offset by the loss of relative spectral resolution (e.g., Figs. 6, 8, and 9: compare 4-channel 0-mm and +5-mm baseline conditions), resulting in poorer performance in spite of the larger bandwidth.

An expansion in the frequency-place mapping could theoretically improve speech recognition by spreading out the critical speech spectral region to a larger range in the cochlea. Echolocating animals have evolved such a strategy to provide better tonotopic resolution in the small frequency region of their echo signal. However, in the present study such expansion conditions mostly resulted in poorer speech recognition. The only exception was for the 4-channel processor with a 25-mm simulated insertion depth. This was likely due to an artifact of band edge placement: the 0-mm condition contained no band division in between 999 and 1843 Hz (see Table II), while the −5-mm expansion condition contained a band division at 1472 Hz, which is an important frequency for distinguishing high from low second formant frequencies. In this particular condition, the contribution from this better frequency partition might have compensated for the loss of bandwidth. With a limited number of bands, the placement of the frequency divisions appears to be more important than the overall frequency range. Alternatively, the slight improvement in performance in the expansion condition could be due to the improved resolution in this condition. The small frequency range of 1168–2864 Hz was represented across a larger cochlear region that would have normally responded to a range of 513–5860 Hz. This expansive mapping may have helped recognition by stimulating a larger neural population with information from the smaller frequency range. Whichever explanation is correct (better band partition or expanded representation) the same effects were not observed with more than 4 bands.

C. Potential effects of learning

One aspect of speech pattern recognition not addressed by the present study is the potential effect of learning. Recent work has demonstrated that NH subjects listening to simulations of cochlear implants can improve their scores on speech recognition with only a modest amount of practice (Rosen *et al.*, 1999). Rosen *et al.* used noise-band vocoders

in which the frequency-place mapping was shifted basally by as much as 6.5 mm. Listeners improved significantly in their ability to recognize phonemes and words with these shifted representations after only a few hours of training. However, their performance after this limited amount of training was still far poorer than their recognition with the unshifted speech. It is not clear if further training would allow complete recovery of performance to the unshifted levels. Fu *et al.* (2002) measured speech recognition in three cochlear implant listeners after a 3-mm apical shift in the frequency-to-electrode assignments. Initially, speech recognition was reduced dramatically. After 10 days of everyday experience with the shifted map there was a significant improvement in recognition, but then only little further improvement was observed over the next 3 months. This result suggests that there may be a limit to the amount of possible relearning. It is not clear if listeners would be able to adapt to a frequency-place compression or expansion over time. In the present experiments the emphasis was solely on speech pattern recognition with no practice or time for accommodation.

D. Implications for speech pattern recognition

The present results illustrate the limitations of central pattern recognition mechanisms for speech, which may provide insights into the critical parameters of the pattern storage and retrieval process of central nervous system. Even though it is known that the patients adapt over time and improve their understanding of speech, it is still unclear how much plasticity exists in these central nervous system mechanisms and whether the ability to accommodate to some types of alterations (e.g., frequency-place shift) might be easier than other types of alterations (e.g., nonlinear frequency-place distortion).

The pattern of results observed in the present experiment, when combined with previous results on frequency-place shifting (Fu and Shannon, 1999; Dorman *et al.*, 1997), warping (Shannon *et al.*, 1998), and frequency lowering (Braida *et al.*, 1979; Reed *et al.*, 1983) suggests that the central pattern recognition of speech is not stored in terms of an abstract pattern, but in terms of an absolute pattern. Speech recognition in healthy acoustic hearing can tolerate a small degree of distortion in this frequency-place pattern, probably to accommodate the natural range of variation in real-world listening conditions, e.g., differences in the gender of the talkers, talking speeds and styles, and different amounts of masking and interference in the listening environment. The results of the present study, combined with the results of previous studies on frequency-place distortions, suggest that speech patterns can tolerate only a relatively small amount of distortion (2–3 mm) in tonotopic space. If the peripheral representation of the pattern of speech information is shifted, warped, expanded, or compressed beyond this tolerated cochlear distance of 2–3 mm, speech recognition will be significantly reduced. (It should be noted that none of these studies gave the subjects the opportunity to adapt to the distorted mappings.)

Figure 10 presents a schematic representation of a vowel spectrum and the various types of distortion that result in a reduction in multitalker vowel recognition to approximately
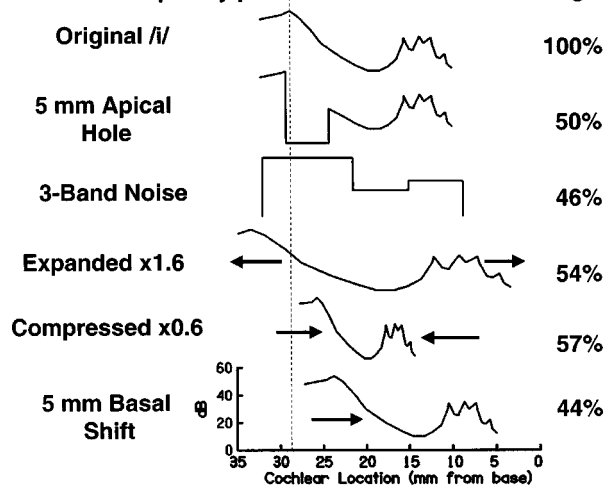

FIG. 10. A review of various frequency-place distortions on vowel recognition. The top curve is the original spectral representation of the vowel /i/ plotted in terms of cochlear distance. The second curve shows the same vowel with a 5-mm apical hole. Next is the vowel represented by a 3-band noise vocoder. The fourth curve is the frequency-place expansion with an expansion factor of 1.6 (10-mm extent expanded to 16 mm), followed by the frequency-place compression with a compression factor of 0.6 (26-mm extent compressed to 16 mm). Finally, the bottom curve shows the spectrum of an /i/ that has been shifted by 5 mm basally in the cochlea.

50% correct. These representations will be altered as they are processed by the cochlea and the central nervous system, but for simplicity they are shown here by their physical spectral representation in terms of distance along the cochlea. The top curve shows the original spectrum of the vowel /i/, presented in terms of mm along the cochlea. In this undistorted representation listeners will generally be able to identify 12 vowels at nearly 100% correct, even with multiple talkers and with the spectral resolution reduced to 16 channels. The second curve shows the same vowel in which spectral information has been removed to create a 5-mm hole in the apical spectral region, resulting in a drop to 50%-correct recognition (Shannon *et al.*, 2001). The third curve shows the vowel represented by a three-band noise vocoder, which allows 46% correct on multitalker vowel recognition (Fu *et al.*, 1998). The fourth curve shows the effect of a frequency-place expansion by a factor of 1.6, which results in 54%-correct vowel recognition with 16 bands (−3-mm condition from Fig. 7). The fifth curve shows the effect of frequency-place compression by a factor of 0.6, which results in 57%-correct recognition (+5-mm condition from Fig. 7). And, the bottom curve shows the spectrum of an /i/ that has been shifted 5 mm basally in the cochlea, resulting in 44%-correct vowel recognition (from Fu and Shannon, 1999). This comparison suggests that the central pattern recognition mechanisms are sensitive to the absolute tonotopic location of the cochlear pattern. If the frequency-place information is in the correct location, the central pattern recognition can tolerate the loss of a full octave of spectral information in the critical low-frequency region, or an extreme loss of spectral resolution—down to three bands. However, if the pattern is distorted by a frequency-place shift or compression or expansion, then speech recognition is impaired even with good spectral resolution. In terms of cochlear implants, even

if an implant patient is able to use many electrodes effectively, their performance might be limited by distortion in the frequency-place mapping.

It is possible that distortion in frequency-place mapping is responsible for at least part of the variability in performance across implant patients. If this is the case, then adjustments to the speech processor to produce a better match in frequency-place mapping may produce improvements in speech recognition.

ACKNOWLEDGMENTS

Advanced Bionics Corporation (**2001**). CII Bionic Ear Programming System.

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Rabinowitz, W. M., and Reed, C. M. (**1979**). "Hearing aids: A review of past research on linear amplification, amplitude compression, and frequency lowering," ASHA Monogr. **19**, 87–108.

Brill, S., Lawson, D., Wolford, R., Wilson, B., and Schatzer, R. (**2001**). Speech Processors for Auditory Prostheses, 11th Quarterly Progress Report, NIH Contract N01-DC-8-2105.

Cochlear Corporation (**1995**). Technical Reference Manual. Englewood, CO.

Cohen, L. T., Busby, P. A., Whitford, L. A., and Clark, G. M. (**1996**). "Cochlear implant place psychophysics. I. Pitch estimation with deeply inserted electrodes," Audiol. Neuro-Otol. **1**, 265–277.

Daniloff, R. G., Shiner, T. H., and Zemlin, W. R. (**1968**). "Intelligibility of vowels altered in duration and frequency," J. Acoust. Soc. Am. **44**, 700–707.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.

Echteler, S. M., Fay, R. R., and Popper, A. N. (**1994**). "Structure of the mammalian cochlea," in *Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 134–171.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Fu, Q.-J., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Gstoettner, W., Franz, P., Hamzavi, J., Plenk, H., Baumgartner, W., and Czerny, C. (**1999**). "Intracochlear position of cochlear implant electrodes," Acta Otolaryngol. (Stockh) **119**, 229–33.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). "*In vivo* measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays," Ann. Otol. Rhinol. Laryngol. Suppl. **175**, 1–16.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Nagafuchi, M. (**1976**). "Intelligibility of distorted speech sounds shifted in frequency and time in normal children," Audiology **15**, 326–337.

National Institute of Standards and Technology (**1990**). DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. Gaithersburg, MD.

Reed, C. M., Hicks, B. L., Braida, L. D., and Durlach, N. I. (**1983**). "Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering," J. Acoust. Soc. Am. **74**, 409–419.

Robert, M. E. (**1998**). CONDOR: *Documentation for Identification Test Program* (House Ear Institute, Los Angeles, CA).

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.

Shannon, R. V., Galvin, J. G. III, and Baskent, D. (**2001**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**(2), 185–199.

Tiffany, W. R., and Bennett, D. A. (**1961**). "Intelligibility of slow-played speech," J. Speech Hear. Res. **4**, 248–258.

Turner, C. W., Fabry, D. A., Barrett, S., and Horowitz, A. R. (**1992**). "Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **35**, 942–949.

Turner, C. W., Chi, S.-L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Hear. Res. **42**(4), 773–784.

Turner, C., and Gantz, B. (**2001**). "Combining acoustic and electric hearing for patients with high frequency hearing loss," Abstracts of the 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, p. 33.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.