The effects of short-term training for spectrally mismatched noise-band speech

Qian-Jie Fu[a)] and John J. Galvin III
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 2 May 2002; accepted for publication 18 November 2002)

The present study examined the effects of short-term perceptual training on normal-hearing listeners' ability to adapt to spectrally altered speech patterns. Using noise-band vocoder processing, acoustic information was spectrally distorted by shifting speech information from one frequency region to another. Six subjects were tested with spectrally shifted sentences after five days of practice with upwardly shifted training sentences. Training with upwardly shifted sentences significantly improved recognition of upwardly shifted speech; recognition of downwardly shifted speech was nearly unchanged. Three subjects were later trained with downwardly shifted speech. Results showed that the mean improvement was comparable to that observed with the upwardly shifted training. In this retrain and retest condition, performance was largely unchanged for upwardly shifted sentence recognition, suggesting that these listeners had retained some of the improved speech perception resulting from the previous training. The results suggest that listeners are able to partially adapt to a spectral shift in acoustic speech patterns over the short-term, given sufficient training. However, the improvement was localized to where the spectral shift was trained, as no change in performance was observed for spectrally altered speech outside of the trained regions. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1537708]

PACS numbers: 43.71.Ky, 43.71.An, 43.66.Ts [CWT]

I. INTRODUCTION

Cochlear implants transform acoustic sounds into electrical signals that directly stimulate remaining auditory nerve fibers, thereby partially restoring hearing sensation to profoundly deaf patients. Multi-channel cochlear implant speech processors divide acoustic signals into several frequency bands, extract the temporal envelope information from each band, convert the acoustic amplitudes into electric currents, and deliver the electric currents to appropriate electrodes situated within the cochlea. To recreate the tonotopic distribution of activity within the normal cochlea, the amplitude envelope from low-frequency bands is delivered to apical electrodes (on an appropriate carrier) and the amplitude envelope from high-frequency bands is delivered to basal electrodes. The frequency-to-electrode mapping provides spectral cues for speech recognition by cochlear implant listeners. The frequency-to-electrode mapping necessarily compresses a wide acoustic frequency range onto the limited cochlear extent of the implanted electrode array. The acoustic frequency ranges mapped to the stimulating electrodes will be shifted as well as compressed. Implant patients may vary in terms of electrode insertion depths and neural populations, and the degree of spectral shifting and compression may be different for individual implant users.

The spectral shift via the frequency-to-electrode mapping can acutely affect implant listeners' speech recognition abilities. Several experiments have explored implant listeners' ability to recognize spectrally shifted speech patterns, given little or no time for adaptation. For example, Fu and Shannon (1999a,b) investigated the interaction between acoustic frequency allocation and electrode location in Nucleus-22 cochlear implant listeners using experimental four-channel continuous interleaved sampling speech processors. In these studies, vowel recognition was measured as a function of ten different frequency allocations and two sets of four-electrode configurations. Each frequency allocation represented the same cochlear extent but different cochlear locations based on Greenwood's frequency-to-place formula (Greenwood, 1990). Results showed that for a given electrode configuration, the best vowel score was obtained within only a narrow range of frequency allocations. When the location of the stimulating electrodes was shifted by 3 mm, the frequency allocation that produced the best vowel recognition also shifted by 3 mm. These results suggest that speech recognition with cochlear implants is highly sensitive to the mapping between frequency allocation and the location of the stimulating electrodes. A severe mismatch between frequency allocation and electrode location could result in a dramatic and immediate deterioration in speech performance. The effect of spectral mismatch was also explored in normal-hearing subjects listening to noise-band simulations of an implant speech processor in the same study. The carrier frequency bands were fixed while the analysis filter bands were systematically shifted. The results showed that the best performance was achieved when the analysis filter bands and the carrier filter bands were matched or closely matched. Performance was unchanged as long as the spectral mismatch between the analysis filters and carrier filters was 3 mm or less (in terms of Greenwood's function). However, the performance dropped steeply for both apically and upwardly shifted speech when the spectral mismatch between

a)Electronic mail: qfu@hei.org

analysis and carrier filters exceeded 3 mm. In conclusion, while some degree of spectral mismatch can be tolerated (~3 mm), speech recognition can suffer acutely if the spectral mismatch is too severe.

While a severe spectral mismatch may cause a significant performance drop under acute testing conditions, it is difficult to gauge the significance of the frequency-to-electrode assignments tested in such studies because subjects were given no time to adapt to the new patterns of electrical stimulation. Several studies have noted improved speech performance by cochlear implant patients after long-term exposure to new electrical stimulation patterns provided by updated speech processors, speech processing strategies, and/or clinical fitting systems (e.g., Wilson *et al.*, 1991; Pelizzone *et al.*, 1999). For example, Pelizzone *et al.* (1999) reported that, initially, vowel identification scores were unchanged for a group of Ineraid users who switched from the compressed analog (CA) to the continuous interleaved sampler (CIS) speech processing strategy. However, after 6 months of experience with the new CIS strategy, these subjects' vowel identification scores were significantly better than those obtained with the previous CA processor. When the subjects were switched back to their previous CA processor, vowel perception scores returned to the same levels measured before switching to the CIS strategy. This example illustrates that ''acute'' experiments may provide different results than those of long-term studies. As cochlear implant users become more experienced with the implant device and parametric changes to the speech processor, they may be able to largely overcome initial difficulties associated with spectral distortions.

To further investigate the importance of frequency-to-electrode assignment on speech performance by cochlear implant users, the longer-term effects of frequency-to-electrode assignment have also been explored in cochlear implant listeners (Fu *et al.*, 2002; McKay and Henshall, 2002; Skinner *et al.*, 1995). Skinner *et al.* (1995) found that six out of seven Nucleus-22 implant patients using the SPEAK strategy performed better in vowel tests with frequency allocation Table 7 (120–8658 Hz) than with frequency Table 9 (150–10 832 Hz). In the SPEAK strategy, the frequency allocation tables determine the frequency-to-electrode assignment. Skinner *et al.* argued that Table 7 assigned more electrodes to important frequency regions below 1 kHz, thereby improving vowel recognition. Similarly, McKay and Henshall (2002) investigated whether selectively increasing the discrimination of low-frequency information by altering the frequency-to-electrode allocation would improve speech perception by cochlear implant patients. Results showed that some subjects were able to adapt to frequency shifts up to ratio changes of 1.33, as well as changes in the distribution of stimulating electrodes, given 2 weeks' experience. Fu *et al.* (2002) measured speech performance over time in three Nucleus-22 cochlear implant subjects who, for a 3-month period, continuously wore experimental speech processors that were altered in terms of the frequency-to-electrode assignment. A large frequency shift was employed in the study in which the frequency boundary assigned to electrodes was lowered by 1 oct in two subjects and 0.68 oct in one subject. Baseline speech performance using each subject's clinically assigned speech processor was measured just prior to implementation of the experimental processor. Results showed that the experimental processor produced significantly lower performance on all measures of speech recognition immediately following implementation, consistent with the results from the previous ''acute'' experiments (Fu and Shannon, 1999a, b). Over the 3-month test period, all measures were significantly higher than those measured immediately post-fitting. These results indicated that long-term exposure to the frequency-shifted speech processor significantly reduced the initial performance deficit. However, even after 3 months' exposure, speech recognition with the experimental processors remained significantly lower than baseline levels measured with the clinically assigned processors, suggesting the detrimental effects of a severe spectral mismatch were indeed long-standing, though not as damaging as were first observed. Similar results have also been reported in normal-hearing subjects listening to an acoustic simulation of a cochlear implant. Rosen *et al.* (1999) examined the effect of short-term training on normal-hearing listeners' perception of spectrally shifted four-channel noise-band speech. Subjects were able to improve from nearly no recognition to identifying correctly nearly 30% of words in sentences in only 3–4 h. However, recognition of spectrally shifted speech remained significantly lower than that of unshifted speech.

These short- and long-term studies suggest that subjects may be able to completely adapt to the new speech patterns as long as the spectral mismatch is moderate (e.g., frequency shifts up to ratio changes of 1.3). However, if the spectral mismatch is severe (e.g., 1-oct frequency shift), complete adaptation may not occur within a somewhat long observation period (e.g., 3 months), though partial adaptation to the new speech patterns may occur. It is also unclear in these studies whether adaptation was restricted to the trained speech patterns or due to subjects' reshaping of internal speech representations. The present study investigated the effects of short-term learning on spectrally altered speech in six normal-hearing subjects listening to noise-band simulations of a cochlear implant speech processor. Sentence recognition scores were measured as a function of the amount of spectral distortion, which was achieved by restricting cochlear stimulation to simulate various implant electrode insertion depths, ranging from a very deep to a very shallow insertion. To simulate shallow electrode insertions, acoustic speech signals were delivered to basal cochlear locations, resulting in upwardly shifted (basally shifted) speech patterns; to simulate deep electrode insertions, acoustic speech signals were delivered to apical cochlear locations, resulting in downwardly shifted (apically shifted) speech patterns. After baseline measures for all simulated electrode locations, subjects were trained using the most upwardly shifted speech and then retested at all simulated electrode locations after the training. Subjects were also trained later using the most downwardly shifted speech and retested at all simulated electrode locations. Several hypotheses concerning training with spectrally shifted speech were explored.

The first hypothesis predicted that training with the most upwardly shifted speech might indeed improve recognition

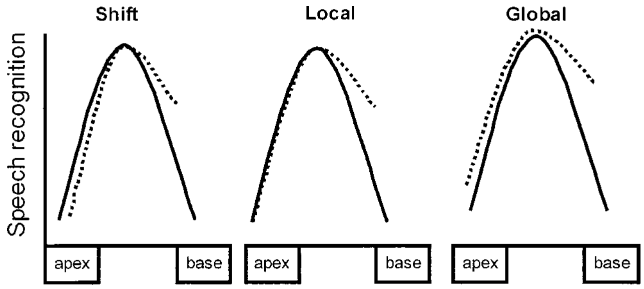

FIG. 1. Three hypothetical outcomes of training with spectrally altered speech. The solid lines represent the baseline data prior to training and the dashed lines show the predicted performance after training.

of upwardly shifted speech, but that recognition of downwardly shifted speech might be reduced; recognition of close-matched speech would not be affected. The underlying mechanism for this ''shift hypothesis'' is a slight adjustment to the internal speech representations due to training with upwardly shifted speech patterns, resulting in a larger spectral mismatch between downwardly shifted speech and the now-reshaped internal representations of speech; subjects would become ''biased'' toward the upwardly shifted speech patterns. However, because subjects would continue to make use of the normal speech patterns in listening conditions beyond the experiment, there would be no change in performance for experimental conditions in which the spectral distortions were closely matched to the long-established, normal speech patterns. The second hypothesis predicted that improved recognition might be restricted to the (upwardly shifted) trained speech patterns, while performance with the untrained (downwardly shifted) speech would not be affected. The underlying mechanism for this ''local adaptation'' hypothesis is that the subjects, in adapting to the (upwardly shifted) trained speech, develop alternative representations of speech while preserving the previous ''internal'' representations. The third hypothesis predicted that training with upwardly shifted speech would improve the recognition of both upwardly and downwardly shifted speech. The underlying mechanism of this ''global adaptation'' hypothesis is that as subjects gradually adapt to the (upwardly shifted) trained speech patterns, they develop an aptitude for adapting to spectrally distorted speech in general, regardless of the shift direction. Figure 1 illustrates the potential outcomes of short-term training based on these three hypotheses.

II. METHODS

A. Subjects

Six normal-hearing listeners aged 25 to 35 participated in the present experiment. All subjects had thresholds better than 15 dB HL at audiometric test frequencies from 250 to 8000 Hz and all were native speakers of American English. All subjects were paid for their efforts.

B. Signal processing

Twenty-channel noise-band speech processors were used to simulate cochlear implant speech processing, and were implemented as follows. Speech signals were band-pass filtered into 20 frequency bands using eighth-order Butterworth filters. To evaluate the effect of acoustic input range, two groups of analysis bands were used: frequency allocation Table 9 used in the Nucleus-22 implant (150–10 831 Hz) and frequency allocation Table 6 used in the Nucleus-24M implant (116–7871 Hz). The temporal envelope of each band was extracted by half-wave rectification and low-pass filtering at 160 Hz. The envelope was used to modulate a wideband noise that was then spectrally limited by a band-pass filter (carrier band). The corner frequencies and bandwidths of the carrier frequency bands were dependent on the simulated electrode insertion depth. The frequency range of the carrier bands was determined by the following equation:

$$p(i) = P_0 + 0.75*i, \quad i = 0,1,...,20, \qquad (1)$$

where P_0 is the most apical carrier band location for a given frequency allocation in mm (from the apex). The corner frequencies of the carrier bands were determined by the following equation, from Greenwood (1990):

$$f(i) = 165.4*(10^{P(i)*0.06} - 0.88). \qquad (2)$$

Note that Eq. (2) assumes a 35-mm-long cochlea; actual cochlea lengths can vary by several mm. Combining Eqs. (1) and (2), all corner frequencies of carrier frequency bands were determined for a given insertion depth. Between adjacent carrier bands, the crossover attenuation was −3 dB. The most apical carrier band location (P_0) varied between 7.75 and 15.25 mm from the apex of the cochlea (or 27.25 to 19.75 mm from the base) to simulate a range of deep to shallow electrode insertion depths. Six carrier band frequency ranges were generated between these endpoints. Table I lists the corner frequencies of the analysis and carrier bands used in the experiment. Schematic diagrams of experimental conditions are shown in Fig. 2.

The outputs of the carrier bands were then summed and presented to listeners seated in a sound-treated booth via one loudspeaker (Tannoy Reveal) at 70 dBA. Figure 3 shows the spectral envelope of the phoneme /i/; spectral envelopes are shown for both the unprocessed speech and for the 20-channel noise-band processed speech at three cochlear locations (simulated insertion depths). The top panel shows the spectral envelopes when NU-22 Table 9 was used for the analysis filters; the bottom panel shows the spectral envelopes when the NU-24 Table 6 was used. Note the different degrees of spectral mismatch caused by the different frequency allocation tables.

C. Speech materials and procedures

Recognition of words in sentences was measured using novel sentences from the IEEE sentence corpus (IEEE, 1969). The sentences were digitized recordings spoken by one male talker (recorded at House Ear Institute). The IEEE sentences were of easy to moderate difficulty. For testing, a list was chosen pseudo-randomly from among 72 lists, and sentences were chosen randomly, without replacement, from the ten sentences within that list; two lists were used for each testing session. Subjects responded by repeating the sentence

TABLE I. The cutoff frequencies of two analysis filters and six carrier filters. CF0 represents the lower cutoff and CF1 represents the upper cutoff for the lowest filter band. Numbers in the table for CF2–CF20 represents the upper cutoff of the frequency band assigned to successively higher frequency bands.

	Analysis filters		Carrier filters					
	NU22	NU24	$P_0=7.75$	$P_0=9.25$	$P_0=10.75$	$P_0=12.25$	$P_0=13.75$	$P_0=15.25$
CF0	150	116	337	448	585	753	960	1214
CF1	350	243	390	513	665	851	1081	1363
CF2	550	393	448	585	753	960	1214	1528
CF3	750	540	513	665	851	1081	1363	1710
CF4	950	687	585	753	960	1214	1528	1913
CF5	1150	833	665	851	1081	1363	1710	2138
CF6	1350	978	753	960	1214	1528	1913	2387
CF7	1550	1125	851	1081	1363	1710	2138	2663
CF8	1768	1285	960	1214	1528	1913	2387	2970
CF9	2031	1477	1081	1363	1710	2138	2663	3310
CF10	2333	1696	1214	1528	1913	2387	2970	3687
CF11	2680	1949	1363	1710	2138	2663	3310	4106
CF12	3079	2238	1528	1913	2387	2970	3687	4570
CF13	3571	2597	1710	2138	2663	3310	4106	5085
CF14	4184	3043	1913	2387	2970	3687	4570	5656
CF15	4903	3565	2138	2663	3310	4106	5085	6289
CF16	5744	4177	2387	2970	3687	4570	5656	6992
CF17	6730	4894	2663	3310	4106	5085	6289	7771
CF18	7885	5734	2970	3687	4570	5656	6992	8635
CF19	9238	6718	3310	4106	5085	6289	7771	9594
CF20	10 823	7871	3687	4570	5656	6992	8635	10 657

as accurately as possible; the experimenter tabulated correctly identified words and sentences. Subjects were trained daily using the DARPA/TIMIT acoustic-phonetic continuous speech corpus (Garofolo *et al.*, 1993). The multi-talker TIMIT sentences were of moderate to extreme difficulty. For training, subjects auditioned a set of 300 sentences processed by the most upwardly shifted noise-band processor (15.25 mm from the apex). Subjects viewed the text of the sentence as the sentence was played, and were allowed to repeat the sentence as often as they liked.

The six subjects were divided into two groups. Three subjects were given speech processors using the Nucleus-22 Table 9 analysis filters and three were given speech processors using the Nucleus-24 Table 6 analysis filters. The testing and training was conducted as follows. Baseline (pretraining) data for all simulated insertion depths were collected on day 1, using the IEEE sentences. Over the next four days, subjects trained with TIMIT sentences processed by the most upwardly shifted noise-band processor (15.25 mm from the apex). On day 5, after a final training session, subjects were retested for all simulated insertion depths. The three subjects who were assigned NU-22 Table 9 were retested and trained using the most downwardly shifted speech (7.75 mm from apex) at a later date (3–10 days after completing testing and training with the most upwardly shifted speech).

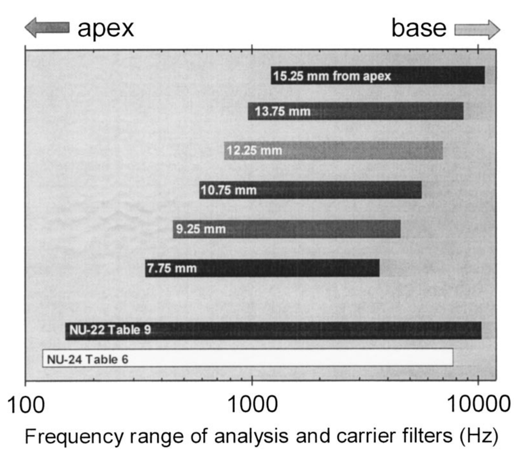

FIG. 2. Frequency ranges for the analysis and carrier filter bands. The bottom two bars show the analysis filter range for frequency allocation Tables 9 (Nucleus-22) and 6 (Nucleus-24). The upper six bars show the carrier band frequency ranges for six simulated electrode insertion depths, ranging from deep (7.75 mm from the apex) to shallow (15.25 mm from apex). The frequency range of the carrier bands was fixed to simulate the cochlear extent of an inserted electrode array.

III. RESULTS

Figure 4 shows the mean IEEE sentence recognition scores for all speech processor conditions before and after subjects were trained with spectrally shifted TIMIT sentences. Figure 4(a) shows data for the three subjects assigned the Nucleus-22 Table 9 analysis filters, before and after training with TIMIT sentences processed by the most upwardly shifted noise-band processor. Figure 4(b) shows data for the three subjects assigned the Nucleus-24 Table 6 analysis filters, before and after training with TIMIT sentences processed by the most upwardly shifted noise-band processor. Figure 4(c) shows data for the three subjects assigned the Nucleus-22 Table 9 analysis filters, before and after training with TIMIT sentences processed by the most downwardly shifted noise-band processor; this additional train and retest condition was performed 3–10 days after subjects had completed the earlier train and test condition. The solid line shows baseline data before the training while the dashed line

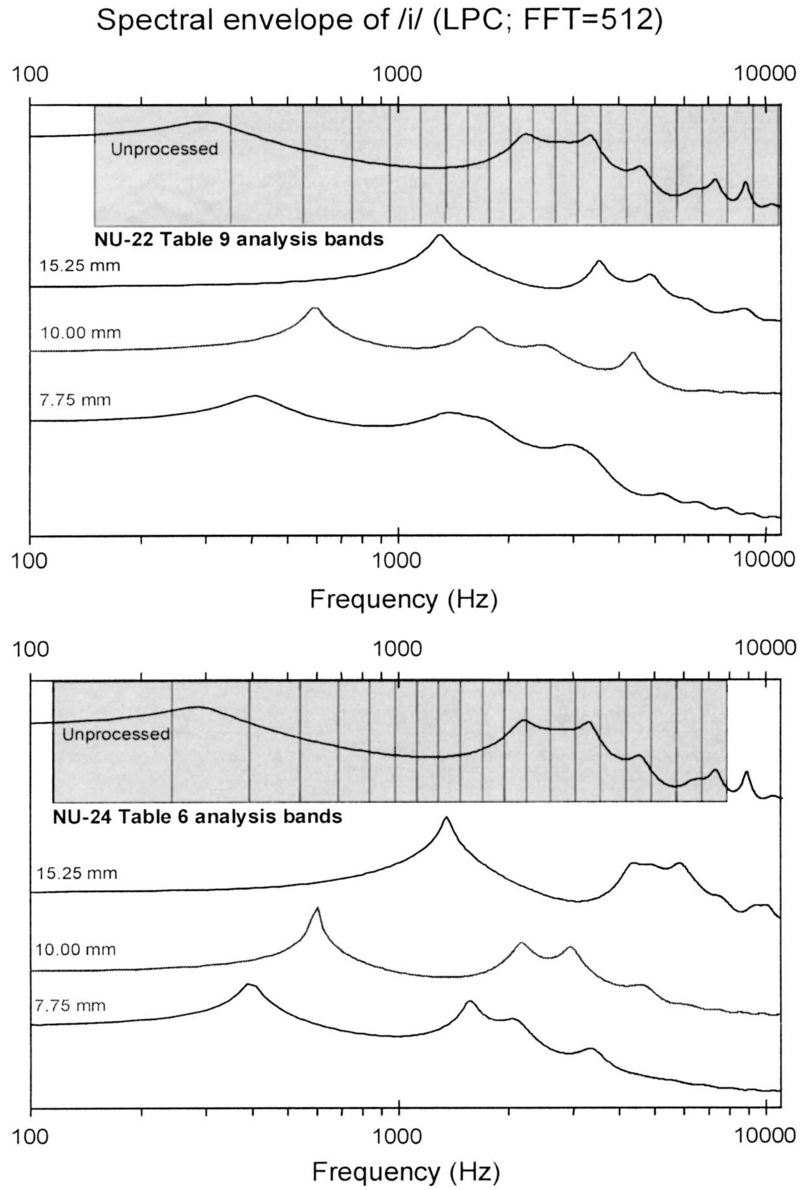

FIG. 3. The effects of analysis and carrier frequency ranges on the spectral envelope of the vowel /i/. The top panel shows the frequency analysis by Nucleus-22 Table 9. The unprocessed spectral envelope is shown in the shaded area, and the individual analysis bands are shown in the shaded area. The signal was analyzed by Table 9 and the amplitude envelope from each filter band was used to modulate a corresponding carrier band, depending on the simulated electrode location. The spectral envelope of /i/ is shown for three carrier band locations, ranging from a simulated deep electrode insertion (7.75 mm from apex) to a simulated shallow insertion. The bottom panel shows the effect of frequency analysis by Nucleus-24 Table 6 on the spectral envelope for three simulated electrode insertion depths.

shows the data after the 5 days of training. The filled symbols show the simulated insertion depth at which subjects were trained.

For subjects assigned the Nucleus-22 Table 9 analysis filters [Fig. 4(a)], baseline measures showed that mean sentence recognition was best with a simulated insertion depth of 12.25 mm from the apex (slightly more basal than the typical insertion depth of 10 mm from apex). Baseline performance was largely unchanged for adjacent simulated electrode insertion depths (10.75 to 13.75 mm from the apex). A large performance drop (~50 percentage points) was observed for the most apically situated carrier bands ($P_0 = 7.75$ mm). A smaller drop in recognition (14 percentage points) was observed for the most basally situated carrier location ($P_0 = 15.25$ mm). After training with upwardly shifted speech, a Student *t*-test revealed no significant difference between pre- and posttraining performance for the apical and mid-cochlea carrier locations ($p > 0.05$). However, a Student *t*-test revealed a significant improvement at the most basally situated carrier location where training had been performed ($p = 0.02$).

For subjects assigned the Nucleus-24 Table 6 analysis filters [Fig. 4(b)], baseline measures showed that peak sentence recognition was found for a range of apical to mid-cochlea carrier band locations (close to the typical electrode insertion depth 10 mm from the apex). Baseline recognition scores were nearly unchanged for simulated electrode insertion depths ranging between 7.75 and 10.75 mm from the apex. However, there was a large performance drop (67 percentage points) for the most basally situated carrier location ($P_0 = 15.25$ mm). After training with upwardly shifted speech, performance at the most basal carrier location improved by more than 20 percentage points ($p = 0.03$). Slight improvement was also observed at nearby basal carrier locations after training. Improvement at apical and middle carrier locations was minimal.

Subjects who were assigned the Nucleus-22 Table 9 analysis filters were also trained with TIMIT sentences pro-

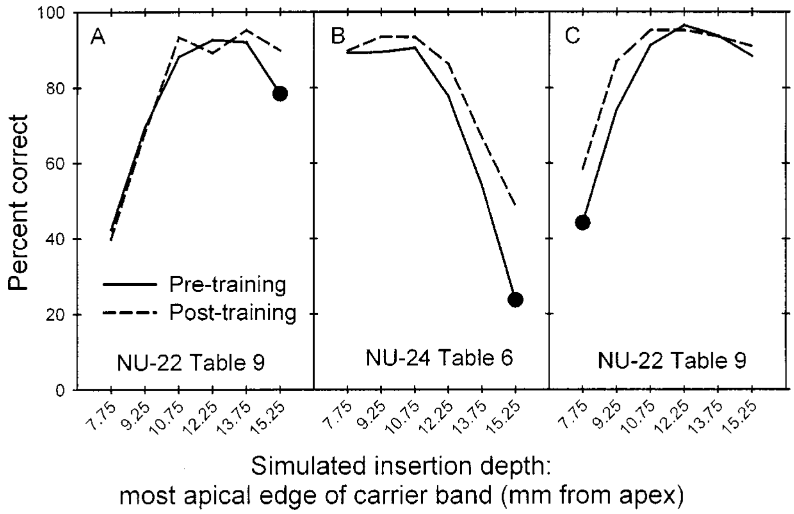

FIG. 4. Baseline pretraining and posttraining results for IEEE sentence recognition with 20-channel noise-band processors. Subjects were trained with either upwardly shifted or downwardly shifted 20-channel speech. Panel (a) shows results for three subjects assigned the frequency analysis filters from Nucleus-22 Table 9 (150–10 823 Hz), trained with upwardly shifted speech. Panel (b) shows results for three subjects assigned the frequency analysis filters from Nucleus-24 Table 6 (116-7871 Hz), trained with upwardly shifted speech. Panel (c) shows results for the three subjects assigned the frequency analysis filters from Nucleus-22 Table 9 [same subjects shown in panel (a)] who were retested and trained with downwardly shifted speech after completing the training with upwardly shifted speech. The *x* axis shows the carrier band locations for various simulated electrode insertion depths; the *y* axis shows percent correct. The solid lines show baseline performance; the dashed lines show performance after five training sessions. The filled circles show the carrier band location where training was performed.

cessed by the most downwardly shifted noise-band processor (7.75 mm from the apex) at a later date. This training and retesting was performed between 3 and 10 days after completing training with upwardly shifted speech. The additional train and retest condition was created because the baseline performance deficit observed at the most basal carrier location was much less than that observed at the most apical carrier location. In fact, the subjects using the Nucleus-22 Table 9 filters analysis filters had relatively high levels of sentence recognition at the most basal carrier location ($P_0 = 15.25$ mm). Note that subjects assigned the Nucleus-24 Table 6 analysis filters did not participate in this additional train and retest condition because near-peak baseline performance was observed at the most apical carrier location, meaning there was little room for improvement. Mean performance for this train and retest condition is shown in Fig. 4(c). Baseline performance for all simulated insertion depths was remeasured before training with downwardly shifted speech, and are shown by the solid line. Similar to the earlier baseline measures, peak performance in the retested baseline measures was found at the simulated electrode insertion depth of 12.25 mm from the apex. In the retested baseline measures, there remained a slight performance drop at the most basal carrier location ($P_0 = 15.25$ mm). However, the performance deficit at this carrier location was less than that observed in the original baseline measures. In fact, the retested baseline scores were quite similar to the posttraining performance of the previous experiment [as shown by the dashed line in the panel 4(a)]. After training with the most downwardly shifted speech, mean recognition scores remained relatively unchanged at the middle and basal carrier locations. However, significantly improved recognition was observed for the trained apical carrier locations ($p < 0.001$).

IV. DISCUSSION

The results demonstrate that short-term training with spectrally altered speech can significantly improve the recognition of the trained speech patterns. However, the improvement was generally restricted to the cochlear location where the training had taken place. The results strongly support the hypothesis of "local adaptation," which asserted that subjects would only adapt to the specific spectral mismatch on which they were trained. The results also suggest that subjects may have developed alternate spectral patterns after training with spectrally altered speech, while retaining previous "internal" representations, at least over the short term.

Training with upwardly shifted speech indeed improved the recognition of upwardly shifted speech, while not affecting speech recognition at the apical and mid-cochlea carrier locations. Similarly, the follow-up training with downwardly shifted speech improved the recognition of downwardly shifted speech and did not affect speech recognition at the basal and mid-cochlea carrier locations. The subjects who were assigned the Nucleus-22 Table 9 analysis filters participated in both training experiments, with downwardly shifted speech training occurring at a later date (3–10 days after completion of upwardly shifted speech training). It was interesting to observe that the improved recognition of upwardly shifted speech (after training with upwardly shifted speech) was largely retained, at least over the short term. The recollected baseline data (before training with downwardly shifted speech) showed that recognition of upwardly shifted speech was similar to the levels of performance after training with upwardly shifted speech. After 5 days of training with and improved recognition of downwardly shifted speech, recognition of upwardly shifted speech remained at the recollected baseline levels (which were the same levels after the earlier training with upwardly shifted speech). These results suggest that subjects may have temporally preserved ("internalized") the upwardly shifted speech patterns while accommodating the newly trained downwardly shifted speech patterns.

The "local adaptation" to spectrally altered speech suggests that listeners are able to accommodate alternate speech patterns, while preserving previously accommodated patterns. If a "global adaptation" had occurred, in which subjects' performance improved at all carrier band locations, subjects may have simply been adapting to the reduced spectral resolution of the noise-band processors. Given enough time and training at various carrier locations, the combined improvements from the localized training might cumulatively amount to a ''global adaptation.'' If a ''shifted adaptation'' adaptation had occurred, in which the improvements due to localized training also produced a deficit at other carrier locations, subjects would be unable to retain previously learned spectral patterns.

There was an interactive effect between the frequency analysis range and the place of stimulation in both baseline and posttraining measures. The baseline performance showed that, acutely measured, frequency allocation Nucleus-22 Table 9 (150–10 823 Hz) was better for basal carrier locations that simulated shallow electrode insertions, while frequency allocation Nucleus-24 Table 6 (116–7871 Hz) was better for middle and apical carrier locations that simulated typical and deep insertion depths. As insertion depths of 10 mm or more from the apex are typical for contemporary implant devices, the frequency allocation used in Nucleus-24 Table 6 should provide the best results, as the degree of spectral mismatch may somewhat reduced. Such reduced spectral mismatch result may have contributed to the improved speech recognition observed in a previous study (Skinner *et al.*, 1995) with Nucleus-22 patients who were reassigned frequency allocation Table 7 (120–8568 Hz) rather than the default frequency allocation Table 9. Skinner *et al.* (1995) argued that the improvement was due to the better spectral resolution in the low-frequency region. In contrast, the long-term studies conducted by Fu *et al.* (2002) revealed that improved spectral resolution in the low-frequency region may not, in and of itself, provide better recognition. In that study, cochlear implant patients continuously used a spectrally shifted speech processor for a 3-month period; subjects were assigned frequency allocation Table I (75–5411 Hz), which mapped the significantly more low-frequency speech information onto the available electrodes than the subjects' clinically assigned frequency allocation (Table 7 or 9). However, vowel recognition remained significantly lower with frequency allocation (Table I) than with the frequency allocations used in subjects' clinically assigned speech processor (Table 7 or 9), even after 3 months of continuous exposure. This suggests that increased spectral resolution of low-frequency speech information may not overcome a severe spectral mismatch, and that the degree of spectral mismatch may be the limiting factor in cochlear implant listeners' utilization of speech spectral cues. McKay and Henshall (2002) also found that, while most subjects were able to accommodate frequency shifts up to ratio changes of 1.3, it was unclear if subjects could have accommodated greater frequency shifts, given enough time or given gradual adaptation (learning small incremental shifts).

These and many other studies have examined listeners' ability to adapt to changes to or deterioration of representations of speech patterns. There would be very little adaptation needed if, for example, the number of perceptual channels and/or spectral resolution were to be increased or the spectral mismatch between the acoustic signal and electrode positions were to be reduced, as speech recognition will most likely be immediately improved. Skinner *et al.* (1995) found this sort of acute improvement by moving Nucleus-22 implant users from frequency allocation Table 9 (150–10 823 Hz) to Table 7 (116–7871 Hz). It is uncertain whether this improvement was due to increased spectral resolution or reduced spectral mismatch, or a combination of both effects, but subjects were able to accommodate a shift from their previously adapted electric map without much practice. McKay and Henshall (2002) found little improvement when configuring implant listeners' speech processors according to patients' electrode discriminability. Such a mapping might have improved the spectral resolution and/or increased the stimulation rate, either of which could have resulted in improved performance. The lack of improvement (indeed, the sometimes reduced performance with perceptual mapping) suggests that a mismatch between the acoustic signal and stimulating electrodes may have proved too difficult to accommodate. The subjects were able to accommodate small shifts in the frequency-to-electrode assignment, reemphasizing the importance of spectral mapping and learning in implant speech recognition.

V. CONCLUSION

The present study examined the effects of short-term learning on normal-hearing listeners' ability to accommodate spectrally altered speech patterns. Results showed that speech recognition with 20-channel noise-band processors was acutely affected by severe spectral mismatch. Initially, subjects could tolerate only a small amount of spectral mismatch. However, after short-term training with severely spectrally altered speech, subjects were able to significantly improve their speech recognition of spectrally shifted speech where the training had been performed. The improvement was restricted to the trained spectral shift, and did not generalize to other spectral shifts distant from the location where the training had occurred. These results strongly suggest that a ''local adaptation'' had occurred, in which listeners adapted only to the specific spectral alteration where training was performed. Such local adaptations were preserved, at least over the short term. Listeners may develop alternate spectral patterns given enough training, while preserving previous ''internal'' representations of speech sounds.

ACKNOWLEDGMENTS

The research was supported by Grants Nos. R03-DC-03861 and R01-DC04993 from NIDCD. We are grateful to Christopher Turner and two anonymous reviewers for useful suggestions on an earlier draft of this paper. We would like to also thank the research participants for their considerable time spent with this experiment.

Fu, Q.-J., and Shannon, R. V. (**1999a**). ''Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus 22 cochlear implant,'' Ear Hear. **20**, 332–344.

Fu, Q.-J., and Shannon, R. V. (**1999b**). ''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 III, 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.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1993**). ''The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM,'' NTIS order number PB91-100354.

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

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronic Engineers, New York).

McKay, C. M., and Henshall, K. R. (**2002**). ''Frequency-to-electrode allocation and speech perception with cochlear implants,'' J. Acoust. Soc. Am. **111**, 1036–1044.

Pelizzone, M., Cosendai, G., and Tinembart, J. (**1999**). ''Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for Ineraid implanted subjects,'' Ear Hear. **20**, 228–237.

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.

Skinner, M. W., Holden, L. K., and Holden, T. A. (**1995**). ''Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 307–311.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''New levels of speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.