JARO 6: 180–189 (2005)
DOI: 10.1007/s10162-005-5061-6

JARO
Journal of the Association for Research in Otolaryngology

Auditory Training with Spectrally Shifted Speech: Implications for Cochlear Implant Patient Auditory Rehabilitation

QIAN-JIE FU, GERALDINE NOGAKI AND JOHN J. GALVIN III

Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA, USA

Received: 8 November 2004; Accepted: 4 March 2005; Online publication: 10 June 2005

ABSTRACT

After implantation, postlingually deafened cochlear implant (CI) patients must adapt to both spectrally reduced and spectrally shifted speech, due to the limited number of electrodes and the limited length of the electrode array. This adaptation generally occurs during the first three to six months of implant use and may continue for many years. To see whether moderate speech training can accelerate this learning process, 16 naïve, normal-hearing listeners were trained with spectrally shifted speech via an eight-channel acoustic simulation of CI speech processing. Baseline vowel and consonant recognition was measured for both spectrally shifted and unshifted speech. Short daily training sessions were conducted over five consecutive days, using four different protocols. For the test-only protocol, no improvement was seen over the five-day period. Similarly, sentence training provided little benefit for vowel recognition. However, after five days of targeted phoneme training, subjects' recognition of spectrally shifted vowels significantly improved in most subjects. This improvement did not generalize to the spectrally unshifted vowel and consonant tokens, suggesting that subjects adapted to the specific spectral shift, rather than to the eight-channel processing in general. Interestingly, significant improvement was also observed for the recognition of spectrally shifted consonants. The largest improvement was observed with targeted vowel contrast training, which did not include any explicit consonant training. These results suggest that targeted phoneme training can accelerate adaptation to spectrally shifted speech. Given these results with normal-hearing listeners, auditory rehabilitation tools that provide targeted phoneme training may be effective in improving the speech recognition performance of adult CI users.

Keywords: auditory rehabilitation, auditory training, cochlear implants, spectrally shifted speech, targeted phoneme training

INTRODUCTION

The cochlear implant (CI) is an electronic device that provides hearing sensation to patients with profound hearing loss. Although the CI can approximately restore the spatial representation of speech signals, the electrically evoked peripheral neural patterns may be dramatically different from normal acoustical patterns. Postlingually deafened CI patients must adapt to these novel peripheral neural patterns. Compared to the central speech pattern templates (which CI patients may have developed during their previous hearing experience), the peripheral neural patterns delivered by the CI will have less spectral detail and will be spectrally distorted. Patients' adaptation to these novel peripheral neural patterns generally occurs during the first three to six months of implant use and may continue for many years (e.g., Tyler et al., 1997).

The spectral mismatch between the peripheral neural patterns and central speech pattern templates may be partially responsible for the poor speech recognition in some CI users. The acute effects of spectral mismatch on speech performance have been well documented (e.g., Fu and Shannon, 1999; Shannon et al., 1998). Recent studies also evaluated the effects

Correspondence to: Qian-Jie Fu · Department of Auditory Implants and Perception · House Ear Institute · 2100 West Third Street, Los Angeles, CA, USA. Telephone: +213-273-8036; fax: +213-413-0950; email: qfu@hei.org

of speech training on NH subjects' recognition of spectrally shifted speech (e.g., Fu and Galvin, 2003; Rosen et al., 1999). The results from these studies suggest that, although recognition performance can be dramatically affected by spectral mismatch, moderate auditory training can alleviate some of the difficulties caused by distortions to the spectral content of speech signals.

Several studies have assessed the benefits of limited training on the speech recognition skills of poor-performing CI users and have shown mixed results (Busby et al., 1991; Dawson and Clark, 1997). The mixed and generally poor outcomes from previous CI speech training studies may well be due to the amount and type of training employed. Prelingually deafened CI users most likely have developed limited auditory-only central speech pattern templates, which are likely much less robust than those developed by NH or even hearing-impaired listeners. As such, while the peripheral neural patterns elicited by electrical stimulation may all sound different, the differences may not be meaningful. Administering phoneme and sentence recognition tests to these patients might not indicate a failure of the implant to provide an adequate peripheral neural patterns, but rather the failure of these patients to develop new central speech pattern templates with their limited experience with the CI device. A more extensive and intensive approach to auditory training that targets phonemic contrasts might yield better results than found in previous CI patient training studies. Fu et al. (2004) explored the effects of moderate auditory training in 10 adult CI patients. In that study, subjects performed moderate auditory training at home using speech stimuli one hour per day and five days per week for a period of one month or longer. Results showed a significant improvement in all patients' speech perception performance after targeted phonemic contrast training.

Besides the amount and frequency of training, the type of training employed may also affect CI patient outcomes. Rosen et al. (1999) used connected discourse tracking (DeFilippo and Scott, 1978) to train listeners' recognition of four-channel, spectrally shifted speech. In the Fu and Galvin (2003) study, NH subjects were trained by listening to 300 spectrally shifted TIMIT sentences each day; subjects were asked to listen to each sentence carefully while reading the text labels shown onscreen. In the study of Fu et al. (2004), CI subjects were trained to discriminate phonemic contrasts, after which they were trained to identify medial vowels. Auditory and visual feedback was also provided which allowed subjects to repeatedly compare their (incorrect) choice to the correct answer. Although all these methods demonstrated promising results for auditory training, the relative effectiveness and optimal time course of these different training approaches remains unclear. The relationship of training procedures and materials to testing procedures and materials may also bear on the perceived effectiveness of the training protocols. More importantly, the potential of a particular training protocol and set of materials to generalize to other speech perception measures is of great interest. In the present study, the effects of three training protocols on the recognition of spectrally reduced and shifted speech were evaluated in 16 NH listeners. Short daily training sessions were conducted over five consecutive days, using three different training protocols and one test-only protocol.

MATERIALS AND METHODS

Subjects

Sixteen NH adults (four males and 12 females), aged 21–47, participated in the study. All subjects had pure tone thresholds better than 15 dB HL at octave frequencies ranging from 125 to 8,000 Hz. All subjects were native English speakers. All subjects were paid for their participation.

Signal processing

CI speech processors using the Continuously Interleaved Sampling (CIS) strategy (Wilson et al., 1991) were acoustically simulated using eight-channel sinewave vocoders. The processors were implemented as follows. The signal was first processed through a preemphasis filter (high-pass with a cut off frequency of 1,200 Hz and a slope of 6 dB/octave). An input frequency range (200–7,000 Hz) was band-passed into eight spectral bands using fourth-order Butterworth filters. The cochlear locations of these two end frequencies were calculated according to Greenwood's (1990) formula. The estimated cochlear distance was evenly divided into eight spectral channels. The estimated cochlear location for each spectral channel was then transformed back into frequency, again using Greenwood's formula. The corner frequencies (3 dB down) of the analysis filters are listed in Table 1. The temporal envelope was extracted from each frequency band by half-wave rectification and low-pass filtering at 160 Hz. For each channel, a sinusoidal carrier was generated; the frequency of the sinewave carrier depended on the experimental condition. For the spectrally unshifted condition, the frequency of the sinewave carrier was equal to the center frequency of the analysis filter. For the spectrally shifted condition, the output carrier bands were upwardly shifted to simulate a

TABLE 1

The corner frequencies of analysis filters and the sinusoidal frequencies of sinewave carriers used in the sinewave cochlear implant simulation

Channel # (apical to basal)	*Analysis band/ unshifted carrier band corner frequencies (Hz)*	*Greenwood distance from cochlear apex (mm)*	*Center frequency of carrier filter (Hz)*	*Shifted carrier band corner frequencies (Hz)*	*Greenwood distance from cochlear apex (mm)*	*Center frequency of carrier filter (Hz)*
1	200–359	5.33–8.07	268	999–1,363	14–16	1,167
2	359–591	8.07–10.81	461	1,363–1,843	16–18	1,585
3	591–930	10.81–13.55	741	1,843–2,476	18–20	2,136
4	930–1,426	13.55–16.30	1,152	2,476–3,310	20–22	2,863
5	1,426–2,149	16.30–19.04	1,751	3,310–4,410	22–24	3,821
6	2,149–3,205	19.04–21.78	2,624	4,410–5,860	24–26	5,084
7	3,205–4,748	21.78–24.52	3,901	5,860–7,771	26–28	6,748
8	4,748–7,000	24.52–27.26	5,765	7,771–10,290	28–30	8,942

shallow insertion of a 16-mm-long, eight-electrode array with 2 mm electrode spacing. The corresponding corner frequencies of the carrier filters for spectrally shifted speech are shown in Table 1; again, the frequency of the sinewave carrier was equal to the center frequency of the carrier filter. Figure 1 shows the relation between the input frequency range and the output frequency range, as well as the degree of spectral shift between the analysis and carrier filter bands. The extracted temporal envelope from each frequency analysis band was used to modulate the corresponding sinusoidal carrier. The modulated carriers of each band were summed and the overall level was adjusted to be the same as the original speech.

Test and training materials

Speech recognition was assessed using multitalker vowel and consonant recognition. Vowel recognition was measured in a 12-alternative identification paradigm. The vowel set included 10 monophthongs (/i I ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and two diphthongs (/o e/), presented in a /h/-vowel-/d/ context. The tokens for vowel recognition test were digitized natural productions from five men and five women drawn from speech samples collected by Hillenbrand et al. (1995). Consonant recognition was measured in a 20-alternative identification paradigm. The consonant set included /b d g p t k m n l r y w f s ʃ v z ð tʃ dʒ/, presented in an /a/-consonant-/a/ context. Consonant tokens consisted of digitized natural productions from five men and five women, for a total of 200 tokens. The tokens for consonant recognition test were digitized natural productions from five men and five women drawn from speech samples collected by Shannon et al. (1999).

For the test token preview protocol, speech materials were the same 12 /h/-vowel-/d/ vowel tokens used in the vowel recognition set, spoken by a novel set of four talkers (two male and two female). For the targeted phonemic contrast training protocol, speech materials included more than 1,000 monosyllable words, spoken by two males and two females (recorded at House Ear Institute). For the sentence training protocol, speech materials included 260 HINT sentences (Nilsson et al., 1994) spoken by a single male talker. Note that for all training materials, talkers in the training protocols were not the same as those used in the recognition tests.

Test and training procedures

For speech testing, each test block included 120 tokens (12 vowels * 10 talkers) for vowel identification and 200 tokens (20 consonants * 10 talkers) for consonant identification. On each trial, a stimulus token was chosen randomly, without replacement, and presented to the subject. Following presentation

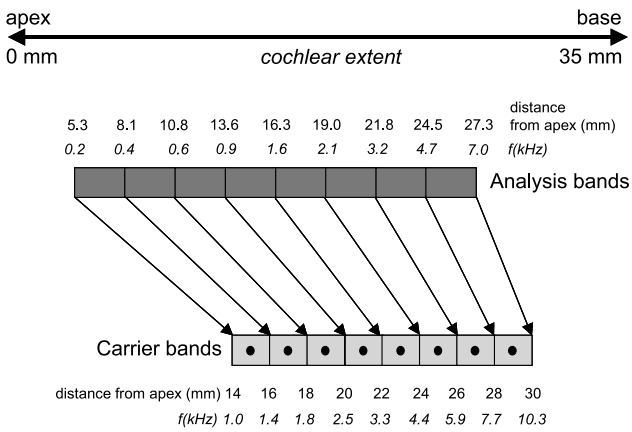

FIG. 1. Frequency allocations of analysis and carrier filter bands for eight-channel acoustic simulation of cochlear implant speech processing.

of each token, the subject responded by pressing one of 12 buttons in the vowel test, or one of 20 buttons in the consonant test, each marked with one of the possible responses. The response buttons were labeled in a /h/-vowel-/d/ context for the vowel recognition task and /a/-consonant-/a/ context for the consonant recognition task. No feedback was provided, and subjects were instructed to guess if they were not sure, although they were cautioned not to provide the same response for each guess.

Baseline phoneme recognition was measured on day 1 of the experiment. Subjects were first tested for recognition of unprocessed speech to ensure nearly perfect recognition of the speech materials before processing and to familiarize subjects with test tokens, labels, and format. Recognition of unprocessed vowels was measured twice, while recognition of unprocessed consonants was measured only once. After this initial testing with unprocessed speech, vowel and consonant recognition were measured for eight-channel, spectrally unshifted sinewave speech. Baseline vowel recognition with the spectrally unshifted speech was measured at least twice (or until performance asymptoted), while consonant recognition was measured only once. Next, vowel and consonant recognition were measured for eight-channel, spectrally shifted sinewave speech. For the shifted speech, baseline vowel and consonant recognition were measured only once.

After baseline measures, training protocols were begun. For each subject, four training sessions per day were conducted for five consecutive days. Table 2 shows the timetable for testing and training for each group. Prior to the initial training session of each day, vowel recognition was remeasured with the shifted speech; after each training session, vowel recognition was again remeasured with the shifted speech.

The sixteen subjects were equally divided into four training groups:

Group 1 (test-only protocol): Subjects in this group received no preview, no feedback, and no training. As an experimental control, vowel recognition with shifted speech was remeasured four times on each of the five days of training (the same number of times as in the other training groups). In this way, any learning strictly due to test exposure could be observed.

Group 2 (preview protocol): Subjects in this group were asked to preview the 12 h/V/d tokens used in the vowel recognition test before each retest with shifted speech. In the preview, speech was processed exactly as for the shifted speech tokens. However, vowels were produced by a different set of talkers (two male and two female) than used in the vowel recognition test. On each of the five training days, before each vowel recognition retest, subjects were asked to preview all vowel tokens spoken by all four preview talkers. During the preview, 48 buttons were shown on the computer screen (12 vowels * 4 talkers); subjects listened to each token by clicking on the desired button. Subjects were asked to listen to each of the 48 tokens and to compare tokens across talkers and within talkers. Subjects spent a minimum of 5 min previewing the vowel tokens, after which the vowel test with the 10 test talkers was administered. By training with the test tokens (although spoken by a different set of talkers), the effects of test-specific speech training could be observed.

Group 3 (targeted vowel contrast training protocol): Subjects in this group were asked to train using custom software (Computer-Assisted Speech Training, or CAST, developed at House Ear Institute). Subjects were trained to identify medial vowels using monosyllable words in a c/V/c context. Note that the training materials were produced by a different talker set than was used for the vowel recognition tests. In the identification training protocol, only the medial vowel differed between response choices (i.e., "seed," "said," "sod," "sued"), allowing subjects to better fo-

TABLE 2

Timetable for the different training protocols

	Group 1 (Test-only protocol)	*Group 2 (Preview protocol)*	*Group 3 (Targeted vowel contrast training protocol)*	*Group 4 (Sentence training protocol)*
Day 1	V, C test: unprocessed	V, C test: unprocessed	V, C test: unprocessed	V, C test: unprocessed
Baseline	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted
	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted
Days 1–5	V Test	V test	V test	V test
Training	V Test	V preview–V test	W train–V test	S train–V test
	V Test	V preview–V test	W train–V test	S train–V test
	V Test	V preview–V test	W train–V test	S train–V test
Day 5	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted	V, C test: 8-ch, shifted
Retest	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted	V, C test: 8-ch, unshifted

V = Vowel, C = consonant, W = word, S = sentence.

cus on differences between medial vowels. Any h/V/d monosyllable words were excluded from the training set to prohibit direct learning of the specific tokens used in the vowel recognition tests. The training materials were processed exactly as for the shifted speech tokens. Initially, subjects chose between two responses that differed greatly in terms of acoustic speech features (i.e., "said," "sued"); as subjects' performance improved, the differences between speech features in the response choices were reduced (i.e., "said," "sad"). The acoustic speech features used to define these levels of difficulty included first and second formant frequencies (F1 and F2) and duration. As subjects continued to improve, the number of response choices was increased (up to a maximum of six choices). Visual feedback was provided as to the correctness of response and auditory feedback was provided in which the subject's (incorrect) response was repeatedly compared to the correct response. Each training block contained 50 trials and subjects completed three blocks each training session; subjects took approximately 15 min to complete each session. After completing the training session, subjects' vowel recognition with shifted speech was immediately retested. By training subjects to identify targeted vowel contrasts (using monosyllable words that were not used in the vowel tests), any generalization of the trained speech to the test speech could be observed.

Group 4 (modified connected discourse/sentence training protocol): Subjects in this group were asked to train using custom software (Speech Test Assessment and Rehabilitation Software, or STARS, developed at House Ear Institute). Subjects were trained using a modified connected discourse method (DeFilippo and Scott, 1978). However, instead of the tester reading phrases aloud and asking the subject to repeat the sentence as accurately as possible, the computer played the sentence and the subject typed the response as accurately as possible into a response window. For each trial of the training exercise, after the sentence was played, a number of empty boxes (equal to the number of words in the sentence) were shown onscreen. The subject typed in as many words as (s)he understood and then pressed the "compare" button; the correctly identified words were revealed in the response boxes. The subject then pressed the "repeat" button to listen to the sentence again, and typed in as many of the remaining words as could be identified. As the subject correctly identified more words in the sentence, more of the response boxes were revealed. A spell-check program helped to reduce typographical errors in the subjects' typed responses. After repeating the sentence a maximum of three times, the subject pressed the "view" button, which revealed all the words in the sentence. Subjects repeated the sentence one more time to compare the text with the sound, then pressed "Next" to move onto the next sentence. HINT sentences, spoken by a single male talker, were used for sentence training. The stimulus set consisted of 26 sentence lists containing 10 sentences each; a sentence was randomly chosen from each list (without replacement) and presented to the subject. The training sentences were processed exactly as for the shifted speech tokens. Each training block contained 10 trials and subjects completed three blocks each training session; subjects took approximately 15 min to complete each session. After completing the training session, subjects' vowel recognition with shifted speech was immediately retested. By training subjects to identify spectrally shifted sentences, any generalization of the trained speech to the test speech could be observed. Also, because most listeners will acquire speech patterns by listening to sentences and phrases, the effect of common listening and learning experiences of CI patients could be approximated.

RESULTS

Figure 2 shows baseline vowel and consonant measures for unprocessed speech, eight-channel spectrally unshifted speech and eight-channel spectrally shifted speech prior to the training. Before statistical analyses, all percentage scores were transformed into rationalized arcsine units (rau; Studebaker, 1985). This had the effect of minimizing the variance that is characteristic of percentage scores, while providing a scoring unit that is similar to percentages. Rau scores were used for statistical analysis; however, the data

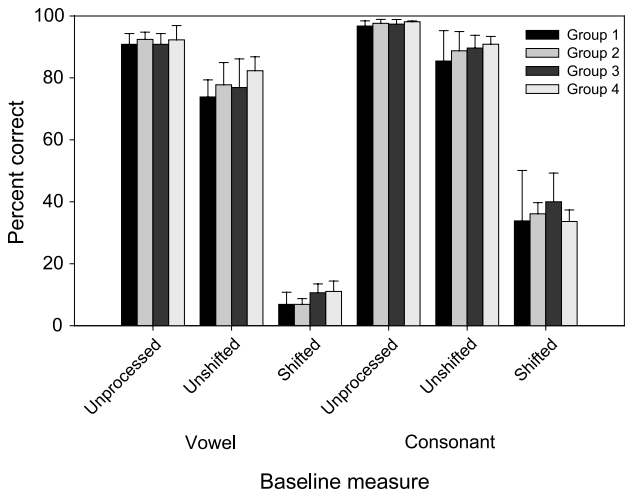

FIG. 2. Baseline vowel and consonant recognition measures for unprocessed speech, eight-channel spectrally unshifted speech and eight-channel spectrally shifted speech. Error bars indicate ±1 standard deviation.

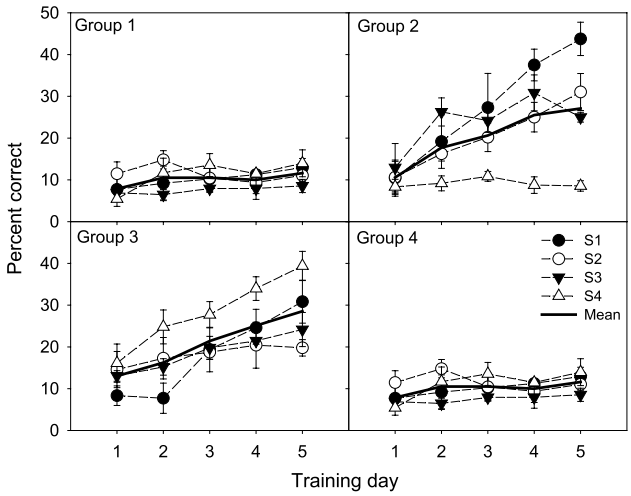

FIG. 3. Mean vowel recognition performance of spectrally shifted speech for the four training groups, as a function of training days. Error bars indicate ±1 standard deviation.

shown in figures and reported in the results are expressed in percent correct, which corresponds closely to rau.

For unprocessed speech, mean vowel recognition was 91.6% correct and mean consonant recognition was 97.5% correct. When the spectral resolution was reduced (baseline eight-channel spectrally unshifted speech), mean vowel recognition dropped to 77.7% correct, while mean consonant recognition dropped to 88.7% correct. When the spectrum of speech signals was further shifted (baseline eight-channel spectrally shifted speech), mean vowel recognition dropped to 8.9% correct and mean consonant recognition dropped to 35.9% correct. Two-way ANOVA tests showed a significant effect of spectral resolution on both vowel ($F_{2,36}$ = 925.93, $p < 0.001$) and consonant recognition ($F_{2,36}$ = 212.71, $p < 0.001$), but no significant difference in baseline performance across training groups on either vowel ($F_{3,36}$ = 1.60, $p = 0.207$) or consonant recognition ($F_{3,36}$ = 1.88, $p = 0.151$). Also, there was no significant interaction between spectral resolution and training groups for either vowel ($F_{6,36}$ = 0.70, $p = 0.604$) or consonant recognition ($F_{6,36}$ = 0.64, $p = 0.695$). Post-hoc Bonferroni *t*-tests revealed that performance with the unprocessed speech was significantly better than that with the eight-channel unshifted speech ($p < 0.001$) and that with eight-channel spectrally shifted speech ($p < 0.001$).

Figure 3 shows mean vowel recognition scores for spectrally shifted speech in the four training groups, as a function of the training days. Dashed lines represent the data from each individual subject and solid lines represent the mean data from all four subjects in each training group. For Group 1 (test-only protocol), vowel recognition scores remained at near-chance level throughout the training period. For Group 2 (preview protocol), vowel recognition scores gradually increased from 10.6% correct (day 1) to 27.1% correct (day 5). For Group 3 (targeted vowel contrast training protocol), mean vowel recognition scores gradually increased from 13.0% to 28.5% correct over the course of the training period. For Group 4 (sentence training protocol), mean vowel recognition score increased slightly from 11.3% to 15.9% after five consecutive days of testing. Two-way ANOVA tests showed that there were significant effects for the training protocol ($F_{3,60}$ = 13.68, $p < 0.001$) and for performance over the training period ($F_{4,60}$ = 5.56, $p < 0.001$); however, there was no interaction between the training protocols and performance over time ($F_{12,60}$ = 0.91, $p = 0.539$). Post-hoc Bonferroni *t*-tests revealed that for Groups 1 and 4, there was no significant change in performance over the five-day training period ($p = 1.00$). For Group 2, vowel recognition scores did not significantly improve until days 4 and 5 ($p < 0.05$). For Group 3, vowel recognition did not significantly improve until day 5 ($p = 0.016$).

Figure 4 shows the mean improvement in vowel and consonant recognition after the five-day training period, as a function of the different training protocols for eight-channel spectrally shifted (panel A) and eight-channel spectrally unshifted speech (panel B). For vowel recognition with eight-channel spectrally shifted speech (panel A), the test-only protocol (Group 1) showed only a 3.8 percentage point increase in mean performance after five consecutive days of retesting. Similarly, the sentence training protocol (Group 4) showed only a 3.2 percentage point increase in mean performance after five consecutive days of retesting. When subjects were allowed to preview the vowel tokens before testing (Group 2), mean vowel recognition with shifted speech significantly improved by 16.5 percentage points over the

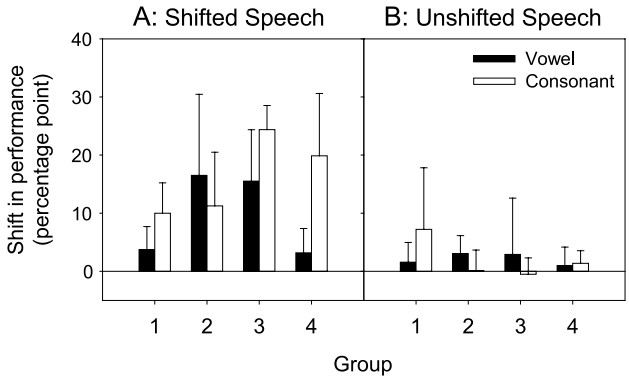

FIG. 4. Mean improvement in vowel and consonant recognition after five days of training, for the different training protocols. **A** Shift in performance for recognition of eight-channel spectrally shifted speech; **B** shift in performance for recognition of eight-channel spectrally unshifted speech. Error bars indicate ±1 standard deviation.

five-day training period. Targeted vowel contrast training (Group 3) produced a 15.5 percentage point improvement in mean vowel recognition with spectrally shifted speech. Three-way ANOVA tests showed that there was a significant effect of training protocols ($F_{3,48} = 4.85$, $p = 0.005$), speech training ($F_{1,48} = 14.23$, $p < 0.001$), as well as spectral resolution ($F_{1,48} = 1206.70$, $p < 0.001$) on vowel recognition scores. Also, there was a significant interaction between speech training and spectral resolution ($F_{1,48} = 6.16$, $p = 0.017$). However, there was no significant interaction among other factors. Post-hoc Bonferroni *t*-tests revealed that there was no significant improvement after the five-day training for either Group 1 ($p = 0.308$) or Group 4 ($p = 0.493$). However, there was a significant improvement after the five-day training for both Group 2 ($p = 0.004$) and Group 3 ($p = 0.008$).

Similar to the vowel recognition results, the test-only protocol (Group 1) produced the smallest improvement in mean consonant recognition scores for spectrally shifted speech. However, for Group 1, consonant recognition improved by 10.0 percentage points; note that Group 1 was exposed to spectrally shifted consonants only on day 1 (baseline) and day 5 (retest), except for "h" and "d" (in "heed," "had," "head," etc.) during the five-day period of vowel retesting. Mean recognition of spectrally shifted consonants improved by 11.3 percentage points for Group 2 (preview protocol) after five days of training; note that, similar to Group 1, Group 2 was exposed to spectrally shifted consonants only on day 1 (baseline) and day 5 (retest), except for "h" and "d" (in "heed," "had," "head," etc.) during the five-day period of vowel preview and retesting. Targeted vowel contrast training (Group 3) provided the best improvement (24.4 percentage points) in mean recognition of spectrally shifted consonants. Note that while consonant recognition was not explicitly trained, Group 3 was exposed to a variety of spectrally shifted consonants in the monosyllable words used for training (e.g., "sad," "said," "bat," "bet," "knack," "neck," etc.). Sentence training (Group 4) also produced better recognition of shifted consonants (19.9 percentage points); note that Group 4 would have been exposed to initial, medial, and final consonants during the course of sentence training. Three-way ANOVA tests showed that there was a significant effect of speech training ($F_{1,48} = 19.00$, $p < 0.001$), spectral resolution ($F_{1,48} = 525.37$, $p < 0.001$), but no significant effect of training protocols [$F_{3,48} = 1.71$, $p = 0.177$) on consonant recognition scores. Also, there was no significant interaction across these factors except a significant interaction between speech training and spectral resolution ($F_{1,48} = 8.75$, $p = 0.005$). Post-hoc Bonferroni *t*-tests revealed that there was no significant improvement after the five-day training for Group 2 ($p = 0.231$). However, there was a significant improvement after the five-day training for Group 1 ($p = 0.019$), Group 3 ($p = 0.012$), and Group 4 ($p = 0.018$). There was no significant difference in performance improvement across different training protocols for either shifted or unshifted speech except that the improvement in Group 3 was significantly higher than that in Group 1 ($p = 0.023$).

DISCUSSION

The results demonstrate that moderate auditory training can significantly improve recognition of spectrally shifted speech, consistent with previous studies' findings (Fu and Galvin, 2003; Rosen et al., 1999). However, the amount of improvement is highly dependent on the training protocols. The results from the present study provide several important findings about the effects of different auditory training protocols on the recognition of spectrally shifted speech.

Not surprisingly, vowel recognition results with Group 1 (test-only protocol; control group) showed that subjects did not benefit from repeated testing. However, despite the absence of explicit test feedback and the fact that only vowel tests with spectrally shifted speech were administered over the five-day training period, subjects' consonant recognition for eight-channel spectrally shifted speech significantly improved after the training period. Previous studies have shown that temporal cues play an important role in consonant recognition (e.g., Van Tasell et al., 1992), and that consonant recognition is generally less susceptible to spectral mismatches introduced by the speech processing than vowel recognition. Because temporal cues may have been well preserved by the eight-channel processing, repeated exposure to the spectrally shifted vowel test stimuli may have helped listeners to better hear out temporal cues. As an experimental control, Group 1 completed the same number of tests as the other training groups. Because vowel recognition with spectrally shifted speech did not improve after five days of repeated testing, no overt "procedural learning" (i.e., test environment, procedures, etc.) or "perceptual learning" (spectrally shifted speech patterns) (Wright and Fitzgerald, 2001; Hawkey et al., 2004) was observed with Group 1, suggesting that any learning observed with the other training groups would be largely "perceptual."

Allowing subjects to preview the spectrally shifted vowel tokens immediately before testing (Group 2) significantly improved recognition of spectrally shifted vowels in three out of four subjects. This result suggests that most subjects were able to learn some of the test-

specific vowels. However, because the preview test talkers (two male and two female) were different from the test talkers (five male and five female), Group 2 subjects did not learn to identify the exact tokens used in the vowel test. One could argue that acoustic differences between talkers may have been substantially reduced by the speech processing, and therefore subjects were largely listening to nearly identical vowel tokens in the preview and test. However, results from a recent study revealed that talker variability was somewhat preserved by spectrally degraded speech (Chang and Fu, unpublished data). For Group 2, recognition of spectrally shifted consonants improved by the end of the training period (11.3 percentage points). This improvement was slightly better than that observed for Group 1 (test-only), but worse than that observed for Group 3 (targeted vowel contrast training) or Group 4 (sentence training), both of whom were exposed to many consonants during the course of training. Thus, repeated preview of and exposure to the spectrally shifted vowel tokens somewhat generalized to improved recognition of spectrally shifted consonants, despite only being exposed to the consonants "h" and "d" during the course of training.

Targeted vowel contrast training with monosyllable words (Group 3) also significantly improved the recognition of spectrally shifted vowels; the amount of improvement was similar to that observed with Group 2 (~16 percentage points). For the training protocol used in Group 3, different stimuli and talkers were used for the training and tests. Recognition of spectrally shifted consonants was most improved with Group 3 (~24 percentage points), suggesting that subjects did benefit from the exposure to the initial and final consonants in the monosyllable training words, even though consonant recognition was not targeted by the training. The targeted vowel contrast training provided the greatest benefit, when recognition of both spectrally shifted vowels and consonants are considered. Because subjects were able to strongly focus on phonemic differences between stimuli, rather than relying on context cues (Group 4) or previewing a limited number of test-specific stimuli (Group 2), the training generalized to improved overall phoneme recognition.

The modified connected discourse/sentence training protocol (Group 4) was meant to mimic the listening and learning conditions that CI patients might experience in their daily life as they adapt to their implant device and speech processing via spoken language. Without overt auditory rehabilitation, CI listeners will adapt to electrically stimulated peripheral neural patterns via extended exposure to daily communication. After time, some speech sounds may be more easily recognized than others (e.g., when spoken by a familiar voice or family member). By training with sentences of easy-to-moderate difficulty, spoken by a single talker, NH listeners experienced to some degree the learning processes experienced by many CI users. Results with Group 4 showed no significant improvement in the recognition of spectrally shifted vowels after five days of sentence training, despite the feedback provided to listeners. Most subjects' performance quickly improved from an initial level of 25–50% correct word-in-sentence recognition to nearly 80% correct by the end of the first day of training. This improvement may have been due to the experimental method (e.g., showing onscreen boxes that represented each word in the sentence, revealing the words that the listeners correctly identified, etc.) or due to the stronger context cues available in a sentence recognition task. However, the improved sentence recognition did not improve subjects' recognition of spectrally shifted vowels. These results are not consistent with those of Rosen et al. (1999), who showed a dramatic improvement in subjects' recognition of intervocalic consonants, medial vowels in monosyllables, and words in sentences. There are significant differences between the speech processing, training methods, and test methods used in the present study and those used by Rosen et al. (1999). In the present study, the overall input frequency range was 200–7,000 Hz and the carrier range was 999–10,290 Hz, while for the Rosen et al. study, the input frequency range was 50–4,000 Hz and the carrier range was 360–10,000 Hz. Thus, in the present study, the spectral shift for the most apical channels was somewhat larger than that used in the Rosen et al. (1999) study, which may have made speech recognition ultimately more difficult. Also, a single female voice was used for training and testing in the Rosen et al. study, whereas a single male talker was used for sentence training and five male and five female talkers were used for vowel and consonant testing. These differences in speech processing, training methods and test procedures may have contributed to differences in the effectiveness of the connected discourse training. However, similar to the Rosen et al. results, mean consonant recognition with spectrally shifted speech improved significantly (~20 percentage points) after the sentence training, significantly better than the improvement observed with Group 1 (test-only) or Group 2 (vowel token preview). Because subjects in Group 4 were exposed to many consonants in the sentence training (as opposed to Group 1 and 2's exposure to "h" and "d" only), subjects were able to better learn spectrally shifted consonants. Thus, while sentence training did not generalize to improved recognition of spectrally shifted vowels, it may have generalized to improved recognition of spectrally shifted consonants.

Retesting vowel and consonant recognition with eight-channel spectrally unshifted speech after training with spectrally shifted speech showed no significant change in performance for any of the training groups. Thus, training with spectrally shifted speech did not seem to generalize to improved performance for frequency carrier ranges other than those used for training, consistent with results from previous studies (Fu and Galvin, 2003). It should be noted that because baseline performance with the eight-channel spectrally unshifted processors was already at such a high level, there was little room for improvement. Given that subjects were trained to listen to speech that was spectrally reduced and shifted, relative to unprocessed speech, it seems unlikely that the training would have resulted in improved perception of spectrally reduced speech only, as this parameter had the smallest effect on performance. It should also be noted that while it is possible to simulate the speech processing experienced by CI listeners, it may not always be possible to simulate the urgency of the learning process experienced by CI patients. There was considerable intersubject variability within most of the training groups in the present study, suggesting that the effects of a given training protocol may well have depended on NH subjects' motivation to learn. As such, the effects of these training methods may be somewhat different with CI listeners. In the previous study by Fu et al. (2004), targeted vowel contrast training significantly improved all CI subjects' phoneme recognition performance. The time course and degree of improvement varied among CI users, suggesting that despite daily, long-term experience with their implant device and a presumably strong motivation to improve their speech recognition skills, considerable intersubject variability remained. The training protocols used in the present study should be tested with CI patients to better evaluate their efficacy. However, because of patient-related and speech processor-related factors, it may be difficult to compare the training methods, as poorer performing subjects may not experience benefits comparable to those for the NH listeners in the present study.

One potential limitation of the present study is the relatively small number of subjects tested in each protocol. Because the experiment required no previous experience with any speech recognition testing or training as well as a time commitment of five consecutive days, it was difficult to find suitable subjects. A larger subject pool might have been helpful in more conclusively interpreting the results. However, even with four subjects per group (16 subjects in all), the results showed significant effects for the training protocol that may hold promise for CI patient rehabilitation protocols.

These results, combined with the results from previous studies (Rosen et al., 1999; Fu and Galvin, 2003; Fu et al., 2004), suggest that moderate amounts of daily training may be an effective approach toward improving CI patients' speech recognition, especially those patients with limited speech recognition abilities. The present study also suggests that the type of training may be an important consideration to effectively and efficiently train CI patients. The data from the sentence training protocol suggest that, while CI patients might eventually adapt via daily experience with spoken language, CI patients may not fully learn novel peripheral neural patterns and will rely more strongly on the context cues available in sentence recognition. Training CI listeners to identify targeted phoneme contrasts may provide a greater benefit, at least for phoneme recognition. This is extremely important for congenitally deafened CI patients, who must develop central speech template via electrical stimulation.

SUMMARY AND CONCLUSION

The results demonstrate that moderate amounts of auditory training can significantly improve NH listeners' recognition of spectrally shifted speech. Testing four different training protocols over a five-day training period revealed the following:

1. Repeated exposure to the test stimuli without training, preview, or feedback did not significantly affect recognition of spectrally shifted vowels.
2. Allowing subjects to preview the spectrally shifted vowel stimuli (spoken by a different set of talkers) significantly improved vowel recognition, and, to a lesser extent, consonant recognition (although subjects were exposed only to the consonants "h" and "d" during the training period).
3. Training that targeted vowel contrasts using monosyllable words provided the best overall phoneme recognition with spectrally shifted speech. Although consonant contrasts were not explicitly trained, subjects' recognition of spectrally shifted consonants was significantly improved, presumably because of exposure to consonants during the training.
4. Sentence recognition using a modified connected discourse method did not significantly improve recognition of spectrally shifted vowels. However, consonant recognition significantly improved, presumably because of exposure to consonants during the training. Sentence training may provide limited benefit for developing phoneme patterns because of strong context cues.

Overall, results from the present study suggest that targeted phoneme training may hold the most pro-

mise for developing CI patients' speech recognition skills, especially those patients with limited speech recognition abilities.

ACKNOWLEDGMENTS

We are grateful to all research participants for their considerable time spent with this experiment. We would also like to thank Dr. Brian Moore and another anonymous reviewer for the useful comments and suggestions. The research was supported by NIDCD grant R01-DC004792.

REFERENCES

Busby PA, Roberts SA, Tong YC, Clark GM. Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. Br. J. Audiol. 25:291–302, 1991.

Dawson PW, Clark GM. Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. Ear Hear. 18:488–501, 1997.

DeFilippo CL, Scott BL. A method for training and evaluation of the reception of on-going speech. J. Acoust. Soc. Am. 63:1186–1192, 1978.

Fu Q J, Galvin JJ. The effects of short-term training for spectrally mismatched noise-band speech. J. Acoust. Soc. Am. 113:1065–1072, 2003.

Fu Q J, Shannon RV. Recognition of spectrally degraded and frequency shifted vowels in acoustic and electric hearing. J. Acoust. Soc. Am. 105:1889–1900, 1999.

Fu Q J, Galvin JJ, Wang X, Nogaki G. Effects of auditory training on adult cochlear implant patients: a preliminary report. Cochlear Implants Int. 5 Suppl 1:84–90, 2004.

Greenwood DD. A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. 87:2592–2605, 1990.

Hawkey DJ, Amitay S, Moore DR. Early and rapid perceptual learning. Nat. Neurosci. 7:1055–1056, 2004.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97:3099–3111, 1995.

Nilsson M, Soli SD, Sullivan JA. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95:1085–1099, 1994.

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

Shannon RV, Zeng FG, Wygonski J. Speech recognition with altered spectral distribution of envelope cues. J. Acoust. Soc. Am. 104:2467–2476, 1998.

Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. J. Acoust. Soc. Am. 106:L71–L74, 1999.

Studebaker G. A "rationalized" arcsine transform. J. Speech Hear. Res. 28:455–462, 1985.

Tyler R, Parkinson AJ, Woodworth GG, Lowder MW, Gantz BJ. Performance over time of adult patients using the Ineraid or Nucleus cochlear implant. J. Acoust. Soc. Am. 102:508–522, 1997.

Van Tasell DJ, Greenfield DG, Logemann JJ, Nelson DA. Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlear implants. J. Acoust. Soc. Am. 92:1247–1257, 1992.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. New levels of speech recognition with cochlear implants. Nature 352:236–238, 1991.

Wright BA, Fitzgerald MB. Different patterns of human discrimination learning for two interaural cues to sound-source location. Proc. Natl. Acad. Sci. 98:12307–12312, 2001.