Moderate auditory training can improve speech performance of adult cochlear implant patients

Qian-Jie Fu, John Galvin, Xiaosong Wang, and Geraldine Nogaki
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057
qfu@hei.org

Abstract: Learning electrically stimulated speech patterns can be a new and difficult experience for many cochlear implant users. In the present study, ten cochlear implant patients participated in an auditory training program using speech stimuli. Training was conducted at home using a personal computer for 1 hour per day, 5 days per week, for a period of 1 month or longer. Results showed a significant improvement in all patients' speech perception performance. These results suggest that moderate auditory training using a computer-based auditory rehabilitation tool can be an effective approach for improving the speech perception performance of cochlear implant patients.

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts [JH]
Date Received: October 13, 2004 **Date Accepted:** April 12, 2005

1. Introduction

The cochlear implant (CI) is an electronic device that provides hearing sensation to patients with total hearing loss. Overall, the speech recognition performance of cochlear implant patients has steadily improved in recent years; however, considerable variability remains in individual CI patient outcomes. Several studies have assessed the effects of limited training on the speech recognition skills of poorer-performing CI users. Busby *et al.* (1991) examined the effects of ten 1-h speech perception and production training sessions that occurred 1–2 times per week. Three prelingually deafened multichannel CI users (two adolescents and one adult) participated in the training sessions. After the training period, there were only minimal changes in these individuals' recognition performance. The authors noted that the subject who exhibited the greatest improvement was implanted at an earlier age than the other two subjects. Dawson and Clark (1997) reported more encouraging results with vowel recognition training assessed for five CI users. Each subject had been deaf for at least 4 years prior to implantation, and none had achieved open-set speech recognition. Training consisted of one 50-min training session per week for 10 weeks. Following training, four of the five subjects showed some measure of improvement; this improvement was retained on subsequent testing 3 weeks after training was completed.

Previous attempts to improve the speech recognition abilities of poor- to moderate-performing CI users have shown only minimal success (e.g., Busby *et al.*, 1991; Dawson and Clark, 1997). Part of the difficulty in training and subsequent testing with these CI users may be due to poorly developed or nonexistent auditory-only central speech patterns. Speech recognition tests designed to measure the implant's ability to represent important speech cues may not be appropriate for these CI users, as there may be no robust central speech pattern with which to compare the electrically evoked speech pattern. These CI users may be quite capable of discriminating between the speech stimuli, but incapable of identifying the correct speech stimulus. Interestingly, these patients may be able to accurately produce these speech tokens, because of extensive speech production training. Thus, for these poorer-performing patients, training may need to include explicit development of central speech patterns. The time course and intensity of such training protocols may need to be much longer than used in previous studies. In NH populations, training has been successfully used to improve speech segment discrimination and identification (Tremblay *et al.*, 1998), and recognition of spectrally-shifted speech (Rosen *et al.*, 1999; Fu and Galvin, 2003). These improvements typically occurred only

Table. 1. The CI patients who participated in the training project.

Subject	Age	Gender	Pre/ post lingual	Duration of use (years)	Implant device	Strategy	Vowels (%)		Consonants		Gender		Sentence	
							Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
S1	46	F	Pre	2	N-24	SPEAK	9.6	26.9			47.9	48.3		
S2	51	M	Pre	3	C-II	CIS	10.0	18.6			55.8	54.6		
S3	25	F	Pre	1	C-II	PPS	11.8	27.8	6.0	11.3				
S4	40	F	Post	2	C-II	MPS/Hi Res	14.1	27.2	15.0	27.8	63.0	64.5		
S5	40	F	Pre	1	N-24	ACE	24.5	35.9	16.1	24.0	85.8	85.0		
S6	48	M	Pre	1	C-II	CIS	26.0	37.5						
S7	40	F	Pre	1	C-II	SAS/HiRes	32.7	56.7	27.5	57.0			0.0	29.7
S8	36	F	Post	1	C-II	HiRes	33.1	60.0	46.5	68.5	89.1	89.5	51.4	81.4
S9	60	M	Post	6	N-22	SPEAK	34.0	47.7	34.0	41.7	85.6	88.3		
S10	38	M	Pre	1	N-24	SPEAK	41.5	56.6	30.4	40.0			32.3	56.4

after the completion of much more intensive training, as compared to the training used in previous studies with CI listeners.

The mixed and generally poor outcomes from previous CI speech training studies may well be due to the amount and type of training employed. A more extensive and intensive approach to pattern training that includes minimal speech contrasts might yield better results. In the present study, two primary hypotheses were investigated regarding speech recognition training with CI patients: (1) CI users' poor speech recognition performance can be improved by intensive auditory training, and (2) Speech recognition training can be performed with minimal supervision using a computer program installed on patients' home computers.

2. Method

2.1 Subjects

Ten CI patients with poor-to-moderate speech recognition ability were recruited to participate in a training program. Table 1 contains relevant information for the ten subjects. Subjects were paid for both their training and testing hours.

2.2 Testing materials and procedure

Speech recognition was assessed using four sets of test materials, including three closed-set identification tasks and one open-set recognition task. The three closed-set identification tasks included multitalker vowel recognition, multitalker consonant recognition, and voice gender discrimination. Vowel recognition was measured in a 12-alternative identification paradigm. The vowel set included ten monophthongs (/i ɪ ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and two diphthongs (/əʊ eɪ/), presented in an /h/–vowel–/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). 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 the consonant recognition test were digitized natural productions from five men and five women drawn from speech samples collected by Shannon *et al.* (1999). Voice gender recognition was measured in a two-alternative identification paradigm. The tokens were the same as those used in the vowel recognition. Open-set auditory-only word-in-sentence recognition was measured only for those subjects whose baseline vowel and consonant recognition scores were greater than 30% correct. Baseline sentence recognition was not

1529-7853/05/6(3)/107/6/$22.50

measured in the remaining poorer-performing patients as, most likely, no recognition was possible given the poor phoneme recognition scores. For the three subjects whose mean phoneme recognition scores were greater than 30% correct subjects, word-in-sentence recognition was measured in quiet using the Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994).

Each test block included 120 tokens (12 vowels×10 talkers) for vowel identification and voice gender 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 of each token, the subject responded by pressing one of 12 buttons in the vowel test, one of 20 buttons in the consonant test, or one of 2 buttons in the voice gender discrimination, each marked with one of the possible responses. The response buttons were labeled in an /h/–vowel–/d/ context for the vowel recognition task, /a/–consonant–/a/ context for the consonant recognition task, and "male or female" for the voice gender discrimination. 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. For HINT sentence recognition, a ten-sentence list was chosen pseudorandomly from among 26 lists, and sentences were chosen randomly, without replacement, from the sentences within that list. The subject responded by repeating the sentence as accurately as possible.

Due to the generally high variability in performance among CI patients, it is difficult to separate within-subject training effects from across-subject variability. To avoid this issue, a within-subject control procedure was adopted instead of the traditional across-subject control group procedure. The within-subject control procedure allowed "perceptual learning" effects to be better separated from "procedural learning" effects (i.e., task familiarization or test experience). For most subjects, baseline speech recognition performance (vowel, consonant, and voice gender recognition) was measured over a 2-week period (three test sessions per week), resulting in a minimum of six runs per test. During the baseline testing, more than 720 vowel and 1200 consonant tokens were presented to each subject. Thus, by the end of baseline testing, subjects were familiar with the test materials and procedures, reducing any influence of "procedural learning" on the subsequent training data. Asymptotic performance was observed in most subjects during the last 2 days of baseline collection; the last two performance measures were averaged and served as the baseline score for each subject. For the poorest-performing CI subjects (who performed at near-chance levels during the baseline collection), baseline performance was only measured for 1 week (three test sessions). Once training began, subjects returned to the lab every 2 weeks for retesting with the baseline speech materials.

2.3 Training materials and procedure

After baseline measures were complete, subjects trained at home, completing targeted training tasks based on their phoneme recognition scores. Speech training materials included more than 1000 monosyllabic and 200 nonsense words, each spoken by two males and two females (recorded at the House Ear Institute). Note that the training stimuli and training talkers were not used in the test stimulus set. Monosyllabic words were the primary stimuli used in the phonetic (vowel and consonant contrasts) training exercises.

Custom software, developed at the House Ear Institute, was used for training. The software was either installed onto subjects' home computers or onto loaner notebook computers. For each subject, baseline recognition results were analyzed and a targeted training program was developed. For the poorest-performing subjects, training began with a 3AFC discrimination protocol, in which three sounds (labeled "sound 1," "sound 2," and "sound 3") were played in sequence; two of the sounds were identical while the third was different. Subjects were asked to choose which sound was different. Initially, there were maximal differences (in terms of acoustic speech features) between the phonemes in the stimuli (i.e., "sit" vs "page"); as subjects' performance improved (i.e., greater than 80% correct for a given level of difficulty), the difference between phonemes in the stimuli was reduced (i.e., "sit" vs "sat"). For vowels, acoustic speech features included first and second formant frequencies ($F1$ and $F2$) and duration; for consonants, speech features included voice, manner, and place of articulation (Miller and

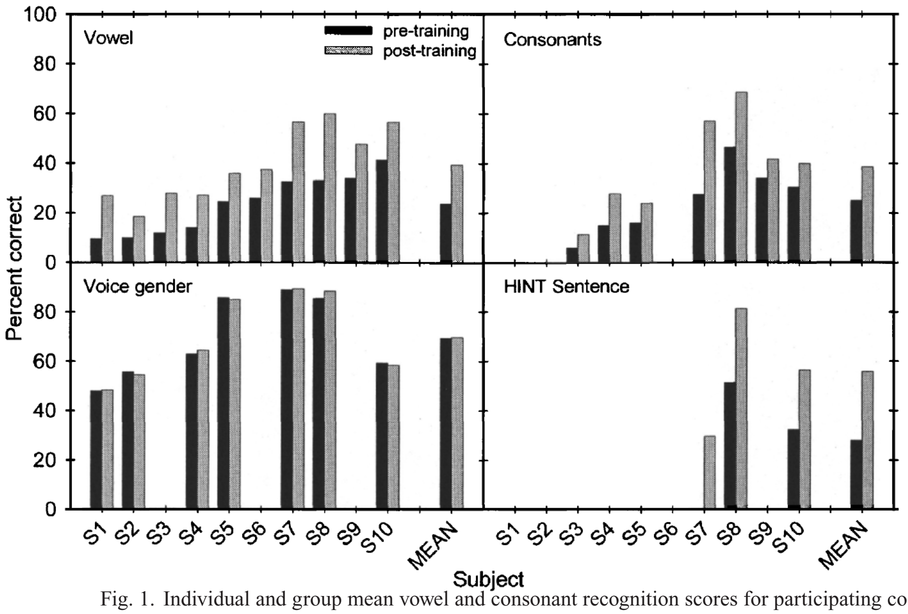

Fig. 1. Individual and group mean vowel and consonant recognition scores for participating cochlear implant subjects, before and after training.

Nicely, 1955). Visual feedback was provided as to the correctness of response, and auditory feedback was provided in which the three sounds, now labeled with the appropriate words (i.e., sit, sit, sat), were replayed in the same sequence. As subjects progressed beyond the 3AFC discrimination task (performance better than 80% correct), subjects were trained to identify medial vowels. In the identification training, only the medial vowel differed between response choices (i.e., "seed," "said," "sod," "sued"); this way, subjects were better able to focus only on differences between the medial vowels. Initially, subjects chose between two responses that differed greatly in terms of speech features (i.e., said, sued); as subjects' performance improved, the differences between speech features in the response choices was reduced (i.e., "said," "sad"). As subjects continued to improve, the number of response choices was increased (up to a maximum of six choices) and the acoustic differences between response choices were reduced. Consonant recognition was similarly trained. Subjects began consonant training with the 3AFC discrimination task. Monosyllabic were used for discrimination training, in which the speech feature differences between initial and final consonants were gradually reduced as subjects' performance improved (i.e., from "can" vs "fan" to "can" vs "pan" or "cap" vs "cat"). As subjects progressed beyond the 3AFC consonant discrimination task (performance better than 80% correct), medial consonant identification was trained using nonsense syllables (phonemes in v/*C*/v format, e.g., "ibi," "idi," "iki," etc.). Again, as subjects' performance improved, the acoustic differences between response choices were reduced and/or the number of response choices was increased.

Subjects were instructed to train for at least 1 hour per day, 5 days per week. The training software logged each subject's training session, including the time for each exercise and the total time spent training. For each subject, after analyzing the results from speech tests conducted in the lab, modifications were made to the training program to target the stimuli that were most confusing. Subjects returned to the lab every 2 weeks for retesting with the baseline speech materials.

3. Results

Figure 1 shows individual vowel, consonant, gender, and sentence recognition scores for the CI patients who participated in the training protocol; both pre- and post-training scores are shown.

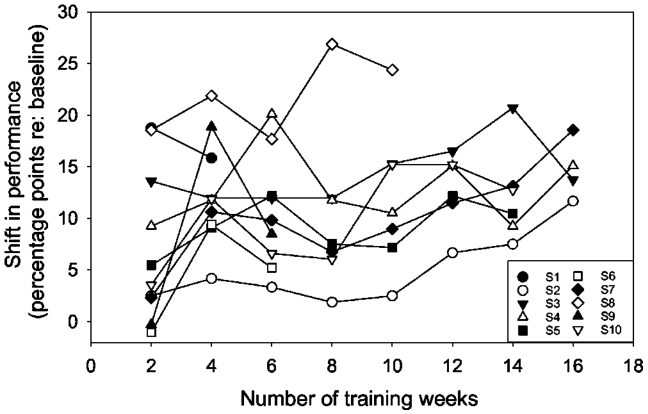

Fig. 2. Vowel recognition performance during the training period for each individual subject.

The results show that all CI patients' vowel and consonant recognition scores significantly improved after training. Mean vowel recognition scores improved from 23.7% to 39.5% correct [paired student t-test: $t(9)=8.72$, $p<0.0001$]. Consonant training and testing was conducted with seven of the ten CI subjects; mean consonant recognition scores improved from 25.1% to 38.6% correct [$t(6)=4.02$, $p<0.005$]. Voice gender recognition was measured in seven of the ten CI patients; mean gender recognition scores did not significantly change with training [69.5% to 69.8% correct; $t(6)=0.62$, $p=0.56$]. Open-set word-in-sentence recognition was measured in three of the ten CI subjects; mean recognition scores improved significantly, from 27.9% to 55.8% correct [$t(2)=14.59$, $p<0.01$].

Figure 2 shows individual subjects' vowel recognition scores for each retest session during the training experiment. There was considerable variability in the amount of improvement, as well as the rate and time course of improvement. Some subjects showed immediate and sharp improvement after only 2 weeks of training (S1, S8), while others required a much longer time course (3 months for S2) before performance significantly improved.

4. Discussion

The results show that there was significant improvement in all CI patients' speech perception performance following 4 weeks of moderate training. However, the amount of improvement was highly variable across subjects. The improvement in vowel recognition ranged from 9 percentage points (S2) to 27 percentage points (S8), while the improvement in consonant recognition ranged from 5 percentage points (S3) to 30 percentage points (S7). The time course and rate of improvement was also highly variable across subjects, with some subjects rapidly improving during the course of training and others much slowly.

Previous studies' failure to show significant benefits from training may reflect the limitations of the training programs, and not necessarily CI patients' potential to improve their speech recognition. In many of these studies, training has been limited to infrequent (1/week), short duration (1-h) sessions. More intensive training programs may result in larger and more consistent improvements in CI users' speech recognition ability. Besides the amount of time spent on training, the design and implementation of a training protocol may also be crucial to the success of that training. A speech contrast training protocol that targeted minimal speech contrasts was used in the present study; a large set of novel, multitalker stimuli was used to train subjects. By using multitalker training materials, subjects may have been able to better learn phoneme patterns than with the single-talker training stimuli used in other studies. The training protocol used in the present study (targeted phoneme training) may have also contributed to the successful outcome; our recent work with NH listeners suggests that training with minimal speech contrasts may best improve phoneme recognition (Fu *et al.*, 2005).

Despite the improvement with training, most subjects' phoneme recognition performance remained in the poor to fair range. Given the difficulty in vowel and consonant recognition even after training, many of these subjects would have great difficulty in recognizing sentences (open set, auditory only). HINT word-in-sentence recognition was tested only for those subjects whose baseline phoneme recognition scores were above 30% correct. For these subjects, sentence recognition performance improved significantly after the targeted phoneme training. For the remaining subjects, the improved phoneme recognition most likely would not have generalized to improved auditory-only sentence recognition. The time course for such an improvement may be years rather than weeks or months. However, it is possible that the gains in vowel and consonant recognition may help these patients' sentence recognition with visual cues (i.e., lip reading).

These results suggest that moderate amounts of daily training (1–2 h per day, 5 days per week) may be an effective approach toward improving CI patients' speech recognition, especially those patients with limited speech recognition abilities. The improved phoneme recognition via targeted training of phonetic contrasts may also generalize to improved recognition of words in sentences. The computer-based auditory training tool used in the present study may be a useful alternative or complement to auditory rehabilitation provided by clinical speech pathologists.

Acknowledgments

We are grateful to all research participants for their considerable time spent with this experiment. We would also like to thank Dr. James Hillenbrand, Dr. Stuart Rosen, and an anonymous reviewer for useful suggestions on an earlier draft of this paper. The research was supported by NIDCD Grant R01-DC004792.

References and links

Busby, P.A., Roberts, S.A., Tong, Y.C., and Clark, G.M. **1991**. "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.

Dawson, P.W., and Clark, G.M. **1997**. "Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant," Ear Hear. **18**, 488–501.

Fu, Q.-J., and Galvin III, J.J. **2003**. "The effects of short-term training for spectrally mismatched noise-band speech," J. Acoust. Soc. Am. **113**(2), 1065–1072.

Fu, Q.-J., Nogaki, G., and Galvin III, J.J. **2005**. "Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation," J. Assoc. Res. Otolaryngol. (in press).

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

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

Nilsson, M., Soli, S.D., and Sullivan, J.A. **1994**. "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.

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., Jensvold, A., Padilla, M., Robert, M.E., and Wang, X. **1999**. "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Tremblay, K., Kraus, N., and McGee, T. **1998**. "The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training," NeuroReport **9**, 3557–3560.

1529-7853/05/6(3)/111/6/$22.50