



Published in final edited form as:

Hear Res. 2008 August ; 242(1-2): 198–208. doi:10.1016/j.heares.2007.11.010.

Maximizing Cochlear Implant Patients' Performance with Advanced Speech Training Procedures

Qian-Jie Fu and John J. Galvin III

Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057

Abstract

Advances in implant technology and speech processing have provided great benefit to many cochlear implant patients. However, some patients receive little benefit from the latest technology, even after many years' experience with the device. Moreover, even the best cochlear implant performers have great difficulty understanding speech in background noise, and music perception and appreciation remain major challenges. Recent studies have shown that targeted auditory training can significantly improve cochlear implant patients' speech recognition performance. Such benefits are not only observed in poorly performing patients, but also in good performers under difficult listening conditions (e.g., speech noise, telephone speech, music, etc.). Targeted auditory training has also been shown to enhance performance gains provided by new implant devices and/or speech processing strategies. These studies suggest that cochlear implantation alone may not fully meet the needs of many patients, and that additional auditory rehabilitation may be needed to maximize the benefits of the implant device. Continuing research will aid in the development of efficient and effective training protocols and materials, thereby minimizing the costs (in terms of time, effort and resources) associated with auditory rehabilitation while maximizing the benefits of cochlear implantation for all recipients.

Keywords

Cochlear Implants; Targeted Auditory Training; Auditory Rehabilitation

I. INTRODUCTION

The cochlear implant (CI) is an electronic device that provides hearing sensation to patients with profound hearing loss. As the science and technology of the cochlear implant have developed over the past 50 years, CI users' overall speech recognition performance has steadily improved. With the most advanced implant technology and speech processing strategies, many patients receive great benefit, and are capable of conversing with friends and family over the telephone. However, considerable variability remains in individual patient outcomes. Some patients receive little benefit from the latest implant technology, even after many years of daily use of the device. Much research has been devoted to exploring the sources of this variability in CI patient outcomes. Some studies have shown that patient-related factors, such as duration

Corresponding Author: Qian-Jie Fu, Ph.D., Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057, Phone: 213-273-8036, Fax: 213-413-0950, Email: qfu@hei.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of deafness, are correlated with speech performance (Eggermont and Ponton, 2003; Kelly et al., 2005). Several psychophysical measures, including electrode discrimination (Donaldson and Nelson, 1999), temporal modulation detection (Cazals et al., 1994; Fu, 2002), and gap detection (Busby and Clark, 1999; Cazals et al., 1991; Muchnik et al., 1994), have also been correlated with speech performance.

Besides the high variability in CI patient outcomes, individual patients also differ in terms of the time course of adaptation to electric hearing. During the initial period of use, post-lingually deafened CI patients must adapt to differences between their previous experience with normal acoustic hearing and the pattern of activation produced by electrical stimulation. Many studies have tracked changes in performance over time in “naïve” or newly implanted CI users. These longitudinal studies showed that most gains in performance occur in the first 3 months of use (George et al., 1995; Gray et al., 1995; Loeb and Kessler, 1995; Spivak and Waltzman, 1990; Waltzman et al., 1986). However, continued improvement has been observed over longer periods for some CI patients (Tyler et al., 1997). Experienced CI users must also adapt to new electrical stimulation patterns provided by updated speech processors, speech processing strategies and/or changes to speech processor parameters. For these patients, the greatest gains in performance also occurred during the first 3–6 months, with little or no improvement beyond 6 months (e.g., Dorman and Loizou, 1997; Pelizzone et al., 1999).

These previous studies suggest that considerable auditory plasticity exists in CI patients. Because of the spectrally degraded speech patterns provided by the implant, “passive” learning via long-term use of the device may not fully engage patients’ capacity to learn novel stimulation patterns. Instead, “active” auditory training may be needed to more fully exploit CI patients’ auditory plasticity and facilitate learning of electrically stimulated speech patterns. Some early studies assessed the benefits of auditory training in poor-performing CI patients. Busby et al. (1991) observed only minimal changes in speech performance after ten 1-hour training sessions; note that the subject with the greatest improvement was implanted at an earlier age, and therefore had a shorter period of deafness. Dawson and Clark (1997) reported more encouraging results for ten 50-minute vowel recognition training sessions in CI users; four out of the five CI subjects showed a modest but significant improvement in vowel recognition, and improvements were retained three weeks after training was stopped.

While there have been relatively few CI training studies, auditory training has been shown to be effective in the rehabilitation of children with central auditory processing disorders (Hesse et al., 2001; Musiek et al., 1999), children with language-learning impairment (Merzenich et al., 1996; Tallal et al., 1996), and hearing aid (HA) users (Sweetow and Palmer, 2005). Studies with normal-hearing (NH) listeners also show positive effects for auditory training. As there are many factors to consider in designing efficient and effective training protocols for CI users (e.g., generalization of training benefits, training method, training stimuli, duration of training, frequency of training, etc.), these studies can provide valuable guidance.

In terms of positive training outcomes, it is desirable that some degree of generalized learning occurs beyond the explicitly trained stimuli and tasks, i.e., some improvement in overall perceptual acuity. The degree of generalization may be influenced by the perceptual task, and the relevant perceptual markers. For example, Fitzgerald and Wright (2005) found that auditory training significantly improved sinusoidally-amplitude-modulation (SAM) frequency discrimination with the trained stimuli (150 Hz SAM). However, the improvement did not generalize to untrained stimuli (i.e., 30 Hz SAM, pure tones, spectral ripples, etc.), suggesting that the learning centered on the acoustic properties of the signal, rather than the general perceptual processes involved in temporal frequency discrimination. In an earlier study, Wright et al. (1997) found that for temporal interval discrimination, improvements for the trained interval (100 ms) generalized to untrained contexts (i.e., different border tones); however, the

improved temporal interval perception did not generalize to other untrained intervals (50, 200, or 500 ms). While these NH studies employed relatively simple stimuli, the results imply that the degree and/or type of generalization may interact with the perceptual task and the training/test stimuli. Some acoustic CI simulation studies with NH listeners have also reported generalized learning effects. For example, Nogaki et al. (2007) found that spectrally-shifted consonant and sentence recognition significantly improved after five training sessions that targeted medial vowel contrasts (i.e., consonant and sentence recognition were not explicitly trained). In contrast, some acoustic CI simulation studies show that auditory training may not fully compensate for the preservation/restoration of the normal acoustic input, especially in terms of frequency-to-cochlear place mapping (e.g., Faulkner, 2006; Smith and Faulkner, 2006). Thus, there may be limited training benefits in the real CI user case. It is unclear whether benefits of training with simple stimuli (e.g., tones, bursts, speech segments, etc.) and simple tasks (e.g., electrode discrimination, modulation detection, phoneme identification, etc.) will extend to complex stimuli and tasks (e.g., open set sentence recognition).

It is also important to consider the most efficient and effective time course of training. How long and/or how often must training be performed to significantly improve performance? Extending their previous psychophysical studies, Wright and Sabin (2007) found that frequency discrimination required a greater amount of training than did temporal interval discrimination, suggesting that some perceptual tasks may require longer periods of training. Nogaki et al. (2007) reported that, for NH subjects listening to severely-shifted acoustic CI simulations, the total amount of training may matter more than the frequency of training. In the Nogaki et al. (2007) study, all subjects completed five one-hour training sessions at the rate of 1, 3, or 5 sessions per week; while more frequent training seemed to provide a small advantage, there was no significant difference between the three training rates.

Different training methods have been shown to affect training outcomes in acoustic CI simulation studies with NH listeners. For example, Fu et al. (2005a) compared training outcomes for severely-shifted speech for 4 different training protocols: test-only (repeated testing equal to the amount of training), preview (direct preview of the test stimuli immediately before testing), vowel contrast (targeted medial vowel training with novel monosyllable words, spoken by novel talkers), and sentence training (modified connected discourse). Medial vowel recognition (tested with phonemes presented in a h/V/d context) did not significantly change during the 5-day training period with the test-only and sentence training protocols. However, vowel recognition significantly improved after training with the preview or vowel contrast protocols. Different protocols may be appropriate for different listening conditions, and may depend on the degree of listening difficulty. For example, Li and Fu (2007) found that, for NH subjects listening to acoustic CI simulations, the degree of spectral mismatch interacted with the different protocols used to train NH subjects' vowel recognition. In that study, subjects were repeatedly tested over the five-day study period while listening to either moderately- or severely shifted speech. During testing, the 12 response boxes were either lexically or non-lexically labeled. With moderately-shifted speech, vowel recognition improved with either the lexical or non-lexical labels, i.e., evidence of "automatic" learning. With severely-shifted speech, vowel recognition improved with the non-lexical labels (i.e., discrimination improved), but not with the lexical labels (i.e., no improvement in identification), suggesting that difficult listening conditions may require explicit training with meaningful feedback. For CI users, these studies imply that different training protocols may be needed to address individual deficits or difficult listening conditions.

Finally, different training materials may influence training outcomes. Is it better to train with a well-known or well-defined group of stimuli, or does variability in the stimulus set provide better adaptation? Will training with relatively difficult stimuli improve recognition of both easy and difficult stimuli, or will enhanced or simplified training stimuli provide better

outcomes? There is positive evidence for both approaches. For example, Nagarajan et al. found that, for language learning-impaired children, modified speech signals (i.e., prolonged duration enhanced envelope modulation) improved recognition of both the modified and unprocessed speech signals during a 4-week training period. The modest benefits of auditory training in CI users observed by Busby et al. (1991) and Dawson and Clark (1997) may have been limited by the small stimulus sets used for training. Intuitively, it seems that more varied training stimuli may provide greater adaptation.

In the following section, we describe some of our recent auditory training studies in CI users. Some sections report preliminary data obtained with one or two subjects, while others summarize previously published work with ten or more subjects. Different from many auditory training studies in NH listeners that use separate experimental control groups, we adopted a “within-subject” control procedure for most of our CI auditory training studies. In CI research, inter-subject variability is a well-known phenomenon. For these sorts of intensive training studies with CI users, it is difficult to establish comparable performance groups to provide appropriate experimental controls. To overcome the high variability of speech performance across CI patients, large numbers of CI subjects are needed to evaluate the benefits of auditory training. A within-subject control procedure provides an alternative method to evaluate the effectiveness of auditory training using a relatively small amount of subjects. In any training study, it is important to determine whether performance improvements are due to perceptual learning or to procedural learning (i.e., improvements that result from learning the response demands of the task). To minimize procedural learning effects, we repeatedly measured baseline performance in each subject (in some cases for 1 month or longer) until achieving asymptotic performance. For each subject, mean asymptotic performance was compared to post-training results to determine the effect of training.

II. RECENT WORK IN AUDITORY TRAINING FOR CI PATIENTS

A) Electrode Discrimination Training

Previous studies have shown a significant correlation between electrode discrimination and speech performance (Donaldson and Nelson, 1999). Training with basic psychophysical contrasts may improve CI users’ speech performance, as improved electrode and/or pitch discrimination may extend to improved sensitivity to the spectral cues contained in speech signals. We conducted a pilot experiment to see whether CI patients’ electrode discrimination abilities could be improved with psychoacoustic training. A pre-lingually deafened 36-year-old male CI user with poor-to-moderate speech performance participated in the experiment. His bilateral hearing loss was identified at 18 months of age and progressed to profound by age 29. At the onset of training, the subject had used the Nucleus-22 implant (SPEAK strategy) for more than four years. Pre-training measures of consonant and vowel recognition were 34.1 and 55.6 percent correct, respectively.

For the electrode discrimination testing and training, all stimuli were delivered via custom research interface (Shannon et al., 1990), which allowed precise control of the pulse trains delivered to the electrodes. All stimuli were charge-balanced biphasic pulse trains, 500 pulses-per-second, 20 ms in duration, with a pulse phase duration of 200 μ sec. The stimulation mode was bipolar plus one (BP+1), i.e., the active and return electrodes were separated by one electrode. For the purposes of this paper, a single electrode pair will be referred to by its active electrode number and the return electrode number will be omitted; thus, electrode pair (8,10) will be referred to as 8. Electrode pair contrasts will be referred to by the probe electrode pair, followed by the reference electrode pair (indicated with an “r”). For example, electrode pair contrast “8,13r” refers to the probe electrode pair 8 and the reference electrode pair 13. In order to reduce the influence of loudness cues on electrode discrimination, all stimulating electrode pairs were loudness-balanced. Before loudness-balancing, threshold and comfortable

stimulation levels (T-and C-levels) were measured to estimate the dynamic range (DR) for each experimental electrode pair. All stimuli were loudness-balanced to 60% of the DR for electrode 10, using a 2-interval forced-choice (2IFC), adaptive, double-staircase procedure (Jesteadt, 1980). During electrode discrimination testing and training, the stimulus amplitude was randomly roved between 79.4 and 100 percent of these loudness-balanced levels to further reduce any effects of loudness cues.

Before training was begun, baseline electrode discrimination was measured between all possible electrode pair contrasts using a 3IFC task. Thus, electrode 1 was compared to all electrodes in the array (1 – 20), electrode 2 was compared to all electrodes in the array (1 – 20), etc. Each trial contained three randomized intervals; two intervals contained the reference electrode pair and one interval contained the probe electrode pair. Blue squares were illuminated on the computer screen in synchrony with the stimulus playback. The subject was instructed to ignore any loudness variations and identify the interval that contained the different pitch (i.e., the probe electrode pair). During baseline testing, no feedback and no training was provided. The percent correct was calculated for each electrode pair from 40 trials per contrast across the entire array of available electrode pairs. The sensitivity index (d') was calculated from the percent correct scores obtained from the cumulative stimulus response matrices. For 100% correct, $d' = 3.62$; for 33.3% correct, $d' = 0.0$.

After baseline electrode discrimination measures were completed, training was begun. The electrode pair contrasts used for training were selected according to the baseline results. Both relatively strong (8, 15r and 7, 13r) and weak (11, 14r and 7, 9r) electrode contrasts were trained. Discrimination of the strong contrasts was trained during the first week of training. The subject trained for approximately 2 hours per day, for 5 consecutive days. Similar to the baseline testing, the amplitude of the stimuli was roved over a small range to reduce any loudness cues on electrode discrimination performance. The presentation of the two contrasts (8, 15r and 7, 13r) was randomized across trials. At the end of the strong contrast training, electrode discrimination was re-measured for all electrode contrasts. Next, the subject similarly trained for 5 consecutive days with the weak electrode contrasts (11, 14r and 7, 9r). After training with weak contrasts, electrode discrimination was again re-measured for all electrode contrasts. Finally, to see whether any improvements of electrode discrimination performance were retained, no training or testing was conducted for 2 weeks. After this hiatus, electrode discrimination was re-measured for all electrode contrasts.

During training, a two-step training approach was used for each trial. Similar to testing, each trial contained three randomized intervals. In Step 1 of the training procedure, each stimulus presentation (interval) was paired with a visual cue: the two intervals containing the reference electrode were paired with red squares and the interval containing the probe electrode was paired with a green square. In Step 2, stimuli were presented without visual cueing (each interval was paired with a blue square, as in the baseline test procedure) in a newly-randomized sequence with newly-roved amplitudes. Visual feedback (correct/incorrect) was provided after each response. For each electrode contrast, daily training consisted of four sets of 100 trials; each trial employed the two-step training procedure.

In addition to electrode discrimination testing and training, closed-set vowel and consonant recognition was evaluated. Speech stimuli were presented to the subject while seated in a sound-treated booth via a single loudspeaker (Tannoy Reveal) at 70 dBA. The subject was tested using his clinically assigned speech processor and sensitivity settings for the duration of the experiment. Vowel recognition was measured using a 12-alternative identification paradigm. The vowel set included 10 monophthongs and 2 diphthongs presented in an /h/vowel/ d/ context. Consonant recognition was measured in a 16-alternative identification paradigm presented in an /a/consonant/a/ context.

Figure 1 shows electrode discrimination performance during the various stages of the experiment. Test Session 1 shows baseline, pre-training performance. Test Session 2 shows performance after 5 days of training with the strong contrasts. Test Session 3 shows performance after 5 days of training with the weak contrasts. Test Session 4 shows follow-up performance 2 weeks after training was stopped. For purposes of presentation, the d' data are displayed as contour surfaces, ranging from 1.25 (approximately 70% correct on a 3IFC task) to 3.62 (100 % correct). Electrode discrimination contrasts that produced higher d' values are represented by “hotter” colors (yellow through red), and lower to moderate d' values are represented by “cooler” colors (blue through green). The contour plots also show the electrode contrasts used for training (8,15r and 7,13r, as shown by the open red circles and 11, 14r and 7, 9r, as shown by the open blue circles).

For the baseline, pre-training measures (Test Session 1), electrode discrimination performance was generally poor. The subject required, on average, a separation of 8 electrodes (or 6 mm) to achieve a d' of 1.25. After training with the strong electrode contrasts (Test Session 2), electrode discrimination markedly improved, as seen by the larger regions of discriminability. The average electrode separation needed to achieve a d' of 1.25 was reduced to 4 electrodes (or 3 mm). The d' values for the trained electrode contrasts improved from 1.16 to 2.09 for contrast 8,15r, and 1.24 to 1.43 for contrast 7,13r. More importantly, this improvement generalized to untrained electrode contrasts, especially for electrodes 18, 19, and 20. After training with the weak electrode contrasts (Test Session 3), electrode discrimination continued to improve, as seen by the expanding regions of discriminability and the distribution of “hot” contours (which indicate strong electrode discrimination). The average electrode separation needed to achieve a d' of 1.25 was reduced to 3 electrodes (or 2.25 mm). However, d' values for the trained electrode contrasts were unchanged after 5 days of training (1.16 for 11,14r, and 0.33 for 7,9r). Interestingly, the d' value for the previously trained contrast 7,13r increased from 1.43 to 2.23 after training with the weak contrasts (d' for 8,15r was unchanged). Post-training follow-up measures (Test Session 4) showed a slight decline in performance, as indicated by the receding regions of discriminability and “hot spots,” relative to Test Sessions 2 and 3. The average electrode separation needed to achieve a d' of 1.25 slightly increased to 5 electrodes (or 3.75 mm). While follow-up measures suggest that some of the electrode discrimination improvements had been reduced, performance was much better than the baseline measures obtained before training (Test Session 1). These results suggest that most of the training benefits had been retained 2 weeks after training had been stopped.

Concurrent with the electrode discrimination training and testing, vowel and consonant recognition were measured on a daily basis. Baseline, pre-training performance was 55.6% correct for vowels and 38.0% correct for consonants. While vowel recognition only slightly improved during the training period (from 55.6% to 62.8% correct), consonant recognition greatly improved (from 38 % to 51 % correct). Post-training follow-up measures showed that the improvement in both vowel and consonant recognition was largely retained 2 weeks after training. The larger improvement in consonant recognition is somewhat surprising, given the assumedly stronger role of spectral envelope cues to vowel recognition (for which good electrode discrimination would be very important).

While only one subject participated in these experiments, this pilot study generated several significant results. First, in an adult CI user with poor-to-moderate electrode discrimination and speech performance despite long-term experience with the device, moderately intensive training significantly improved both electrode discrimination and speech performance. Second, the improvements in electrode discrimination were not restricted to the trained electrode contrasts, but rather generalized to other, untrained contrasts. Third, the improvements in electrode discrimination and phoneme recognition appeared to be largely retained after training

was stopped, at least over the short-term. While these results are encouraging, a greater number of subjects are needed to confirm these findings.

B) Targeted Speech Contrast Training in Quiet

While electrode discrimination training may improve the basic auditory resolution (and to some degree, speech performance), recent studies suggest that speech-specific training may be necessary to improve identification of spectrally-degraded and -distorted speech (Li and Fu, 2007). Passive adaptation may be adequate if the spectral resolution is adequate (i.e., > 8 functional channels) and the degree of spectral mismatch is small (i.e., < 3 mm, on average). For listeners with poorer spectral resolution or greater degrees of spectral mismatch, active training may be needed.

We recently studied the effect of targeted phonetic contrast on English speech recognition in ten CI users (Fu et al., 2005b). After extensive baseline measures, subjects trained at home using custom software (i.e., Computer-Assisted Speech Training, or CAST, developed at the House Ear Institute). Figure 2 shows examples of the extended baseline measures used for the within-subject experimental controls. The gray symbols represent performance influenced by procedural learning effects. The black symbols show asymptotic performance levels, which were averaged as the mean baseline performance for each subject. Subjects were asked to train 1 hour per day, 5 days per week for a period of 1 month (some subjects trained longer than 1 month). Training stimuli included more than 1000 monosyllable words, produced by 2 female and 2 male talkers; the training stimuli or talkers were not used in the test stimulus set. Vowel recognition was trained by targeting listeners' attention to medial vowels. During training, the level of difficulty was automatically adjusted in terms of the number of response choices and/or the degree of acoustic similarity among response choices. Each training run contained 25 trials; subjects typically completed 4–5 trials per day. During training, auditory and visual feedback were provided. Subjects returned to the lab every two weeks for re-testing of baseline measures. Results showed that, for all subjects, both vowel and consonant recognition significantly improved after training. Mean vowel recognition improved by 15.8 percentage points (paired t-test: $p < 0.0001$), while mean consonant recognition improved by 13.5 percentage points ($p < 0.005$). HINT sentence recognition (Nilsson et al., 1994) in quiet was also tested in three subjects; mean performance improved by 28.8 percentage points ($p < 0.01$). Thus, improved vowel and consonant recognition generalized to improved sentence recognition, to some degree. While performance improved for all subjects after four or more weeks of training, there was significant variability across subjects, in terms of the amount and time course of improvement. For some subjects, performance quickly improved after only a few hours of training, while others required a much longer time course to show any significant improvements.

We also recently studied the effect of targeted phonetic contrast training on Mandarin-speaking CI users' Chinese speech performance (Wu et al., 2007). In tonal languages such as Mandarin Chinese, the tonality of a syllable is lexically important (Lin, 1998; Wang, 1989). Tone recognition has been strongly correlated with Chinese sentence recognition (Fu et al., 1998a, 2004). In general, fundamental frequency (F0) information contributes most strongly to tone recognition (Lin, 1998). In CIs, F0 is only weakly coded, because of the limited spectral resolution and limits to temporal processing. However, other cues that co-vary with tonal patterns (e.g., amplitude contour and periodicity fluctuations) contribute to tone recognition, especially when F0 cues are reduced or unavailable (Fu et al., 1998a, 2004; Fu and Zeng, 2000). These cues are generally weaker than F0, and CI users may require targeted training to attend to these cues. Wu et al. (2007) studied the effect of moderate auditory training on Chinese speech perception in 10 hearing-impaired Mandarin-speaking pediatric subjects (7 CI users and 3 hearing aid users). After measuring baseline Chinese vowel, consonant and tone

recognition, subjects trained at home using the CAST program. Training stimuli included more than 1300 novel Chinese monosyllable words spoken by 4 novel talkers (i.e., the training stimuli and talkers were different from the test stimuli and talkers). The training protocol was similar to the Fu et al. (2005a) study. Subjects spent equal amounts of time training with vowel, consonant and tone contrasts. Results showed that mean vowel recognition improved by 21.7 percentage points (paired t-test: $p=0.006$) at the end of the 10-week training period. Similarly, mean consonant recognition improved by 19.5 percentage points ($p<0.001$) and Chinese tone recognition scores improved by 15.1 percentage points ($p=0.007$). Vowel, consonant, and tone recognition was re-measured 1, 2, 4 and 8 weeks after training was completed. Follow-up measures remained better than pre-training baseline performance for vowel ($p=0.018$), consonant ($p=0.002$), and tone recognition ($p=0.009$), suggesting that the positive effects of training were retained well after training had been stopped.

C) Melodic Contour Identification Training

While many CI patients are capable of good speech understanding in quiet, music perception and appreciation remains challenging. The relatively coarse spectral resolution, while adequate for speech perceptual performance, is not sufficient to code complex musical pitch (e.g., McDermott and McKay, 1995; Smith et al., 2002; Shannon et al., 2004), which is important for melody recognition, timbre perception and segregation/streaming of instruments and melodies. Previous studies have shown that CI listeners receive only limited melodic information (Gfeller et al., 2000; Kong et al., 2004). However, music perception may also depend on CI users' musical experience before and after implantation. Auditory training may be especially beneficial for music perception, enabling CI users to better attend to the limited spectro-temporal details provided by the implant. Structured music training has been shown to improve CI patients' musical timbre perception (Gfeller et al. 2002).

We recently investigated CI users' melodic contour identification (MCI) and familiar melody identification [FMI; as in Kong et al. (2004)], and explored whether targeted auditory training could improve CI listeners' MCI and FMI performance (Galvin et al., 2007). In the closed-set MCI test, CI subjects were asked to identify simple 5-note melodic sequences, in which the pitch contour (i.e., "Rising," "Falling," "Flat," "Rising-Falling," "Falling-Rising," etc.), pitch range (i.e., 220–698 Hz, 440–1397 Hz, 880–2794 Hz) and pitch intonation (i.e., the number of semitones between successive notes in the sequence, from 1 to 5 semitones) was systematically varied. MCI test results with 11 CI subjects showed large inter-subject variability in both MCI and FMI performance. The top performers were able to identify nearly all contours when there were 2 semitones between successive notes, while the poorer performers were able to identify only 30 percent of the contours with 5 semitones between notes. There was only a slight difference in mean performance among the three pitch ranges.

Six of the 11 CI subjects were trained to identify melodic contours, using pitch ranges not used in the baseline measures. Similar to the Fu et al. (2005a) and Wu et al. (2007) training protocols, subjects trained for 1 hour per day, 5 days per week, for a period of 1 month or longer. Training was performed using different frequency ranges from those used for testing, and auditory/visual feedback was provided, allowing participants to compare their (incorrect) response to the correct response. The level of difficulty was automatically adjusted according to subject performance by reducing the number of semitones between notes and/or increasing the number of response choices. Results showed that mean MCI performance improved by 28.3 percentage points ($p=0.004$); the amount of improvement ranged from 15.5 to 45.4 percentage points. FMI performance (with and without rhythm cues) was measured in 4 of the 6 CI subjects, before and after training. Mean FMI performance with rhythm cues improved by 11.1 percentage points, while mean FMI performance without rhythm cues improved by 20.7 percentage points ($p=0.02$). Note that FMI was not explicitly trained. Participants were trained only to identify

melodic contours, suggesting that some generalized learning occurred. Some subjects reported that their music perception and appreciation generally improved after MCI training (e.g., better separation of the singer's voice from the background music).

D) Speech in Noise Training

Background noise is a nearly constant presence for all listeners, whether from the environment (e.g., traffic noise, wind, rain, etc.), appliances (e.g., fans, motors, etc.), or even other people (e.g., "cocktail party effect," competing speech, etc.). Hearing-impaired listeners are generally more susceptible to interference from noise, and CI users are especially vulnerable (e.g., Dowell et al., 1987; Fu et al., 1998b; Hochberg et al., 1992; Kiefer et al., 1996; Nelson and Jin, 2004; Qin and Oxenham, 2003; Skinner et al., 1994). Indeed, CI users are more susceptible to background noise than normal-hearing (NH) subjects listening to comparable CI simulations (Friesen et al., 2001), especially when the noise is dynamic or competing speech (Fu and Nogaki, 2005; Müller-Deiler et al., 1995; Nelson et al., 2003). CI users' increased susceptibility to noise is most likely due to the limited spectral resolution and the high degree of spectral smearing associated with channel interaction (Fu and Nogaki, 2005). Much recent research and development have been aimed at increasing the spectral resolution and/or reducing channel interactions (e.g., virtual channels, current focusing, tri-polar stimulation, etc.). High stimulation rates have also been proposed to improve temporal sampling and induce more stochastic-like nerve firing. However, these approaches have yet to show consistent benefits to speech recognition in noise. The additional cues provided by different stimulation modes and processing strategies may be weaker than the more dominant coarse spectral envelope cues and low-frequency temporal envelope cues. Auditory training may be necessary to allow CI users to hear out these relatively weak acoustic cues in adverse listening environments. Individualized training programs have been shown to improve hearing-aid users' speech performance in noise (Sweetow and Palmer, 2005). However, the benefit of auditory training (alone or in conjunction with emergent CI technologies) has yet to be assessed for CI users' speech performance in noise.

We recently conducted a pilot study examining the effect of auditory training on CI users' speech perception in noise. Baseline, pre-training performance was measured for multi-talker vowels (Hillenbrand et al., 1994), consonants (Shannon et al., 1999) and IEEE sentences (1969) in the presence of steady speech-shaped noise and multi-talker speech babble at three signal-to-noise ratios (SNRs): 0, 5, and 10 dB SNR. After completing baseline measures, subjects trained at home using their personal computers for 1 hour per day, three days per week, for a period of 4 weeks or more. One CI subject was trained using a "phoneme-in-noise" protocol (similar to the phonetic contrast protocol described in Fu et al., 2005a), while another CI subject was trained using a "keyword-in-sentence" protocol. For the phoneme-in-noise training, stimuli consisted of multi-talker monosyllable words (recorded at House Ear Institute as part of the CAST program); for the keyword-in-sentence training, stimuli consisted of multi-talker TIMIT sentences (Garofolo et al., 1993). In the phoneme-in-noise training, a monosyllable word was played in the presence of multi-talker speech babble, and the subject responded by clicking on one of four choices shown onscreen; response choices differed in terms of only one phoneme (i.e., initial or final consonants, medial vowels). The SNR was automatically adjusted according to subject response. If the subject answered correctly, visual feedback was provided and the SNR was reduced by 2 dB for the next trial. If the subject answered incorrectly, auditory and visual feedback was provided (allowing subjects to compare their response to the correct response) and the SNR was increased by 2 dB. In the keyword-in-sentence training, a TIMIT sentence was played in the presence of multi-talker speech babble, and the subject responded by clicking on one of four keyword response choices shown onscreen. As in the phoneme-in-noise training, auditory and visual feedback was provided and the SNR was adjusted according to subject response.

Figure 3 shows the mean improvement (averaged across the three SNRs) for vowel, consonant, and sentence recognition in steady speech-shaped noise and speech babble, after training with speech babble. With the phoneme-in-noise training (left panel), mean vowel, consonant and IEEE sentence recognition in speech babble all improved by ~8 percentage points. With the keyword-in-sentence training, mean vowel and consonant recognition in speech babble improved by 12 and 6 percentage points, respectively; however, mean IEEE sentence recognition improved by 17 percentage points. For both protocols, there was a greater improvement in speech performance in steady noise than in speech babble (note that subjects were trained only using speech babble). While the two training protocols produced comparable improvements in phoneme recognition, the keyword-in-sentence training provided greater improvement in sentence recognition, for both noise types. These preliminary results suggest that different training protocols and materials may generalize more readily to a variety of listening conditions. Again, these are preliminary results with two subjects, and more subjects are needed to fully evaluate the efficacy of these training protocols.

E) Telephone Speech Training

Successful telephone use would certainly improve CI users' quality of life and sense of independence, as they would rely less on TTY, relay service, and/or assistance from NH people. For CI users who are able to understand natural speech (i.e., "broadband speech") without difficulty, the limited bandwidth (typically 300–3400 Hz) associated with telephone speech can make communication more difficult; although the SNR is typically good in telephone systems, some CI users may also be susceptible to telephone line noise (depending on the microphone sensitivity and gain control settings in the speech processor). Many CI users are capable of some degree of telephone communication (Cohen et al., 1989), but understanding of telephone speech is generally worse than that of broadband speech (Hornig et al., 2007; Ito et al., 1999; Milchard and Cullington, 2004). Hornig et al. (2007) measured vowel, consonant, Chinese tone and voice gender recognition in 16 Chinese-speaking CI subjects, for both broadband and simulated telephone speech. Results showed no significant difference between broadband and telephone speech for Chinese tone recognition. However, mean vowel, consonant and voice gender recognition was significantly poorer with telephone speech. Fu and Galvin (2006) also found that mean consonant and sentence recognition was significantly poorer with telephone speech in 10 English-speaking CI patients. These studies suggest that limited bandwidth significantly reduces CI patients' telephone communication. The large inter-subject variability in telephone speech performance in these studies suggests that some CI listeners may make better use of (or rely more strongly on) the high-frequency speech cues available with broadband speech.

We recently conducted a pilot study to investigate the effects of auditory training on CI users' understanding of telephone speech. One subject from the previous telephone speech experiment (Fu and Galvin, 2006) participated in the experiment. Because baseline measures showed that only consonant and sentence recognition were significantly poorer with telephone speech, only consonant recognition was trained. Using training software and protocols similar to Fu et al. (2005a), the subject trained for 1 hour per day, five days per week, for a period of 2 weeks. The subject trained with medial consonant contrasts that were not used for testing (i.e., "ubu," "udu," "uku," ... "ebe," "ede," "eke," etc.). All training stimuli were bandpass-filtered (300–3400 Hz) to simulate telephone speech. Similar to Fu et al. (2005a), the level of difficulty was adjusted according to subject performance, by changing the number of response choices and/or by changing the degree of acoustic similarity among response choices. Audio and visual feedback were provided, allowing the subject to compare his response to the correct response.

Baseline performance with simulated telephone speech was 0.83, 13.3, and 27.1 percentage points lower than that with broadband speech for vowel, consonant, and sentence recognition, respectively. The generally lower performance with telephone speech suggests that this subject relied strongly on high frequency speech cues, which were not preserved in simulated telephone speech. Figure 4 shows the performance deficit with telephone speech relative to broadband speech, before and after training. After training, the deficit in consonant recognition with telephone speech was reduced from 13 to 3 percent percentage points, and the deficit in sentence recognition was reduced from 27 to 9 percentage points. These preliminary results suggest that moderate auditory training may improve CI patients' recognition of telephone speech, and that training with band-limited consonants may generalize to improved recognition of band-limited sentences.

F) Interaction between Signal Processing Strategies and Auditory Training

After first receiving the implant, all CI users undergo a period of adaptation to the novel patterns of electric stimulation. Post-lingually deafened CI patients adapt to these new patterns in relation to previously learned auditory patterns experienced during normal hearing, prior to deafness. Pre-lingually deafened patients generally adapt to the electric stimulation patterns without these central patterns in place; thus, central patterns are developed with the implant only. When a CI patient is re-implanted or assigned a new speech processing strategy, a second adaptation process occurs. For post-lingually deafened patients, this second adaptation occurs in relation to the original acoustically-based central patterns, as well as the initial electrically-based central patterns. Because CI patients may benefit from new processing strategies, it is important to know the limits and time course of adaptation.

In Fu et al. (2002), we studied CI users' (Nucleus-22) passive adaptation to an extreme change in speech processing (i.e., a large shift in the place of stimulation). CI subjects' frequency allocation was abruptly changed from the clinical assignment (Table 9 or 7) to Table 1, resulting in ~ 1-octave up-shift in the channel frequency assignments; this frequency assignment effectively mapped acoustic information <1500 Hz onto a greater number of electrodes, at the expense of severe spectral shifting. No explicit training was provided, and subjects continuously wore the experimental maps for a three-month study period. Results showed that subjects were able to overcome the initial performance deficit with the experimental processors during the study period. However, performance with the experimental processors never equaled or surpassed performance with the clinically-assigned processors (with which subjects had years of familiarity). In a follow-up study with one of the subjects (Fu and Galvin, 2007), the same large spectral shift was gradually introduced. Gradual introduction of extreme changes in speech processing might allow for greater adaptation and reduce the stress associated with adaptation (e.g., Svirsky et al., 2001). The subject's clinically assigned frequency allocation (Table 7) was changed every three months until reaching Table 1 fifteen months later; no explicit training was provided, and the subject continuously wore these experimental processors during each 3-month study period. While the gradual introduction provided better adaptation for some measures (e.g., sentence recognition in quiet and noise), other measures (e.g., vowel and consonant recognition) remained poorer than those with the clinically assigned allocation. These studies suggest that passive learning may not be adequate to adapt to major changes in electrical stimulation patterns, and that explicit auditory training may be needed to provide more complete adaptation or accelerate the adaptation process.

We studied the effects of both passive and active learning on adaptation to changes in speech processing in one CI subject. The subject was Clarion II implant user with limited speech recognition capabilities, and previously participated in the Fu et al. (2005a) auditory training study. The subject was originally fit with an 8-channel CIS processor (Wilson et al., 1991; stimulation rate: 813 Hz /channel). Baseline (pre-training) multi-talker vowel and consonant

recognition was measured over a three-month period with the 8-channel CIS processor. No explicit training was provided during this baseline measurement period. Initial vowel and consonant recognition performance were 16.7% and 12.0% correct, respectively. During the three-month baseline period, performance improved until reaching asymptotic performance (32.7% correct for vowels, 27.5% correct for consonants). After completing baseline measures, the subject trained at home using the CAST program 2–3 times per week for 6 months; the subject was trained to identify medial vowel contrasts using monosyllable words in a *c/V/c* context. After training, mean vowel and consonant recognition with the 8-channel CIS processor improved by 12 and 17 percentage points, respectively. Note that medial consonant contrasts were not explicitly trained, although the subject was exposed to many initial and final consonants during the medial vowel contrast training. After completing the six-month training period with the 8-channel CIS processor, the participant was fit with the 16-channel HiRes strategy (stimulation rate: 5616 Hz/ channel). Baseline performance with the HiRes processor was measured for the next three months; again no explicit training was provided during this period of passive adaptation to the HiRes processor. Initial vowel and consonant recognition scores with the HiRes processor were 15.0% and 13.0% correct, respectively. At the end of the adaptation period, vowel and consonant recognition with the HiRes processor was not significantly different from that with the 8-channel CIS processor (post-training); vowel recognition was 45.3% correct and consonant recognition was 45.1% correct. After completing this second set of baseline measures with the HiRes processor, the subject again trained with medial vowel contrasts at home, 2–3 times per week for next three months. At the end of the training period, vowel and consonant recognition improved by an additional 10 and 12 percentage points, respectively. While this long-term experiment was conducted with only one CI user, these preliminary results demonstrate that active auditory training may benefit performance, even after extensive periods of passive adaptation to novel speech processors. The results also suggest auditory training may be necessary for CI users to access the additional spectral and temporal cues provided by advanced speech-processing strategies.

III. REMAINING CHALLENGES IN AUDITORY TRAINING FOR CI USERS

The results from these studies demonstrate that auditory training can significantly improve CI users' speech and music perception. The benefits of training extended not only to poor-performing patients, but also to good performers listening to difficult conditions (e.g., speech in noise, telephone speech, music, etc.). While auditory training generally improved performance in the targeted listening task, the improvement often generalized to auditory tasks that were not explicitly trained (e.g., improved sentence recognition after training with phonetic contrasts, or improved familiar melody identification after training with simple melodic sequences). More importantly, the training benefits seem to have been retained, as post-training performance levels generally remained higher than pre-training baseline levels 1–2 months after training was completed. Because most CI subjects had long-term experience with their device prior to training (i.e., at least one year of "passive" learning), the results suggest that "active" learning may be needed to receive the full benefit of their implant.

The generalized improvements observed in our research are in agreement with previous studies showing that auditory training produced changes in behavioral and neuro-physiological responses for stimuli not used in training (Tremblay et al., 1997). Tremblay et al. (1998) also reported that training-associated changes in neural activity may precede behavioral learning. In Fu et al. (2005a), the degree and time course of improvement varied significantly across CI subjects, most likely due to patient-related factors (e.g., the number of implanted electrodes, the insertion depth of the electrode array, duration of deafness, etc.). Some CI users may require much more auditory training, or different training approaches to improve performance. It may be useful to use objective neuro-physiological measures to monitor the progress of training

early on, and determine whether a particular training protocol should be continued or adjusted for individual CI patients.

The present results with CI users are largely in agreement with NH training studies with acoustic stimuli. Wright and colleagues (e.g., Wright et al., 1997; Wright, 2001; Fitzgerald and Wright, 2005; Wright and Fitzgerald, 2005; Wright and Sabin, 2007) have extensively explored fundamental aspects of learning using relatively simple stimuli and perceptual tasks, allowing for good experimental control. Of course, speech perception by the impaired auditory system involves greater complexity, both in terms of stimuli and perception. While it is difficult to directly compare the present CI studies to these psychophysical studies, some common aspects of training effects were observed. For example, in the present CI training studies, generalized improvements were observed from vowel training to consonant and sentence recognition (Fu et al., 2005a), or from training with one frequency range to testing with another (Galvin et al., 2007), or from training with one type of noise to testing with another type. Similarly, generalized improvements were observed in Wright et al. (1997) for untrained temporal interval discrimination contexts; however, improvements did not generalize to untrained temporal intervals. Further, Fitzgerald and Wright (2005) found that trained improvements in SAM frequency discrimination were limited to the training stimuli, and did not generalize to untrained SAM frequencies or stimuli. Again, it is difficult to compare these studies, as speech and music perception generally involve complex stimuli that may invoke different perceptual mechanisms than those used in more straightforward psychophysical tasks with relatively simple stimuli. And while it is important to understand the fundamentals of perceptual learning, it is encouraging that speech and music training have begun to show significant benefits for CI users.

In designing effective and efficient training programs for CI users, it is important to properly evaluate training outcomes, i.e., experiments and stimuli should be properly designed and selected. At a minimum, training studies should use different stimuli for testing and training (i.e., stimuli used for baseline measures should not be used for training). This requires adequate number of novel stimuli for testing and training. In our research, baseline speech measures were collected using standard testing databases [e.g., vowel stimuli recorded by Hillendbrand et al. (1994), consonant stimuli recorded by Shannon et al. (1999), etc.]. Training was conducted using novel monosyllable and/or nonsense words produced by different talkers, which were not included in the test database (more than 4000 stimuli in total). Similarly, for the MCI training, the melodic contours used for training were of different frequency ranges than those used for testing, and familiar melodies were only used for testing. It may be debatable whether the training stimuli in these studies deviate sufficiently to demonstrate true generalization effects. However, many post-training improvements were observed for stimuli that may target different speech processes (e.g., “bottom-up” segmental cue training generalizing to “top-down” sentence recognition). It is also important to collect baseline measures that may be responsive to changes in CI processing, and to train using stimuli and methods that may elicit improvements in these listening tasks. For example, if a novel processing scheme that may provide additional spectral channels is to be evaluated before and after auditory training, it is important to collect baseline measures that will be sensitive to these additional spectral cues (e.g., speech in noise, melodic contour identification, speaker identification) rather than those that may be less sensitive (sentence recognition in quiet, consonant recognition in quiet). Similarly, the training should target contrasts or listening tasks that might be enhanced by the additional spectral cues. Finally, it is important to minimize the expense of auditory rehabilitation, as well CI users’ time and effort. Our research with NH subjects listening to CI simulations suggests that the frequency of training may be less important than the total amount of training completed (Nogaki et al., 2007). The benefits of auditory training in our CI studies generally involved 2–3 hours per week over a 5-day period, for one month or longer. Computer-based auditory training, such as the CAST software, or

other commercial products (e.g., Cochlear's "Sound and Beyond," Advanced Bionic's "Hearing your Life," etc.) provides CI users with affordable auditory rehabilitation that can be performed at patients' convenience. The time, expense and effort associated with training may have a strong influence on training outcomes.

"Real-world" benefits are, of course, the ultimate goal of auditory training. However, these can be difficult to gauge, as listening conditions constantly change and will vary from patient to patient. Subjective evaluations (e.g., quality of "listening life" questionnaires) may provide insight to real-world benefits, and may be an important complement to more "objective" laboratory results. In our research, anecdotal reports suggest that training benefits may have extended to CI subjects' everyday listening experiences outside the lab (e.g., better music perception when listening to radio in the car). As many CI users do quite well in speech tests administered under optimal listening conditions, it is important that auditory training and testing address the more difficult listening environments typically encountered by CI users.

One limitation for many of these CI training studies is the relatively small number of subjects and/or the lack of a separate control group. Even as preliminary studies, a greater number of subjects are needed to evaluate the benefits of psychophysical training for speech recognition, or the improvements associated with passive and active learning when adapting to novel speech processors. Longitudinal adaptation studies (e.g., George et al., 1995; Gray et al., 1995; Loeb and Kessler, 1995; Spivak and Waltzman, 1990; Waltzman et al., 1986) have generally included a far greater number of CI patients. However, these studies were not sensitive to individual subject differences in terms of learning experience (i.e., there was no standard auditory rehabilitation provided) or the degree of improvement. As CI technology has improved, and as the criteria for implantation has changed, newly implanted CI users may experience a different time course and degree of adaptation. This variability in the CI population Different patient-related factors (e.g., duration of deafness, etiology of deafness, experience with the device, device type, etc.) may contribute to similar levels of performance, making it difficult to establish comparable and meaningful experimental control groups for CI training studies. As discussed earlier, most of the previously described CI training studies used "within-subject" controls, in which baseline measures were extensively collected until achieving asymptotic performance. As most subjects had more than 1 year of experience with their device, asymptotic performance was obtained within 3–4 test sessions. Note that it is difficult to completely rule out that long-term experience with the test stimuli and methods may have contributed to the observed training effects. Given the extensive baseline measures for the within-subject controls, the relative change in performance seems to be a meaningful and relevant measure. In general, better experimental controls would help to clarify learning effects and ultimately give credence to the benefits of auditory rehabilitation.

The improvements in speech performance with auditory training were comparable to (and often surpassed) performance gains reported with recent advances in CI technology and signal processing. Indeed, beyond the initial benefit of restoring hearing sensation via implantation, very few changes in CI signal processing have produced the gains in performance (~15–20 percentage points, on average) observed in our auditory training studies. This is not to suggest diminishing returns for continued development of CI devices and processors. Instead, auditory training may be a necessary complement to new device and processor technologies, allowing patients to access the additional, but subtle spectro-temporal cues. Ultimately, it is important to recognize that cochlear implantation alone does not fully meet the needs of many patients, and that auditory rehabilitation may be a vital component for patients to realize the full benefits their implant device.

ACKNOWLEDGMENT

Research was partially 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* 1991;25(5):291–302. [PubMed: 1742555]
- Busby PA, Clark GM. Gap detection by early-deafened cochlear-implant subjects. *J. Acoust. Soc. Am* 1999;105(3):1841–1852. [PubMed: 10089607]
- Cazals Y, Pelizzone M, Kasper A, Montandon P. Indication of a relation between speech perception and temporal resolution for cochlear implantees. *Ann. Otol. Rhinol. Laryngol* 1991;100(11):893–895. [PubMed: 1746822]
- Cazals Y, Pelizzone M, Saudan O, Boex C. Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants. *J. Acoust. Soc. Am* 1994;96(4):2048–2054. [PubMed: 7963020]
- Cohen NL, Waltzman SB, Shapiro WH. Telephone speech comprehension with use of the nucleus cochlear implant. *Ann. Otol. Rhinol. Laryngol. Suppl* 1989 Aug;142:8–11. [PubMed: 2504098]
- 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* 1997;18(7):488–501. [PubMed: 9416451]
- Donaldson GS, Nelson DA. Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies. *J. Acoust. Soc. Am* 1999;107(8):1645–1658. [PubMed: 10738818]
- Dorman MF, Loizou PC. Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with continuous interleaved sampling (CIS) processors. *Ear. Hear* 1997;18(2):147–155. [PubMed: 9099564]
- Dowell RC, Seligman PM, Blamey PJ, Clark GM. Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise. *Acta Otolaryngol* 1987;104(5–6):439–446. [PubMed: 3434265]
- Eggermont JJ, Ponton CW. Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. *Acta Otolaryngol* 2003;123(2):249–252. [PubMed: 12701751]
- Faulkner A. Adaptation to distorted frequency-to-place maps: implications for cochlear implants and electro-acoustic stimulation. *Audiol. Neurootol. Suppl* 2006;(1):21–26.
- Fitzgerald MB, Wright BA. A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *J. Acoust. Soc. Am* 2005;118(6):3794–3803. [PubMed: 16419824]
- Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am* 2001;110(2):1150–1163. [PubMed: 11519582]
- Fu Q-J, Zeng F-G, Shannon RV, Soli SD. Importance of tonal envelope cues in Chinese speech recognition. *J. Acoust. Soc. Am* 1998a;104(1):505–510. [PubMed: 9670541]
- Fu Q-J, Shannon RV, Wang X. Effects of noise and number of channels on vowel and consonant recognition: Acoustic and electric hearing. *J. Acoust. Soc. Am* 1998b;104(6):3586–3596. [PubMed: 9857517]
- Fu Q-J, Zeng F-G. Effects of envelope cues on Mandarin Chinese tone recognition. *Asia-Pacific J. Speech Lang. and Hear* 2000;5:45–57.
- Fu Q-J. Temporal processing and speech recognition in cochlear implant users. *Neuroreport* 2002;13(13):1635–1640. [PubMed: 12352617]
- Fu QJ, Shannon RV, Galvin JJ III. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J. Acoust. Soc. Am* 2002;112(4):1664–1674. [PubMed: 12398471]

- Fu Q-J, Hsu C-J, Horng M-J. Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant patients. *Ear, Hear* 2004;25(5):501–508. [PubMed: 15599196]
- Fu Q-J, Nogaki G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J. Assoc. Res. in Otolaryngol* 2005;6(2):19–27. [PubMed: 15735937]
- Fu Q-J, Galvin JJ III. Recognition of simulated telephone speech by cochlear implant patients. *Am. J. Audiol* 2006;15(2):127–132. [PubMed: 17182877]
- Fu Q-J, Galvin JJ III. Perceptual learning and auditory training in cochlear implant patients. *Trends Amplif* 2007;11(3):193–205. [PubMed: 17709574]
- Fu Q-JJ, Nogaki G, Galvin JJ III. Auditory training with spectrally shifted speech: an implication for cochlear implant users' auditory rehabilitation. *J. Assoc. Res. Otolaryngol* 2005a;6(2):180–189. [PubMed: 15952053]
- Fu Q-J, Galvin JJ III, Wang X, Nogaki G. Moderate auditory training can improve speech performance of adult cochlear implant users. *J. Acoust. Soc. Am* 2005b;6(ARLO):106–111.
- Galvin JJ III, Fu Q-J, Nogaki G. Melodic Contour Identification in Cochlear Implants. *Ear Hear* 2007;28(3):302–319. [PubMed: 17485980]
- Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS, Dahlgren NL. The DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM. 1993NTIS order number PB91-100354
- Gfeller K, Christ A, Knutson JF, Witt S, Murray KT, Tyler RS. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *J. Am. Acad. Audiol* 2000;11(7):390–406. [PubMed: 10976500]
- Gfeller K, Witt S, Adamek M, Mehr M, Rogers J, Stordahl J, Ringgenberg S. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *J. Am. Acad. Audiol* 2002;13(3):132–145. [PubMed: 11936169]
- George CR, Cafarelli Dees D, Sheridan C, Haacke N. Preliminary findings of the new Spectra 22 speech processor with first-time cochlear implant users. *Ann. Otol. Rhinol. Laryngol Suppl* 1995 Sep; 166:272–275. [PubMed: 7668666]
- Gray RF, Quinn SJ, Court I, Vanat Z, Baguley DM. Patient performance over eighteen months with the Ineraid intracochlear implant. *Ann. Otol. Rhinol. Laryngol. Suppl* 1995 Sep;166:275–277. [PubMed: 7668667]
- Hesse G, Nelting M, Mohrmann B, Laubert A, Ptok M. Intensive inpatient therapy of auditory processing and perceptual disorders in childhood. *HNO* 2001;49(8):636–641. [PubMed: 11544885]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am* 1995;97(5 Pt 1):3099–3111. [PubMed: 7759650]
- Hochberg I, Boothroyd A, Weiss M, Hellman S. Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear Hear* 1992;13(4):263–271. [PubMed: 1397769]
- Horng M-J, Chen H-C, Hsu C-J, Fu Q-J. Telephone Speech Perception by Mandarin-Speaking Cochlear Implantees. *Ear Hear* 2007;28(2 Suppl):66S–69S. [PubMed: 17496651]
- IEEE Subcommittee. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics* 1969;Vol. AU-17(3):225–246.
- Ito J, Nakatake M, Fujita S. Hearing ability by telephone of patients with cochlear implants. *Otolaryngol. Head Neck Surg* 1999;121(6):802–804. [PubMed: 10580241]
- Jesteadt W. An adaptive procedure for subjective judgments. *Percept. Psychophys* 1980;28(1):85–88. [PubMed: 7413416]
- Kelly AS, Purdy SC, Thorne PR. Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clin. Neurophysiol* 2005;116(6):1235–1246. [PubMed: 15978485]
- Kiefer J, Müller J, Pfennigdorf T, Schön F, Helms J, von Ilberg C, Gstöttner W, Ehrenberger K, Arnold W, Stephan K, Thumfart W, Baur S. Speech understanding in quiet and in noise with the CIS speech coding strategy (Med-El Combi-40) compared to the Multipeak and Spectral peak strategies (Nucleus). *ORL J. Otorhinolaryngol. Relat. Spec* 1996;58(3):127–135. [PubMed: 8797215]
- Kong YY, Cruz R, Jones JA, Zeng F-G. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear* 2004;25(4):173–185. [PubMed: 15064662]

- Loeb GE, Kessler DK. Speech recognition performance over time with the Clarion cochlear prosthesis. *Ann. Otol. Rhinol. Laryngol. Suppl* 1995 Sep;166:290–292. [PubMed: 7668675]
- Li TH, Fu Q-J. Perceptual adaptation to spectrally shifted vowels: the effects of lexical and non-lexical labeling. *J. Assoc. Res. Otolaryngol* 2007;8(1):32–41. [PubMed: 17131213]
- Lin M-C. The acoustic characteristics and perceptual cues of tones in standard Chinese. *Chinese Yuwen* 1988;204:182–193.
- McDermott HJ, McKay CM. Musical pitch perception with electrical stimulation of the cochlea. *J. Acoust. Soc. Am* 1997;101(3):1622–1631. [PubMed: 9069629]
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 1996;271(5245):77–81. [PubMed: 8539603]
- Milchard AJ, Cullington HE. An investigation into the effect of limiting the frequency bandwidth of speech on speech recognition in adult cochlear implant users. *Int. J. Audiol* 2004;43(6):356–362. [PubMed: 15457818]
- Muchnik C, Taitelbaum R, Tene S, Hildesheimer M. Auditory temporal resolution and open speech recognition in cochlear implant recipients. *Scand. Audiol* 1994;23(2):105–109. [PubMed: 8085109]
- Müller-Deiler J, Schmidt BJ, Rudert H. Effects of noise on speech discrimination in cochlear implant patients. *Ann. Otol. Rhinol. Laryngol. Suppl* 1995 Sep;166:303–306. [PubMed: 7668681]
- Musiek FE, Baran JA, Schochat E. Selected management approaches to central auditory processing disorders. *Scand. Audiol Suppl* 1999;51:63–76. [PubMed: 10803915]
- Nelson PB, Jin SH, Carney AE, Nelson DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am* 2003;113(2):961–968. [PubMed: 12597189]
- Nelson PB, Jin SH. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am* 2004;115(5 Pt 1):2286–2294. [PubMed: 15139640]
- 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* 1994;95(2):1085–1099. [PubMed: 8132902]
- Nogaki G, Fu Q-J, Galvin JJ III. The effect of training rate on recognition of spectrally shifted speech. *Ear Hear* 2007;28(2):132–140. [PubMed: 17496666]
- Pelizzone M, Cosendai G, Tinembart J. Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for Ineraid implanted subjects. *Ear Hear* 1999;20(3):228–237. [PubMed: 10386849]
- Qin MK, Oxenham AJ. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J. Acoust. Soc. Am* 2003;114(1):446–454. [PubMed: 12880055]
- Shannon RV, Adams DD, Ferrel RL, Palumbo RL, Grandgenett M. A computer interface for psychophysical and speech research with the Nucleus cochlear implant. *J. Acoust. Soc. Am* 1990;87(2):905–907. [PubMed: 2307784]
- Shannon RV, Jansvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. *J. Acoust. Soc. Am* 1999;106(6):L71–L74. [PubMed: 10615713]
- Shannon RV, Fu Q-J, Galvin JJ III. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol. Suppl* 2004;552:50–54. [PubMed: 15219048]
- Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Menapace CM, Arndt PL, et al. Evaluation of a new spectral peak coding strategy for the Nucleus-22 channel cochlear implant system. *Am. J. Otol* 1994;15:15–27. [PubMed: 8572106]
- Smith M, Faulkner A. Perceptual adaptation by normally hearing listeners to a simulated 'hole' in hearing. *J. Acoust. Soc. Am* 2006;120(6):4019–4030. [PubMed: 17225428]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 2002;416(6876):87–90. [PubMed: 11882898]
- Spivak LG, Waltzman SB. Performance of cochlear implant patients as a function of time. *J. Speech Hear. Res* 1990;33(3):511–519. [PubMed: 2232769]

- Svirsky MA, Silveira A, Suarez H, Neuburger H, Lai TT, Simmons PM. Auditory learning and adaptation after cochlear implantation: a preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Otolaryngol* 2001;121(2):262–265. [PubMed: 11349792]
- Sweetow R, Palmer CV. Efficacy of individual auditory training in adults: a systematic review of the evidence. *J. Am. Acad. Audiol* 2005;16(7):494–504. [PubMed: 16295236]
- Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* 1996;271(5245):81–84. [PubMed: 8539604]
- Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am* 1997;102(2):3762–3773. [PubMed: 9407668]
- Tremblay K, Kraus N, McGee T. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport* 1998;9(16):3557–3560. [PubMed: 9858359]
- Tyler RS, Gantz BJ, Woodworth GG, Fryauf-Bertschy H, Kelsay DM. Performance of 2- and 3-year-old children and prediction of 4-year from 1-year performance. *Am. J. Otol. Suppl* 1997;18(6):157–159.
- Waltzman SB, Cohen NL, Shapiro WH. Long-term effects of multichannel cochlear implant usage. *Laryngoscope* 1986;96(10):1083–1087. [PubMed: 3762284]
- Wang, R-H. Chap. 3, Chinese phonetics. In: Chen, Y-B.; Wang, R-H., editors. *Speech Signal Processing*. University of Science and Technology of China Press; 1989. p. 37-64.
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. *Nature* 1991;352(6332):236–238. [PubMed: 1857418]
- Wright BA, Buonomano DV, Mahncke HW, Merzenich MM. Learning and generalization of auditory temporal-interval discrimination in humans. *J. Neurosci* 1997;17(10):3956–3963. [PubMed: 9133413]
- Wright BA. Why and how we study human learning on basic auditory tasks. *Audiol. NeuroOtol* 2001;6(4):207–210. [PubMed: 11694729]
- Wright, BA.; Fitzgerald, MB. Learning and generalization of five auditory discrimination tasks as assessed by threshold changes. In: Pressnitzer, D.; de Cheveigne, A.; McAdams, S.; Collet, L., editors. *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*. New York: Springer; 2005.
- Wright BA, Sabin AT. Perceptual learning: how much daily training is enough? *Exp. Brain Res* 2007;180(4):727–736. [PubMed: 17333009]
- Wu J-L, Yang H-M, Lin Y-H, Fu Q-J. Effects of computer-assisted speech training on Mandarin-speaking hearing impaired children. *Audiol. Neurotol* 2007;12(5):31–36.

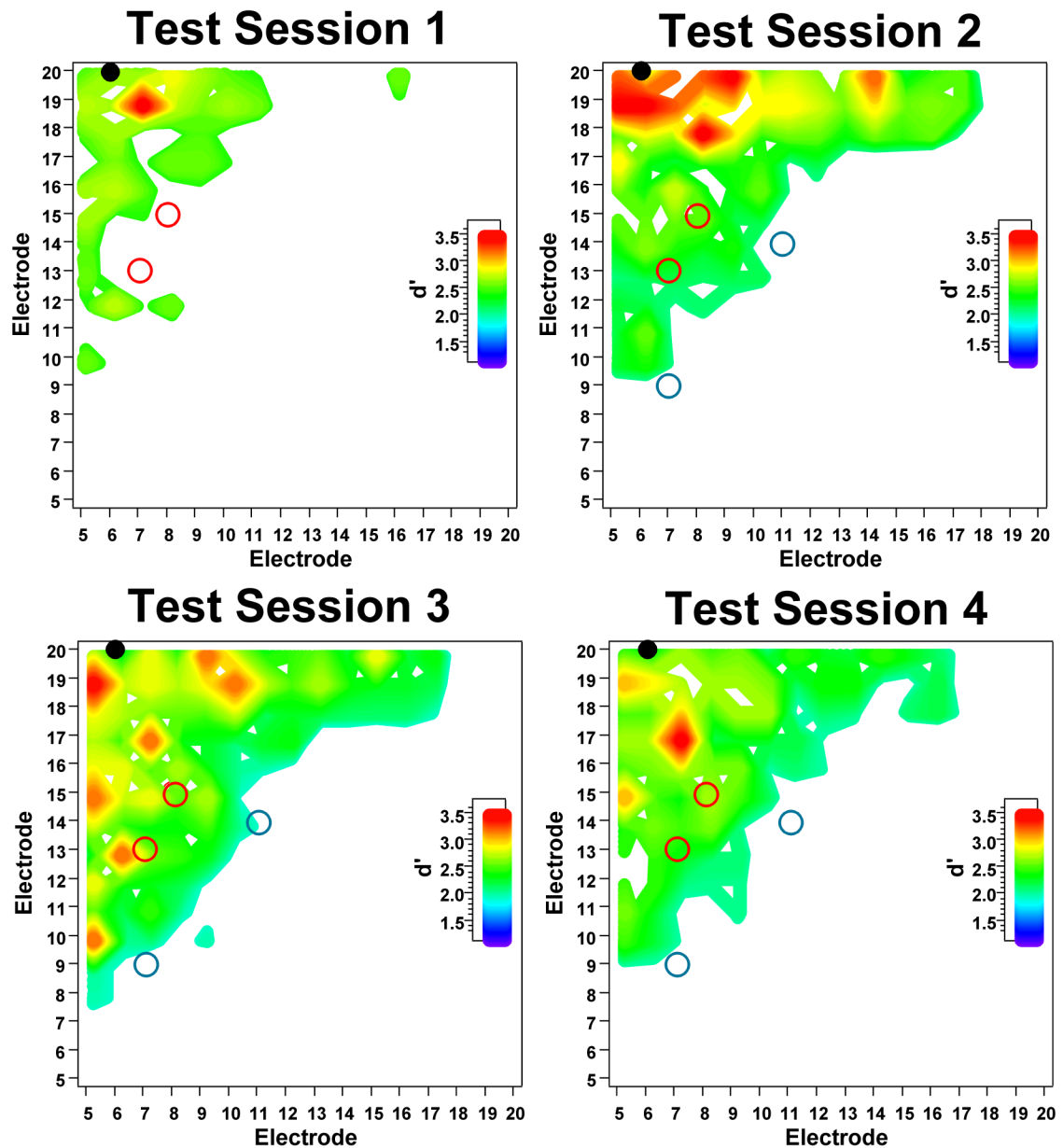


Figure 1.

Electrode discrimination data from an adult cochlear implant user. The contour plots show discrimination performance in terms of d' value. Electrode discrimination test data collected before (Test Session 1), during (Test Sessions 2 & 3), and after (Test Session 4) electrode discrimination training with the strong (8, 15r and 7, 13r – open red circles) and weak electrode contrast separations (11, 14r and 7, 9r – open blue circles).

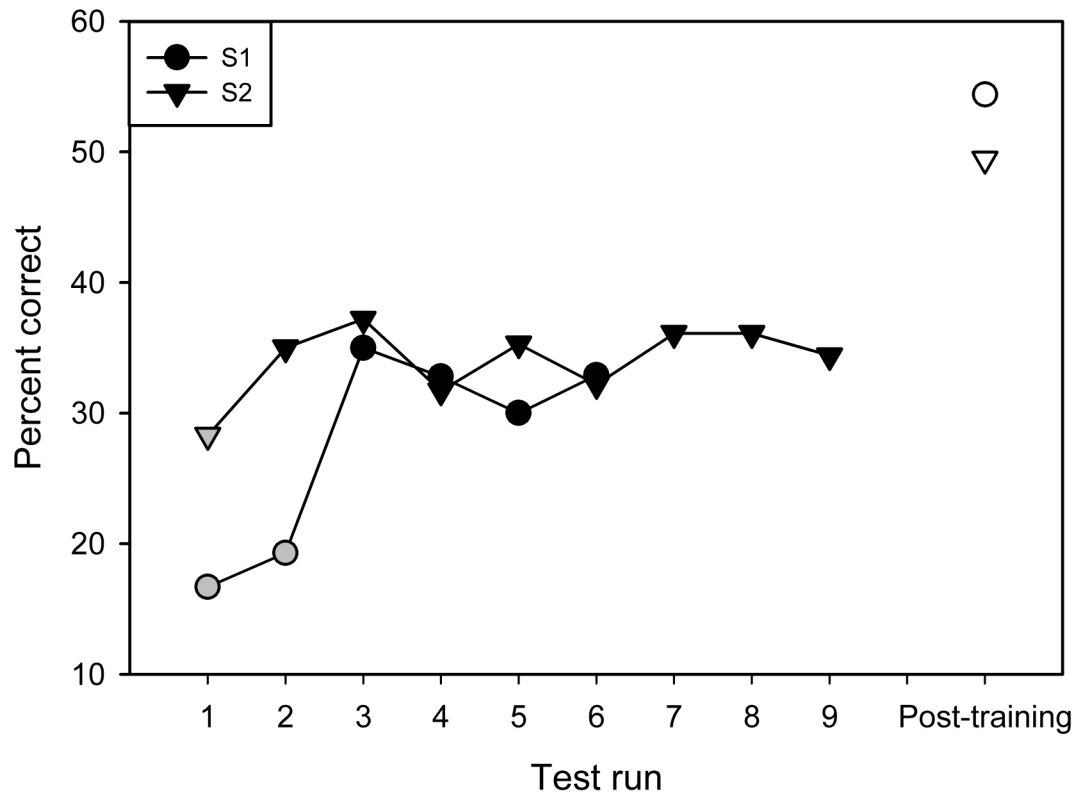


Figure 2.

Baseline vowel recognition performance for two subjects from Fu et al. (2005b). The figure shows an example of the extended baseline measures used for the within-subject experimental controls. The black filled symbols show asymptotic performance; mean asymptotic performance was used as baseline performance for each subject. The gray symbols show data excluded from baseline performance, as procedural learning effects were evident. The white symbols show mean post-training performance.

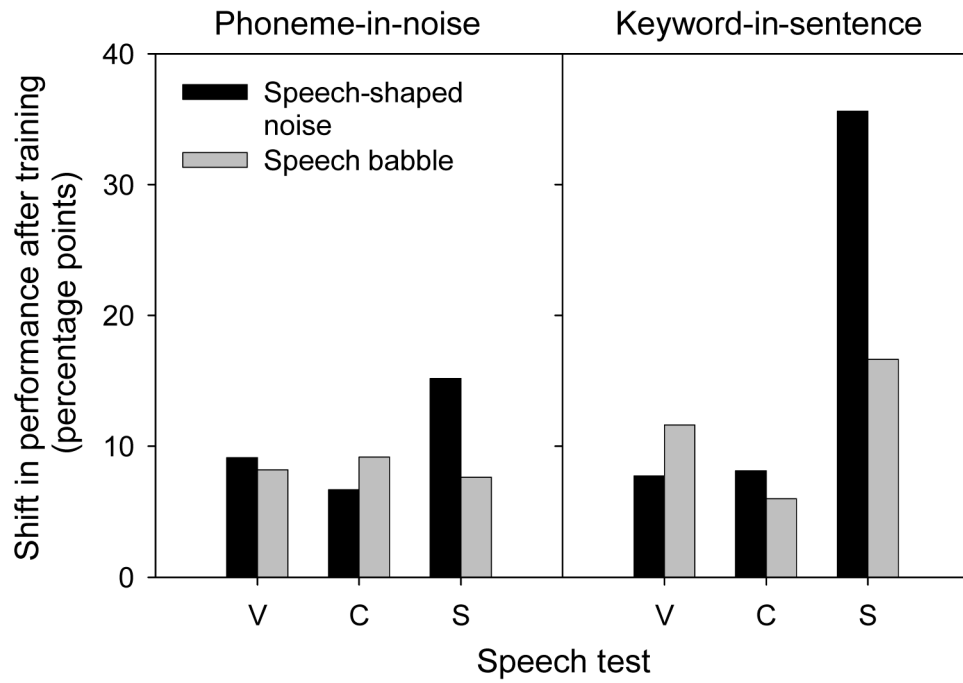


Figure 3. Change in performance after training with the phoneme-in-noise and keyword-in-sentence protocols, for vowels (V), consonants (C), and sentences (S).

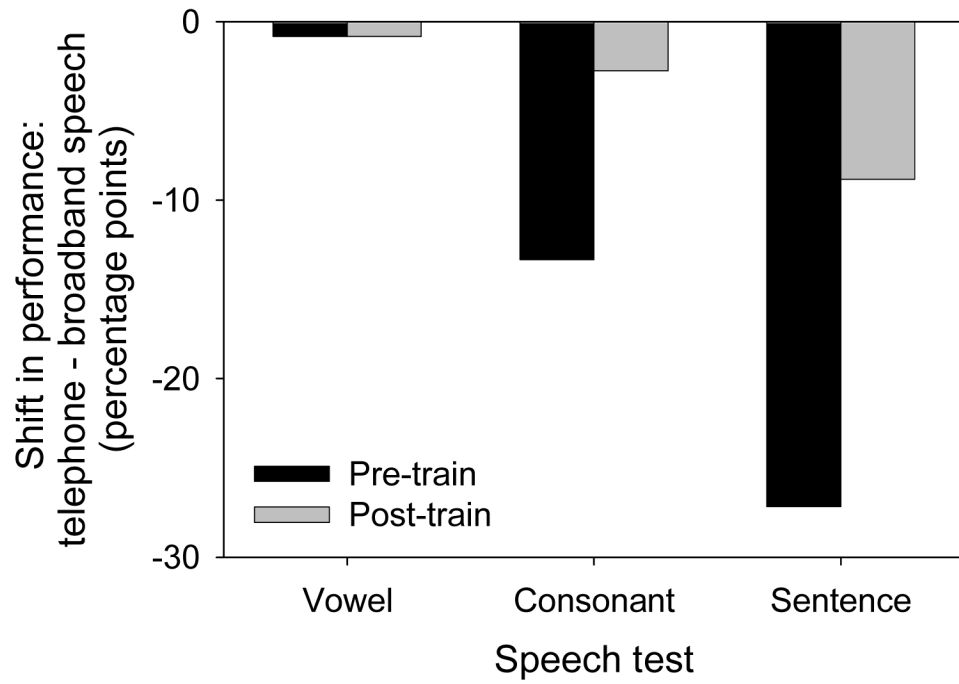


Figure 4. Performance deficit (relative to broadband speech) with telephone speech, before and after training, for one CI subject.