Perceptual Learning and Auditory Training in Cochlear Implant Recipients

Qian-Jie Fu, PhD, and John J. Galvin III

Learning electrically stimulated speech patterns can be a new and difficult experience for cochlear implant (CI) recipients. Recent studies have shown that most implant recipients at least partially adapt to these new patterns via passive, daily-listening experiences. Gradually introducing a speech processor parameter (e.g., the degree of spectral mismatch) may provide for more complete and less stressful adaptation. Although the implant device restores hearing sensation and the continued use of the implant provides some degree of adaptation, active auditory rehabilitation may be necessary to maximize the benefit of implantation for CI recipients. Currently, there are scant resources for auditory rehabilitation for adult, postlingually deafened CI recipients. We recently developed a computer-assisted speech-training program to provide the means to conduct auditory rehabilitation at home. The training software targets important acoustic contrasts among speech stimuli, provides auditory and visual feedback, and incorporates progressive training techniques, thereby maintaining recipients’ interest during the auditory training exercises. Our recent studies demonstrate the effectiveness of targeted auditory training in improving CI recipients’ speech and music perception. Provided with an inexpensive and effective auditory training program, CI recipients may find the motivation and momentum to get the most from the implant device.

Keywords: perceptual learning; auditory rehabilitation; computer-assisted speech training; cochlear implants

The cochlear implant (CI) is an electronic device that provides hearing sensation to listeners with profound hearing loss. As the science and technology of the CI has developed over the past 50 years, the overall speech recognition of CI recipients has steadily improved. With the most advanced implant technology and speech-processing strategies, many recipients receive great benefit and are capable of conversing with friends and family in face-to-face situations and even over the telephone; however, considerable variability remains in individual patient outcomes. Some implant recipients receive little benefit from the latest technology, even after many years of daily use of the device. Much research has been devoted to exploring the sources of variability in CI patient outcomes. Some studies have shown that patient-related factors, such as duration of deafness, are correlated with speech perception performance.1–3 Furthermore, auditory evoked potentials (AEPs) have been correlated with experienced CI recipients’ speech perception performance.2 Several psychophysical measures, including electrode discrimination,4 temporal modulation detection,5,6 and gap detection,7–9 have also been correlated with speech perception performance. The correlation between duration of deafness and speech perception performance suggests that it is important to implant deaf patients as early as possible.10,11 The correlation between AEPs and speech perception performance provides objective evidence of central auditory processing differences across experienced CI users. Correlations between psychophysical measures and speech perception performance suggest that CI recipients’ speech perception is likely limited by their psychophysical capabilities. Despite these correlations, there is as yet no metric with which to predict CI patient outcomes accurately.

Much research has been directed at optimizing speech processor parameters to provide each patient with the best mapping possible, within the limits of each implant recipient’s psychophysical capabilities.
Previous studies have shown dramatic improvements in CI recipients’ speech recognition performance with advances in speech processing strategies. For example, CI performance significantly improved when the “compressed analog” strategy was replaced by the “continuous interleaved sampling” (CIS) strategy.\textsuperscript{12,13} Although speech processor optimization may provide immediate benefits to individual CI recipients, some parameter changes (even those that might transmit additional speech cues) may result in no change or even a temporary deficit in performance. The benefit of speech processor optimization may also depend on CI recipients’ ability to adapt to changes in speech processing via daily “passive” listening.

**Passive Learning and Perceptual Adaptation in CI Recipients**

In CI speech processing, patient performance is most strongly influenced by parameters that affect spectral cues. All CI recipients experience some degree of spectral mismatch between the acoustic input and electrode location in the cochlea, as the acoustic frequency range must be compressed onto the limited extent of the electrode array. The acoustic-to-electric frequency allocation and frequency compression generally result in a spectral mismatch between the peripheral neural patterns and central speech pattern templates acquired during normal hearing (NH). Acutely measured speech perception performance typically worsens with changes in frequency-to-electrode mapping, relative to the frequency allocation with which participants have prior experience.\textsuperscript{14,15} Some studies, however, have shown acute benefits to changes in frequency allocation.\textsuperscript{16} In general, if the spectral mismatch is not too severe, recipients largely adapt over the first 3 to 6 months of implant use; however, if the mismatch is severe, the time course of adaptation may be much longer, and recipients may never fully adapt.

**Passive Adaptation to an Abruptly Introduced Spectral Shift**

Although longitudinal studies are an important measure of adaptation and passive learning, it is difficult to gauge CI users’ auditory plasticity because of patient- and processor-related factors. Differences in CI recipients’ initial listening experiences and motivation may also influence adaptation. In some ways, it may be experimentally preferable to measure CI recipients’ adaptation to changes in speech processing. This way, at least the processor-related factors may be controlled. We recently studied CI users’ adaptation to a severe spectral mismatch over an extended learning period.\textsuperscript{17} Three Nucleus-22 users, each with several years of experience with their clinically assigned speech processors, continuously wore experimental speech processors over a 3-month period. In the experimental processors, the frequency allocation was shifted from the clinical assignment (either Table 7 or Table 9, depending on the research participant) to Table 1. The input frequency ranges for Table 7 and Table 9 were 120-8658 Hz and 150-10,823 Hz, respectively; the input frequency range for Table 1 was 75-5411 Hz. By shifting the frequency allocation to Table 1, the input signal was upshifted by approximately one octave. Although the spectral resolution below 1500 Hz was increased with Table 1, there was severe spectral shifting. After measuring baseline performance with research participants’ clinically assigned processors, Table 1 was abruptly introduced.

Figure 1 shows the shift in mean vowel and sentence recognition relative to baseline performance with participants’ clinically assigned allocations, as a function of experience with Table 1. Acutely measured mean vowel and sentence recognition dropped by 31% and 41%, respectively. Performance was regularly remeasured throughout the 3-month study period. As participants became more familiar with Table 1, speech recognition gradually improved. At the end of the study, vowel and sentence recognition scores with Table 1, although significantly better than acutely measured performance, remained significantly lower than with participants’ clinically assigned allocations. After reintroducing participants’ clinically assigned frequency allocations, acutely measured performance was no different from baseline performance. These results suggest that CI recipients may only partially adapt to a large spectral shift with passive, incidental learning, even after several months of exposure.

**Passive Adaptation to a Gradually Introduced Spectral Shift**

Although CI recipients may only partially adapt to a large spectral shift when it is abruptly introduced, previous studies have shown that gradual exposure to a parameter change may allow for greater adaptation and/or a less stressful adaptation period.\textsuperscript{18,19} To evaluate whether CI recipients could better adapt...
with gradual exposure to a large spectral shift, we systematically shifted the frequency allocation from Table 7 to Table 1 over an 18-month period. One of the participants who participated in the previous 3-month study described previously here17 participated in this study. Beginning with the participant’s clinically assigned frequency allocation (Table 7, 120-8650 Hz), the Frequency Allocation Table was changed every 3 months, until reaching Table 1 (75-5411 Hz), thereby gradually introducing the 0.68-octave frequency shift. Baseline performance was measured for all experimental frequency allocations before beginning the 18-month study; baseline performance with all allocations was also remeasured at the end of each 3-month adaptation period, allowing gradual changes to be observed across all allocations. During each 3-month adaptation period, the participant continuously wore the assigned experimental allocation and returned to the laboratory every 2 weeks for retesting.

Figure 2 shows speech recognition performance as a function of Frequency Allocation Table before, during, and after the adaptation period; results are shown for the following: (Figure 2A) HINT20 sentence recognition thresholds in steady, speech-shaped noise; (Figure 2B) IEEE21 sentence recognition in quiet; (Figure 2C) multitalker vowel recognition in quiet; and (Figure 2D) multitalker consonant recognition in quiet. The dashed lines show baseline performance acutely measured at the beginning of the experiment. The different symbols show the mean performance for each experimental allocation during the 3-month study period in which it was continuously worn. The solid line shows mean performance for all experimental allocations at the end of the 18-month study. Results showed that the Frequency Allocation Table significantly affected all speech measures. Best performance was generally observed with Table 7 (the clinical assignment), and acutely measured baseline performance (dashed lines) generally worsened as the frequency allocation was shifted from Table 7 to Table 1. After 3 months of experience with each experimental allocation (symbols), performance was generally improved (relative to the initial, acute baseline measures), especially for the most spectrally shifted allocations (Tables 3-1). At the end of the 18-month study (solid lines), vowel and sentence recognition were much improved for all experimental allocations. In particular, sentence recognition with Table
1 was comparable to that with the clinical frequency assignment allocation (Table 7); however, vowel recognition with Table 1, although improved relative to baseline measures, remained significantly poorer than with Table 7. These results suggest that gradual exposure to changes in speech processing may allow CI recipients to adapt better to the parameter changes. The results also suggest that the degree of adaptation may greatly depend on the speech tests used to measure performance. Nearly complete adaptation was observed for sentences in quiet or noise, whereas only partial adaptation was observed for vowels and consonants. It is possible that passive learning may depend more on top-down processes (ie, the contextual, linguistic cues provided in sentences), whereas recognition of spectrally shifted phonemes may depend more on bottom-up processes (ie, phonetic contrast discrimination) and may require more active auditory training.

Overall, the results of these 2 experiments suggest that passive adaptation may depend on the degree of spectral shift, how a shift is introduced, and which test materials are used to evaluate performance. Although gradual exposure may provide a more complete (and less stressful) adaptation, the relatively long adaptation period is a concern. It is unclear whether 18 months of continuous exposure to Table 1 would provide a similar outcome or
whether outcomes would differ strongly between CI recipients. Unfortunately, this sort of study is difficult to conduct, given subjects’ disinclination to wear a speech processor that initially sounds so much poorer than their clinically assigned processor. These results also suggest that passive learning via daily exposure and conversation (ie, at the linguistic and context-heavy sentence level) may be insufficient to improve recognition of spectrally shifted phonemes. Nevertheless, these studies demonstrate considerable auditory plasticity in CI recipients, even after years of experience with their device.

Active Learning and Perceptual Adaptation in CI Recipients

For postlingually deafened implant recipients, the ideal CI device would require only minimal, passive adaptation, as it would faithfully reproduce the speech patterns previously experienced during acoustic hearing. As yet, CI devices only provide a coarse representation of the input acoustic patterns, thus necessitating some period of adaptation. Even with years of experience with the device, performance remains below that of NH listeners, especially for difficult listening conditions (eg, speech in noise) and listening tasks that require perception of spectrotemporal fine structure cues (eg, music appreciation, and vocal emotion perception). Passive adaptation with current CI technology is clearly not enough. It is unclear whether “active” auditory training may improve the degree of adaptation, or at least accelerate adaptation. Unfortunately, relatively few studies have explored the effects of auditory training on CI users’ speech perception. In general, there are fewer auditory rehabilitation resources for adult, postlingually deafened CI recipients than for children, whether due to a lack of convincing evidence for the benefits of training, limited financial resources to support training, or a failure to prioritize training as a necessary part of the CI experience.

Auditory training has been recently shown to improve hearing-impaired (HI) listeners’ speech comprehension and communication. Some early studies also assessed the effects of auditory training on speech recognition by poor-performing CI recipients. Busby et al examined the effects of 10 1-hour speech perception training sessions (1 to 2 sessions per week); 3 prelingually deafened CI users (2 adolescents and 1 adult) participated in the experiment. After the training period, there were only minimal changes in speech perception performance; the participant with the greatest improvement was implanted at an earlier age and therefore had a shorter period of deafness. Dawson and Clark reported more encouraging results for vowel recognition training in 5 CI users. Each participant had been deaf for at least 4 years before implantation, and none had achieved open-set speech recognition. Training consisted of one 50-minute training session per week for 10 weeks. After training, 4 of the 5 participants showed improvement in some measures; this improvement was retained on subsequent testing 3 weeks after training was completed. These early results did not suggest much promise for auditory training for CI users, and few studies have emerged since. We have recently studied the effects of auditory training on English speech recognition, Chinese speech recognition, and music perception. Our preliminary results suggest that auditory training may benefit CI users greatly, providing as much (and often greater) improvements in performance than observed with recent advances in CI technology.

Auditory Training and Speech Perception in CI Recipients

We recently studied the effect of targeted phonemic contrast on English speech recognition by experienced CI users; 10 adult CI recipients with limited speech recognition capabilities participated in the study. Baseline (pretraining) multitalker phoneme recognition performance was measured for at least 2 weeks or until performance asymptoted. After baseline measures were complete, participants trained at home using the Computer-Assisted Speech Training (CAST) program developed at House Ear Institute (loaded onto their personal computers or loaner laptops). Participants were instructed to train at home 1 hour per day, 5 days per week, for a period of 1 month or longer. During the training, production-based contrasts were targeted (ie, second formant differences, and duration); auditory and visual feedback were provided, allowing participants to compare their (incorrect) response to the correct response. Participants returned to the laboratory every 2 weeks for retesting (same tests as baseline measures). Results showed that both vowel and consonant recognition significantly improved for all participants after training. The mean vowel recognition significantly improved by 15.8% (paired t test, \( P < .0001 \)). The mean consonant recognition significantly improved by 13.5% (\( P < \))
For a subset of CI participants (only 3 participants were tested), mean sentence recognition significantly improved by 28.8% ($P < .01$). Mean results are shown in the left panel of Figure 3. Only vowel and consonant contrasts were trained, using monosyllable words (ie, “seed” versus “said” versus “sad” versus “sawed”). Thus, the improved vowel and consonant recognition somewhat generalized to improved sentence recognition. Although performance significantly improved for all participants after 4 weeks or more of moderate training, there was significant intersubject variability in terms of the amount and time course of improvement. For some participants, performance significantly improved after only a few hours of training, whereas others required a much longer time course.

We also recently studied the effect of targeted phonemic contrast on Chinese speech recognition (Mandarin) by CI users. In tonal languages such as Mandarin Chinese, the tonality of a syllable is lexically important, and Chinese sentence recognition is highly correlated with tone recognition. Although the fundamental frequency (F0) contour contributes most strongly to tone recognition, other temporal cues that co-vary with tonal patterns (eg, amplitude contour and periodicity fluctuations) also contribute to tone recognition, especially when F0 cues are reduced or unavailable. In the Wu et al study, we investigated whether moderate auditory training could improve recognition of Chinese vowels, consonants, and tones; 10 Mandarin-speaking children (7 CI users and 3 HA users) participated in the study. After completing baseline measures, participants were trained at home using the CAST program. Training stimuli included more than 1300 novel Chinese monosyllable words, spoken by 4 novel talkers (ie, the training stimuli and talkers were different from the test stimuli and talkers). Participants were trained for 1 hour per day, 5 days per week, for a period of 10 weeks. Participants spent an equal amount of time training with vowel, consonant, and tone contrasts. Results showed that mean vowel recognition significantly improved by 21.7% (paired t test, $P = .006$) after the 10-week training period. Similarly, mean consonant recognition significantly improved by 19.5% ($P < .001$), and Chinese tone recognition scores significantly improved by 15.1% ($P = .007$). Results are shown in the middle panel of Figure 3. Vowel, consonant, and tone recognition performance was remeasured after training was completed; follow-up performance was measured at 1, 2, 4, and 8 weeks after training was completed. Follow-up measures remained significantly higher than pretraining baseline measured for vowel ($P = .018$), consonant ($P = .002$), and tone recognition ($P = .009$), suggesting that the improved performance with training was retained well after training had stopped.

**Auditory Training and Music Perception in CI Recipients**

Music perception and appreciation are particularly difficult for CI users, given the limited spectrotemporal resolution provided by the implant device.
Unlike speech perception, music perception relies more strongly on pitch perception, and implant listeners have great difficulty with complex pitch perception. We recently studied CI users’ melodic contour identification (MCI) and familiar melody identification (FMI; as in Kong et al) and explored whether targeted auditory training (similar to that in Fu et al and Wu et al) could improve MCI and FMI performance. During the MCI test, CI participants were asked to identify 1 of 9 simple 5-note melodic sequences, in which the pitch contour was systematically varied (ie, “rising,” “falling,” “flat,” “rising-falling,” “falling-rising”). The pitch range was varied over 3 octaves, and the interval between successive notes was systematically varied between 1 and 5 semitones to test the intonation provided by the CI device. MCI and FMI performance (without training) was evaluated in 11 CI recipients. Results showed large intersubject variability in MCI performance. The best performers correctly identified more than 90% of the melodic contours when there were 2 semitones between notes in a sequence, whereas poor performers correctly identified less than 40% of the sequences with 5 semitones between notes. A subset of 6 CI participants was trained using targeted melodic contour contrasts. Participants 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. Results are shown in the right panel of Figure 3. Mean MCI performance significantly improved by 28.3% (P = .004); the amount of improvement ranged from 15.5% to 45.4%. FMI performance (with and without rhythm cues) was measured in 4 of the 6 CI participants before and after training. Mean FMI performance with rhythm cues improved by 11.1%, whereas mean FMI performance without rhythm cues significantly improved by 20.7% (P = .02). FMI was not explicitly trained; participants were trained only in the MCI task. Again, there was some generalization from one trained listening task (MCI) to another untrained task (FMI). Anecdotal reports suggested that CI participants’ music perception and appreciation generally improved after MCI training. For example, some participants reported that they were better able to separate the singer’s voice from the background music while listening to music in the car.

**Passive and Active Learning in Adaptation to Changes in CI Speech Processing**

In Fu et al, we studied CI users’ passive adaptation to an abrupt change in speech processing (ie, a large spectral shift due to a change in frequency allocation); in a follow-up study with one of the participants, we studied passive adaptation to the same spectral shift when it was gradually introduced. Although gradual introduction provided better adaptation for some measures (ie, sentence recognition in quiet and noise), other measures remained poorer than with the clinically assigned speech processor. It is possible that passive learning was not adequate and that active auditory training may provide for more complete adaptation to changes in speech processing.

We studied the effects of both passive and active learning on adaptation to changes in speech processing in one of the participants who participated in the Fu et al study; the participant was a Clarion II implant user with limited speech recognition capabilities. The participant was originally fit with an 8-channel CIS processor (stimulation rate, 813 Hz per channel). Baseline (pretraining) multitalker phoneme recognition performance was measured 6 times over a 3-month period, by which time, vowel and consonant recognition reached asymptotic performance levels. After completing baseline measures, the participant trained at home using the CAST program 2 to 3 times per week for 6 months (as in Fu et al); the participant was trained to identify medial vowel contrasts. After training, mean vowel and consonant recognition with the 8-channel CIS processor improved by 12% and 17%, respectively. Asymptotic performance was achieved 3 months into the 6-month training period. After completing the 6-month training period with the 8-channel CIS processor, the participant was fit with the 16-channel HiRes strategy (stimulation rate, 5616 Hz per channel). Baseline performance with the HiRes processor was measured for the next 3 months. Performance was unchanged over this 3-month period (approximately 45% correct, ie, the same as at the end of training with the 8-channel CIS processor). After completing this second set of baseline measures with the HiRes processor, the participant trained at home 2 to 3 times per week for 3 months. Vowel and consonant recognition improved by an additional 10% and 12%, respectively. Figure 4 shows pretraining and posttraining performance for the
8-channel CIS and 16-channel HiRes processors. These preliminary results show that active, targeted auditory training can significantly improve phoneme recognition in quiet, even after an extensive period of passive adaptation to novel speech processors. The results also suggest that auditory training may aid in the reception of additional spectral and temporal cues provided by advanced speech-processing strategies, such as HiRes.

### Effects of Training Protocols, Training Materials, and Training Frequency

Although active auditory training may provide greater and/or more accelerated adaptation than passive learning, it is as yet unclear which training protocols, training materials, and training rates may be most effective. The results from our auditory training studies with CI users showed some degree of generalization (i.e., performance improved for several measures after training with one type of stimulus). Some studies of NH persons listening to acoustic CI simulations also showed generalized improvements, suggesting that the improved performance may be less due to targeting specific processes (e.g., segmental speech contrasts) than to increasing listeners’ overall attention to speech signals. We have recently conducted some experiments with NH participants listening to CI simulations in which we compared training outcomes with different training protocols, training materials, and training rates.

### Effect of Training Protocols

Although the results from our recent studies with the CAST program demonstrate successful training outcomes with CI recipients, earlier studies have shown somewhat mixed results. The better training outcomes with the CAST program may have been due to differences in training protocols. For example, Dawson and Clark used one-to-one personal training that targeted a variety of phonetic contrasts, whereas the CAST program used in our research studies targets minimal phonetic contrasts.

There are some common elements to many training studies. For example, most training protocols require active, rather than passive, participation from the listener. The level of difficulty is typically adjusted to maintain the listener’s interest and motivation during training. Immediate trial-by-trial auditory and/or visual feedback is often provided; however, there can be fundamental differences in training protocols. In general, auditory training involves 2 approaches: a bottom-up approach or a top-down approach. A bottom-up approach aims to improve the efficiency of peripheral auditory processing. For CI listeners, a “true” bottom-up approach might involve electrode discrimination, modulation detection, or rate discrimination. The phonetic contrast training protocol used in Fu et al and the Fast ForWord training software used in Cohen et al are more bottom up in approach, as listeners are trained to attend to small acoustic differences in speech stimuli (e.g., formant frequencies and voice onset times). A top-down approach aims to improve the efficiency of central processing; participants are trained to develop active listening strategies (e.g., attention to lexical or contextual cues). Top-down training targets higher levels of auditory processing and selectively uses lower-level auditory processing as needed. For example, the connected discourse tracking used in Rosen et al is more top-down in approach.

The effectiveness and efficiency of these approaches may illuminate the relative contribution of top-down and bottom-up auditory processing to speech understanding. For example, Gentner and Margoliash found that both top-down and bottom-up processes modify the tuning properties of neurons during learning and give rise to “plastic” (i.e., malleable) recognition patterns for auditory objects. Rather than invoking either top-down or bottom-up
processes, auditory processing most likely involves both mechanisms to relative degrees that may change according to the listening demands (ie, “interactive processing”). Interactive processing is probably compensatory (ie, one type of processing will become more dominant when there is difficulty with the other type). For example, when the quality of the stimulus is good (eg, speech in quiet), bottom-up processing may dominate; when stimulus quality deteriorates (eg, speech in noise), top-down processing may compensate. The opposite may also be true: attention to segmental speech cues may be more important for speech in noise, whereas speech in quiet requires less attention to spectrotemporal details. In automatic speech recognition, combining bottom-up and top-down constraints has been shown to improve speech recognition in the presence of nonstationary noise.

We recently evaluated the effects of 4 different training protocols on training outcomes in NH participants listening to spectrally shifted vowels. The test-only protocol provided an experimental control for procedural learning. The preview protocol (direct preview of the test stimuli with novel talkers, immediately before testing) and the vowel contrast protocol (extended training with novel monosyllable words, spoken by novel talkers, targeting acoustic differences between medial vowels) were more bottom up in approach, as participants were trained to identify only small acoustic differences between stimuli (without linguistic or contextual cues). The sentence training protocol (modified connected discourse) was more top down in approach, as participants were provided with contextual cues in the sentence stimuli. With the test-only and sentence training protocols, there was no significant change in vowel recognition during the 5-day training period; however, vowel recognition significantly improved after 5 days’ training with either the preview protocol or the vowel contrast training protocol. Although a bottom-up approach (ie, phonetic contrast training) may be more effective in improving CI recipients’ phoneme recognition in quiet, it is unclear how training protocols might affect outcomes in difficult listening conditions (eg, background noise and competing talkers), where top-down processes may be more dominant. For example, there is some evidence that top-down training can improve HA users’ sentence recognition in noise. It is also possible that different training protocols may be required for individual CI recipients, depending on their initial speech recognition ability.

Effect of Training Materials

Besides different training protocols, different training materials (ie, speech stimuli) may also influence training outcomes. A variety of speech materials have been used in previous studies (eg, natural speech, synthetic speech, noisy speech, acoustically modified speech); these materials may include meaningful or nonsense speech. Recipients with different disabilities may require different training materials. For example, training with acoustically modified speech significantly improved the speech perception performance of recipients with specific language impairment and dyslexia. Nagarajan et al modified speech signals to prolong their duration and enhance the envelope modulation; modified speech was used to train language learning–impaired children. During the 4-week training period, recognition of both modified and normal speech gradually increased until reaching near age-appropriate levels. These results suggest that acoustically modified speech can be useful in training and that improved recognition of modified speech may generalize to normal speech.

CI recipients are often trained and tested using “clear speech” (ie, carefully articulated words and sentences produced at slower speaking rates). However, CI recipients typically encounter “conversational speech” (ie, more variable pronunciation at faster speaking rates). Recently, Liu et al measured recognition of both clear and conversational speech in noise. They found that the speech reception threshold was 3 to 4 dB higher for conversational speech for both CI participants and NH participants listening to an acoustic CI simulation. Again, differences in CI recipients’ initial performance may dictate which training materials may be most appropriate. For poor performers (for whom speech recognition in noise is nearly impossible), phonemic contrast training may best improve speech perception performance in quiet. For moderate-to-good performers, speech training in noise (whether with sentences or phonemes) may improve performance in noise and in quiet (depending on initial performance levels). Of course, it is most desirable to identify the training materials that will be most effective and most easily generalize to a variety of listening conditions.

Effect of Training Frequency

The better training outcomes in Fu et al may have also been due to differences in terms of the frequency
of training, that is, how often the training was conducted during the training period. For example, in the Busby et al. study, participants were trained for 10 1-hour sessions, 1 to 2 sessions per week. In the Dawson and Clark study, participants were trained for ten 50-minute sessions once per week. In the Fu et al. study, participants trained 30 to 60 minutes per day, 5 days per week, for a period of 1 month or longer. We recently evaluated the effect of training frequency on NH listeners’ adaptation to 8-channel spectrally shifted speech. Participants were trained over 5 sessions, using a phonetic contrast training protocol. Eighteen participants were divided into 3 groups, in which the 1-hour training sessions were provided either 1, 3, or 5 times per week. The results suggested that although more frequent training seemed to provide better adaptation over the 5 training sessions, the frequency of training did not significantly affect training outcomes (at least for the training frequencies that were studied). Thus, it may be more important to complete an adequate number of training sessions over a reasonable time period. These simulation results suggest that CI recipients may significantly benefit from auditory training, even when infrequently performed. The results further suggest that the difference in CI training outcomes between our recent studies and previous studies may be less due to time-intensive training and more due to differences in training protocols and (perhaps) training stimuli. CAST makes use of a far greater number of multitalker speech stimuli (more than 4000 stimuli) than used in previous CI speech-training studies. The results suggest that it may be more important to develop effective training protocols and stimuli than to simply increase recipients’ time commitment.

Conclusion

Many challenges remain in designing effective aural rehabilitation programs for CI recipients. It is important to develop training protocols and materials that will provide rapid improvement over a relatively short time period and to determine that the training provided will generalize to the many listening environments that CI recipients encounter in the real world. It is also important to ensure that training programs are efficient and effective so as to minimize the time commitment required of the CI recipient while maximizing training outcomes.

The most important issue in auditory training is generalization of improvement to different speech tests and listening environments. Thus, it is necessary that training studies use different stimuli for testing and training. In our research studies, stimuli used for baseline measures are never used for training. For speech testing, baseline measures were collected using standard speech test databases (eg, the vowel stimuli recorded by Hillenbrand et al. and consonant stimuli recorded by Shannon et al.). Training was conducted using novel monosyllable and/or nonsense words produced by different talkers. Similarly, in the music perception training, the melodic contours used for training were of different frequency ranges than those used for testing, and familiar melodies were used only for testing. It may be debatable whether the training stimuli in these studies deviate sufficiently to demonstrate true generalization effects, but generalized improvements were observed for stimuli that may target different speech processes (eg, bottom-up segmental cue training generalizing to top-down sentence recognition). The large multitalker database used in the CAST software (more than 4000 stimuli) may also provide some advantage over smaller, more limited databases.

The generalized improvements in performance observed in our research agree with neurophysiological studies, which show that both behavioral performance and neurophysiological changes after auditory training generalized to stimuli not used in the training, thus demonstrating behavioral “transfer of learning” and plasticity in underlying physiologic processes. Tremblay et al. reported results further indicating that training-associated changes in neural activity may precede behavioral learning. The results in Fu et al. also indicated that the time course of improvement varied significantly across CI participants. Depending on patient-related factors (eg, the number of implanted electrodes, the insertion depth of the electrode array, and duration of deafness), some CI recipients may require much more auditory training to noticeably improve performance. Objective neurophysiological measures may provide useful information about the progression of training, that is, whether a particular training protocol should be continued or whether the training protocol and/or training materials should be adjusted. These objective measures may allow for the development of more efficient training protocols for CI recipients; fortunately, the flexibility in computer-assisted auditory training (as in CAST) allows such changes to be easily implemented.
Real-world benefits are, of course, the ultimate goal of auditory training. These can be difficult to gauge, as listening conditions constantly change. Subjective evaluations may provide insight to real-world benefits and may be used to supplement the more objective laboratory results. In our research, participants’ anecdotal reports suggest that these benefits may also be realized in their everyday listening experiences outside the laboratory (ie, better music perception when listening to radio).

Besides generalization and real-world benefits, a major concern regarding auditory training for CI recipients is the financial cost incurred with auditory training, which can be addressed by cost-effectiveness analyses. As pointed out by Barton et al., cochlear implantation is cost-effective for profoundly deafened people, even with the high cost of the implant device (about $30,000). Although the CI provides the basic (but indispensable) benefit of hearing for deaf people, the ultimate benefit of cochlear implantation may depend on several factors. One factor is speech processor optimization. Previous studies have demonstrated improved performance with advances in speech processing strategies. For example, CI performance significantly improved when the compressed analog strategy was replaced by the CIS strategy. More recently, promising results have been reported for speech processors that employ high-rate conditioner pulses. Another factor may be auditory training. For example, in our studies, some CI recipients were able to improve their baseline speech and music performance by 100% or more with moderate auditory training. These are dramatic gains, considering that the commercially available versions of the CAST software (currently) cost less than 1% of the cost of the CI device itself. Auditory training may also be a requisite complement to speech processor optimization and technological advances. The data mentioned earlier suggest that the full benefit of a high-rate speech processor may be realized only after targeted speech training. The costs of auditory training and/or speech processor optimization are relatively low compared with the cost of the device itself.

Besides the financial cost, another concern is the effort associated with CI recipients’ auditory training. How long must recipients train before realizing tangible returns on their efforts? Our recent research suggests that moderate amounts of training provide significant benefits and that the frequency of training may be less important than the total amount of training completed. It is important to develop efficient and effective training protocols and materials to minimize the time commitment while maximizing training outcomes. It is important to develop training protocols and materials that will provide rapid improvement over a relatively short time period, and it is important that training with one protocol/material will generalize to the many listening environments that CI recipients encounter in the real world.

In conclusion, our recent findings suggest great promise for auditory training as part of the aural rehabilitation of adult, postlingually deafened CI recipients. Indeed, after the initial benefit of restoring hearing by implantation, very few changes in speech processing have produced the gains in performance (approximately 20%, on average) we have observed with targeted auditory training. The training has also resulted in improvements in music perception. Provided with an inexpensive and effective auditory training program, implant recipients may be able to obtain greater benefit and satisfaction from their device.

Acknowledgment

CAST is a research tool developed at the House Ear Institute for the purpose of speech perception testing and training. Three commercially available products have been developed for cochlear implant recipients based on the CAST technology: Sound and Beyond (distributed by Cochlear Americas), Hearing Your Life (distributed by Advanced Bionics Corp, a division of Boston Scientific), and Melody Speech Training Software (distributed by Melody Medical Instruments Corp). Both authors have some financial interests (royalty income) in the products based on the CAST technology.

References

3. UK Cochlear Implant Study Group. Criteria of candidacy for unilateral cochlear implantation in postlin-


