Published in final edited form as:

Ear Hear. 2007 April ; 28(2): 132–140. doi:10.1097/AUD.0b013e3180312669.

The Effect of Training Rate on Recognition of Spectrally Shifted Speech

Geraldine Nogaki, B.S., Qian-Jie Fu, Ph.D., and John J Galvin III, B.A.

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

Abstract

OBJECTIVE—Previous studies have shown that the protocol used for auditory training may significantly affect the outcome of training. However, it is unclear how often training should be performed to maximize its benefit. The present study investigated how the frequency of training contributed to normal-hearing listeners' adaptation to spectrally shifted speech.

METHODS—Eighteen normal-hearing listeners were trained with spectrally shifted and compressed speech via an 8-channel acoustic simulation of cochlear implant speech processing. Five short training sessions (1 hour per session) were completed by each subject; subjects were trained at one of three training rates: 5 sessions per week, 3 sessions per week or 1 session per week. Subjects were trained to identify medial vowels presented in a cVc format; depending on the level of difficulty, the number of response choices was increased and/or the acoustic differences between vowels were reduced. Vowel and consonant recognition was measured before and after training, as well as at regular intervals during the training period. Sentence recognition was measured before and after training only.

RESULTS—Results showed that pre-training vowel recognition scores were poor (14.0 % correct, on average) for all subjects, due to the severe spectral shift. After five sessions of targeted vowel contrast training, there was a significant improvement of shifted vowel recognition for most subjects. The mean improvement was comparable (~ 15 percentage points) across the three training rate conditions, despite significant inter-subject variability in pre- and post-training baseline performance. There was no significant difference in training outcomes among the three training rates. Spectrally shifted consonant and sentence recognition also improved by ~ 20 percentage points after training, even though consonants and sentences were not explicitly trained. Similar to vowel recognition, there was no significant difference in training outcomes among the three training rates for shifted consonant and sentence recognition.

CONCLUSION—The results demonstrated that the training rate had little effect on normal-hearing listeners' adaptation to spectrally shifted speech, at least for the training periods (ranging from one to five weeks) used in the present study. The outcome of auditory training may depend more strongly on the amount of training (i.e., total number of training sessions), rather than the frequency of training (i.e., daily or once a week). While more frequent training may accelerate listeners' adaptation to spectrally shifted speech, there may be significant benefits from training as little as one session per week. The results of the present study suggest that appropriate training schedules can be developed to optimize the effectiveness, efficiency and effort associated with hearing-impaired patients' auditory rehabilitation.

I. INTRODUCTION

With advances in cochlear implant (CI) technology, the overall speech recognition of CI patients has steadily improved. With the most advanced implant device and speech processor, many CI 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 CI technology, even after many years of daily use of the device. Although considerable efforts have been made to develop and optimize speech processing strategies for poorly performing patients, auditory training is also an important approach toward improving CI patients' speech recognition performance (Fu et al., 2005a).

Previous studies have shown mixed results with auditory training for poorly performing CI patients. Busby et al. (1991) measured the effect of auditory training on the speech recognition performance of three prelingually deafened CI users (two adults and one adolescent). Auditory training consisted of ten one-hour sessions (1 – 2 sessions per week). After training, there were only minimal changes in these subjects' speech performance; the subject who improved the most was implanted at an earlier age than the other two subjects and therefore had a shorter period of deafness. Dawson and Clark (1997) later investigated the effects of auditory training on the vowel recognition performance of five congenitally deafened patients (three children, one adolescent, and one young adult). Training was specifically focused on improving vowel perception. Training was provided once per week for 10 weeks, for a total of 10 training sessions; each training session lasted approximately 50 minutes. Results showed that after training, two children showed significant gains on a number of tests; however, there were only minimal improvements for the remaining three CI subjects. Recently, Fu et al. (2005a) investigated the effects of auditory training on the speech recognition performance in seven pre-lingually deafened and three post-lingually deafened adult CI patients who have limited speech recognition abilities. Subjects were trained with their home computers using speech stimuli and custom training software; subjects trained one hour per day, five days per week, for a period of one month or longer. Using monosyllabic words, subjects were trained to identify medial vowels. Auditory and visual feedback was provided, allowing subjects to repeatedly compare their (incorrect) choice to the correct response. Results showed a significant improvement in all subjects' speech perception performance after this moderate but regular training.

The type of training protocol (i.e., phonetic contrast training) used in Fu et al. (2005a) may have contributed to the better training outcomes than those observed with previous studies. Fu et al. (2005b) investigated the effect of different training protocols on 16 normal-hearing (NH) subjects' ability to learn spectrally shifted speech; all training and testing was conducted using spectrally shifted speech. Short daily training sessions were conducted over 5 consecutive days, using 3 different training protocols and one test-only protocol. Subjects in the test-only protocol received no preview, no feedback and no training. Subjects in the "preview" protocol were asked to preview the 12 hVd tokens used in the vowel recognition test before each test. Subjects in the "vowel contrast training" protocol were trained to identify medial vowels using monosyllabic words in a cVc context. Subjects in the "sentence training" protocol were trained using modified connected discourse tracking (DeFilippo and Scott, 1978), similar to methods used in previous training studies (Fu and Galvin, 2003; Rosen et al., 1999). Results showed that recognition of spectrally shifted vowels was significantly improved by training with the preview and vowel protocols; no significant improvement in vowel recognition was observed with the test-only or sentence training protocols. These results suggest that training protocols may significantly contribute to auditory training outcomes.

Another factor that may contribute to differences in auditory training outcomes is the amount of training. For example in the previous Busby (1991) and Dawson and Clark (1997) studies, subjects completed 10 training sessions; in the Busby study, subjects were trained 1 – 2 times per week, while in the Dawson and Clark study, subjects were trained once per week. In Fu et al. (2005a), subjects trained for one hour per day, five days per week for a period of one month or longer, resulting in a minimum of 20 training sessions. Recently, Wright and her colleagues explored patterns of learning and generalization on a variety of basic auditory tasks provided by different amounts of training (Fitzgerald and Wright, 2000, Reference note 1; Ortiz et al., Reference note 2). They found that less than 1 hour of training appeared to yield less learning, but more generalization than did multi-hour training. Rosen et al. (1999) used connected discourse tracking (DeFilippo and Scott, 1978) to train listeners' recognition of 4-channel, spectrally shifted speech. They found that performance improved significantly for vowel, consonant, and sentence recognition after just nine 20-min sessions (~ 3 hours) of connected discourse tracking with the shifted simulation. In addition, differences between previous studies in terms of the total number of training sessions/hours and the frequency of training may have contributed to differences in training outcomes.

As both a practical and theoretical consideration, it is important to understand the effects of the frequency of training on auditory training outcomes to design appropriate training schemes that CI patients may integrate into their daily lives. For some busy patients, this may mean committing to the fewest number of training sessions per week possible while still getting some benefit. The present study explored the effects of training rate on auditory training outcomes. NH subjects were trained and tested with spectrally shifted speech while listening to an acoustic CI simulation similar to that used by Fu et al. (2005b). Subjects were trained using the vowel contrast protocol used in Fu et al. (2005a, b). The total number of training sessions was fixed (five hour-long sessions). Depending on the subject group, the frequency of training was one, three, or five sessions per week - training rates that are typical of those used in previous studies.

II. MATERIALS AND METHODS

A. Subjects

18 NH adults (10 females and 8 males), aged 21–39, participated in the study. All subjects had pure tone thresholds better than 20 dB HL at octave frequencies ranging from 125 to 8000 Hz. All subjects were native speakers of American English. All subjects were paid for their participation.

B. Signal Processing

NH subjects were trained and tested while listening to 8-channel acoustic simulations of CI speech processing implemented with the Continuously Interleaved Sampling (CIS) strategy (Wilson et al., 1991). The sine-wave vocoders used in the CI simulations were implemented as follows. The signal was first processed through a pre-emphasis filter (high-pass with a cutoff frequency of 1200 Hz and a slope of 6 dB/octave). The input frequency range (200 – 7000 Hz) was bandpass-filtered into 8 spectral bands using 4th-order Butterworth filters. The corner frequencies of the bandpass filters were calculated according to Greenwood's (1990) formula; thus, each bandpass filter was comparable in terms of cochlear extent. The corner frequencies (3 dB down) of the analysis filters are listed in Table 1. The temporal envelope was extracted from each frequency band by half-wave rectification and lowpass filtering at 160 Hz. The extracted envelopes were used to modulate sinusoidal carriers. The modulated carriers of each band were summed and the overall level was adjusted to be the same as the original speech. The frequencies of the carriers depended on the experimental condition. For the spectrally unshifted condition, the frequencies of the sinewave carriers

were equal to the center frequencies of the analysis filters. For the spectrally shifted condition, the carrier frequency bands were upwardly shifted to simulate a shallow insertion of a 16-mm-long, 8-electrode array with 2-mm electrode spacing; sinewave carrier frequencies were equal to the center frequencies of the shifted carrier bands. The analysis and carrier filters, sinewave carrier frequencies and cochlear distance from the apex (according to Greenwood, 1990) are shown in Table 1, for both unshifted and shifted speech. Note that for the shifted speech condition, the output signal was both spectrally shifted and compressed to simulate two aspects of spectral distortion typically associated with CI devices and speech processing. Figure 1 illustrates the distribution of analysis and carrier bands for both the unshifted and shifted speech conditions.

C. Test and training materials

Speech recognition was assessed using multi-talker vowel, consonant and sentence recognition. The vowel test stimuli included 12 medial vowel tokens presented in a/h/-vowel-/d/context (“heed”, “hid”, “hayed”, “head”, “had”, “hod”, “hawed”, “hoed”, “hood”, “who’d”, “hud”, “heard”). Vowel tokens were digitized natural productions from 5 male and 5 female talkers, drawn from speech samples collected by Hillenbrand et al. (1995). Consonant stimuli included 20 medial consonant tokens presented in an/a/-consonant-/a/context (“aba”, “ada”, “aga”, “apa”, “ata”, “aka”, “ala”, “ara”, “aya”, “awa”, “ama”, “ana”, “afa”, “asa”, “asha”, “ava”, “aza”, “atha”, “acha”, “aja”). Consonant tokens were digitized natural productions from 5 male and 5 female talkers (recorded by Shannon et al., 1999). Sentences were digitized natural productions from the IEEE database (1969) from 1 male and 1 female talker, recorded at House Ear Institute; 72 lists of 10 sentences each were available for testing.

Training stimuli included more than 1,000 monosyllabic words, and were digitized natural productions from 2 male and 2 female talkers (Fu et al., 2005a). The talkers used for the training stimuli were not the same as those used for the test stimuli, with the exception of one male talker, who was used for both the training stimuli and the IEEE sentence stimuli.

D. Test and training procedures

Test and training materials were presented at 65 dBA in free field in a double-walled, soundproof booth through a Tannoy monitor. Vowel and consonant recognition was measured prior to, during and after training. Sentence recognition was measured prior to and after training. Vowel recognition was measured in a 12-alternative identification paradigm. Consonant recognition was measured in a 20-alternative identification paradigm. Sentence recognition was measured in an open set recognition paradigm. For phoneme testing, vowel stimuli included 120 tokens (12 vowels * 10 talkers); consonant stimuli included 200 tokens (20 consonants * 10 talkers). During each trial of a phoneme recognition test, a stimulus token was chosen randomly, without replacement, and presented to the subject. The subject responded by clicking on one of the response buttons shown onscreen (12 response buttons for the vowel test, 20 response buttons for the consonant test). The response buttons were labeled in a/h/-vowel-/d/context for vowel recognition and a/a/-consonant-/a/context for consonant recognition. No feedback was provided, and subjects were instructed to guess if they were not sure, although they were cautioned not to provide the same response for each guess. For sentence testing, two lists were chosen from among the 72 sentence lists. Sentence recognition was measured for each list, spoken by a different talker, for a total of 20 sentences; thus, sentence List 1 would be tested for male Talker 1 and sentence List 2 would be tested with female Talker 2. During each trial of the sentence recognition test, a sentence would be chosen randomly, without replacement, from among the 10 sentences in the test list. The subject responded by repeating as many words as possible, and the

experimenter scored the number of correctly identified words in each sentence. No feedback was provided.

Baseline phoneme and sentence recognition was measured before training was begun. Phoneme and sentence recognition was first tested using unprocessed speech to familiarize subjects with the test tokens, labels, and formats and to ensure that subjects were capable of near-perfect recognition of the unprocessed speech stimuli. After baseline testing with unprocessed speech, vowel, consonant and sentence recognition was tested for 8-channel, spectrally unshifted sinewave speech. Baseline measures with spectrally unshifted speech were repeated three times for vowel recognition (or until performance asymptoted), and two times for consonant and sentence recognition. After testing with the 8-channel unshifted speech, vowel, consonant and sentence recognition was tested for 8-channel, spectrally shifted sinewave speech. Baseline measures with spectrally shifted speech were repeated three times for vowel recognition (or until performance asymptoted), two times for consonant recognition and three times for sentence recognition. After baseline measures were obtained, subjects completed five training sessions. Immediately before and after each training session, vowel and consonant recognition with shifted speech was retested. Each training session lasted one hour. At the end of the fifth and final training session, vowel, consonant and sentence recognition with shifted speech was retested. As with the baseline measures, post-training measures with shifted speech were repeated three times for vowel recognition, two times for consonant recognition, and three times for sentence recognition. Vowel, consonant and sentence recognition was also retested with unshifted speech to verify that any improvement in performance due to training was due to learning the spectral shift and compression, rather than simply learning 8-channel sinewave processing.

The eighteen subjects were divided into three groups of six subjects each. Each group was trained at one of three rates: one session per week (1x group), three sessions per week (3x group), or five sessions per week (5x group). Note that in the 5x group, data for 4 of the 6 subjects were previously reported in Fu et al. (2005b). There were no sentence recognition data for those 4 subjects. There were 5 total training sessions for each group. No control group was used due to previous results in Fu et al. (2005b), which showed no significant improvement in vowel recognition for subjects who received no training, but repeated vowel tests daily for 5 consecutive days.

Targeted vowel contrast training was conducted using the protocol described in Fu et al. (2005a, b). Subjects were trained using custom software (Computer-Assisted Speech Training, or CAST, developed at House Ear Institute) and monosyllabic cVc words; the training stimuli were produced by a different set of talkers than used for test stimuli. Training stimuli were processed exactly the same as the shifted speech test stimuli. During training, a stimulus was presented to the subject. Depending on the level of difficulty, there were 2, 4 or 6 response choices; only the medial vowel differed between response choices (i.e., “seed,” “said”), allowing subjects to focus on medial vowel differences. Initially, the response choices differed greatly in terms of acoustic speech features (i.e., “said,” “sued”); as subjects’ performance improved, the difference in speech features among the response choices was reduced (i.e., “said,” “sad”). The acoustic speech features used to define these levels of difficulty included tongue height (which is associated with the acoustic frequency of the vowel first formant, or, F1), tongue position (from front to back of the oral cavity, which is associated with the frequency difference between F2 and F1) and vowel duration. As subjects’ performance improved beyond a criterion level (80 % correct), the number of response choices was increased and/or the acoustic feature differences between the response choices was reduced. For example, subjects started with two choices, and vowel feature contrasts were decreased from high to low in three steps; the next three levels had four choices, with vowel feature contrasts decreasing from high to low. Audio and visual

feedback was provided. If the subject responded correctly, visual feedback was provided and a new stimulus was selected. If the subject responded incorrectly, auditory and visual feedback was provided; the correct response and the subjects' (incorrect) response were played in sequence repeatedly, allowing subjects to directly compare the two choices. Each training block contained 50 trials. Subjects completed as many training blocks as they could within each one-hour session.

III. RESULTS

Note that for all figures, and within the text, data are reported as percent correct, or as the shift in performance in percentage points. For statistical analyses, subject scores were transformed to rationalized arcsine units (rau) (Studebaker, 1985) to correct for floor and ceiling effects in subjects' performance. Table 2 summarizes the results for all measures, listed in percent correct.

Near-perfect recognition scores were obtained for all subjects in the preliminary vowel, consonant, and sentence recognition tests with unprocessed speech. When the spectral resolution was reduced (8-channel unshifted speech), mean vowel recognition dropped 11 percentage points, mean consonant recognition dropped 7 percentage points, and mean sentence recognition only dropped about 3 percentage points. When the signal was both spectrally degraded and shifted (8-channel, spectrally-shifted speech), mean vowel recognition dropped almost 80 percentage points from the unprocessed speech scores to about chance level, mean consonant recognition dropped 44 percentage points, and mean sentence recognition dropped 45 percentage points.

Figure 2 shows mean vowel, consonant and sentence recognition scores with 8-channel spectrally shifted speech for the three training groups, before and after training. For all groups, and for all speech measures, mean performance improved with the vowel contrast training. A two-way ANOVA, with training and training rate as factors, showed a significant main effect of training on vowel recognition [$F(1,30)=25.585$, $p<0.001$], consonant recognition [$F(1,30)=30.004$, $p<0.001$], and sentence recognition [$F(1,22)=21.695$, $p<0.001$]. However, there was no significant effect of training rate on either vowel recognition [$F(2,30)=0.039$, $p=0.962$], consonant recognition [$F(2,30)=0.279$, $p=0.759$], or sentence recognition [$F(2,22)=1.194$, $p=0.322$]. While the trend in Figure 2 shows greater training effects for the 3x and 5 x groups than for the 1x group, statistical analysis revealed that for training rate effects, power (with $\alpha=0.05$) was only 0.05 for vowels and consonants, and 0.0746 for sentences. The power analysis further revealed that the required difference between groups to achieve sufficient statistical power ($\text{power}>0.8$) was approximately 19% for vowels, 21% for consonants, and 24% for sentences.

Figure 3 shows mean and individual shift in vowel recognition performance (in percentage points), relative to pre-training performance, as a function of training session. Note that for the 5x group, the data shown for subjects S13–S16 were previously reported in Fu et al. (2005b). Although there seems to be a trend that shows more accelerated improvement with increased training rate, there were no significant differences between the three training rates. A two-way ANOVA (performed on the raw data, including baseline performance), with training rate and training session as factors, showed a significant main effect of training session [$F(5, 90)=5.509$, $p<0.001$], but no significant effect of training rate [$F(2,90)=0.0911$, $p=0.913$]. Post-hoc Tukey pair-wise comparisons showed that there was no significant improvement in performance until the third training session.

Performance with 8-channel, unshifted speech was re-tested after training was completed. Since there was no significant difference across the training groups, all data from the different training groups were grouped together (Table 2). One-way ANOVA tests showed

no significant effect of training on unshifted vowel recognition [$F(1,34)=0.480$, $p=0.493$], consonant recognition [$F(1,34)=0.055$, $p=0.817$], or sentence recognition [$F(1,26)=0.163$, $p=0.690$].

IV. DISCUSSION

The results of the present study demonstrate that moderate amounts of auditory training can significantly improve recognition of spectrally shifted speech, consistent with previous studies (Fu and Galvin, 2003; Fu et al., 2005b; Rosen et al., 1999). The results showed no significant difference in training outcomes when subjects trained 1 – 5 times a week, when a total of five training sessions was completed. It should be noted that the designation of five training sessions was arbitrary. It is not implied that performance will plateau after five training sessions, nor that difference between training rates may become apparent over a longer training period. The present study also offers several interesting findings regarding effects of auditory training on the recognition of spectrally shifted speech.

Although auditory training significantly improved mean vowel recognition scores, there was a large inter-subject variability in training outcomes for all three training groups. Individual subjects improved from a minimum of 3 to a maximum of 34 percentage points after completing five training sessions. There was also large inter-subject variability in terms of the time course of improvement. Some subjects improved incrementally after each training session, while others showed no improvement during the first few training sessions. The large inter-subject variability observed in the present study may be explained in terms of NH subjects' motivation to learn. Because of the extreme spectral shift and compression, most subjects could understand little, if any, of the speech presented prior to training. For most people, this was very discouraging. NH subjects were only exposed to the CI simulations during the training and test sessions, after which they would return to their normally-hearing lives. Whether or not they improved their performance with the shifted speech, they were still paid at the end of the experiment. On the other hand, CI listeners will never return to a normal hearing world. If they can improve their understanding of the spectrally distorted speech, then they will be able to communicate with others and perceive what's going on in their environment. Since communication in the real world is at stake, CI subjects may be more uniformly motivated to learn spectrally shifted speech. It is also possible that the vowel contrast training protocol used in the present study may only be suitable for some subjects. Different training protocols may be required for different subjects to effectively adapt to spectrally shifted speech.

Surprisingly, the frequency of training had no significant effect on training outcomes, at least within the experimental training period. For the 1x group, the improvement in vowel recognition ranged from 7.6 – 24.7 percentage points, with a mean of 13.4 percentage points. For the 3x group, the improvement in vowel recognition ranged from 5.6 to 26.1 percentage points, with a mean of 14.8 percentage points. For the 5x group, the improvement ranged from 2.6 – 33.6 percentage points, with a mean of 18.0 percentage points. Due to the large inter-subject variability, there was no significant difference among the training groups in terms of the time course or amount of improvement. However, careful examination of the data shows some potential trends in terms of training rate. First, the mean improvement for the 5x group was ~ 5 percentage points higher relative to the 1x group, and ~3 pts higher relative to the 3x group. Second, for the 5x group, 3 of the 6 subjects improved more than 20 percentage points after completing five training sessions; only one subject in the 1x group and two subjects in the 3x group improved by 20 percentage points or more. Again, subject motivation may have contributed to these trends in results. If all NH subjects experienced the urgency of learning that is likely experienced by CI patients, it is possible

that the frequency of training may have significantly affected the training outcomes in the present study.

Another interesting finding is that the improved vowel recognition with the vowel contrast training protocol generalized to improved consonant and sentence recognition, for all three training groups. Note that different stimuli and talkers were used for training and testing, and that consonant and sentence recognition was not explicitly trained; training was performed using more than 1000 monosyllabic words in a cVc format. All 20 of the consonants in the consonant recognition tests were present in the cVc training words as initial and final consonants, though only those combinations of consonants and vowels that could create commonly used words were used in the training word database. The improvement in consonant and sentence recognition was comparable among the three training groups. The improved recognition of spectrally shifted consonants suggests that subjects benefited from exposure to the initial and final consonants in the monosyllabic training words, consistent with results from a previous training study (Fu et al., 2005b). In that Fu et al. study, all subjects, who were exposed to the shifted consonants during consonant tests, improved their shifted consonant recognition; subjects who had additional exposure to consonants, through targeted vowel training with the monosyllabic cVc words or through sentence training, improved their shifted consonant recognition even more. Although consonant discrimination or recognition was not directly trained in the targeted vowel training with monosyllabic words, the exposure to the consonant sound in conjunction with the visual word labels helped improve shifted consonant recognition.

After training with spectrally shifted speech was completed, re-testing with 8-channel spectrally unshifted speech showed no significant difference in vowel, consonant, and sentence recognition from baseline measures. Thus, training with spectrally shifted speech did not seem to generalize to improved performance for frequency carrier ranges other than those used for training, consistent with results from previous studies (Fu and Galvin, 2003; Fu et al., 2005b). However, baseline performance with 8-channel unshifted speech was already at a high level, leaving little room for improvement. Given that subjects were trained to listen to speech that was both spectrally reduced and shifted (relative to unprocessed speech), it seems unlikely that training would have improved recognition of spectrally reduced speech, as this parameter had the smallest effect on baseline performance. Rosen et al. (1999) found a slight improvement in unshifted 4-channel speech after training with shifted 4-channel speech, but the change was small compared to the improvement with shifted 4-channel speech. This small increase in performance for unshifted 4-channel speech may have been related to “procedural” learning as opposed to training. To avoid as much of this effect as possible in our study, baseline tests for each condition were repeated until performance asymptoted, as opposed to a fixed number of runs of initial baseline tests (two, in the case of Rosen et al. (1999) for spectrally reduced 4-channel speech). Also, it is unlikely that subjects in the present experiment experienced other “procedural” types of learning. In Fu et al. (2005b), performance from a control “test only” group was compared to that of the “vowel contrast training” group; note that data from the 4 subjects in vowel training group were included in the present study (5x group). There was no significant difference in performance for the test-only group after 5 consecutive days of testing, suggesting that the improved vowel recognition in the vowel training group was due to “perceptual” learning, rather than procedural learning (Wright and Fitzgerald, 2001; Hawkey et al., 2004).

These results, combined with those from previous studies (Rosen et al., 1999; Fu and Galvin, 2003; Fu et al., 2005a,b), suggest that auditory training may be an effective approach toward improving CI patients’ speech recognition. The present study also suggests that the amount of training, rather than the frequency of training, may strongly influence CI

patients' training outcomes. Given that the present study was conducted with NH subjects listening to CI simulations, it should also be noted that, while it is possible to simulate CI speech processing, it may not always be possible to simulate the urgency of the learning process experienced by CI patients. As such, the effect of training rate may be somewhat different with CI patients.

V. SUMMARY AND CONCLUSION

The present study showed that moderate amounts of auditory training significantly improved NH listeners' recognition of spectrally shifted speech. Completing five training sessions at three different training frequencies revealed:

1. Targeted vowel contrast training using monosyllabic words significantly improved recognition of spectrally shifted vowels, even with only one training session per week.
2. There was no significant difference in training outcomes when subjects trained 1, 3 or 5 times a week, when five training sessions in total were completed.
3. For spectrally shifted speech, the improved vowel recognition performance with the vowel contrast training protocol generalized to improved recognition of consonants and sentences.
4. There was large inter-subject variability in terms of the amount and time course of improvement with training, suggesting that individualized training protocols may be appropriate for different subjects.

Acknowledgments

We are grateful to all subjects for their participation in our experiments. Research was supported by NIDCD grant R01-DC004792.

References

- Busby PA, Roberts SA, Tong YC, Clark GM. Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. *Br J Audiol.* 1991; 25:291–302. [PubMed: 1742555]
- Dawson PW, Clark GM. Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear and Hearing.* 1997; 18:488–501. [PubMed: 9416451]
- DeFilippo CL, Scott BL. A method for training and evaluation of the reception of on-going speech. *J Acoust Soc Am.* 1978; 63:1186–1192. [PubMed: 649877]
- Fu QJ, Galvin JJ. The effects of short-term training for spectrally mismatched noise-band speech. *J Acoust Soc Am.* 2003; 113:1065–1072. [PubMed: 12597199]
- Fu QJ, Galvin JJ III, Wang X, Nogaki G. Moderate auditory training can improve speech performance of adult cochlear implant users. *Acoustics Research Letters Online.* 2005a; 6(3):106–111.
- Fu QJ, Nogaki G, Galvin JJ. Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *J Assoc Res Otolaryngology.* 2005b; 6:180–189.
- Greenwood DD. A cochlear frequency-position function for several species – 29 years later. *J Acoust Soc Am.* 1990; 87:2592–2605. [PubMed: 2373794]
- Hawkey DJ, Amitay S, Moore DR. Early and rapid perceptual learning. *Nature Neuroscience.* 2004; 7:1055–1056.
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am.* 1995; 97:3099–3111. [PubMed: 7759650]
- IEEE. IEEE recommended practice for speech quality measurements. Institute of Electrical and Electronic Engineers; New York: 1969.

- Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am*. 1999; 106:3629–3636. [PubMed: 10615701]
- Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. *J Acoust Soc Am*. 1999; 106:L71–L74. [PubMed: 10615713]
- Studebaker GA. A “rationalized” arcsine transform. *J Speech Hear Res*. 1985; 28:455–462. [PubMed: 4046587]
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. New levels of speech recognition with cochlear implants. *Nature*. 1991; 352:236–238. [PubMed: 1857418]
- Wright BA. Why and how we study human learning on basic auditory tasks. *Audiol Neurotol*. 2001; 6:207–210. [PubMed: 11694729]
- Wright BA, Fitzgerald MB. Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proc Natl Acad Sci*. 2001; 98:12307–12312. [PubMed: 11593048]

REFERENCE NOTES

1. Fitzgerald MB, Wright BA. Specificity of learning for the discrimination of sinusoidal-amplitude-modulation rate. *J Acoust Soc Am*. 2000; 107:2916(A).
2. Ortiz JA, Wright BA, Fitzgerald MB, Pillai J. Rapid improvements on interaural-time-difference discrimination: Evidence for three types of learning. *J Acoust Soc Am*. 2001; 109:2289(A).

Figure 1. Frequency allocations of analysis and carrier filter bands for 8-channel acoustic simulations of cochlear implant speech processing.

Figure 2.

Mean pre- and post-training vowel, consonant recognition and sentence recognition scores for 8-channel, spectrally shifted speech. From top to bottom, results are shown for three training rates: one session per week (1x), three sessions per week (3x) and five sessions per week (5x). The asterisks indicate a significant difference between pre- and post-training performance. Error bars indicate ± 1 standard deviation. It should be noted that sentence recognition data from the 5x group were limited to only 2 out of the 6 subjects.

Figure 3.

Shift in vowel recognition performance (in percentage points), relative to pre-training performance, as a function of training session. From top to bottom, results are shown for three training rates: one session per week (1x), three sessions per week (3x) and five sessions per week (5x). Individual data are shown by different symbols; mean data are shown by the solid lines.

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

Table 1

Channel# (apex to base)	Analysis band unshifted carrier band corner frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Center frequency of carrier filter (Hz)	Shifted carrier band corner frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Center frequency of carrier filter (Hz)
1	200–359	5.3–8.1	268	999–1,363	14–16	1,167
2	359–591	8.1–10.8	461	1,363–1,843	16–18	1,585
3	591–930	10.8–13.6	741	1,843–2,476	18–20	2,136
4	930–1,426	13.6–16.3	1,152	2,476–3,310	20–22	2,863
5	1,426–2,149	16.3–19.0	1,751	3,310–4,410	22–24	3,821
6	2,149–3,205	19.0–21.8	2,624	4,410–5,860	24–26	5,084
7	3,205–4,748	21.8–24.5	3,901	5,860–7,771	26–28	6,748
8	4,748–7,000	24.5–27.3	5,765	7,771–10,290	28–30	8,942

Table 2

Mean vowel, consonant and sentence recognition scores before and after training with shifted speech

Vowel (% correct)		Consonant (% correct)		Sentence (% correct)	
pre-train	post-train	pre-train	post-train	pre-train	post-train
<i>Unprocessed speech</i>					
All subjects	93.4	97.7		100.0	
1x/week only	95.2	98.1		100.0	
3x/week only	93.1	97.1		100.0	
5x/week only	91.9	97.9		99.7	
<i>8-channel unshifted speech</i>					
All subjects	82.4	90.9	91.0	96.8	96.4
1x/week only	87.5	90.7	90.5	97.4	96.3
3x/week only	79.7	91.0	92.1	95.5	95.9
5x/week only	81.9	90.9	90.4	98.7	97.9
<i>8-channel shifted speech</i>					
All subjects	14.0	44.4	64.6	30.2	54.7
1x/week only	14.9	46.5	63.1	32.5	55.8
3x/week only	13.7	41.9	63.7	26.3	51.3
5x/week only	13.3	44.9	67.0	35.4	61.5