Speech Processors for Auditory Prostheses

NIH Contract NO1-DC-92100

QPR #11: July-August-September 2001

The Effects of Short-Term Training for Spectrally Mismatched Noise-Band Speech: Implications for Cochlear Implants

Submitted by

Qian-Jie Fu, John J. Galvin III, Robert V. Shannon, Lendra Friesen, John Wygonski, Monita Chatterjee, Mark Robert

Department of Auditory Implants and Perception
House Ear Institute
2100 W. Third St.
Los Angeles, CA 90057

And

Fan-Gang Zeng
Department of Otolaryngology
University of California, Irvine

26 November 2001

Table of Contents: **PAGE**

Experimental Report: **The Effects of Short-Term Training for Spectrally Mismatched Noise-Band Speech: Implications for Cochlear Implants**

Abstract 3

Introduction 4

Methods 7

Results 11

Discussion 15

References 16

Publications and Presentations in this Quarter 18

Plans for the Next Quarter 19

ABSTRACT

The present study examined the effects of short-term learning on listeners' ability to accommodate spectrally distorted speech patterns. Speech signals were band-pass filtered into twenty frequency bands; the cutoff frequencies were matched to those of either frequency Table 9 for Nucleus-22 implant (150-10823 Hz) or Table 6 for Nucleus-24M implant (116-7871 Hz). Noise carrier bands were spectrally limited to simulate an electrode array at various cochlear locations, ranging from shallow to deep insertion depths. Six subjects were tested using novel sentences from the IEEE corpus over a five-day period. Baseline data for all insertion depths were collected on Day 1. Over the next four days, all subjects were trained on the most basally-shifted processor using sentences from the DARPA/TIMIT corpus. On Day 5, after a final training session, subjects were re-tested for all insertion depths. Baseline measures showed that speech recognition was significantly affected by the frequency range of the analysis filters, the spectral resolution of the carrier filters, and the place of stimulation. The NU-24 Table 6 filters produced better performance for deep to middle insertion depths, while the NU-22 Table 9 filters provided better performance in the middle to basal insertions. For both groups, training on the most basally-shifted processor significantly improved recognition of basally-shifted speech while the recognition of apically-shifted speech was nearly unchanged. The three subjects who used the NU-22 Table 9 filters were also trained on the apically-shifted speech at a later date; mean improvement was comparable to that observed with the basally-shifted training. In this re-train and re-test condition, performance was largely unchanged at the most basal location, suggesting that these listeners had retained some of the improved speech perception as a result of the previous training. The differential effects caused by the analysis frequency range and resolution, as well as some inter-subject variability suggest that electrode insertion depth, spectral mismatch and the capacity to learn spectrally-shifted speech all interact and should be well-considered in the clinical fitting of speech processors.

INTRODUCTION

Cochlear implants transform acoustic sounds into electrical signals that directly stimulate remaining auditory nerve fibers, thereby partially restoring hearing sensation to profoundly deaf patients. Multi-channel cochlear implant speech processors divide acoustic signals into several frequency bands; extract the temporal envelope information from each band; convert the acoustic amplitudes into electric currents; and deliver the electric currents to appropriate electrodes situated within the cochlea. To recreate the tonotopic distribution of activity within the normal cochlea, the envelope cues from low frequency bands are delivered to apical electrodes and the envelope cues from high frequency bands are delivered to basal electrodes.

One major factor affecting the utilization of speech spectral cues by cochlear implant (CI) users is the number of electrodes/channels. Many studies have examined the effect of number of electrodes on speech recognition in CI patients (e.g. Fishman et al., 1997; Friesen et al., 2001) and the number of spectral channels in normal-hearing (NH) listeners (Dorman et al., 1997; Friesen et al., 2001; Fu, 1997; Loizou et al., 1999; Shannon et al., 1995). Friesen et al. (2001) measured speech recognition as a function of spectral resolution (number of spectral channels) and speech-to-noise ratio in NH listeners and CI patients. They found that for CI patients, vowel and consonant recognition with the SPEAK speech processor did not improve with more than seven electrodes, while for NH listeners, performance continued to increase up to at least 20 channels. The average implant score on all processing strategies was poorer than scores of NH listeners with similar processing. However, the scores of the best CI patient were similar to the mean scores of NH listeners for the same condition (up to seven channels). CI patients with the highest performance level increased in performance as the number of electrodes increased up to seven, while CI patients with low levels of speech recognition did not increase in performance as the number of electrodes was increased beyond four. These results suggest that most CI subjects are not able to fully utilize the spectral information provided by the number of electrodes used in their implant.

The cochlear insertion depth and limited spatial extent of the electrode array causes a mismatch between the frequency content of speech signals and the place of stimulation along the cochlea. The frequency-to-electrode mapping may be a determining factor in implant users’ ability to utilize spectral cues, and may be responsible for the high variability in speech performance among implant users.

The effect of frequency-to-electrode mapping on implant users’ speech performance has been explored in several "acute" experiments. Fu and Shannon (1999a,b) investigated the potential interaction between frequency allocation and electrode location in Nucleus-22 cochlear implant listeners using experimental 4-channel speech processors with the continuous interleaved sampling (CIS) strategy. In these studies, vowel recognition was measured as a function of ten different frequency allocations and two sets of 4-electrode configurations. Each frequency allocation represented the same cochlear extent but different cochlear locations based on Greenwood’s frequency-to-place formula (Greenwood, 1990). Experimental 4-channel CIS speech processors were implemented through a custom research interface (Shannon et al., 1990) and subjects were given no time to adapt to the new patterns of electrical stimulation. Results showed that for a given electrode configuration, the best vowel score was obtained with only a narrow range of frequency allocations. When electrode locations were shifted by 3 mm, the frequency allocation that produced the best vowel recognition also shifted by 3 mm. These results suggested that speech recognition with cochlear implants is highly sensitive to the mapping between frequency allocation and electrode location. A severe mismatch between frequency allocation and electrode location could result in a dramatic and immediate deterioration in speech performance, at least for "acute" measurements.

Similarly, the effect of spectral mismatch caused by frequency-to-electrode mapping is also explored in NH listeners with noise-band speech. The carrier frequency bands were fixed while the analysis filter bands were systematically shifted. The results showed that the best performance was achieved when the analysis filter bands and the carrier filter bands were same or similar. Also performance remains constant when the analysis filters and carrier filters were 3 mm or smaller in terms of Greenwood's function. However, the performance drops steeply in both direction when the distance between analysis and carrier filters was 3 mm or more in the "acute" study where no training (practice) is provided. In conclusion, some degree of mismatch can be tolerated (~3 mm) even in the "acute" measures.

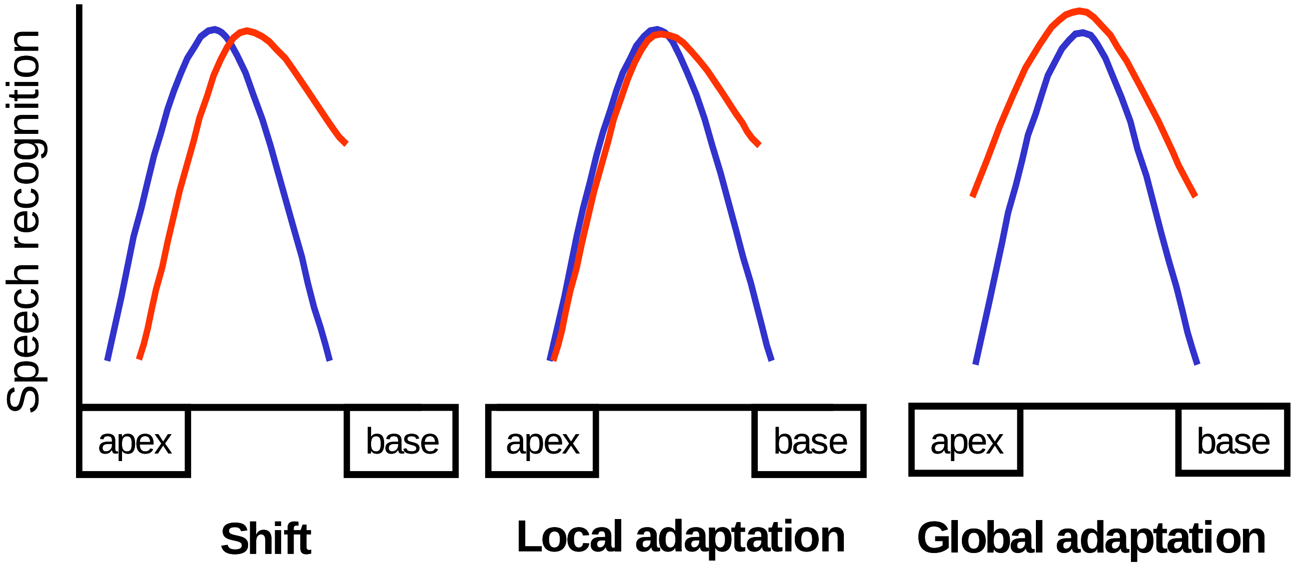
These results indicated that spectral mismatch might cause a significant performance drop at least under "acute" situation. Therefore, it is possible that the performance deficit of CI patients in higher number of electrodes may be, at least partially caused by the spectral mismatch between the frequency information and the location of electrode pairs. However, one major question in these previous studies was the learning issue. In general, it is difficult to gauge the importance of the frequency-to-electrode assignment found in many previous studies because subjects had either no time or limited time to adapt to the new patterns of electrical stimulation produced by changes in those parameters. Several studies have noted that the speech performance of CI users improved significantly after long-term exposure to the new electrical stimulation patterns provided by updated speech processors, speech processing strategies and/or clinical fitting systems (e.g. Wilson et al., 1991; Pelizzone et al., 1999). For example, Pelizzone et al. (1999) reported that, initially, vowel identification scores were unchanged for a group of Ineraid users who switched from the compressed analog (CA) to the CIS speech processing strategy. However, after 6 months of experience with the new CIS strategy, these subjects' vowel identification scores were significantly better than those obtained with the previous CA processor. When the subjects were switched back to their previous CA processor, vowel perception scores returned to the same levels measured before switching to the CIS strategy. This example illustrates that "acute" experiments may provide different results than those of long-term studies. These results suggested that as they become more experienced with the implant device, CI listeners may be able to largely overcome the initial speech perception difficulties associated with such spectral distortions.

To further investigate the importance of frequency-to-electrode assignment on speech performance by CI users, the long-term effects of frequency-to-electrode assignment have also been explored in CI listeners (Fu et al., 2001; Skinner et al., 1995) as well as in NH subjects listening to acoustic simulation of a cochlear implant (Rosen et al., 1999). Fu et al. (2001) measured speech performance over time in three Nucleus-22 CI subjects who, for a three-month period, continuously wore experimental speech processors that were purposely mismatched in terms of the frequency-to-electrode assignment. The frequency boundary assigned to electrodes was lowered by 1 octave in two subjects and 0.68 octaves in one subject. Baseline speech performance using each subject's clinically assigned speech processor was measured just prior to implementation of the experimental processor. Results showed that the experimental processor produced significantly lower performance on all measures of speech recognition immediately following implementation, consistent with the results from the previous "acute" experiments (Fu and Shannon, 1999ab). Over the three-month test period, all measures were significantly higher than those measured immediately post-fitting. These results indicated that long-term exposure to the frequency-shifted speech processor significantly reduced the speech performance deficit. However, after three months' experience with the experimental processors, speech recognition for all subjects remained significantly lower than the baseline levels measured with the clinically assigned processors. Previous studies also explored the short-term training in NH listeners with noise-band speech. In the Rosen et al study (1999), normal-hearing listeners needed only 3-4 hours to improve performance with 4-channel noise-band speech processors from chance-level to nearly 30 % correct of words in sentences.

While CI users were able to somewhat adapt to the new spectral patterns over the long term, the adaptation was not complete over the observed period. However, it is unclear whether subjects adapted by reshaping their internal representation of speech, or by training on specific speech tests and stimuli. Also, it is unclear whether the learning of spectrally-shifted speech is affected by place of cochlear stimulation, acoustic input range, or other factors relating to the electrically stimulated auditory system.

The present study investigated the effects of short-term learning on spectrally mismatched speech in six normal-hearing subjects. Sentence recognition scores were measured as a function of different simulated electrode insertion depths, ranging from very deep to very shallow insertions. Subjects were trained using speech processed to simulate typically shallow insertions in CI patients; the spectral information was delivered to the basal cochlear location, resulting in upwardly shifted speech. Several hypotheses concerning training spectrally mismatched speech were explored.

Figure 1: Three potential outcomes of short-term training



The first hypothesis predicts that training with basally-shifted speech may improve the recognition of trained speech, but may also reduce the recognition of apically-shifted speech. The underlying mechanism for this "shift" hypothesis is that the "internal" representation of speech patterns are reshaped to accommodate the basally-shifted patterns during training, and results in a larger spectral mismatch between apically-shifted speech and the reshaped "internal" representation. The second hypothesis predicts that improvement may occur only with the trained (basally-shifted) speech; performance with the untrained (apically-shifted) speech would not be changed. The underlying mechanism for this "local adaptation" hypothesis is that the subjects adapt only to a specific spectral mismatch; subjects may develop alternative speech patterns for basally-shifted speech while retaining the previous "internal" representations. The third hypothesis predicts that recognition of both apically-shifted and basally-shifted speech will be improved because of training with basally-shifted speech. The underlying mechanism of this "global adaptation" hypothesis is that subjects will gradually adapt to or become familiar with spectrally-mismatched speech in and of itself, resulting in increased recognition of all the spectrally-shifted speech. Figure 1 illustrates the potential outcomes of short-term training based on these three hypotheses.

I. METHODS

1. Subjects

Six normal-hearing (NH) listeners aged 25 to 35 participated in the present experiment. All NH subjects had thresholds better than 15 dB HL at audiometric test frequencies from 250 to 8000 Hz and all were native speakers of American English. All subjects were paid for their efforts.

2. Signal Processing

Both 20- and 4-channel noise-band speech processors were used in the present study. The implementation of these processors is shown as follows. For the 20-channel noise band processors, speech signals were band-pass filtered into 20 frequency bands using 8th order Butterworth filters. To evaluate the effect of acoustic input range, two groups of analysis bands were used: frequency allocation Table 9 for Nucleus-22 implant (150-10,831 Hz) and frequency allocation Table 6 for Nuclues-24M implant (116-7,871 Hz). The temporal envelope of each band was extracted by half-wave rectification and low-pass filtering at 160 Hz. The envelope of each band was then used to modulate a wideband noise that was spectrally limited by a bandpass filter (carrier). The corner frequencies and bandwidths of the carrier frequency bands were dependent on the simulated insertion depth. The simulated tonotopic locations of carrier frequency bands were determined by the following equation:

$$P(i) = P_0+0.75*i \qquad i = 0, 1, \ldots, 20 \qquad (1)$$

where P_0 is the most apical location for a given frequency allocation in mm (from the apex). The corner frequencies of bands were determined by the following equation, from Greenwood (1990):

$$F(i) = 165.4 * (10^{P(i)*0.06}-0.88) \qquad (2)$$

Note that Eq. 2 refers to a 35 mm long cochlea and that actual lengths can vary by several mm. Combining Eqs. 1 and 2, the corner frequencies of all carrier frequency bands were determined for a given insertion depth. The attenuation at the cross-over point of adjacent filter bands was –3 dB. P_0 was varied from 7.75 mm to 15.25 mm from the apex (or equivalently, 27.25 mm to 19.75 mm from the base) to simulate a range of shallow to deep electrode insertion depths. [I think you originally had the ‘from the base #s reversed…I changed them] 6 carrier frequency allocations were generated between these endpoints. Schematic diagrams of experimental conditions are shown in Figure 2.

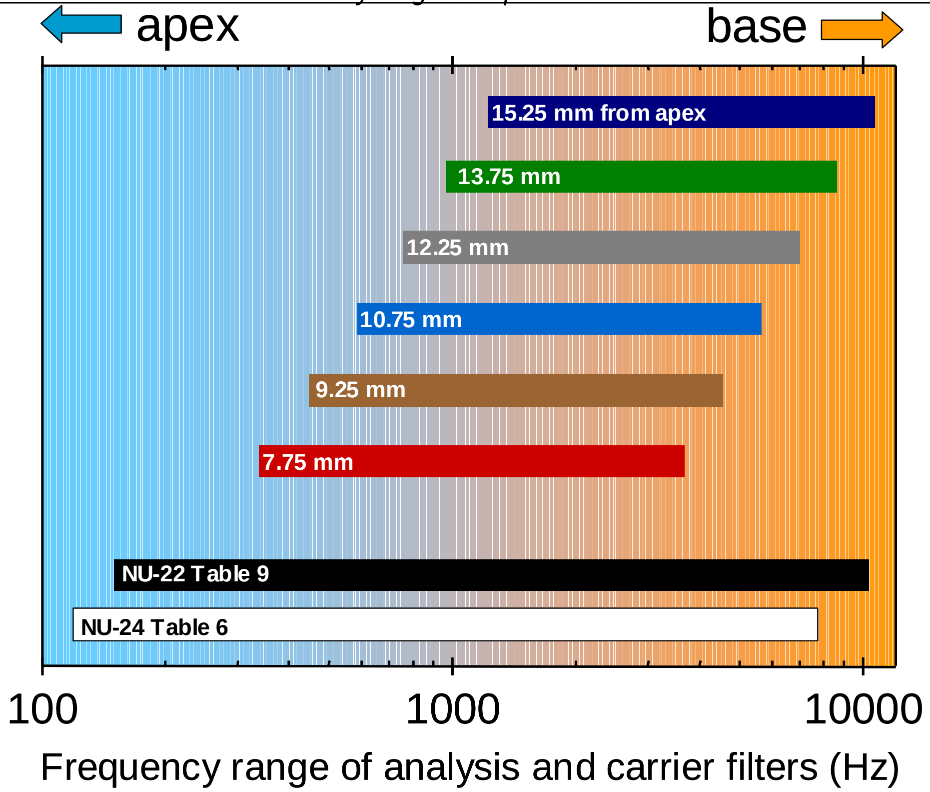


Figure 2:Frequency range of analysis (lower black, white bars) and carrier band (upper bars) experimental noise-band processors

The output from each band was then summed and presented to the listeners seated in a sound-treated booth via one loudspeaker (Tannoy Reveal monitors) at 70 dBA. Figure 3 shows the spectral envelope of the original speech /i/ and 20-channel noise-band speech /i/ at three different cochlear locations (insertion depths). The left panel used NU-22 frequency allocation Table 9 for the analysis filters and right panel used NU-24 Table 6.

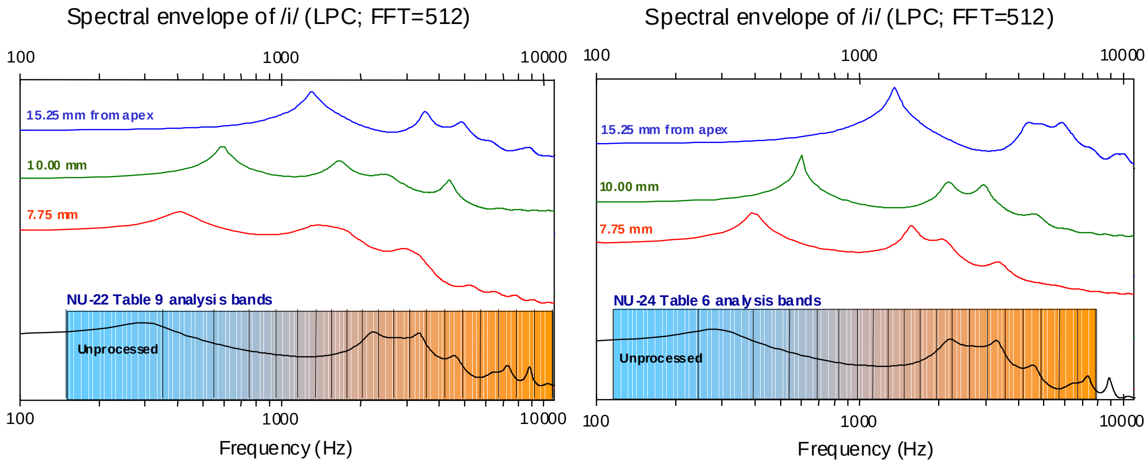


Figure 3: Spectral envelope of the original speech stimulus /i/ and 20-channel noise-band speech /i/ at three different cochlear locations (insertion depths)

For the 4-channel noise band processors, the extraction of the temporal envelope was the same as that used in the 20-channel speech processors except that the output of 20 analysis bands was assigned to 4 carrier bands distributed evenly across bandwidths used for 20-channel carriers. A schematic diagram of the 4-channel speech processor is shown in Figure 4.

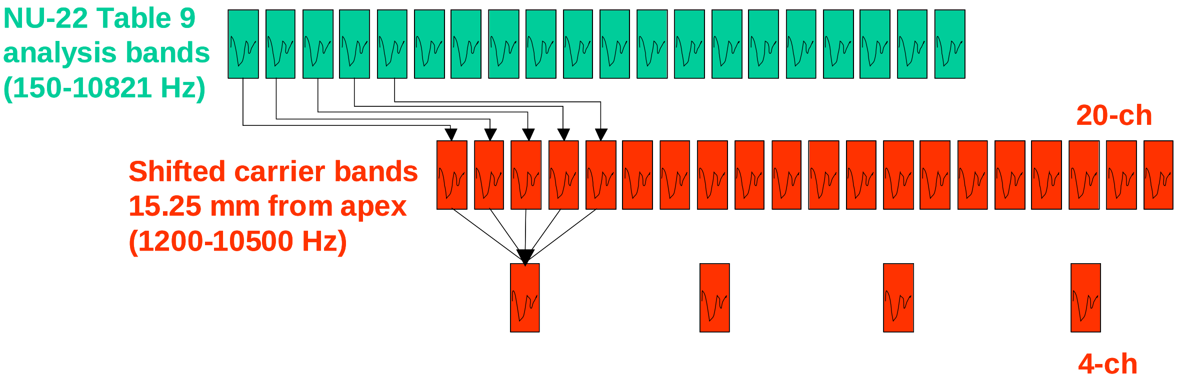


Figure 4 Schematic diagram for 4-channel speech processor

3. Speech Materials and Procedures

Two open-set recognition tasks included two sentence tests. Recognition of words in sentences was measured using novel sentences from the IEEE corpus (IEEE, 1969). The single-talker IEEE sentences were of moderate difficulty. Subjects were trained was using the DARPA/TIMIT acoustic-phonetic continuous speech corpus (Garofolo et al., 1993). The multi-talker TIMIT sentences were of moderate to extreme difficulty. Baseline (pre-training) data for all speech processor conditions (20- and 4-channel processors at 6 carrier locations) were collected on Day 1. Over the next four days, all subjects were trained on the most basally-shifted processor (15.25 mm from the apex) using 300 sentences from the DARPA/TIMIT corpus; performance was tested daily for the trained insertion depth only. On Day 5, after a final training session, subjects were re-tested for all speech processor conditions. The three subjects who were tested using the NU-22 Table 9 analysis bands were further trained with an apically-shifted processor (7.75 mm from apex) at a later date (3 - 10 days after completing the training on basally-shifted speech).

For the IEEE sentence recognition, a list was chosen pseudo-randomly from among 72 lists, and sentences were chosen randomly, without replacement, from the 10 sentences within that list. The subject responded by repeating the sentence as accurately as possible; the experimenter tabulated correctly identified words and sentences. Two lists of sentences were used for each test condition. For TIMIT sentence training, sentences were presented sequentially, without replacement, from a set of 300 sentences.

II. RESULTS

Figure 5 shows the mean word-in-sentence recognition scores for the 20-channel speech processors before and after subjects were trained with 20-channel basally-shifted speech, as a function of noise-band carrier location. NUC-22 frequency allocation Table 9 was used for the analysis bands for the results in Panel A, while NUC-24 Table 6 was used for Panel B. The blue line represents the baseline data before training, and the red line represents the data after the 5 days of training with basally-shifted speech.

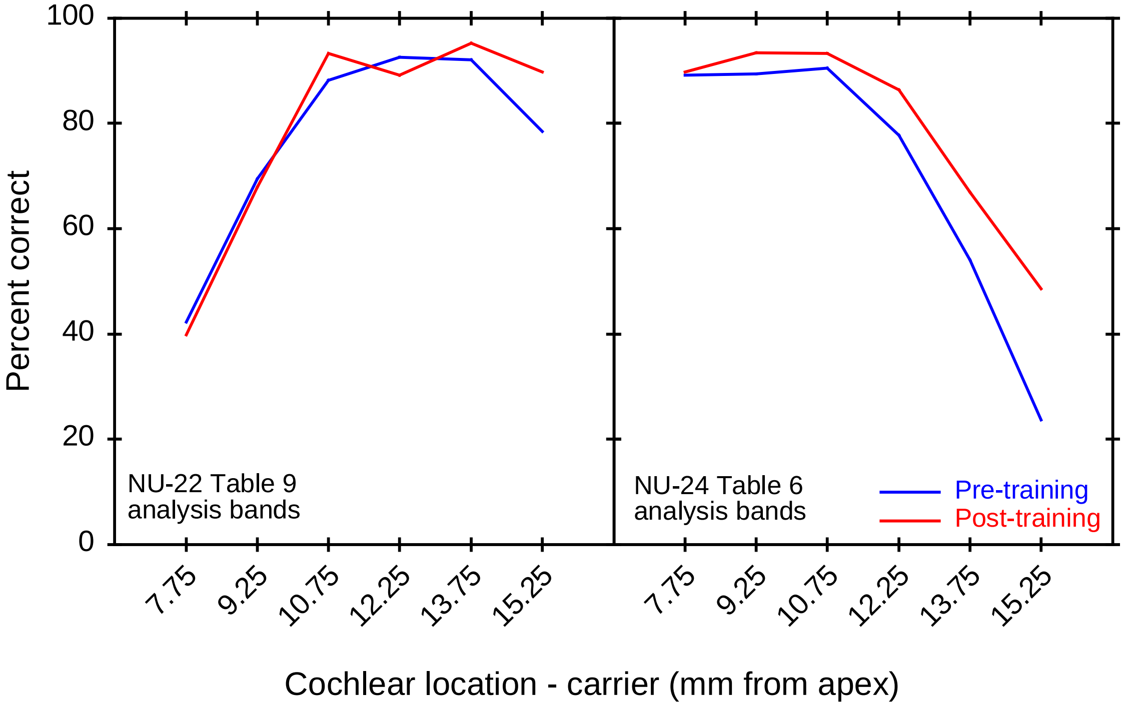


Figure 5: 20-channel word-in-sentence recognition scores before and after the 5-day training sequence with 20-channel basally-shifted speech. (A) NUC-22 frequency allocation Table 9 used as the analysis filters. (B) NUC-24 frequency allocation Table 6 used as the analysis filters.

For the pre-training measures obtained with the NUC-22 Table 9 analysis filters, the mean recognition scores peaked at a carrier location slightly more basal than a typical electrode insertion depth (10 mm from apex). Mean baseline performance remained relatively constant for moderately basally-shifted speech (10.75 mm to 13.75 mm from the apex). For the most basal carrier location (15.25 mm from the apex), mean performance dropped 14 percentage points relative to the peak mean performance. The largest performance drop (50 percentage points relative to the peak mean performance) was observed for the most apically-shifted carrier location (7.75 mm from the apex). Training with the basally-shifted speech did not significantly change performance in the apical and middle carrier locations. However, significant improvement was observed at the most basal carrier location where the training had occurred.

For the pre-training measures obtained with the NUC-24 Table 6 analysis filters, the mean recognition scores peaked over a range of carrier locations close to a typical insertion depth (10 mm from apex). Mean recognition scores were nearly unchanged for carrier locations ranging from 10.75 mm to 7.75 mm from the apex. However, there was a large drop in performance (67 percentage points) for the most basal carrier location (15.25 mm from the apex). Training with basally-shifted speech had little effect at the apical to middle carrier locations, probably because of ceiling effects associated with the already high performance obtained with the baseline measures. However, recognition of extremely basally-shifted speech improved by more than 20 percentage points after training at the most basal location (15.25 mm from the apex). Slight improvement was also observed for moderately basally-shifted speech.

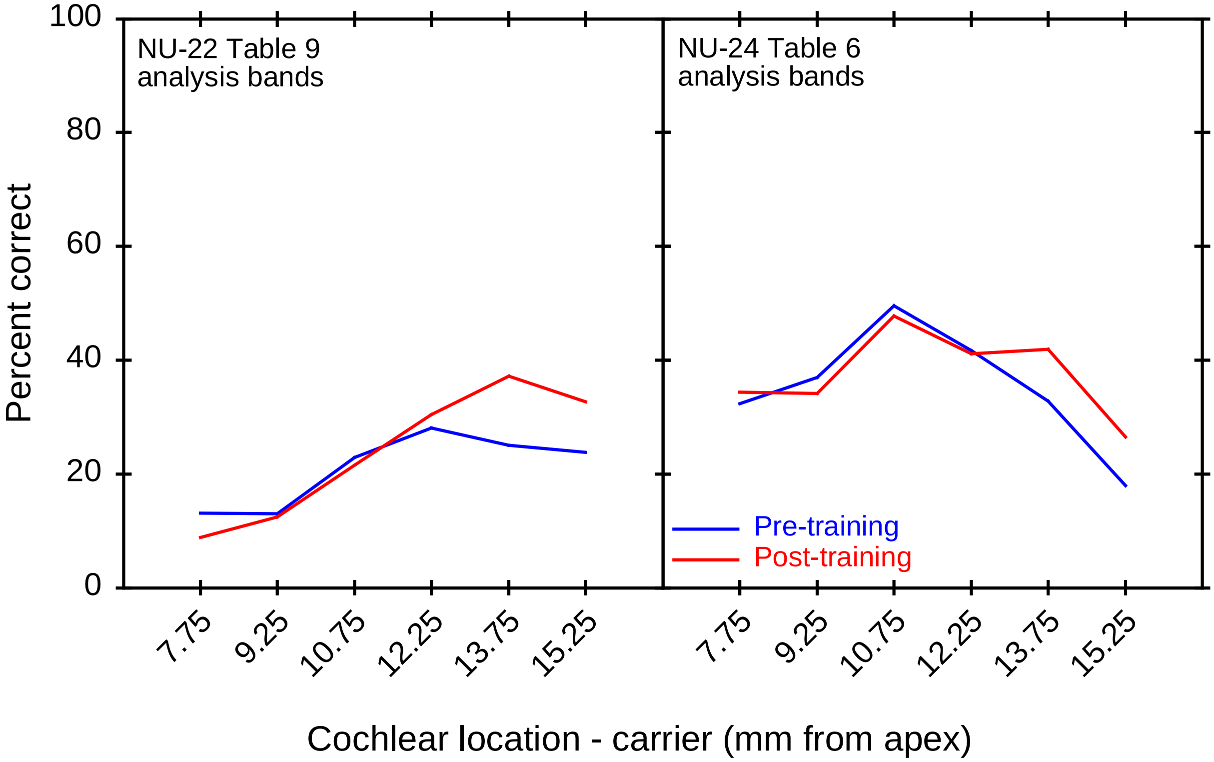


Figure 6: 4-channel word-in-sentence recognition scores before and after the 5-day training sequence with 20-channel basally-shifted speech. (A) NUC-22 frequency allocation Table 9 used as the analysis filters. (B) NUC-24 frequency allocation Table 6 used as the analysis filters.

Figure 6 shows the mean word-in-sentence recognition scores for the 4-channel speech processors before and after subjects were trained with 20-channel basally-shifted speech, as a function of noise-band carrier location. NUC-22 frequency allocation Table 9 was used for the analysis bands for the results in Panel A, while NUC-24 Table 6 was used for Panel B. The blue line represents the baseline data before training, and the red line represents the data after the 5 days of training with basally-shifted speech.

For the pre-training measures obtained with the NUC-22 Table 9 analysis filters, the mean recognition scores peaked at a simulated insertion depth of 12.25 mm, a carrier location more basal than a typical electrode insertion depth (10 mm from apex). Similar to performance with the 20-channel speech processor, there was a relatively large drop in recognition (20 percentage points relative to the peak performance location) for the most apical carrier location, but only a small drop (2 percentage points relative to the peak performance location) for the most basal carrier location. Training with the 20-channel basally-shifted speech caused the peak recognition to shift to a more basal carrier location (13.75 mm from the apex). Performance at this location and at the most basal carrier location (15.25 mm from apex) was significantly improved by the training with basally-shifted speech. Performance was unaffected by the training at the middle and apical carrier locations.

For the pre-training measures obtained with the NUC-24 Table 6 analysis filters, the mean recognition scores sharply peaked at a middle carrier location (10.75 mm from the apex) close to a typical electrode insertion depth. Similar to performance with the 20-channel speech processor, there was a large drop in recognition (20 percentage points relative to the peak performance location) at the most basal carrier location. However, in contrast to the 20-channel performance, there was significant drop in recognition with the 4-channel processors (10 percentage points relative to the peak performance location) at the most apical carrier location. Training with the 20-channel basally-shifted speech did not cause the peak recognition to shift from the middle carrier location. Performance at the extreme basal carrier location, as well as at the moderately basal location, was significantly improved by the training with basally-shifted speech. Performance was unaffected by the training at all other

Subjects using the NU-22 Table 9 analysis filters were also trained at a later date (3- 10 days later) with apically-shifted speech after completing the training/testing scheldule with basally-shifted speech. Figure 7 shows the mean word-in-sentence recognition scores with 20-channel speech processors before and after training with spectrally-shifted speech, as a function of noise-band carrier location. For all speech processing, NU-22 frequency allocation Table 9 was used for the analysis filter bandwidths. Panel A shows the performance before and after training with basally-shifted speech (as shown previously in Figure 5) and Panel B shows performance before and after training with apically-shifted speech. The blue line represents the baseline data before training, and the red line represents the data after the 5 days of training with spectrally-shifted speech.

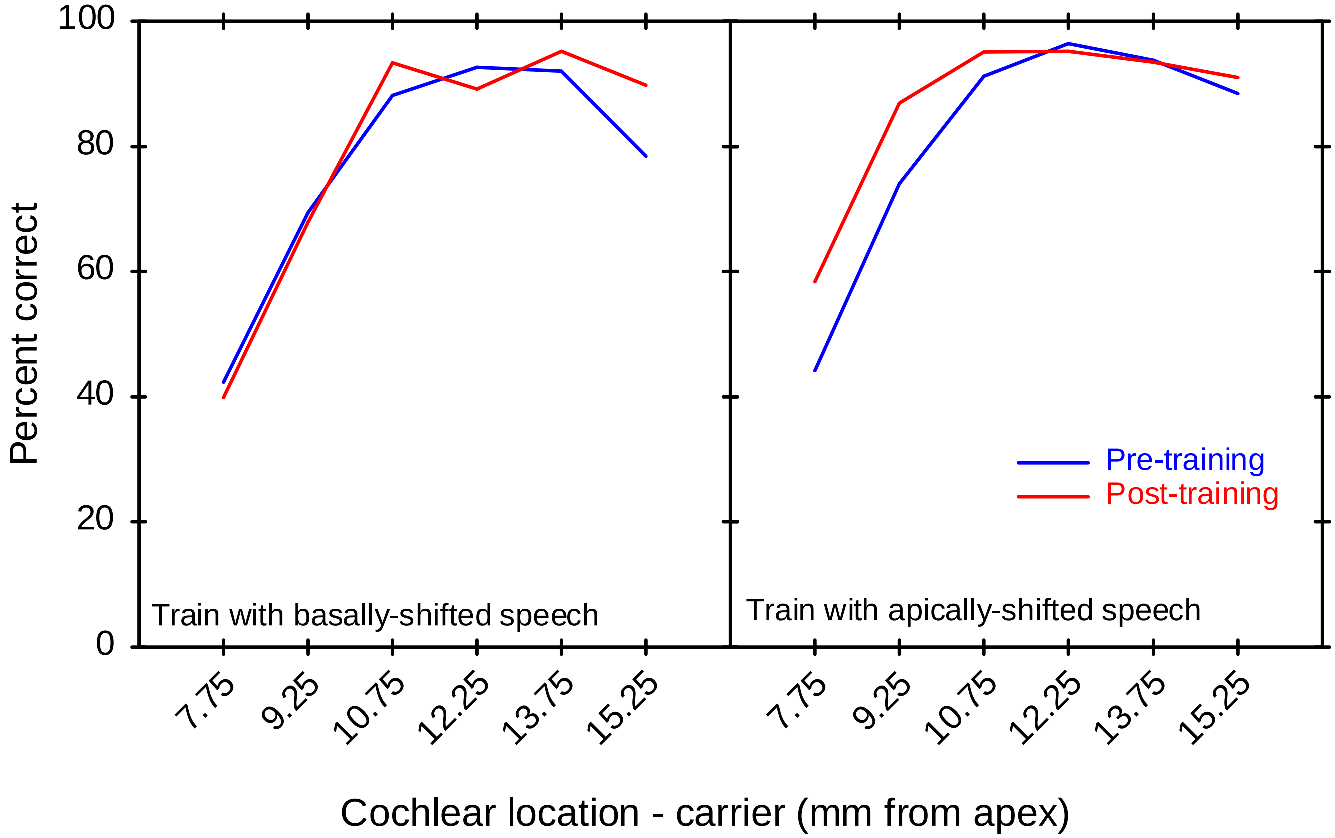


Figure 7: 20-channel word-in-sentence recognition scores before and after the 5-day training sequence with 20-channel spectrally-shifted speech. NUC-22 frequency allocation Table 9 was used for the analysis filter bands. (A) Training with basally-shifted speech (B) Training with apically-shifted speech

Pre-training baseline measures for both training/testing experiments show the peak mean recognition scores at a middle carrier location. In Panel A (training with basally-shifted speech), pre-training recognition at the most basal carrier location was 14 percentage points lower than the peak recognition performance; pre-training recognition at the most apical carrier location was 50 percentage points lower than the peak recognition score. In Panel B (training with apically-shifted speech), pre-training recognition at the most basal carrier location was less than 10 percentage points lower than the peak recognition performance; pre-training recognition at the most apical carrier location was again 50 percentage points lower than the peak recognition score. Training with spectrally-shifted speech showed differential effects on word-in-sentence recognition, depending on which extreme carrier location was used for training. As shown earlier, training with basally-shifted speech improved recognition at the moderate and extreme basal carrier locations, with no change in performance at the apical and middle carrier locations. As shown in Panel B, training with apically-shifted speech improved recognition at the apical and middle carrier locations, with no change in performance at the upper middle and basal carrier locations. Note also that the pre-training baseline measures shown in Panel B (blue line) are comparable to the post-training measures shown in Panel A (red line), suggesting that some of the effects of training with basally-shifted speech had been retained during the interim between the training/testing experiments.

III. DISCUSSION

The results of the present study suggest that the hypothesis of “local adpatation”, in which subjects adapt only to a specific spectral mismatch between the acoustic attributes of a signal and the place of stimulation along the cochlea. There was no “shift” in performance (in which performance for carrier locations distant from the training location would deteriorate) or “global” adaptation (in which performance improvements would have generalized to all carrier locations). Improved recognition was restricted to the carrier frequency regions where training was performed. Because performance was unchanged at cochlear locations where training did not occur, subjects may have been able to develop alternate spectral patterns for the trained, spectrally-shifted speech while retaining previous “internal” representations of speech, at least over the short term. The spectral distortions to these dominant internal representations were accommodated only by intensive training.

The differential effects of the frequency analysis range, as shown by the results obtained with the NU-22/Table 9 and the NU-24/Table 6 analysis filter banks, have important clinical implications. For the “acute”, pre-training baseline measures, NU-22/Table 9 (150-10823 Hz) provided better performance for shallow electrode insertion depths, while NU-24/Table 6 (117-7871 Hz) provided better performance for typical to deep insertion depths. With typical electrode insertions of 10 mm or less from the apex for the latest cochlear implant devices, the frequency range provided by NU-24/Table 6 should produce less spectral mismatch and better speech recognition results. The interactive effects between the analysis frequency range and the place of stimulation suggest that the electrode insertion depth and the acoustic frequency-to-electrode allocation range should be well-considered in the clinical fitting of speech processors.

The results also show that short-term training significantly improved recognition of shifted speech patterns with either fine (20 channels) or rough (4 channels) spectral resolution. Despite being trained with only 20-channel spectrally-shifted speech, subjects were able to also improve 4-channel word-in-sentence recognition at the carrier location where the training occurred. Theoretically, increasing the number of electrodes should be able to convey more spectral information to CI users, resulting in better speech recognition. Unfortunately, previous studies have shown that, for CI patients, phoneme recognition did not significantly improve with more than 7 electrodes, while for NH listeners, performance continued to increase up to at least 20 channels (e.g. Friesen et al., 2001). The results from the present study suggest that the effect of spectral mismatch between the acoustic frequency range and the place of stimulation was independent of the number of electrodes (or in this case, carrier bands).

Training with basally-shifted speech improved the recognition of basally-shifted speech only, leaving speech recognition for apical or middle carrier locations largely unchanged. Follow-up training with apically-shifted speech improved the recognition of apically-shifted speech only, leaving speech recognition for basal or middle carrier locations largely unchanged. These results are consistent with the “local adaptation” hypothesis. However, it is interesting to note that the improvement achieved with training at the basal carrier location was retained in the follow-up training experiment at the apical carrier location. After a 3 – 7 day interim, and even after the 5 days of training with apically-shifted speech, the improved performance at the basal carrier locations remained steady. These results strongly suggest that subjects may have been able to develop alternate representation for the trained speech patterns while preserving previous “internal” representations, even if they were just recently learned, and that these alternate representations can be preserved for at least the short term. Although the time courses are much different, these results are consistent with Fu et al. (2001), who observed that, after three months of listening to upwardly shifted speech and gradually improving performance, CI listeners’ baseline measures with their clinical processors was unchanged. Further, follow-up tests in which speech recognition was acutely measured with the shifted speech processors showed that subjects were able to largely retain the improvement they had achieved in the three-month period

IV. CONCLUSION

The present study examined the effects of short-term learning on normal-hearing listeners’ ability to acclimate to distorted speech patterns. Results showed that performance with 20- and 4-channel processors was significantly affected by the place of stimulation. Depending on the frequency analysis filter range used, there were differential effects due to the degree of spectral mismatch; however, the best performance was achieved at carrier locations most closely-matched with the acoustic frequency range. For both the baseline, pre-training and post-training measures, the frequency range associated with NU-22 frequency allocation Table 9 provided the best performance for shallow electrode insertion depths; the frequency range associate with NU-24 frequency allocation Table 6 provided the best performance for typically deep and deeper insertion depths. After training for 5 days with 20-channel basally-shifted speech, subjects were able to significantly improve word-in-sentence recognition at the basal carrier location where the training had occurred. However, this improvement did not generalize carrier locations remote from the basal training location. Performance at these remote carrier locations was not adversely affected, suggesting that only a local adaptation had occurred, specific to the spectral shift that was trained. A follow-up train and re-test experiment with apically-shifted speech showed a similar local adaptation. After training with 20-channel spectrally-shifted speech, subjects were also able to improve the recognition of 4-channel spectrally-shifted speech at the carrier location where the training had occurred, suggesting that subjects were accommodating an alternate representation of frequency space, regardless of the spectral resolution. There was also some retention of the trained shifted speech patterns, suggesting that these alternate representations may be developed while preserving previous internal representations, at least over the short term.

REFERENCES

Bilger, R.C. (**1984**). "Speech recognition test development," in *Speech Recognition by the Hearing Impaired* (ASHA Reports **14**), pp. 2-7.

Dorman, M.F., Loizou, P.C. and Rainey, D. (**1997a**). “Speech understanding as a function of the number of channels of stimulation for processors using sine-wave and noise-band outputs”, J. Acoust. Soc. Am. **102**, 2403-2411.

Dorman, M.F., Loizou, P.C. and Rainey, D. (**1997b**). "Simulating the effect of cochlear-implant electrode insertion-depth on speech understanding", J. Acoust. Soc. Am. **102**, 2993-2996.

Fishman K., Shannon R.V. and Slattery W.H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech Hear Res. **40**, 1201-1215.

Friesen, LM, Shannon, RV, Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150-1163.

Fu, Q.-J. (**1997**). *Speech perception in acoustic and electric hearing*, Ph.D. dissertation, University of Southern California, Los Angeles.

Fu, Q.-J. and Shannon, R.V (**1999a**). "Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus 22 cochlear implant," Ear Hear **20**, 332-344.

Fu, Q.-J. and Shannon, R.V. (**1999b**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889-1900.

Fu, Q.-J., Shannon, R.V., and Galvin, J.J. III (**2001**). "Adaptation of Nucleus-22 cochlear implant listeners to a speech processor with an apically-shifted frequency-to-electrode assignment," J. Acoust. Soc. Am., Re-submitted, July, 2001

Galvin, J.J. III and Fu, Q.-J. (2000). "Effects of spectral mismatch on sentence recognition for noise-band processors", J. Acoust. Soc. Am. 108(5), 2602

Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., and Dahlgren, N.L. (**1993**). "The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM", NTIS order number PB91-100354.

Greenwood, D.D. (**1990**). "A cochlear frequency-position function for several species - 29 years later," J. Acoust. Soc. Am. **87**, 2592-2605.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," Institute of Electrical and Electronic Engineers, New York, **1969**.

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

Shannon, R.V., Adams, D.D., Ferrel, R.L., Palumbo, R.L., and Grantgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," J. Acoust. Soc. Am. **87**, 905-907.

Shannon, R.V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M (**1995**). "Speech recognition with primarily temporal cues," Science, **270**, 303-304.

Skinner, M.W., Holden L.K., and Holden, T.A. (**1995**). "Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy," Ann Otol Rhinol Laryngol Suppl **166**, 307-311.

Wilson. B. S., Finley, C. C, Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (**1991**). "New levels of speech recognition with cochlear implants," Nature **352**, 236-238.

Peer-Reviewed Publications in This Quarter:

Chatterjee, M. and Robert, M.E. (2001). Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?, Journal of the Association for Research in Otolaryngology, 2(2)159-171.

Friesen, L., Shannon, R.V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants, Journal of the Acoustical Society of America, 110(2), 1150-1163.

Manuscript Submitted this Quarter:

Shannon, R.V. (2002). The Auditory Brainstem Implant, in MIT Encyclopedia of Communication Disorders, R.D. Kent (Ed.), MIT Press, Cambridge, MA. (Invited chapter)

Presentations this quarter:

Invited Presentations:

Chatterjee, M. Auditory proccessing with Cochlear Implants. Smith-Kettlewell Eye Research Institute Colloquium, August 15 2001.

Fu, Q.-J., Ponton, C., and Kwong, B. (2001). Auditory resolution and speech performance in cochlear implant users, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (invited oral presentation)

Shannon, R.V. (2001). Pattern recognition of speech: Implications for cochlear implants, Cochlear Implant Association International Convention, Minneapolis, July 13. (invited oral presentation)

Shannon, R.V. (2001). Cochlear implants: Respecting the past – Inspiring the future, Cochlear Implant Association International Convention, Minneapolis, July 14 (invited KEYNOTE address).

Shannon, R.V. (2001). 2001 Convention: Futures Panel, Cochlear Implant Association International Convention, Minneapolis, July 14 (Moderator, invited panel).

Shannon, R.V. (2001). Frequency-place alignment is critical for speech recognition. Starkey Corporation, Eden Prairie, MN, July 13. (invited oral presentation)

Shannon, R.V. (2001). Frequency-place compression and expansion, Departments of Psychology and Otolaryngology, University of Minnesota, July 16. (invited oral presentation)

Shannon, R.V., Fu, Q.-J., Baskent, D. and Galvin, J.J. (2001). Frequency-place alignment is critical for speech recognition, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (invited oral presentation)

Presentations:

Baskent, D. and Shannon, R.V. (2001). Speech recognition under conditions of frequency-place compression and expansion, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Burns, E.M., Sanborn, E.S., Shannon, R.V., and Fu, Q.-J. (2001). Perception of familiar melodies by cochelar implant users, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Chatterjee, M. (2001). Noise-induced release from modulation masking in Nucleus-22 cochlear implant listeners, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Chatterjee, M. and Galvin, J.J. (2001). Auditory streaming in cochlear implant listeners, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Friesen, L. and Shannon, R.V. (2001) Effect of stimulation rate on speech recognition in CIS processors , 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Galvin, J.J. and Fu, Q.-J. (2001). The effects of short-term learning for spectrally-mismatched noise-band speech processors, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Kuchta, J. , Otto, S., Shannon, R. (2001). Pitch range and number of electrodes as predictors of speech in an auditory brainstem implant, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Otto, S.R., Shannon, R.V., Friesen, L.M., Brackmann, D.E., and Hitselberger, W.E. (2001). The multichjannel auditory brainstem implant: Overview of results of from more than 100 patients, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Padilla, M. and Shannon, R.V. (2001), Effects of English experience and spectral resolution on English phoneme and word recognition by non-native English speakers, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Stickney, G., Loizou, P., Assman, P., Shannon, R.V., and Opie, J. (2001). Electrode interaction and speech intelligibility in multichannel cochlear implants, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Wang, X.-S., Fu, Q.-J and Galvin, J.J.. (2001). Consonant intelligibility with adjusted consonant-vowel intensity ratios by cochlear implant users, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Wygonski, J.J., Faltys, M., Shannon, R.V., Tateyama, T., and Alabashayan, J. (2001). Research interface for Clarion CII cochlear implant system, 2001 Conference on Implantable Auditory Prostheses, Asilomar, CA, August 19-24. (poster)

Plans for the next Quarter:

We continue to work on cochlear implant research interface hardware and software, which would enable experiments similar to the one described in this report to be performed with implant patients.

Advanced Bionics. The hardware and basic software for the Clarion CII research interface (CRI-2) is available. Final testing of basic software is underway and new software to implement our own experimental paradigms are being modified to address this interface. This interface will allow flexibility to provide high rate stimulation on up to 16 simultaneous electrodes (31 in “virtual channel mode”). We have received software from Advanced Bionics that will allow testing of CII patients with stimulation rates in excess of 5000 pps/electrode for 8 and 16 channel CIS processors. These high rates should allow us to measure any improvements in performance due to these high rates possibly putting the electrically stimulated auditory nerve into a more stochastic mode of firing. Some frequency-to-electrode remapping experiments are also possible with this software. Patients are already being tested with these research systems.

Nucleus-24. We have developed our own custom designed hardware interface for the Nucleus-24 cochlear implant system. This system is in use in the laboratory and allows high-rate stimulation under experimental stimulus control, as well as frequency-place remapping in a laboratory setting. We have ordered a SPEAR3 portable research speech processor from the developers of this hardware in Melbourne, Australia, but have not yet received the research devices. We anticipate receiving this hardware in the next quarter and beginning software development. Since the SPEAR3 uses the same DSP chip as our research interface, the real-time DSP code should be relatively easy to port to the new device. This portable research processor will allow short-term and long-term learning experiments with different processor parameter adjustments. We expect these experiments to begin in the first quarter of 2002.

Presentations and Publications. We are preparing a talk and several posters to be presented at the annual NIH Neural Prosthesis Workshop in Bethesda on October 17-19. We will submit several posters based on the present work for the ARO Midwinter Research Meeting, to be held in St. Petersburg, Florida on January 27-31, 2002. Bob Shannon will give the Keynote Speech at the New York University Symposium “Next Generation Cochlear Implants”, to be held November 30 to Dec 1. Quotes from Shannon appear in a web site (http://www.pbs.org/saf/1205) for a show about cochlear implants that will air November 13 on “Scientific American Frontiers”, a science magazine TV show on public television. Sound clips on the program and on the web site demonstrate the effects of the number of channels on speech understanding.

We are preparing manuscripts for submission on our work quantifying the effect on speech recognition performance of stimulation pulse rate, pulse phase duration, and of frequency-place compression and expansion. We anticipate submission of these papers in the final quarter of 2001. We are also preparing a book chapter on speech recognition in cochlear implants for a book titled “Auditory Prostheses” that will be edited by Fan-Gang Zeng, Richard Fay, and Art Popper, and will be published as part of the Springer Handbook of Auditory Research series.