Speech Processors for Auditory Prostheses

NIH Contract N01-DC-92100

Quarterly Progress Report #8: Oct-Nov-Dec 2000

Speech Recognition under Conditions of Frequency-Place Expansion and Compression

Submitted by

Robert V. Shannon, Deniz Baskent, John Wygonski, Mark Robert,
Qian-Jie Fu, John J. Galvin III, Lendra Friesen, and Monita Chatterjee
House Ear Institute
Los Angeles, CA 90057

31 January 2001

Table of Contents: **PAGE**

ABSTRACT

In this quarter we completed hardware and software development of the research interface for the Nucleus-24 device in SEMA protocol. This interface is now in use for data collection with Nucleus-22 and Nucles-24 patients.

Hardware and software development continued for the Clarion Research Interface 2. Prototypes are now working in the lab and we expect the interface to be available in the next quarter for working with Clarion CII patients.

In this report we present speech recognition results as a function of frequency-place compression and expansion. Previous experiments have shown that the mapping of frequency information onto the correct cochlear place is critical for speech recognition. In cochlear implants, the cochlear tonotopic range represented by the electrode array is smaller than the acoustic frequency range used in the speech processor. While this condition utilizes a wide range of speech frequency information, it results in a compression of the tonotopic pattern of speech information delivered to the brain. An alternative approach is to take the core frequency range necessary for speech and expand its representation in the cochlea, like an “acoustic fovea.”

This study examined the effect of linear (in mm) frequency-place compression and expansion on speech intelligibility for various cochlear locations and number of spectral channels. These conditions were presented to normal-hearing listeners using a noise band vocoder. The cochlear tonotopic range was held constant by employing the same carrier bands for each condition, while the frequency range of the analysis bands was changed. For each condition, the result was compared to that of the perfect tonotopic match, where the carrier and the analysis bands were perfectly matched. Speech recognition in frequency-place expansion and compression was always equal to or poorer than the matched condition.

In the next quarter we will present some of our experimental results at the Association for Research in Otolaryngology (ARO).

CI Research Interface Development

Nucleus-22 and -24 SEMA Research Interfaces

We have completed software and hardware development of the research interface for the Nucleus-22 and 24 cochlear implants. Both hardware and software have been fully debugged and verified under the most extreme stimulation conditions. These new interfaces are in use now in our lab using the SEMA protocol. For the Nucleus-22 device this new interface can stream data from the host computer, allowing delivery of long-duration stimuli, like sentence materials and musical stimuli. The higher clock speed for transcutaneous transmission in the Nucleus-24 device limits the delivery of stimuli to 4000 pulses at present. Further optimization is required in the DSP code to allow full streaming of data from PC to interface for the Nucleus-24. We anticipate that this software development will begin in the next quarter.

Additional DSP programming is required to implement the Embedded Stimulation Protocol. We project that software development to implement the Embedded protocol will begin in the next quarter and may be complete in the second quarter of 2001.

Clarion CII Research Interface

The research interface for the Clarion CII device is nearing completion. A workshop has been scheduled in the first quarter of 2001 to introduce the new interface to qualified research groups. It is anticipated that the new CRI-2 interface will be available for use in patients with the new Clarion CII implant in the second quarter of 2001.

SPEAR Portable Processors for the Nucleus-24

We have not yet received a firm date for delivery of the SPEAR portable processors from the University of Melbourne. Experiments on long-term learning of new processing strategies, and experiments with binaural implants are dependent on the SPEAR processor.

Experiment Report: Frequency-Place Expansion and Compression

INTRODUCTION

Cochlear implant electrodes can be inserted as far as 25-30 mm into the round window and the active stimulation region is typically 15-16 mm in length. The characteristic frequency region corresponding to the stimulated electrodes is 500-5000 Hz, according to Greenwood’s (1990) frequency-to-place equation. However, implant speech processors assign the entire acoustic frequency range of 150Hz-10kHz to the electrodes. This range of acoustic frequency would normally occupy a characteristic frequency range in the cochlea of 25 mm. Thus, mapping the larger frequency range onto the electrode length results in compression in the frequency-to-place mapping. The present experiment evaluated the effect of such frequency-place compression on speech recognition in normal-hearing listeners. Frequency-place expansion was also simulated. In this condition we took the most significant frequency region for speech (1100-2900 Hz) and expanded its representation in the cochlea, effectively increasing the resolution within this frequency range. This frequency-place expansion is analogous to the “acoustic fovea” in bats, where a large portion of the bat cochlea is devoted to the small frequency region used for echolocation. For the data in the present report, both frequency expansion and compression were accomplished in normal-hearing listeners using a noise-band vocoder technique (Shannon *et al.*, 1995). In the next quarter similar frequency-place manipulations will be evaluated in cochlear implant listeners.

METHOD

Subjects.

Five normal hearing listeners between ages 26-34 participated in this experiment. All subjects were native speakers of American English and had hearing thresholds better than 10 dB HL at standard audiometric frequencies.

Speech materials

Speech perception tests used to evaluate the experimental settings were all presented without lip-reading (sound only). The tests consisted of medial vowel and consonant discrimination, and sentence recognition. All stimuli were presented monaurally over headphones at a level of 70 dB SPL.

Vowel stimuli were taken from materials recorded by Hillenbrand et al. (1995) and were presented to the listeners with custom software (Robert, 1998). Ten presentations (5 male and 5 female talkers) each of twelve medial vowels in a h/V/d context including 10 monophthongs (/i ɪ ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and 2 diphthongs (/o e/) presented in a /h/-vowel-/d/ context (heed, hawed, head, who’d, hid, hood, hud, had, heard, hoed, hod, hayed). Chance level on this test was 8.33% correct and the 95% confidence level was 11.8% correct.

Consonant stimuli (5 male and 5 female talkers) were taken from materials created by Shannon *et al.* (1999). Consonant confusion matrices were compiled from 12 presentations of each of 14 medial consonants /b d g p t k l m n f s ʃ v z j θ/, presented in an /a/-consonant-/a/ context. Tokens were presented in random order by custom software (Robert, 1998) and the confusion matrices were analyzed for information received on the production based categories of voicing, manner, and place of articulation (Miller and Nicely, 1955). Chance performance level for this test was 7.14% correct, and the 95% confidence level was 10% correct.

Sentence materials were taken from the TIMIT database. These sentences are of varying difficulty and each sentence was spoken by a different talker. Scores reflect the number of key words identified. Twenty sentences were presented for each condition.

Speech Processor Conditions:

All speech materials were processed through noise-band vocoders (Shannon et al., 1995) with 4, 8, and 16 channels. This type of processing removes spectral fine-structure information and restricts the spectral information to 4, 8, or 16 bands. The target frequency range was divided into 4, 8, or 16 equal segments in mm using the frequency-to-place formula of Greenwood (1990). The envelope was extracted from each band by half-wave rectification and low-pass filtering at 160 Hz. The envelope from each band was used to modulate a band of noise. The bandwidth of each noise band was determined according to the experimental condition. For the conditions that simulated implant electrode locations, the center frequency of each band was computed from Greenwood's (1990) formula for either 25-mm or 20-mm simulated insertion depth. The electrode length was always assumed to be 16 mm. Thus, the electrodes were assumed to lie between 4 and 20 mm from the round window for the 20 mm simulated insertion depth and between 9 and 25 mm from the round window for the 25 mm simulated insertion depth. From Greenwood's formula the frequency ranges corresponding to these insertion depths are 1200 Hz – 11.8 kHz for the 20 mm depth and 510-5800 Hz for the 25 mm depth. The 16-mm electrode length was divided into 4, 8, or 16 equal parts and the frequencies corresponding to those divisions were computed from the Greenwood formula and used to specify the noise bands.

Analysis frequency bands were calculated in a similar fashion, except that the overall range was varied. The cochlear distance corresponding to the analysis filters was varied by –5, -1, 0, +1, +3, and +5 mm on both apical and basal ends relative to the simulated electrode location. The actual frequency ranges corresponding to these conditions are presented in Table 1. The cochlear locations of the bands are given in mm from the round window according to the Greenwood formula.

TABLE 1: Frequency ranges for the analysis bands

Simulated 25 mm insertion:

condition	location	frequency range
-5 mm	20-14 mm	1100-2900 Hz
-3 mm	22-12 mm	850-3800 Hz
-1 mm	24-10 mm	610-5100 Hz
0 mm	25- 9 mm	510-5800 Hz
1 mm	26- 8 mm	430-6800 Hz
3 mm	28- 6 mm	290-9000 Hz
5 mm	30- 4 mm	180-11800 Hz

Simulated 20 mm insertion:

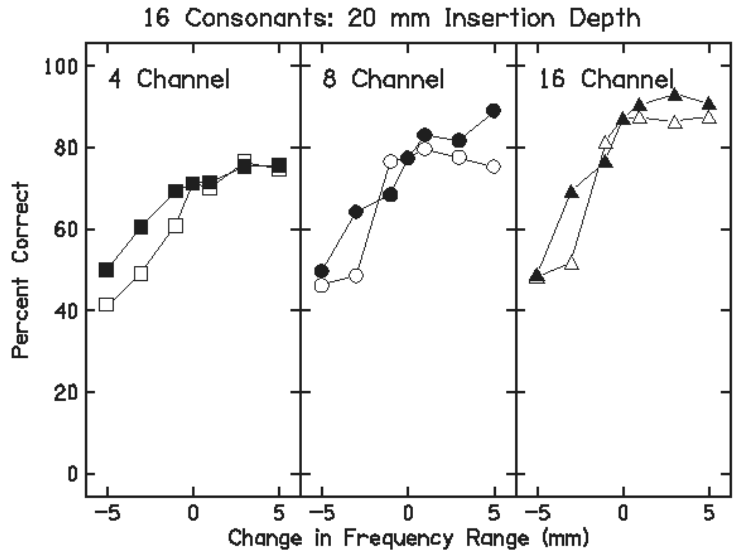
condition	location	frequency range
-5 mm	15-9 mm	2500-5900 Hz
-3 mm	17-7 mm	1800-7800 Hz
-1 mm	19-5 mm	1400-10300 Hz
0 mm	20-4 mm	1200-11800 Hz
1 mm	21-3 mm	1000-13600 Hz
3 mm	23-1 mm	720-18000 Hz
5 mm	25-0 mm	510-20000 Hz

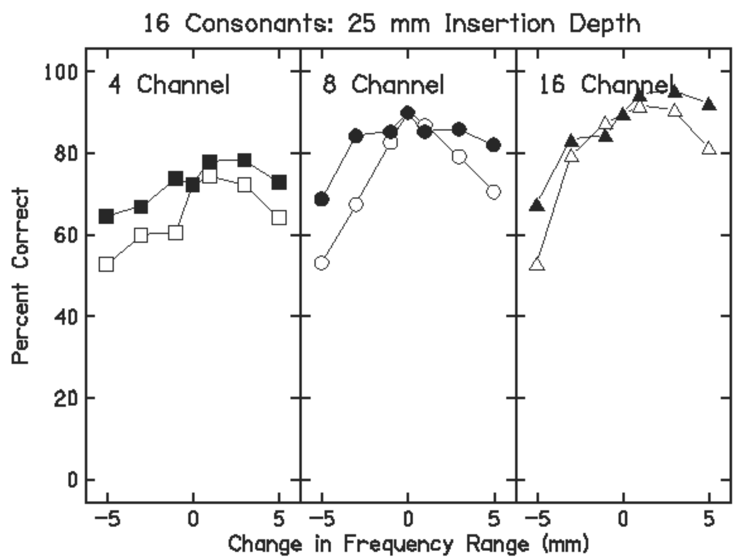
In the baseline conditions the analysis filters and electrode locations were matched in frequency-place. These baseline conditions do not simulate any implant condition, because the noise carrier bands were not intended to simulate any electrode insertion depth and spacing. Rather, this baseline condition simulates the effect of changing the acoustic range used in the processor. If an implant electrode could be designed that would match the acoustic frequency range, the baseline condition would indicate the best performance that could be obtained. Performance in this baseline condition indicates the effect of the gain or loss of acoustic information resulting from the expansion or truncation of the analysis frequency range.

RESULTS

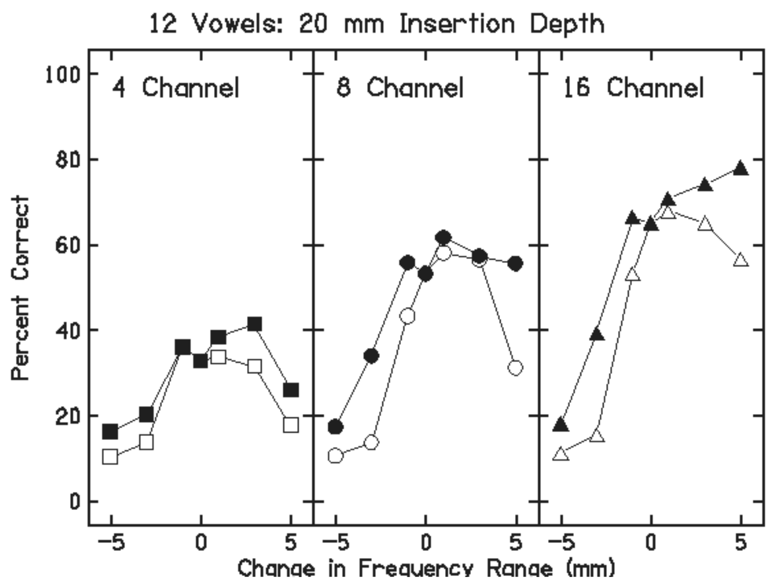
<u>Consonants</u>

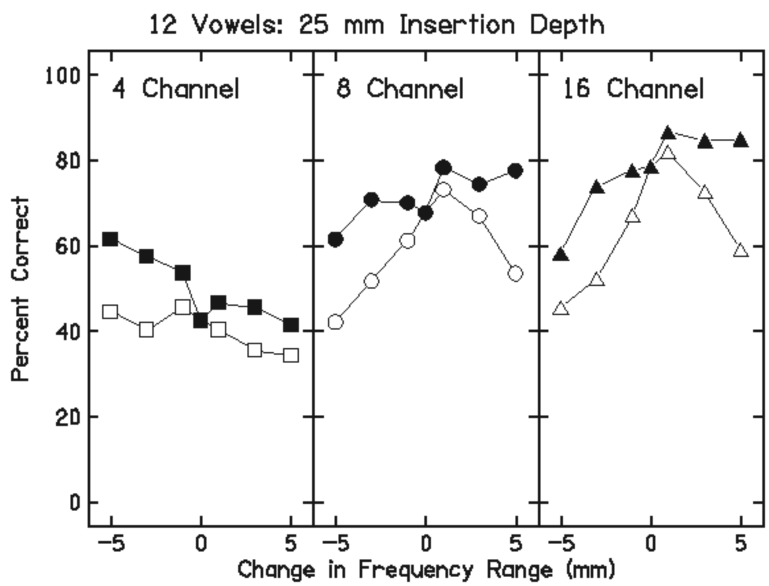
Figures 1 and 2 present the results for consonant recognition. The three panels in each figure present results from 4, 8, and 16 channel processors, respectively. Figure 1 presents the results from a 20 mm simulated insertion depth and Figure 2 presents results from a 25 mm simulated insertion depth. Within each panel the consonant recognition scores are presented for the baseline conditions in which the analysis and carrier bands were always matched (filled symbols) and for conditions in which the analysis bands varied while the carrier bands were fixed to simulate a fixed electrode location (open symbols). For the matched conditions the overall shape of the curves shows reduced performance as the analysis bands are reduced in frequency range. This simply illustrates the reduction in performance due to the loss of acoustic information, and so constitutes the baseline condition. Performance in the mismatched conditions was always equal to or poorer than the baseline condition. Note that the frequency expansion conditions (-5 mm and –3 mm) always resulted in a large reduction in performance. A substantial amount of this reduction was due to the loss of acoustic information, as indicated by the filled symbols, and an additional reduction in performance was observed when the frequency-place mapping was expanded (open symbols). Thus, even though the cochlear representation of the spectral information was expanded, presumably resulting in improved spectral resolution, performance was always poorer





than the matched condition. This result suggests that improved resolution in the spectral domain does not necessarily improve speech recognition. At least without time to adapt to the new mapping, changes in frequency-place always resulted in poorer consonant recognition. For extreme conditions of frequency-place compression (+5mm) or expansion (-5 mm) the reduction in performance was 10-15% relative to the baseline condition. There was no clear difference between the pattern of results for 4, 8, and 16 channels other than the overall better performance with more channels. Thus, it appears that additional spectral resolution does not mitigate the deleterious effects of a frequency-place mismatch.



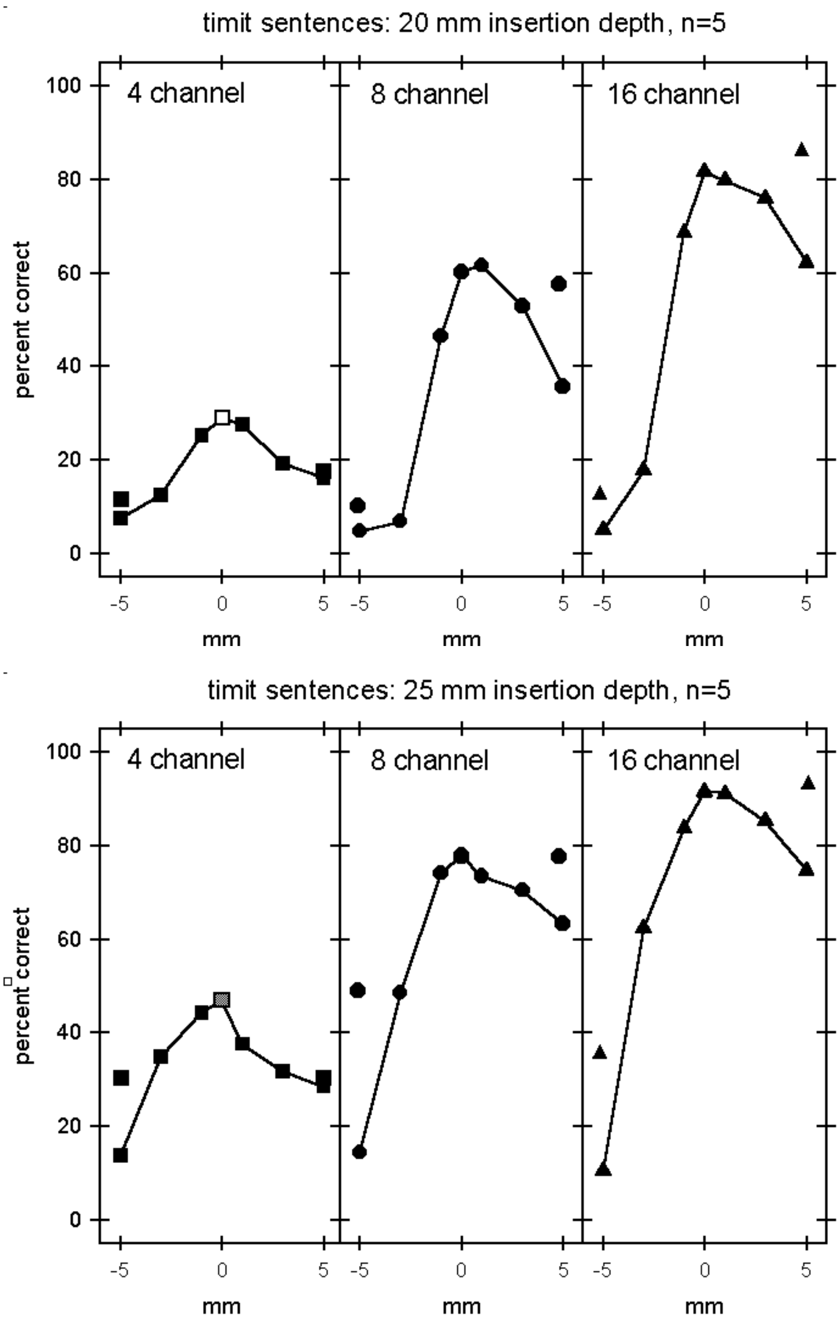


<u>Vowels</u>

Figures 3 and 4 present the results for vowel recognition for the simulated 20 mm and 25 mm simulated insertion depths, respectively. The overall pattern of results is similar to that shown for consonants in Figures 1 and 2. For the 20 mm simulated electrode insertion depth frequency-place expansion always resulted in a substantial reduction in performance. However, for the 25 mm simulated insertion depth frequency-place expansion resulted in improved performance for the four-channel matched condition. Because a smaller frequency range was used, the –5 mm baseline condition resulted in better frequency resolution because the smaller frequency range was still divided into four bands. This produced better performance only for the four-channel condition. For the four-channel condition there was also little decrease in performance for the –5 mm expansion (compared to the 0 mm baseline condition), suggesting that the improved frequency resolution was sufficient to offset the deleterious effects of the frequency-place expansion. For 8 and 16 channel processors both frequency-place expansion and compression resulted in 20-30% drop in recognition compared to the baseline/matched conditions. Note that normal cochlear implant processors typically use a frequency-place assignment that is similar to the +5 mm compression condition, which produced a 20-30% reduction in vowel recognition.

TIMIT Sentences

Figures 5 and 6 present TIMIT sentence recognition with simulated electrode insertion depths of 20 and 25 mm, respectively. Due to the limited number of sentence sets, the matched baseline performance was only measured for the –5, 0, and +5 mm conditions. The matched data points are shown as isolated filled symbols. The frequency-place compression and expansion results are shown as filled symbols connected by a solid line. Note that the drop in performance for frequency-place expansion (-5 mm) is dramatic. For all three processors performance at the –5 mm condition drops to almost chance levels (we assume chance performance on sentence recognition is 5-10% when scored by words correct). For the 16-channel processor this drop is 80%. A smaller drop in performance is observed for frequency-place compression, but the decrease is still 20% or greater between the matched baseline condition and +5 mm compression, except for the 4-channel processor. This condition is the typical frequency-place assignment used in most commercial cochlear implants.



DISCUSSION

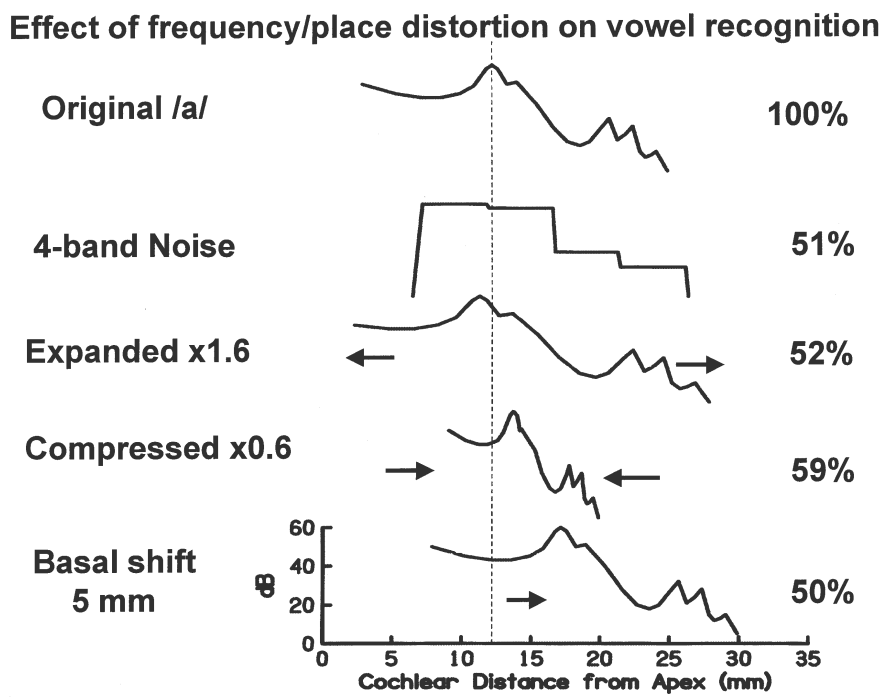
These results in normal-hearing listeners simulating frequency-place conditions that might occur in implants suggest that careful attention should be paid to matching the frequency range to the cochlear range of the electrodes. In almost all cases, the best speech recognition performance was observed in the conditions in which the frequencies were matched to the normal acoustic cochlear place for those frequencies. Previous studies have shown that a reduction in speech recognition results when the frequencies are mapped to cochlear locations that are shifted apical or basal from their normal acoustic locations (Fu and Shannon, 1999; Dorman *et al.*, 1997), and when the frequency-place mapping is warped (Shannon et al., 1998). The present study extends those results to show that speech recognition is also reduced when the frequency range is larger or smaller than the cochlear range.

A 20-30% reduction in speech recognition was observed when a frequency range was mapped onto a cochlear range that was smaller by 2 octaves (+5 mm compression condition). In this case a large acoustic range was compressed into a smaller cochlear range. Even though more acoustic information was represented, performance was reduced due to the compression in the frequency-place assignment. Unfortunately, this condition is similar to the mapping used in most cochlear implant systems, in which the full acoustic frequency range of 150 Hz to 10 kHz is typically mapped onto electrodes that span the normal acoustic range of only 500-5000 Hz. This result implies that speech recognition performance in cochlear implants might be improved by as much as 20% if the frequency range for each electrode could be mapped according to the normal acoustic characteristic frequency of that cochlear location.

An expansion in the frequency-place mapping could theoretically improve speech recognition by spreading out the critical speech spectral region to a larger range in the cochlea, thus improving the resolution within a given frequency range. Echolocating animals have evolved such a strategy to provide better spectral resolution in the small frequency region of their echo signal. However, in the present results such expansion conditions mostly resulted in poorer speech recognition. The only exception was for the 4-channel processor with a 25 mm simulated insertion depth. This result suggests that frequency-place expansion could be beneficial in conditions in which tonotopic resolution is minimal. In such a case the frequency-place expansion may be more beneficial than the detrimental effects of losing part of the spectral range.

In implant listeners there is additional uncertainty in the exact location of the electrodes and further uncertainty in the location of the stimulated neurons. Recent advances in imaging technology allow sufficient resolution to evaluate the depth of electrode insertion and o detect the presence of any kinks or abnormalities in the electrode carrier (Ketten et al., 2000). However, even these high-resolution images do not allow determination of the location of the electrode in terms of the medial-lateral position within the scala tympani, and the images do not indicate the location of the neurons stimulated by each electrode. It is possible that the actual range of characteristic frequencies stimulated will be larger than the estimates based on Greenwood’s formula because the stimulation actually takes place at the spiral ganglion. The spiral ganglion are arranged tonotopically in Rosenthal’s canal, which is inside the cochlear spiral and so has a shorter total length than the cochlea measured at the center of the basilar membrane. So the computation of frequency-place along the axis of the spiral ganglion would be different from that along the axis of the basilar membrane. These factors produce additional uncertainties regarding the appropriate frequency-place mapping in implant patients. The data presented above control for these factors by making the measures in normal-hearing listeners, in whom the actual stimulation locations can be controlled. Cochlear implant listeners may show best speech recognition when the frequency-place mapping is compressed relative to the normal-hearing results.

The pattern of results observed in the present experiment, when combined with previous results on frequency-place shifting (Fu and Shannon, 1999; Dorman *et al.*, 1997) and warping (Shannon *et al.*, 1998) suggests that the central pattern recognition of speech is not stored in terms of an abstract pattern, but in terms of an absolute pattern. Speech patterns are not "relocatable". If the peripheral representation of the pattern of speech information is shifted, warped, expanded or compressed, the result is almost always a decrement in speech recognition. Figure 7 presents a schematic representation of a vowel spectrum and the various types of distortion that result in a reduction in vowel recognition to approximately 50%. The top curve shows the original undistorted spectrum of the vowel /a/. The second curve shows the same vowel represented by a four-band noise vocoder, which allows 51% correct on multi-talker vowel recognition. The middle curve shows the effect of a frequency-place expansion by a factor of 1.6 (10 mm extent expanded to 16 mm), which results in 52% correct vowel recognition. The fourth trace shows the effect of frequency-place compression by a factor of 0.6 (26 mm extent compressed to 16 mm), which results in 59% correct. And the bottom curve shows the spectrum of an /a/ that has been shifted by 5mm basally in the cochlea, resulting in 50% correct vowel recognition (based on Fu and Shannon, 1999). Note that the shifted and expanded representations look less distorted than the four-band and compressed representations, yet all produce a similar level of performance. This comparison suggests that the central pattern recognition mechanisms are sensitive to the absolute frequency place. If the frequency-place information is in the correct location, the central pattern recognition can tolerate an extreme loss of spectral resolution – down to four bands. However, speech recognition



is impaired even with good spectral resolution if the frequency-place pattern is distorted. This result suggests that a possible cause of poor performance in cochlear implants is not necessarily the small number of electrodes or number of effective spectral channels. Even if an implant patient is able to use many electrodes effectively, their performance might be limited by the distortion in the frequency-place mapping. If this is the case, then efforts to produce a better match in frequency-place mapping may produce improvements in speech recognition.

In the next quarter we will initiate experiments to evaluate the effect of frequency-place expansion and compression in implant listeners.

Plans for the Next Quarter

In the next quarter (January-March 2001) we will continue hardware and software development on the Nucleus-24 and Clarion S-2 research interfaces. We will present several posters based on recent work at the Midwinter meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida, Feb4-8, 2001.

Experimental work in the next quarter will include:

1. Completion of channel interaction measures and speech recognition measures in Clarion implant subjects with the original Clarion spiral electrode, the original electrode plus the Positioner (EPS), and the new Hi-Focus electrode plus Positioner (HF+EPS).
2. Experimental manipulations that expand or compress the frequency-place mapping in implant patients with the Nucleus-22 and –24 device.
3. Completion of measurements of the effect of stimulation rate on speech recognition with the Clarion and Nucleus-24 implants.
4. Continuing psychophysical measures of temporal processing (Modulation detection and discrimination) in good and poor implant users.

Publications and Presentations in this Quarter

Publications: NONE

Manuscripts Submitted this Quarter:

Shannon, R.V. (2001). Speech Perception, McGraw-Hill 2002 Yearbook of Science and Technology, McGraw-Hill, New York. (Invited chapter submitted 4 Dec 00)

Shannon, R.V. (2001). Overview of cochlear implant technology and research, in proceedings of “Nonsyndromic deafness: Clinical issues and research opportunities”, Short Course at the 2001 Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL., Feb 4-8. (submitted 1 Dec 00)

Presentations:

Baskent, D.J. and Shannon, R.V. (2000). Speech recognition under conditions of frequency-place expansion and contraction, 140th Meeting of the Acoustical

Society of America, Newport Beach, CA, 3-8 December, J. Acoust. Soc. Amer, 108(5), Pt. 2, 2571. (poster)

Chatterjee, M. (2000). The form of the expansive nonlinearity in cochlear implant stimulation, 140th Meeting of the Acoustical Society of America, Newport Beach, CA, 3-8 December, J. Acoust. Soc. Amer, 108(5), Pt. 2, 2571. (poster)

Fu, Q.-J. (2000). Auditory temporal resolution and speech performance in cochlear implant users, 140th Meeting of the Acoustical Society of America, Newport Beach, CA, 3-8 December, J. Acoust. Soc. Amer, 108(5), Pt. 2, 2600. (poster)

Galvin, J.J III, and Fu, Q.-J. (2000). Effects of spectral mismatch on sentence recognition for noise-band speech processors, 140th Meeting of the Acoustical Society of America, Newport Beach, CA, 3-8 December, J. Acoust. Soc. Amer, 108(5), Pt. 2, 2603. (poster)

Stickney, G.S., Loizou, P., Mishra, L., Assmann, P.F., Shannon, R.V., and Opie, J.M. (2000). Channel interaction and speech processing strategies for cochlear implants, 140th Meeting of the Acoustical Society of America, Newport Beach, CA, 3-8 December, J. Acoust. Soc. Amer, 108(5), Pt. 2, 2601. (poster)

References

Dorman, M.F., Loizou, P.C. and Rainey, D. (1997). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding", J. Acoust. Soc. Amer., 102(5), 2993-2996.

Fu, Q.-J. and Shannon, R.V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing, Journal of the Acoustical Society of America, 105(3), 1889-1900.

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

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (1995). Acoustic characteristics of American English vowels, J. Acoust. Soc. Am., 97, 3099-3111.

Ketten, D.R., Skinner, M.W., Wang, G., Vannier, M.W., Gates, G.A., and Neely, J.G. (1998). In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays, Ann. Otol. Rhinol. Laryngol., Suppl. 175, 1-16.

Miller, G. & Nicely, P. (1955). An analysis of perceptual confusions among some english consonants. Journal of Acoustical Society of America**,** 27, 338-352.

Robert, ME. (1998). *CONDOR: Documentation for Identification Test Program*. Los Angeles; House Ear Institute.

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

Shannon, R.V., Zeng, F.-G., and Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues, J. Acoust. Soc. Amer., 104(4), 2467-2476.

Shannon, R.V., Jensvold, A., Padilla, M., Robert, M., and Wang, X. (1999). Consonant recordings for speech testing, Journal of the Acoustical Society of America (ARLO), 106(6), L71-L74.