Speech Processors for Auditory Prostheses

NIH Contract N01-DC-92100

Quarterly Progress Report #6: April-May-June 2000

Recognition of Spectrally Asynchronous Speech By Normal-Hearing Listeners and Nucleus-22 Cochlear Implant Users

Submitted by

Qian-Jie Fu, John J. Galvin III, Robert V. Shannon,
Mark Robert, and John Wygonski,
House Ear Institute
Los Angeles, CA 90057

31 July 2000

Table of Contents: **PAGE**

ABSTRACT

In this quarter we continued hardware and software development of research interfaces for the Nucleus-24 and Clarion S-2 implant devices. The Nucleus-24 interface was checked for accuracy at high pulse rates and long pulse phase durations. A custom circuit board was designed to minimize the chip count and power consumption of the interface design. In the next quarter we will have several boards printed, populated, and tested.

In this report we present sentence recognition results as a function of spectral resolution and spectrally asynchrony by normal-hearing and cochlear-implant listeners. Sentence recognition was measured in six normal-hearing listeners with either full-spectrum or noise-band processors and five Nucleus-22 cochlear implant users with a 4-channel continuous interleaved sampler (CIS) processor. For full-spectrum processors, the speech signals were divided into either 4 or 16 frequency bands. A time delay was then added to the output of each band, varying the maximum delay across bands from 0-240 ms (in 40 ms steps). Within each delay condition, delays across bands were generated systematically to ensure a maximum delay between adjacent spectral bands while avoiding local pockets of channel synchrony. For noise-band processors, after band-pass filtering into 4 or 16 bands, the envelope of each channel was extracted and used to modulate noise of the same bandwidth as the analysis band, thus eliminating the spectral fine structure available in the full-spectrum processors. For 4-channel CIS processors, the amplitude envelopes extracted from 4 bands were transformed to electric currents by a power function and the resulting electric currents were used to modulate pulse trains delivered to four electrode pairs. Results show no significant difference between the 4- and 16-band full-spectrum speech processors for normal-hearing listeners. Scores dropped significantly only when the maximum delay reached 200 ms for 4-channel processor and 240 ms for 16-channel processor. When spectral fine structure was removed in the noise-band processors, sentence recognition dropped significantly when the maximum delay was 160 ms for the 16-band processor and 40 ms for the 4-band processor. There was no significant difference between implant listeners using the 4-channel CIS processor and normal-hearing listeners using the 4-channel noise-band processors. The results imply that when spectral fine structure is not available, as in the implant listener's case, increased spectral resolution is very important in overcoming any spectral asynchrony in speech signals.

In the next quarter we will present some of our experimental results at the International Symposium on Hearing, in the Netherlands (August 4-9) and at the International Hearing Aid Research Conference at Lake Tahoe (IHCON 2000, August 23-29).

CI Research Interface Development

Nucleus-22 and -24 SEMA Research Interfaces

We have designed and constructed a research interface for the Nucleus-22 and –24 implants based on a Motorola 56k series DSP processor. Each pulse is specified by a packet of 8 words, which are transmitted to the DSP from the PC via a high-speed (enhanced) parallel port. The existing prototype is based on an evaluation model (EVM-DSP) from Domain Technologies. Because this EVM contains additional hardware elements that are not necessary for the present application, we designed a new printed circuit board that contains only the chips that are necessary for the interface. The prototype interface hardware and software are working and in the last quarter we began extensive validation testing on the interface to insure that the output of the device was exactly as specified in the software. At the present time the DSP code is only able to deliver the SEMA (sync-electrode-mode-amplitude) code sequence. Once the SEMA transmission is fully operational and validated we will program the new embedded transmission protocol, which will allow higher pulse rates. The same PC and DSP software allows stimulation of the Nucleus-22 system (2.5 MHz) and the Nucleus-24 system (5 MHz). The clock rate is selectable with a software switch.

To validate the output of the interface we connect the coil output to an “implant-in-a-box”, which is a model of the implanted receiver/stimulator in a plastic housing. We have implants-in-a-box for both the Nucleus 22 and Nucleus 24 systems. We calibrated these devices as a preliminary stage of the validation. Simulated electrode impedance loads can be placed at the output of the implant-in-a-box and the actual voltage and current measured on an oscilloscope. We established a testing protocol that measures pulse current amplitude, pulse phase duration, and pulse rate as a function of electrode pair. The protocol varies each of these parameters over the full range of possible values while holding other parameters at typical standard values (electrode pair (10,12), 316 µA, 200 µs/phase, 20 µs interphase duration, 500 pps, 200 ms burst duration). The test protocol includes measuring the breakdown points for each parameter by increasing (or decreasing) the parameter value until the output fails. Combinations of extreme parameter values are tested as appropriate.

Results from the first stage of testing show that the interface produces the desired output specified in software within the following working ranges:

1. For active electrodes 1-22 for Monopolar 1 mode (Nuc-24) and selected electrodes (1,10,20) in BP+1 mode for Nuc-22 and Nuc-24
2. Pulse phase duration: 10 us to 1000 us/phase
3. Pulse Rate*: 20-4000 pps for Nuc-22, 20-8000 pps for Nuc-24
4. Current levels: measured 10-20 uA to more than 1880 uA for the range of device-specific amplitude “units” (1-238 and 1-255)

Due to the multitasking nature of the Windows operating system, we also plan to test the interface for failure due to multiple Windows interrupts. Specifically, the protocol calls for at least one application requiring frequent hard disk access running in the background while the test application delivers data to the interface to produce the highest stimulation rate. While this is a later stage of the protocol, our work on the current stage with no Windows programs running in the background shows no failure

for large pulse trains at fairly high rates (4096 pulses at 4 kHz), leading us to believe that our software and hardware buffering strategies are working properly.

Clarion S-2 Research Interface

We are working with Advanced Bionics Corporation (ABC) to develop a research interface for the next generation of Clarion cochlear implants (S-2). The software microcode, firmware, and operating system are being defined. The development of the Clarion Research Interface (CRI) is not funded by this contract.

Experiment Report: Spectral Asynchrony

INTRODUCTION

It has been widely assumed that a detailed auditory analysis of the short-term acoustic spectrum is essential for understanding spoken language. However, results from several perceptual experiments with normal-hearing adults indicated that a detailed spectral-temporal analysis of the speech signal might not be required. A study by Remez *et al.* (1983) showed that phonetic information can be conveyed by sinewave replicas of speech signals. In their study, the tonal patterns were made of three sinusoids whose frequency and amplitude were equal to the respective peaks of the first three formants of natural-speech utterances. Unlike natural and most synthetic speech, the spectrum of sinewave speech contains neither harmonics nor broadband formants, and can be described as sounding grossly unnatural in voice timbre. Despite the marked alteration of the short-time speech spectrum that disrupted the spectro-temporal properties, listeners were able to perceive the phonetic content. They argued that phonetic perception might then depend on properties of coherent spectrum variation (a second-order property of the acoustic signal) rather than any particular set of acoustic elements present in speech signals.

Warren *et al.* (1995) measured the intelligibility of simple sentences when heard through narrow spectral slits. They found that very little spectral information was required to identify the key words in the "everyday speech" sentences. Near-perfect intelligibility was obtained for a single 1/3-octave band with a center frequency in the vicinity of 1500 Hz. Greenberg et al. (1998) also investigated the contribution of specific spectral bands to speech intelligibility. In their study, each sentence was spectrally partitioned into 14 1/3-octave bands ("slits") and the stimulus for any single presentation consisted of four spectral slits presented concurrently. The passbands of the four spectral slits were 298-375 Hz, 750-945 Hz, 1890-2381 Hz, and 4762-6000 Hz, respectively. They found that this sparse spectral representation was sufficient for accurate identification of the majority of words in spoken sentences, at least under ideal listening situations. They argued that a detailed spectral-temporal analysis of the speech signal was not required to understand spoken language. A more likely basis for their speech perception results was the amplitude and phase components of the modulation spectrum distributed across the frequency spectrum.

Recently, Shannon *et al.* (1995) measured speech recognition performance as a function of spectral resolution. In their approach, speech was divided into several frequency bands. The temporal envelope was extracted by half-wave rectification and low-pass filtering and then modulated with noise that was spectrally shaped by the

same filters as those used in the analysis bands. The spectral resolution was systematically changed by manipulating the number of bands. Their results showed near-perfect sentence recognition when only four frequency bands were available, suggesting that speech provides enough redundancy in the acoustic spectrum to overcome considerable spectral degradation. Similar results have been reported by other studies (Fu *et al.*, 1998; Dorman *et al.*, 1998).

The necessity of a detailed auditory analysis of the short-term acoustic spectrum has also been challenged by studies involving cross-channel spectrally asynchronous speech, a condition reminiscent of acoustic reverberation. Greenberg and his colleagues (Greenberg *et al.*, 1998; Arai and Greenberg, 1998) measured speech intelligibility in the presence of cross-channel spectral asynchrony. In their study, the spectrum of speech signals was partitioned into 1/4-octave channels and the onset of each channel shifted in time relative to the others so as to desynchronize spectral information across frequency. (NOTE: A visual analogy of this condition is presented in the pixilated picture of Abraham Lincoln on the front page of this progress report, which was “desynchronized” by adding a random horizontal displacement to each line of pixels.) In this case, although high spectral resolution is preserved, the significant alteration in cross-channel spectral synchrony may disrupt the decoding process and thereby degrade the speech intelligibility. They found that speech intelligibility was highly tolerant of cross-channel spectral asynchrony when the full spectrum was available. However, when only four of the narrow spectral slits were available, intelligibility was seriously degraded when the slits were desynchronized by more than 25ms. The four slits presented synchronously provided high levels of sentence recognition, suggesting that the reduced spectral information was not the limiting factor in speech recognition. They argued that the amplitude and phase components of the modulation spectrum were highly important when listening to speech with limited spectral resolution, and that listeners' sensitivity to the modulation phase was generally "masked" by the redundancy contained in full-spectrum speech.

These results indicate that although a detailed auditory analysis of the short-term acoustic spectrum might not be required to understand speech, the redundancy contained in full-spectrum speech plays a vital role in understanding spectrally distorted (e.g. asynchronous) speech. Unfortunately, full-spectrum speech may not be available for some listeners, such as cochlear implant users.

Modern multi-channel cochlear implant devices divide speech sounds 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 electrodes located in the different places within the cochlea. To recreate the tonotopic distribution of activity within the normal cochlea, the envelope cues from low frequency bands are delivered to electrodes located near the apex and the envelope cues from high frequency bands are delivered to basal electrodes.

Although this approach can preserve the temporal envelope within each frequency band or the approximate spectral envelope across frequency bands, the detailed fine structure inherent in each band is lost. Even without the fine spectral cues, many cochlear implant users still achieve high levels of speech recognition. Similar results have also been reported in normal-hearing subjects listening to spectrally degraded speech with noise-band speech processors (Shannon *et al.*, 1995; Fu *et al.*, 1998; Dorman *et al.*, 1998). However, many cochlear implant users have difficulty

understanding speech in adverse listening environments. Noisy or reverberant environments are made more challenging because of the limited spectral resolution of the implant device. Results from previous studies confirm that the limited spectral resolution in cochlear implants is the major factor causing a rapid deterioration of speech recognition in noisy environments (Fu *et al.*, 1998; Dorman *et al.*, 1998).

While noisy environments or multiple-talker listening situations are one aspect of the challenging listening conditions which implant users regularly face, reverberant environments also present implant listeners with difficulty in speech recognition. In noisy or multiple-talker situations, implant listeners make use of the spectral resolution available to them to distinguish speech from noise or one talker from many. In reverberant environments, implant listeners use this limited spectral resolution to reconstruct a speech signal whose spectro-temporal properties have been desynchronized. Little is known about the effect of spectral asynchrony on speech recognition when the spectral resolution of speech signals is severely limited.

The present study investigates the effect of cross-channel spectral asynchrony on speech recognition with the absence of fine spectral cues. Both normal-hearing listeners and cochlear implant users participated in the present experiment. Sentence recognition was measured as a function of the amount of cross-channel asynchrony in six normal-hearing listeners with full-spectrum or noise-band processors (Shannon et al., 1995) and five cochlear implant users using a 4-channel speech processor with continuous interleaved sampler (CIS) strategy.

METHODS

Subjects

Six normal-hearing (NH) listeners aged 25 to 35 and five Nucleus-22 cochlear implant (CI) users aged 40 to 60 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 implant subjects had at least five years experience utilizing the SPEAK speech processing strategy (Nucleus-22 device) and all were native speakers of American English. The Nucleus-22 speech processor with the SPEAK strategy divides an input acoustic signal into 20 frequency bands, extracts the amplitude envelope from each band, and stimulates the electrodes corresponding to the 6 to 10 bands with the maximum amplitudes (Seligman and McDermott, 1993). The frequency allocation table specifies the acoustic frequency range covered by the speech processor. Three subjects (N4, N7, and N9) used frequency allocation table 9 (150-10,823 Hz) in their clinical implant processor, and two subjects (N3 and N19) used frequency allocation table 7 (120 Hz - 8,658 Hz). All implant subjects had 20 active electrodes available for use. Table 1 contains relevant information for the five CI subjects. All subjects were paid for their efforts and all provided informed consent before proceeding with the experiment.

Subject	Age	Gender	Cause of Deafness	Duration of use	Insertion Depth	Frequency Table
N3	55	M	Trauma	6 years	3 rings out	7
N4	39	M	Trauma	4 years	4 rings out	9
N7	54	M	Unknown	4 years	0 rings out	9
N9	55	F	Hereditary	7 years	4 rings out	9
N19	68	M	Noise-Induced	3 years	6 rings out	7

Table 1: Subject information for three Nucleus-22 cochlear implant listeners who participated in the present study. Frequency table refers to the frequency allocation used by the listener in their clinically assigned processor. Frequency table 7 has a frequency range of 120 to 8658 Hz while frequency table 9 has a range of 150 to 10823 Hz. Frequency table 9 is intended to be an approximate tonotopic map to the electrode locations for a full electrode insertion. Insertion depth is reported as the number of stiffening rings outside the round window from the surgical report. A full insertion would be 0 rings out.

Test Materials and Procedure

Recognition of words in sentences was measured using the Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994). For HINT sentence recognition, a list was chosen randomly from among 26 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. The order of the seven time conditions were randomized and counterbalanced across subjects.

Stimuli

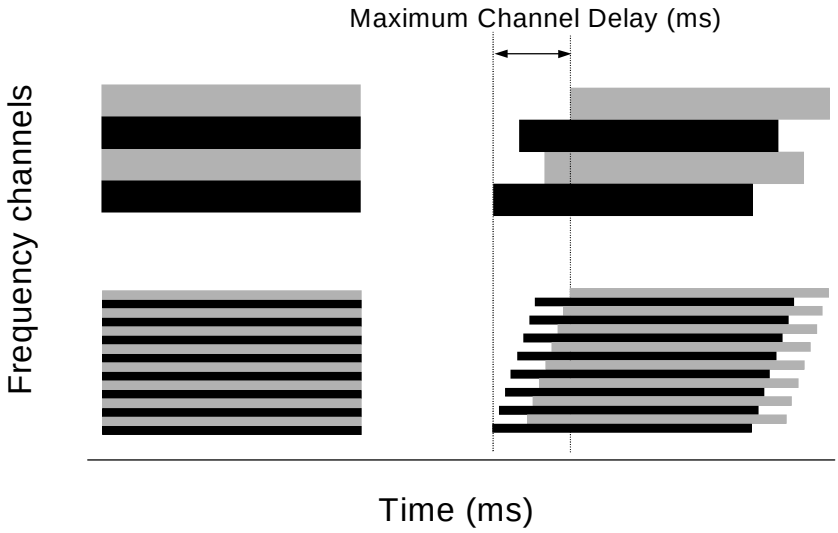
Two kinds of speech processors (full-spectrum and noise-band) were created for normal-hearing listeners in this experiment. For the full-spectrum processors, the speech signal was band-pass filtered into either four or sixteen frequency bands using 8^{th}-order Butterworth filters. The corner frequencies of the bands were 300 Hz, 713 Hz, 1509 Hz, 3043 Hz, and 6000 Hz for the 4-channel full-spectrum processor and at 300, 379, 473, 583, 713, 866, 1046, 1259, 1509, 1804, 2152, 2561, 3043, 3612, 4281, 5070, and 6000 Hz for the 16-channel full-spectrum processor. The output of each channel was then time-shifted, varying the maximum channel delay from 0-240 ms (in 40 ms steps). Channel delays were also generated to ensure a maximum delay between adjacent channels, thereby avoiding local regions of spectral synchrony. The delay sequences used to desynchronize channels of spectral information were generated as follows. For odd-numbered channels, the amount of delay for a particular channel can be represented as:

$$D_i = \frac{i-1}{N-1} \times \frac{D_{\max}}{2} \qquad \text{i = 1, 3, 5, … , N-1} \qquad (1)$$

where D_i is the amount of delay in i^{th} channel, D_{max} is the maximum channel delay, and N is the number of channels (either 4 or 16). Similarly, for even-numbered channels, the amount of delay for a particular channel can be represented as:

$$D_i = \frac{N + i - 2}{N - 1} \times \frac{D_{\max}}{2} \qquad \text{i = 2, 4, 6, … , N} \qquad (2)$$

For example, for a 4-channel processor with a maximum delay of 240 ms, the channel delays were 0, 160, 80, 240 for channel 1, 2, 3, 4, respectively. In this way, a minimum of 80 ms delay was placed between adjacent channels, with 160 ms between channels 1 and 2; these maximal delays between adjacent channels disrupted the formant transitions whose channel synchrony is important for speech recognition. Figure 1 shows the delay patterns across channels in both the 4-channel processor (top) and 16-channel processor (bottom). Figure 2 shows the waveform and spectrograph representation of the 4-channel full-spectrum processor and the 4-channel noise-band processor at 240 ms maximum delay condition. Note that the pixilated picture of Lincoln on the cover page of this QPR has been randomly "desynchronized" in spatial terms as a visual analogy of this manipulation.



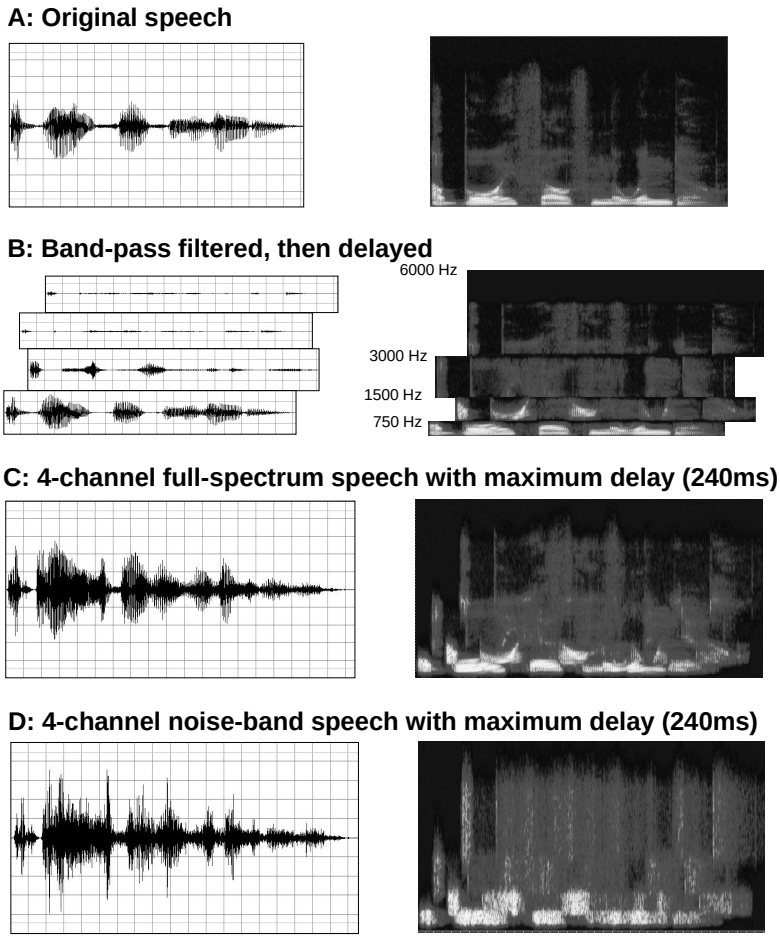


Figure 2

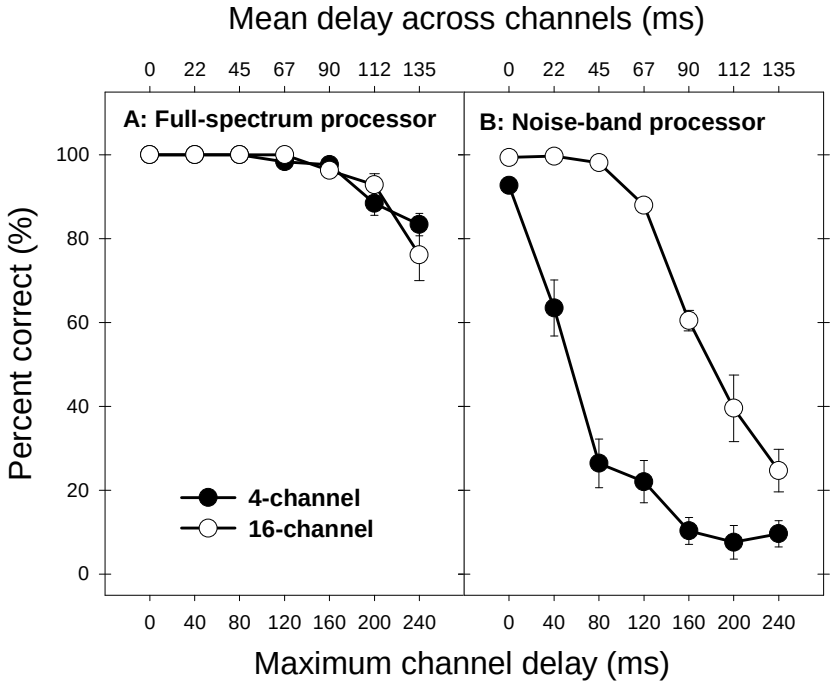
For the noise-band processors, after band-pass filtering into 4 or 16 channels, the envelope of each channel was extracted and used to modulate noise of the same bandwidth as the analysis band, thus eliminating the spectral fine structure available in the full-spectrum processors. The same method was

used to generate channel delay sequences as in the full-spectrum processors. To examine the frequency-specific effects of spectral asynchrony, one additional delay sequence was created for the 4-channel noise-band processor. In this condition, only one channel was delayed at 120 ms while the other three channels were not delayed.

For the implant listeners, the 4-channel CIS processor was implemented through a custom research interface (Shannon *et al.*, 1990), thereby bypassing the subject's Spectra-22 speech processor. The signal was first pre-emphasized using a first-order Butterworth high-pass filter with a cutoff frequency of 1200 Hz, and then band-pass filtered into four broad frequency bands using 8th-order Butterworth filters. The corner frequencies of the bands were 300 Hz, 713 Hz, 1509 Hz, 3043 Hz, and 6000 Hz. The envelope of the signal in each band was extracted by half-wave rectification and low-pass filtering at 160 Hz. The acoustic amplitude (40-dB range) was transformed into electric amplitude by a power-law function with an exponent of 0.2 ($E = A^{0.2}$; Fu and Shannon, 1998) between each subject's threshold (T-level) and upper level of loudness (C-level). This transformed amplitude was then used to modulate the amplitude of a continuous biphasic pulse train with a 100 □s/phase pulse duration, and delivered to four electrode pairs interleaved in time: (18,22), (13,17), (8,12), and (3,7). Note that a relatively broad stimulation mode (BP+3) was used in the present study because several subjects were unable to reach an upper level of loudness (C-level) on the apical electrode pairs with BP+1 stimulation mode. The same method was used to generate channel delay sequences as in the full-spectrum and noise-band processors. The additional delay sequence was also used to examine any frequency-specific effects of spectral asynchrony.

RESULTS

Figure 3 shows the percent of words in sentences correctly identified as a function of the maximum channel delay by normal-hearing listeners. Panel A shows the recognition scores for the full-spectrum processors. For the 16-channel full-spectrum processor, there was no significant drop in performance until the maximum channel delay was 240 ms; even at this extreme delay, average performance only dropped about 20 percentage points. For the 4-channel full-spectrum processor, there was no significant drop in performance until the maximum channel delay was 200 ms. Again, at this



extreme delay, average performance remained high at around 80% correct. A one-way, repeated-measures analysis of variance (ANOVA) showed a significant effect of spectral asynchrony on sentence recognition for both the 4-channel full-spectrum processor [$F(6,35)=17.71$, $p<0.001$] and the 16-channel full-spectrum processor [$F(6,35)=11.27$, $p<0.001$].

Figure 3B shows the recognition scores for the noise-band processors. For the 16-channel noise-band processor, performance dropped significantly with a maximum channel delay of 160 ms. Average performance continued to steadily decline to 24% correct at the 240-ms maximum channel delay. For the 4-channel noise-band processor, there was an immediate drop in performance, even with a maximum channel delay of only 40 ms. Average performance continued to drop quickly, nearing chance level at 160 ms. A one-way ANOVA showed a significant effect of spectral asynchrony on sentence recognition for both the 4-channel noise-band processor [$F(6,35)=51.73$, $p<0.001$] and the 16-channel noise-band processor [$F(6,35)=69.22$, $p<0.001$].

A two-way ANOVA revealed no significant interaction between the 4- and 16-channel full-spectrum processors [$F(6,70)=1.35$, $p=0.247$]. However, recognition scores between the 16-channel full-spectrum and noise-band processors began to significantly differ when the maximum channel delay was 120 ms. The difference is even more evident when comparing the 4-channel full-spectrum and noise-band processors. The upper limit of performance with the 4-channel noise-band processor was at 0 ms maximum channel delay; here, average performance dropped only 7 percentage points from the other conditions. While the 4-channel full-spectrum processor performance remained high even at 240 ms, the 4-channel noise-band processor performance was immediately and severely degraded at only 40 ms. A two-way ANOVA revealed a significant interaction between 4- and 16-channel noise-band processors [$F(6,70)=17.28$, $p<0.001$], as well as a significant interaction between 4-channel full-spectrum and noise-band processors [$F(6,70)=37.01$, $p<0.001$] and 16-channel [$F(6,70)=27.57$, $p<0.001$].

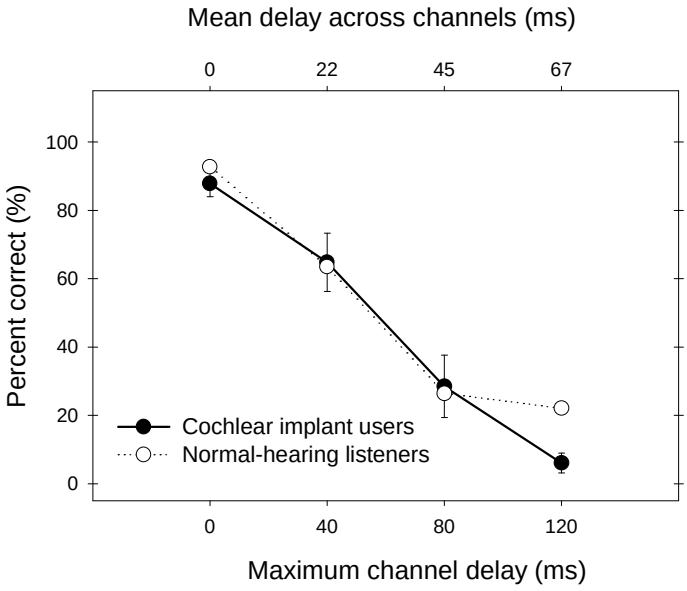
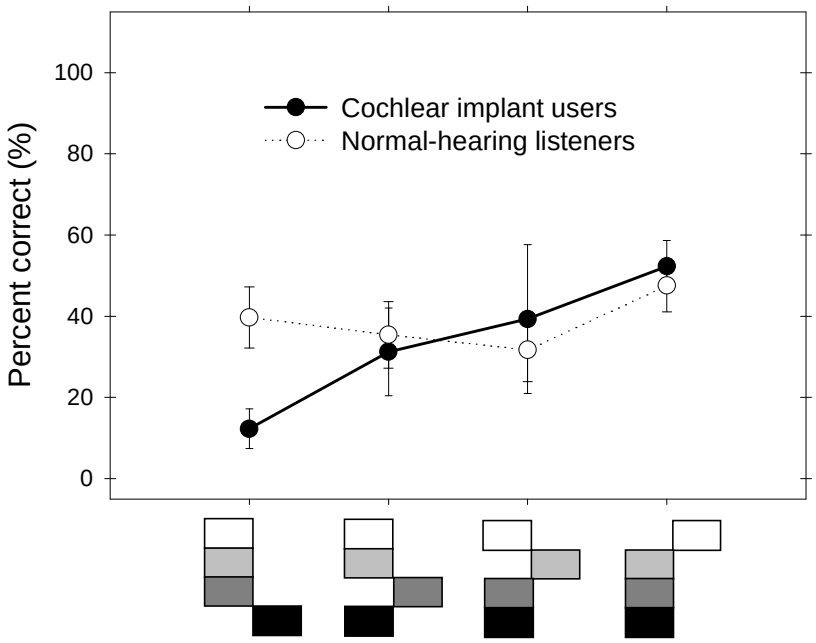


Figure 4 shows the percent of words in sentences correctly identified by cochlear implant listeners using the 4-channel CIS processor and normal-hearing listeners using the 4-channel noise-band processors, as a function of the maximum channel delay. The solid line represents the mean recognition score from five cochlear implant listeners and the dotted line represents the mean recognition score from six normal-hearing listeners. Implant and normal-hearing performance was very comparable for these delay conditions. A one-way ANOVA revealed a significant effect of spectral asynchrony on sentence recognition for cochlear implant listeners [$F(3,16)=29.82$, $p<0.001$]. A two-way ANOVA also revealed no significant interaction between cochlear implant users with 4-channel CIS processor

and normal-hearing listeners with 4-channel noise-band processor [F(3,16)=1.01, p=0.402]. At 120 ms maximum delay, average implant listeners' performance was only 6% correct while normal-hearing listeners performed about 22% correct. A student t-test did reveal a slightly significant difference between normal-hearing listeners and cochlear implant users at this 120 ms maximum delay condition.

Figure 5 shows the results for frequency-specific conditions conducted with normal-hearing listeners using the 4-channel noise-band processor and cochlear implant listeners using the 4-channel CIS processor. The pictures under the axis tics show the delay sequence; the lowest bar represents the lowest frequency band. On average, delaying the highest frequency channel (3-6kHz) had the least effect, although performance did drop about 40 percentage points when compared to the 4-channel noise-band processor with no delay. While there is significant inter-subject variability in this data set, the overall pattern seems consistent across subjects. A one-way ANOVA showed no significant effect of delay sequence for both cochlear implant listeners [F(3,20)=2.66, p=0.084] and normal-hearing listeners [F(3,20)=0.984, p=0.42] as well as no significant interaction between cochlear implant users and normal-hearing listeners [F(3,36)=1.76, p=0.172]. A student t-test did reveal a slightly significant difference between normal-hearing listeners and cochlear implant users when delaying the lowest frequency channel.



Delay sequence (maximum channel delay=120 ms)

DISCUSSION

The present results clearly demonstrate that speech intelligibility of full-spectrum speech is highly tolerant of cross-channel spectral asynchrony, in agreement with previous studies' results (Arai and Greenberg, 1998). There are, however, slight differences between the present study's results and those of the previous studies. In the present study, average recognition scores were about 80% correct, even when the maximum channel delay reached 240 ms. However, in the previously cited studies, recognition scores were only about 75% correct for 140 ms of maximum channel delay; intelligibility was further reduced to 50% correct when the cross-channel asynchrony exceeded 200 ms of maximum channel delay. One explanation for this difference may be the different speech materials used in the two studies. In the present study,

recognition of words in sentences was measured using the Hearing in Noise Test (HINT) sentences (Nilsson et al., 1994); in the studies by Greenberg and his colleagues (Greenberg et al., 1998; Arai and Greenberg, 1998), the DARPA TIMIT acoustic-phonetic continuous speech corpus was used. The sentences are of easy-to-moderate difficulty for HINT and of moderate-to-hard difficulty for TIMIT. Besides a difference in sentence difficulty, the mean sentence duration for these two speech corpuses is also significantly different. The average duration per phoneme is about 100 ms for the HINT sentences, while the mean duration per phoneme for the TIMIT sentences is about 72 ms.

The loss of fine spectral cues had a dramatic effect on speech recognition in the presence of cross-channel asynchrony (even though the loss of these fine spectral cues had only a small effect at the 0 ms maximum delay condition). The difference in performance between the full-spectrum and noise-band processors indicates that the fine spectral cues might provide the redundant spectro-temporal information necessary for speech intelligibility in adverse listening environments such as noisy backgrounds or reverberant rooms. As shown in the figures, performance between the 16-channel full-spectrum and noise-band processors began to significantly differ at 120 ms of maximum channel delay, indicating that the spectro-temporal fine structure contributed greatly in overcoming spectral asynchrony in speech. This contribution is even more evident when comparing the 4-channel full spectrum and noise-band processors. While performance with the 4-channel full-spectrum processor remained high even at 240 ms of maximum channel delay, performance with the 4-channel noise-band processor was immediately and severely degraded at only 40 ms. Furthermore, when the spectral resolution is reduced, the fine harmonic structure becomes very important in overcoming cross-channel asynchrony. The contribution of increased spectral resolution is most apparent when comparing performance between the 4- and 16-channel noise-band processors. Again, the 4-channel noise-band processor showed an immediate decline in performance at only 40 ms of maximum channel delay. However, performance with the 16-channel noise-band processor began to drop significantly at the much longer 160 ms condition; at the extreme delay of 240 ms, average performance was about 30% correct, well above the chance level performance exhibited with the 4-channel noise-band processor.

The importance of modulation spectrum, as defined by Greenberg and colleagues in previous studies, is also put into question because of the marked differences between full-spectrum and noise-band processors. Theoretically, the modulation spectrum (between 3 and 6 Hz) should be exactly the same for full-spectrum and noise-band speech because the noise-band processor preserves all temporal envelope cues within each frequency band. If the intelligibility of speech depended on the integrity of the modulation spectrum as suggested (Arai and Greenberg, 1998), the difference between the full-spectrum and noise-band processors for these spectrally asynchronous conditions would be much smaller. These results suggest that the fine spectral cues, rather than the modulation spectrum, may provide the spectro-temporal redundancy within speech signals necessary for the high tolerance of cross-channel spectral asynchrony.

A remarkable similarity was observed between the results of normal-hearing listeners using the 4-channel noise-band processor and cochlear implant listeners using the 4-channel CIS processor. This indicates a common mechanism may underlie speech recognition in the presence of cross-channel asynchrony. This result also

suggests that the cochlear implant users are likely to be highly susceptible to cross-channel asynchrony due to the loss of fine spectral resolution. However, despite these similarities, there was a significant difference between normal-hearing and cochlear implant listeners. Figure 5 shows the effect of frequency-specific delays on speech recognition. For normal-hearing listeners, a delay of 120 ms in any one-frequency band had an equally detrimental effect on speech intelligibility. However, the lowest frequency band showed a much more detrimental effect on speech intelligibility for cochlear implant listeners than any of the other frequency channels. This drop in performance may have been caused by masking effects introduced by the delayed channel; the overall amplitude in the low-frequency band also may have been slightly too high, relative to the other three channels. This relatively high amplitude in the low-frequency band may not have any effect on speech intelligibility under ideal listening conditions; however, it may have a considerable effect in adverse listening environments.

SUMMARY AND CONCLUSION

A detailed auditory analysis of the short-term spectrum is not required to understand spoken language. A spectral representation that lacks fine spectro-temporal cues is sufficient for speech recognition in ideal listening conditions. However, the loss of fine spectral cues has a marked detrimental effect on speech intelligibility in the presence of cross-channel spectral asynchrony. The results indicate that the redundant information in speech signals may be contained in the fine spectro-temporal cues rather than the modulation spectrum; this redundancy within speech signals is important in overcoming adverse listening environments. The recognition pattern is remarkably similar between normal-hearing listeners using the 4-channel noise-band processor and cochlear implant listeners using the 4-channel CIS processor. The results imply that when the spectral fine structure is not available, as in the implant listener's case, increased spectral resolution is very important in overcoming any spectral asynchrony in speech signals.

REFERENCES

Arai, T. and Greenberg, S. (**1998**). "Speech intelligibility in the presence of cross-channel spectral asynchrony," IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, pp933-936.

Dorman, M.F., Loizou, P.C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583-3585.

Fu, Q.-J. and Shannon, R.V. (**1998**). "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **104**, 2570-2577.

Fu, Q.-J., Shannon, R.V., and Wang, X (**1998**). "Effects of noise and number of channels on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586-3596.

Greenberg, S., Arai, T. and Silipo, R. (**1998**). "Speech Intelligibility derived from exceedingly sparse spectral information," Proceedings of the International Conference of Spoken Language Processing, Sydney, December 1-4.

Seligman, P.M. and McDermott, H.J. (1995). “Architecture of the Spectra-22 speech processor”, Ann. Otol. Rhinol. Laryngol., 104, Suppl. 166, 139-141.

Nilsson, M, Soli, S.D., Sullivan, J.A. (**1994**). “Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am. **95**, 1085-1099.

Remez, R.E., Rubin, P.E., and Pisoni, D.B. (**1983**). "Coding of the speech spectrum in three time-varying sinusoids," Ann N Y Acad Sci. **405**, 485-489.

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.

Warren, R.M., Riener, K.R., Bashford, J.A., Jr., and Brubaker, B.S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Perception and Psychophysics **57**, 175-182.

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

Plans for the Next Quarter

In the next quarter (July-September 2000) we will continue hardware and software development on the Nucleus-24 and Clarion S-2 research interfaces. We anticipate that the SEMA protocol for the Nucleus-22 and -24 will be fully validated and debugged in the next quarter and will be integrated into our experimental stimulus delivery software. We have received rf transmitter parts from Cochlear Corp. and will design a PC board to contain the rf modulator/transmitter circuitry.

We will continue development of the research interface for the Clarion S-2 implant system (not funded by this contract). We anticipate that the software microcode, firmware, and operating system will be defined and partially implemented in the next quarter.

Wearable research processors (SPEAR) are not yet available from their developers at the University of Melbourne, but we hope that they will be available in the next quarter. These processors will be able to present coordinated binaural signals to two Nucleus-24 implants.

We will give an invited presentation at the NIH-sponsored AG Bell Association Research Symposium on Biotechnology and the Cochlea, two presentations at the International Symposium on Hearing in the Netherlands, August 4-9, and an invited presentation at the International Hearing Aid Conference (IHCON 2000) at Lake Tahoe August 23-27.

Experimental work in the next quarter will include:

1. Comparison of channel interaction measures and speech recognition for the original Clarion electrode, the original electrode plus the Positioner, and the new Hi-Focus electrode plus Positioner.

2. Measures of electrode interaction and speech recognition in Nucleus-24 patients with the original electrode and with the new Contour electrode.
3. Experimental manipulations that expand or compress the frequency-place mapping.
4. Measurements of the effect of stimulation rate on speech recognition with the Clarion and Nucleus-24 implants.
5. Psychophysical measures of temporal processing in good and poor implant users.

Publications and Presentations in this Quarter

Peer-Reviewed Publications:

Eisenberg, L., Shannon, R.V., Martinez, A.S., Wygonski, J., and Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age, Journal of the Acoustical Society of America, 107(5), 2704-2710.

Fu, Q.-J., and Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition, Asia Pacific Journal of Speech, Language, and Hearing, 5, 45-57.

Fu, Q.-J. and Shannon, R.V. (2000). Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users, Ear & Hearing, 21(3), 227-235.

Non-peer-reviewed Publications:

Baskent, D. and Shannon, R.V. (2000). Acoustic simulation of a spectral hole in cochlear implants, Proceedings of the Grodins Graduate Research Symposium, USC Biomedical Engineering Department.

Padilla, M. and Shannon, R.V. (2000). English phoneme and word recognition by nonnative English speakers as a function of English experience. Proceedings of the Grodins Graduate Research Symposium, USC Biomedical Engineering Department.

Submitted this Quarter:

Chatterjee, M., Shannon. R.V., Galvin, J.J. and Fu, Q.-J. (2001). Spread of excitation and its influence on auditory perception with cochlear implants, Physiological and Psychological bases of Auditory Function: Proceedings of the 12th International Symposium on Hearing, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven (Eds.), Shaker Publishing BV, Maastricht, NL.

Fu, Q-J. and Galvin, J. Spectrally asynchronous speech recognition by normal-hearing listeners and Nucleus-22 cochlear implant users, J. Acoust. Soc. Amer., submitted June 2000.

Fu, Q.-J., Galvin, J., and Wang, X. (2000). Time-altered sentence recognition by normal-hearing listeners and Nucleus-22 cochlear implant listeners, J. Acoust. Soc. Amer, Submitted June 2000.

Fu, Q.-J. and Shannon, R.V. Frequency mapping in cochlear implants, Ear and Hearing, submitted June 00.

Shannon, R.V. Fu, Q.-J., Wang, X., Galvin, J. and Wygonski, J. (2001). Critical cues for auditory pattern recognition in speech: Implications for cochlear implant speech processor design, Physiological and Psychological bases of Auditory Function: Proceedings of the 12th International Symposium on Hearing, A.J.M.

Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven (Eds.), Shaker Publishing BV, Maastricht, NL.

Shannon, R.V. Auditory Pattern recognition: Implications for hearing aids and cochlear implants, Proceedings of AG Bell Association Research Symposium on Biotechnology and the Cochlea, Philadelphia, July 9, 2000.

Invited Presentations:

Shannon, R.V. (2000). How the ear and brain work together to recognize speech: Lessons from cochlear implant research, Northeastern University, April 13-14, 2000. (Distinguished Neuroscience Lecturer)

Presentations:

Baskent, D. and Shannon, R.V. (2000). Acoustic simulation of a spectral hole in cochlear implants, Grodins Graduate Research Symposium, USC Biomedical Engineering Department, May 1.

Padilla, M. and Shannon, R.V. (2000). English phoneme and word recognition by nonnative English speakers as a function of English experience. Grodins Graduate Research Symposium, USC Biomedical Engineering Department, May 1.