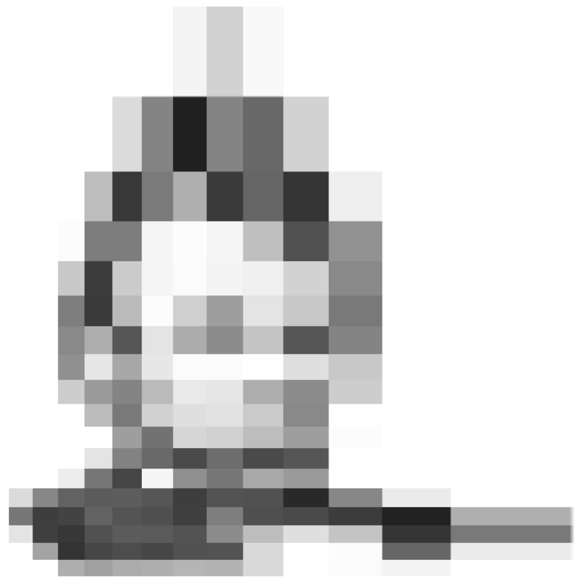
Speech Processors for Auditory Prostheses

NIH Contract NO1-DC-92100

QPR #5: Jan-Feb-Mar 2000



Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants

Submitted by

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

28 April 2000

Table of Contents: **PAGE**

ABSTRACT

In this quarterly progress report, we report on hardware and software development for laboratory interfaces for the Clarion and Nucleus-24 implant systems, and present the final results from an experiment on speech recognition as a function of the number of channels in noise.

A custom DSP interface has been developed to send user-specified sequences of pulses to patients with the Nucleus-22 and Nucleus-24 implants. The DSP code and basic PC software library has been developed to implement the SEMA (Sync-Electrode-Mode-Amplitude) transmission protocol in both the 2.5 MHz Nucleus-22 device and the 5 MHz Nucleus-24 device. Future software development will include implementation of the embedded transmission protocol. Prior to freezing the hardware design we evaluate the possibility of including the capability of recording the back-telemetry values for CAP measurement.

We are presently developing the hardware and software for a research interface for the next generation Clarion implant in collaboration with Advanced Bionics Corp.

A new portable speech processor for the Nucleus-24 device has been developed by the University of Melbourne. They have agreed to make these research processors available to other research groups. Since the portable system uses a Motorola 56309 DSP chip, much of our laboratory signal processing code should be compatible.

A major experiment was recently completed to investigate the differences between the SPEAK, CIS, and SAS speech processing strategies. These three strategies were implemented in the Nucleus-22 and Clarion commercial implant devices. The number of electrodes was varied as the independent variable in each strategy from 2 to the maximum number available. Recognition of vowels, consonants, monosyllable words, and sentences was measured as a function of the number of electrodes and as a function of the speech-to-noise ratio. Results show no significant differences across the three processing strategies for any condition. This equivalence suggests that the large differences between commercial devices and speech processing strategies were not important in terms of speech recognition. Speech recognition was strongly effected by the number of electrodes used in the processors, increasing with the number of electrodes up to 7 or 8, and did not increase as the number of electrodes was increased above 8. Performance of the best cochlear implant patients were similar to normal-hearing listeners with the same signal processing (up to 8 electrodes). This suggests that all three signal processing strategies are capable of conveying a similar amount of speech information and that some implant listeners are capable of using that information optimally. The main results were (1) the number of electrodes was the main factor determining performance, (2) there were no significant differences between SPEAK, CIS, and SAS processing strategies, or between Nucleus and Clarion devices, and (3) the best implant performance was similar to that of normal hearing listeners with the same processing.

Cochlear Implant Hardware Interface Development

Nucleus-22 Interface

As reported previously, we have developed a custom interface for the Nucleus-22 cochlear implant based on a Motorola 56309 DSP board. This interface accepts command streams of bytes from an extended parallel port (EPP) and generates the appropriate modulation on a 2.5 MHz carrier in SEMA transmission protocol. A software library has been completed to interface with PC programs in the form of a DLL module for Windows 95 and 98 programs and in the form of a LIB module for DOS programs. The hardware and software has been checked and debugged and is ready for integration with existing experimental software in the lab. The output of the interface has been checked with a "implant-in-a-box" to ensure that the appropriate current amplitude, pulse duration and electrode selection is achieved as specified. A full quality control check of the full range of parameters will be done in the next quarter.

Nucleus-24 SEMA Interface

The same 56309-based hardware interface described above is capable of sending commands in SEMA mode to the Nucleus-24 device. Under software control the rf modulation frequency can be switched from 2.5 to 5 MHz. The hardware and software have been checked, and a full quality control check of the full range of parameters will be done in the next quarter.

Nucleus-24 Embedded Protocol

The same hardware developed for the Nucleus-22 and Nucleus-24 SEMA transmission protocol is capable of transmitting higher pulse rates in the "Embedded Protocol (EP)". Software to implement the EP will be developed in future quarters. The priority will be to integrate the SEMA protocol into the existing laboratory software and to produce multiple versions of the hardware. The SEMA protocol at 5 MHz is capable of pulse rates up to 8-9 kHz, which should be adequate for initial experiments on pulse rate. A printed circuit board will be developed to complement the Motorola 56309 evaluation module (56k EVM). This board will contain the logic, power supply and rf circuitry to complement the 56k EVM board. Emphasis in the next quarter will be to finalize the printed board layout and produce multiple interface hardware units.

Clarion Research Interfaces

The Clarion Research Interface (CRI) is fully operational and we expect the initial experiments to begin in the next quarter. We have initiated design of a hardware/software interface for the next generation Clarion implant (CRI-2), in collaboration with Advanced Bionics Corp. This development is still in an early stage. We anticipate that considerable development effort will be made on this interface in the next quarter, although that effort will not be billed as part of this contract.

Portable Speech Processors

We had been planning to package our 56k EVM interface for portable use, but developments in the last quarter have changed those plans. Researchers at the University of Melbourne, Australia and the Cooperative Research Center(CRC) have developed a portable implant processor based on a Motorola 56309 DSP chip. This processor is packaged in the existing SPRINT housing, and so is highly portable and cosmetically acceptable. Since our existing interface already uses a 56309 DSP, our software should be easily modified to run on this platform. The University of Melbourne is willing to sell these portable processors to qualified research groups and the tentative date of availability is in the next quarter. The next generation of Clarion implants will also be programmable by research groups under the direction of the company as well. The availability of portable processors that are researcher programmable for these two implant devices, which make up the vast majority of implants worldwide, makes it uneconomical to develop our own portable processors.

Experiment Report: Speech Recognition in Noise as a Function of the Number of Spectral Channels

Introduction

Previous work with cochlear implants has demonstrated that speech recognition dramatically increases with increasing number of electrodes (Fishman *et al.,* 1997). Fishman *et al.* demonstrated that, in quiet listening conditions, recognition increased up to 4-7 electrodes, but no significant improvement in performance was observed when the number of electrodes was increased beyond seven. However, most everyday listening situations contain background noise, which can reduce intelligibility even for individuals with normal hearing. In noisy listening conditions even 16 spectral channels do not allow the same performance level as full-spectrum speech in normal-hearing listeners (Fu *et al.*, 1998). The present study measured speech recognition in noise for normal-hearing listeners as a function of the number of spectral channels and in cochlear implant listeners as a function of the number of electrodes used in their speech processors.

Studies with acoustic hearing have demonstrated that speech recognition is reduced when the spectral resolution is degraded by spectral smearing or hearing impairment (Stelmachowicz *et al.*, 1985; Dubno and Dorman, 1987; Horst, 1987; ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993, 1994; Turner and Henn, 1989' Turner *et al.*, 1999; Shannon *et al.*, 1995; Boothroyd *et al.*, 1996; Nejime and Moore, 1997). In general, these studies found that speech recognition in quiet listening conditions was quite resistant to spectral smearing, with significant decreases in performance occurring only when the spectrum was smeared over 1000 Hz, or reduced to less than 4 spectral channels. Speech recognition was more susceptible to spectral smearing in the presence of added noise.

Studies with cochlear implants (Lawson *et al.*, 1993, 1996; Brill *et al.*, 1997; Fishman *et al.*, 1997; Dorman and Loizou, 1997, 1998, Dorman *et al.*, 1997; Fu *et al*., 1998) have demonstrated that speech recognition improves with the number of electrodes in quiet listening conditions, at least up to 7 or 8 electrodes. Fu *et al.* (1998) measured recognition of vowels and consonants as a function of signal-to-noise ratio in three cochlear implant listeners and in four normal-hearing listeners in conditions simulating cochlear implants with both CIS and SPEAK-like strategies. Recognition scores for vowels and consonants decreased as the S/N level worsened in all conditions. Recognition of vowels and consonants was further measured in Nucleus-22 cochlear implant users using either their normal SPEAK speech processor or a custom processor with a four-channel CIS strategy. The best cochlear implant users showed similar performance with the CIS strategy in quiet and in noise to that of normal-hearing listeners when listening to correspondingly spectrally degraded speech, suggesting that the noise susceptibility of cochlear implant users is at least partly due to the limited spectral resolution.

Several studies have examined the effect of noise on speech recognition in cochlear implants (Dowell *et al.*, 1987; Hochberg *et al.*, 1992; Keifer *et al.*, 1996; M ller-Deiler *et al.*, 1995; Skinner *et al*., 1994). These studies all used the full number of electrodes available in the implant and the standard speech processing strategies. In general, findings indicate that speech processing strategies that produce higher levels of performance in quiet also allow better performance in noise. Several simple techniques for pre-processing the signal prior to the speech processing have also been shown to improve performance in noise (Weiss, 1993; Fu and Shannon, 1999b; Kompis and Dillier, 1994; van Hoesel and Clark, 1995a,b).

In the present experiment, speech recognition was measured as a function of the number of electrodes in various levels of noise for three processing strategies: SPEAK, CIS, and SAS. Speech recognition was also measured in normal-hearing listeners with noise-band processors (Shannon et al., 1995) as a function of the number of bands.

METHODS

<u>Listeners</u>

Ten adults (18 years and older) utilizing the Nucleus 22 cochlear implant with the SPEAK speech processing strategy and 9 adults using the Clarion cochlear implant device, each having at least six months experience, participated in this study. Five of the Clarion patients used the continuous interleaved sampler (CIS) processor (Wilson *et al*, 1991) and four used the simultaneous analog stimulation (SAS) processor (Eddington *et al.*, 1978; Battmer *et al.*, 1999). All implant subjects were postlingually deafened and native speakers of American English. General demographic information for the 19 subjects is presented in Table 1 and Table 2. All Nucleus 22 listeners had 20 active electrodes available for use, while Clarion users had either 7 or eight, depending on the speech processing strategy used: SAS users had 7 available electrode pairs and CIS users had 8 available electrode pairs. Five normal-hearing listeners, ranging in age from 25 to 53 years, were recruited as controls.

Table 1: Nucleus 22 Listeners

Listener	Age	Gender	CI Ear	Etiology	Age of HL Onset		Age of Profound HL Onset		Hearing Aid Usage		CI Use
					L	R	L	R	L	R	(years)
N3	56	M	R	Trauma	45	10	45	45	N	N	7
N4	40	M	R	Trauma	35	35	35	35	N	N	5
N5	81	M	R	Meningitis	52	52	52	52	N	N	7
N6	65	F	R	Ototoxicity	54	54	54	54	Y	Y	7
N7	55	M	R	Unknown	20	20	47	44	Y	N	2
N9	55	F	L	Hereditary	8	8	38	38	Y	Y	7
N14	63	M	R	Unknown	37	37	47	61	N	Y	1
N17	71	F	R	Unknown	41	41	68	68	Y	Y	1
N18	77	F	R	Otosclerosis	40	40	45	45	Y	Y	1
N19	7	M	L	Unknown	40	40	62	56	Y	N	1

Table 2. Clarion Listeners.

Sub	Speech Process Strategy	Age (Yrs)	Gen	CI Ear	Etiology	Age Onset of HL		Age Onset Profound		Hearing Aid Usage		Duration of CI Use
						L	R	L	R	L	R	(yrs)
C1	CIS	66	F	L	Otosclerosis	32	32	45	45	Y	N	1
C3	CIS	56	M	R	Trauma/Unk	18	0	18	45	N	N	3
C4	CIS	51	F	L	Meningitis	1.5	1.5	47	47	Y	N	2
C5	CIS	38	M	L	Unknown	3	3	28	22	Y	Y	2.5
C9	CIS	46	F	R	Ototoxicity	43	43	45	45	Y	Y	0.5
C2	SAS	72	M	R	C. Otoscler.	30	30	69	69	Y	N	2
C6	SAS	61	F	R	Menieres	22	33	57	57	Y	Y	1
C7	SAS	82	M	R	Unknown	15	15	63	63	N	Y	2
C8	SAS	76	M	R	Unknown	18	64	75	64	Y	N	0.5

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, monosyllable word recognition and sentence recognition.

Vowel stimuli were taken from materials recorded by Hillenbrand *et al.* (1995) and were presented to the listeners with the Auditory Implant Perception System Software - Identification (1997). Ten presentations (5 male and 5 female talkers) each of twelve medial vowels including 10 monophthongs (/i I E Q u U A Ã • ÎÕ/) and 2 diphthongs (/o e/) were 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 (3 male and 3 female talkers in a /a/C/a/ context) were taken from materials created by Turner *et al.* (1999) and Fu *et al.* (1998). Consonant confusion matrices were compiled from 12 presentations of each of 14 medial consonants /b d g p t k m n f s ʃ v z θ/. Tokens were presented in random order by custom software (Robert, 1997; Shannon *et al.*, 1999) and the confusion matrices were analyzed for percent correct on the production based categories of voicing, manner, and place of articulation. Chance performance level for this test was 7.14% correct, and the 95% confidence level was 10% correct.

The CNC Word Test from the Minimum Speech Test Battery for Adult Cochlear Implant Users CD was used to evaluate open-set phoneme and word recognition (House Ear Institute and Cochlear Corporation, 1996). The CD contains 10 lists of 50 monosyllabic words containing 150 phonemes. Listener responses were scored separately for words and phonemes correctly identified. Because there were more test conditions (25) than lists of words (10), the word lists used in the conditions with the poorest scores were repeated.

Recognition of words in sentences was measured using the Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994) from the Minimum Speech Test Battery for Adult Cochlear Implant Users CD (House Ear Institute and Cochlear Corporation, 1996). For each condition, data was collected for 10 sentences of varying lengths from each listener. The sentences were of easy-to-moderate difficulty, presented with no context and no feedback, and no sentences were repeated to an individual listener.

Experimental speech processor conditions

Each listener was tested with 5 experimental speech processors immediately after receiving them (no practice). Each of the 5 experimental processors was tested in quiet and with four different signal-to-noise ratios (S/N) of +15, +10, +5, and 0 for a total of 25 conditions. The Nucleus-22 SPEAK processing strategy divides speech into 20 contiguous frequency bands and normally assigns the output of each band to one electrode pair. The listeners' original frequency band divisions were used. In the present experiment, processors were created with 2, 4, 7, 10, and 20 activated electrodes by assigning the output of more than one band to a single electrode pair. In the normal 20-electrode processor the output of analysis bands 1, 2, 3, 4, and 5 would normally be assigned to active electrodes 1, 2, 3, 4, and 5, respectively.

In the present 4-electrode experimental processor the outputs of all five bands were assigned to active electrode 3 only. In this case active electrodes 1, 2, 4, and 5 received no stimulation. When this assignment pattern was repeated along the entire electrode array the outputs of the 20 analysis filters were presented to only 4 active electrodes. In similar fashion, analysis filters were summed to create processors with 10, 7, 4, and 2 active electrodes. In the 7-electrode condition the basal-most electrode pair was assigned only 2 frequency bands instead of 3. (see Fishman *et al.,* 1997 for more details)

In the normal SPEAK processing strategy the acoustic signal is analyzed into 20 frequency bands and between six and ten frequency bands with the highest energy are selected for stimulation approximately every 4 msec (McDermott *et al.*, 1992; Seligman and McDermott, 1995). The electrodes assigned to those bands receive pulses. The average pulse rate per electrode was higher in the experimental processors, because the activated electrodes received the output from more than one analysis filter band. For example, if an electrode pair was assigned to receive the output of three contiguous analysis bands (7-electrode processor condition) then that electrode pair received a stimulation pulse if any of the three filter bands was selected for stimulation. If all three filter bands were selected for stimulation, the electrode pair would receive three pulses in that stimulus frame. Thus, as the number of electrodes was reduced, the effective stimulation pulse rate on each electrode pair was increased. “Stimulus level” coding was used, which changes the electrical stimulation level by changing both pulse amplitude and pulse phase duration (Cochlear Corp., 1995). At high loudness levels the pulse duration is longer, which results in a lower overall pulse rate.

With the Clarion SAS and CIS speech processing strategies, 7 or 8 frequency bands are normally directed to 7 or 8 electrode pairs (Clarion by Advanced Bionics, 1998). With a reduction in the number of electrode pairs, the total frequency range remains the same, but the range for each electrode is broadened, with the exception of the 2-electrode processor. With the two-channel processor only high and low frequency bands are transmitted, and the mid-frequency information is left out (Breeuwer and Plomp, 1984). Five electrode conditions were created where all 7 or 8 electrode pairs were utilized initially and then reduced to, 6, 4, 3 and 2 pairs.

Normal hearing listeners were tested on the same materials on conditions with CIS-like processing (see Shannon *et al.*, 1995). Acoustic processors were designed to divide the speech spectrum from 100-6kHz into tonotopic bands of equal width in mm, according to the cochlear tonotopic formula of Greenwood (1990). The envelope was extracted from each band by half-wave rectification and low-pass filtering at 160 Hz. This envelope signal was then used to modulate band-pass filtered white noise, filtered with the same filter set as was used on the original speech signal. The modulated noise bands were then summed and presented via a calibrated loudspeaker in a sound treated room (IAC). The speech-shaped masking noise was added to the speech signals at the desired speech-to-noise level prior to processing.

Procedure

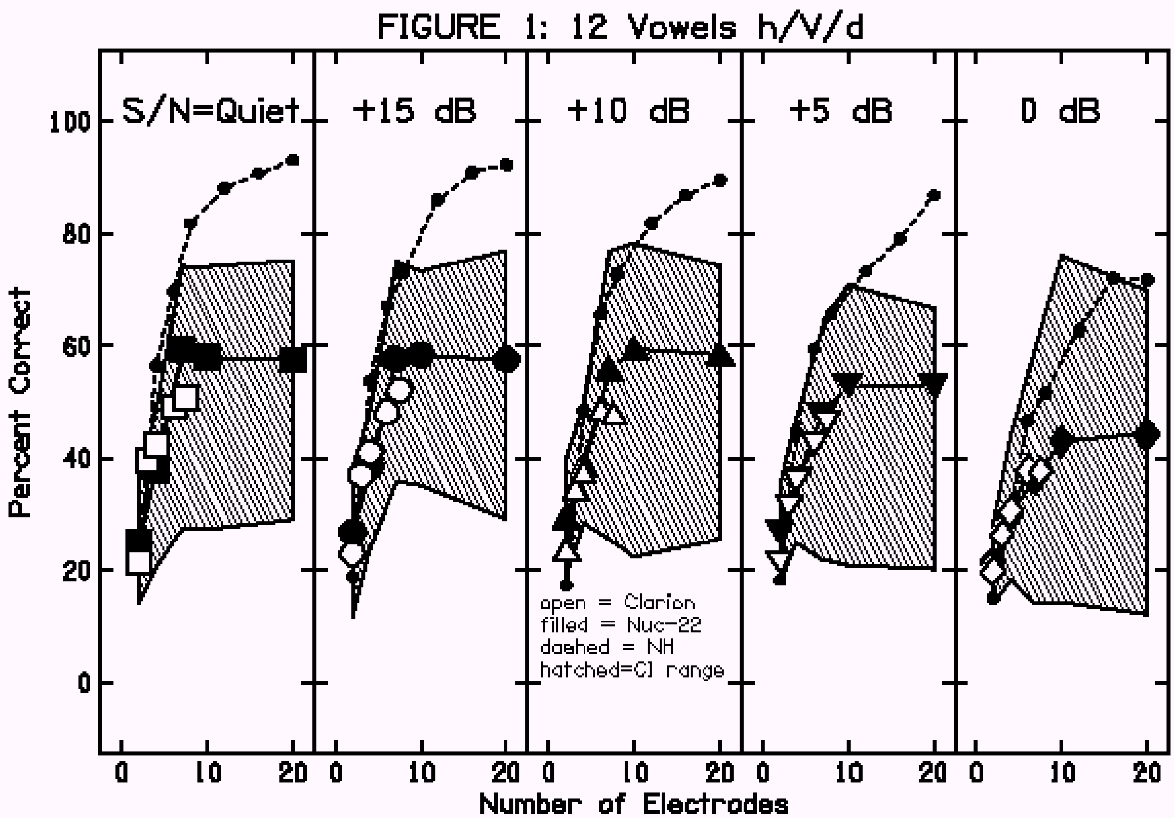
During all testing the listener was seated one meter in front of a loudspeaker (Grason-Stadler audio monitors) in a sound treated room (IAC). The presentation level was 65 dB SPL for all speech perception testing, as measured by a B&K one inch microphone (Model 4144) at the location of the listener's head. All speech materials were recorded. A computer with a sound card (Turtle Beach Fiji), CD player, and a GSI audiometer (Model 16) were used to present the test items. The GSI 16 audiometer generated the speech-shaped noise used during the vowel, consonant and word tests for the implant listeners. The CD utilized for presenting the HINT sentence materials provided the speech-shaped noise for that test.

Threshold (T) and comfort (C) or most comfortable (M) loudness levels were measured separately for each experimental condition. The five experimental processors were presented to each listener in random order. Within each of the five experimental processor conditions, the four noise conditions were presented in random order, following the condition in quiet, which was always presented first. The battery of speech tests was administered to each listener immediately after they were given the experimental processor (no practice). The listener's normal settings were restored to the speech processor after each testing session until the listener returned for the next experimental condition, typically one week later. After the T and C level adjustments, Nucleus 22 listeners' were told to set their sensitivity level to the most comfortable position with the function switch set to Normal (N). This setting was used during all the test conditions for that particular processor. All listeners were programmed in the bipolar-plus-one mode electrode pairing (BP+1) for both their normal processors and for all experimental conditions.

For the Nucleus-22 device electrical thresholds (T) and maximum acceptable loudness (C) levels were obtained using the Nucleus diagnostic and programming system with a personal computer and a dual processor interface with Cochlear Corporation 6.100 software. For obtaining T- and C- levels the stimulus was a 250 Hz pulse train of 500 msec duration. Threshold levels were estimated by a standard clinical bracketing procedure. One to five pulse bursts were presented and the listener was instructed to count the number of bursts heard. The T level used in the processor was the level at which the listener counted the number of bursts correctly 100% of the time. To obtain C levels the experimenter increased the electrical level until the listener felt the loudness was at the maximum acceptable level. Adjacent electrodes were balanced for loudness at C level for each electrode.

For the Clarion device electrical thresholds (T) and most comfortable loudness (M) levels were obtained using the SCLIN for Windows software, Clinician's Programming Interface (CPI), and power supply with a personal computer. The IDR (Input Dynamic Range) was set to –60 dB for all conditions. All other parameters were set as in the listener's original processor. In the CIS processing strategy, threshold levels were estimated by a standard clinical bracketing procedure. Initially, all the electrodes were screened for threshold level and the patient was instructed to identify when they first heard the sound. Then, going back to the first electrode, one to five pulse bursts were presented and the listener was instructed to count the number heard. The T level used in the processor was the level at which the listener counted the number of bursts correctly 50% of the time. To obtain M levels the experimenter increased the electrical level until the listener felt the loudness was at the most comfortable loudness level (the level where they heard the sound at a normal conversational level and could listen to it for a long time without discomfort). Adjacent electrodes were balanced for loudness at M level for each electrode.

In the SAS strategy, measurement procedures were identical to CIS except for threshold levels which were not measured as per the CLARION device fitting manual. (Clarion, 1998)



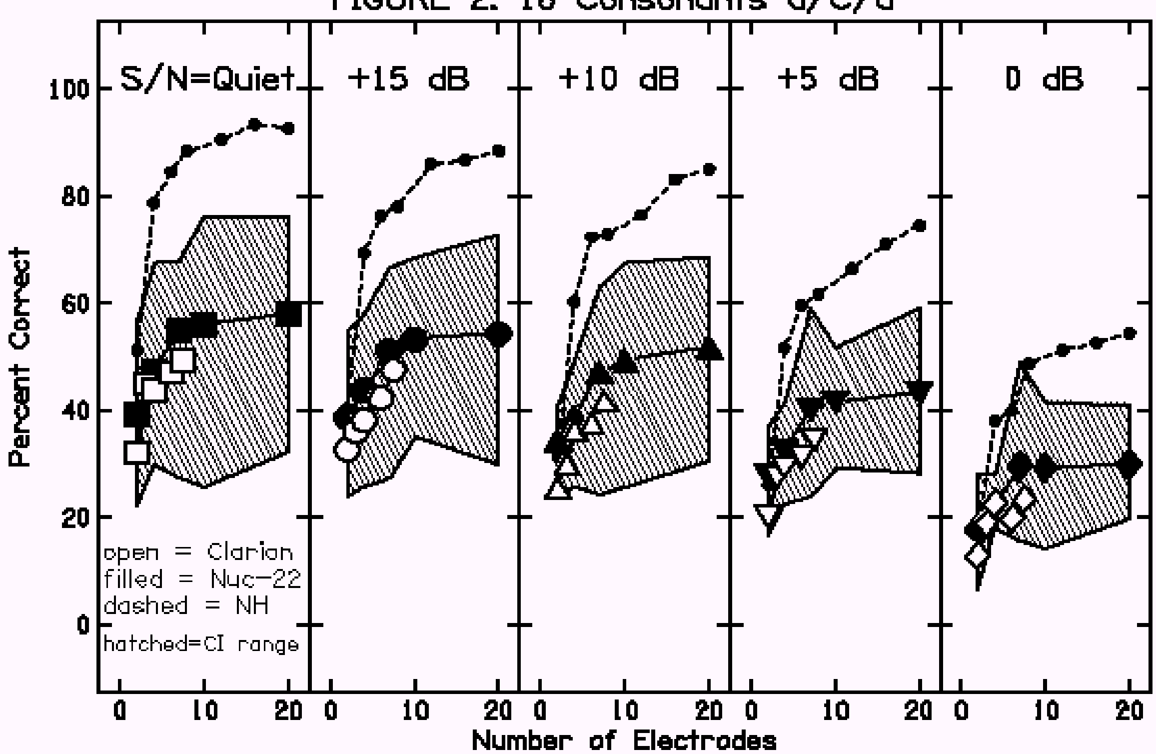
RESULTS

Figures 1-4 present the results for vowels, consonants, CNC words, and HINT sentences, respectively. Within each figure, the panels present recognition performance for quiet listening conditions and signal-to-noise levels of +15, +10, +5, and 0 dB, respectively, from left to right. In each panel the open symbols present data from subjects with the Clarion device, filled symbols present data from subjects with the Nucleus-22 device, and the dashed line presents results from normal-hearing listeners with a noise-band processor. Average standard deviations for the three types of listeners on the four sets of test materials are given in Table 3. Note that the variability was similar for the two sets of implant listeners, while the variability across normal hearing listeners was generally about half that observed in the implant listeners.. The hatched area in Figures 1 and 2 outlines the entire range of performance across all 19 implant subjects.

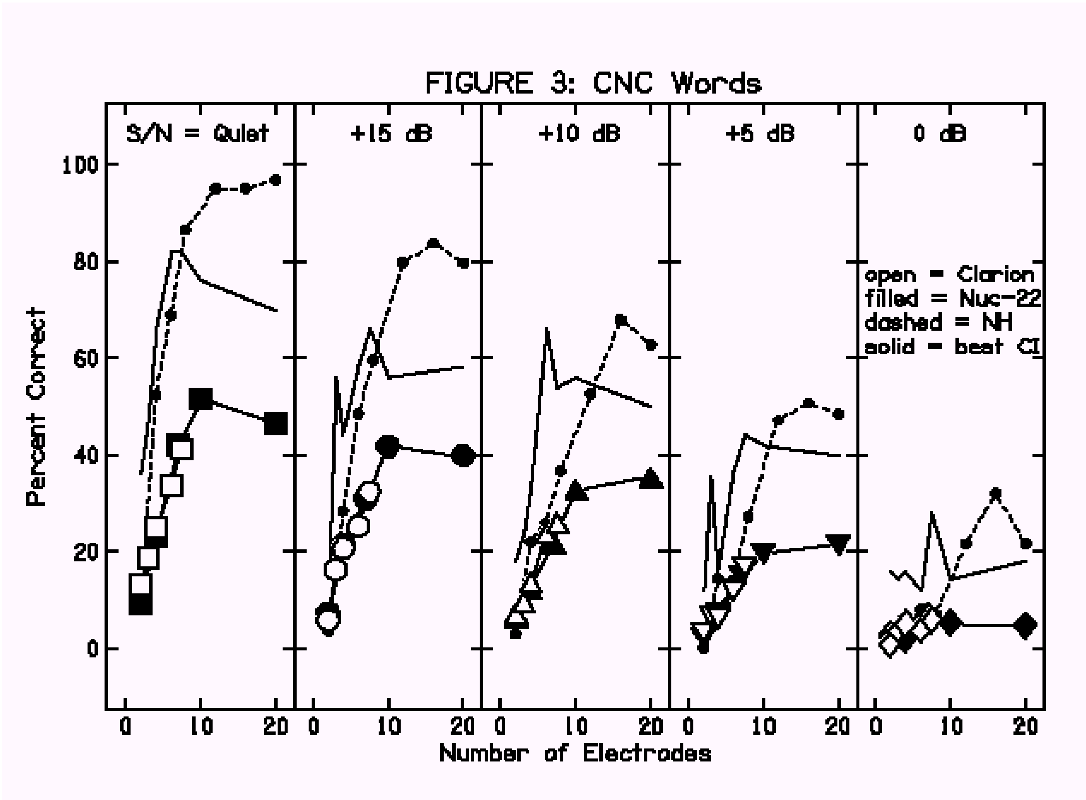
Table 3: Average standard deviations for each condition and listener type

	Normal-Hearing	Clarion	Nucleus-22
Vowels	4.87	11.50	10.95
Consonants	3.58	10.81	9.67
CNC Words	7.50	11.86	10.32
HINT Sentences	7.24	19.74	15.21

FIGURE 2: 16 Consonants a/C/a



For Clarion listeners, a repeated measures ANOVA revealed no difference in performance between the CIS and SAS patients ($F = 0.451$, $df = 1$, $p = 0.511$) at all numbers of electrodes and at all noise levels, although the number of patients with each processor was quite small. The two groups of Clarion listeners were then grouped together for comparison with Nucleus listeners. A

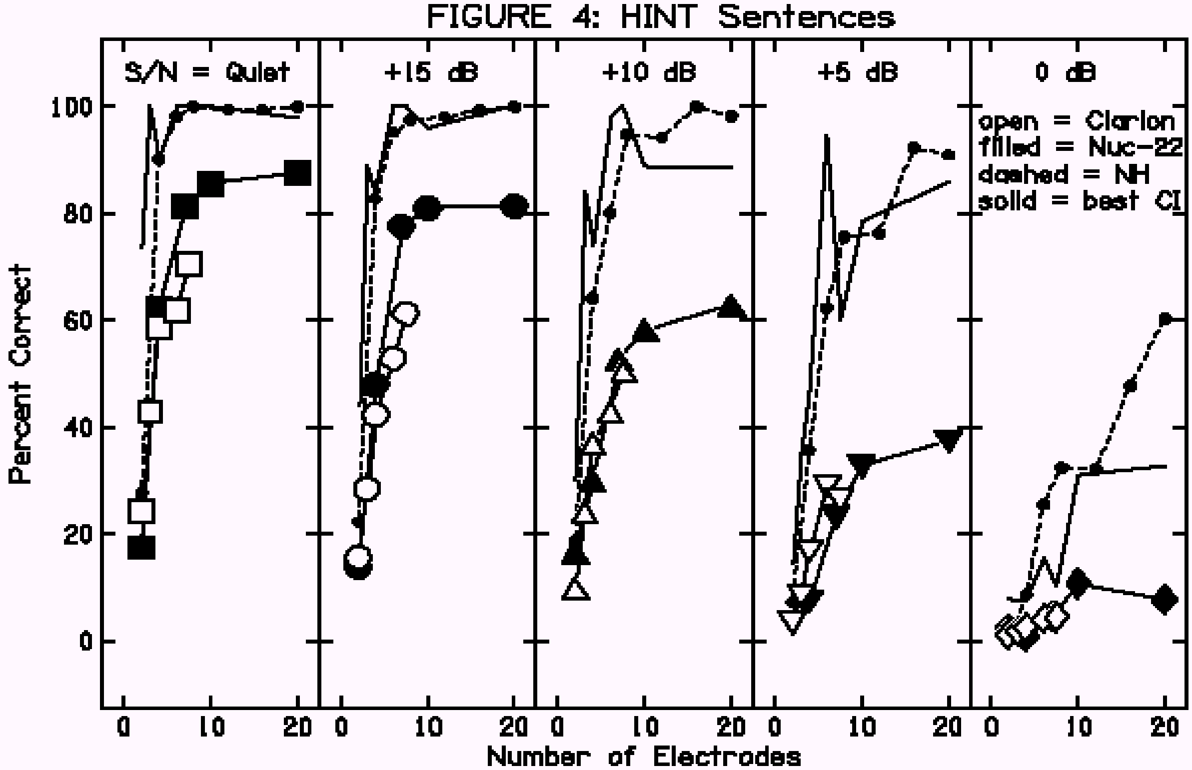


repeated measures ANOVA revealed no significant difference in performance between listeners using the two implants for all conditions ($F = 0.632$, $df = 1$, $p = 0.453$). Performance with both the Clarion and Nucleus-22 processors improved as the number of electrodes increased (up to 7 or 8) for all conditions. Even though more electrodes were available with the Nucleus-22 speech processor, performance did not improve significantly as the number of electrodes was increased from 7 to 10 to 20 for all test conditions.

The performance of normal-hearing listeners was significantly better than the average implant listeners for all conditions (F=12.73, df = 2, p = 0.000). In addition, scores of the normal hearing listeners continued to increase up to 20 channels with similar signal processing conditions. For vowel recognition (Figure 1), performance by the best cochlear implant listeners (top edge of the hatched area) was similar to normal hearing listeners, but only up to 8 channels. For more than 8 channels, performance for the implant listeners remained constant, while for normal hearing listeners, performance continued to increase with the number of channels. For consonant recognition (Figure 2), normal-hearing listeners scored consistently higher than the best implant listeners for all numbers of electrodes, particularly at high signal-to-noise levels. For CNC word recognition (Figure 3) and HINT sentence recognition (Figure 4) only the best performance by cochlear implant patients is presented because the poorest performance level was near zero for all conditions. As with the phoneme results, the best performance level with cochlear implants was similar to that of normal-hearing listeners with the same processing, up to 7-8 channels. As the number of channels/electrodes was increased above 7-8, word and sentence recognition continued to increase in normal-hearing listeners, but not in implanted listeners. The line representing the best implant score is somewhat erratic because it represents a single score.

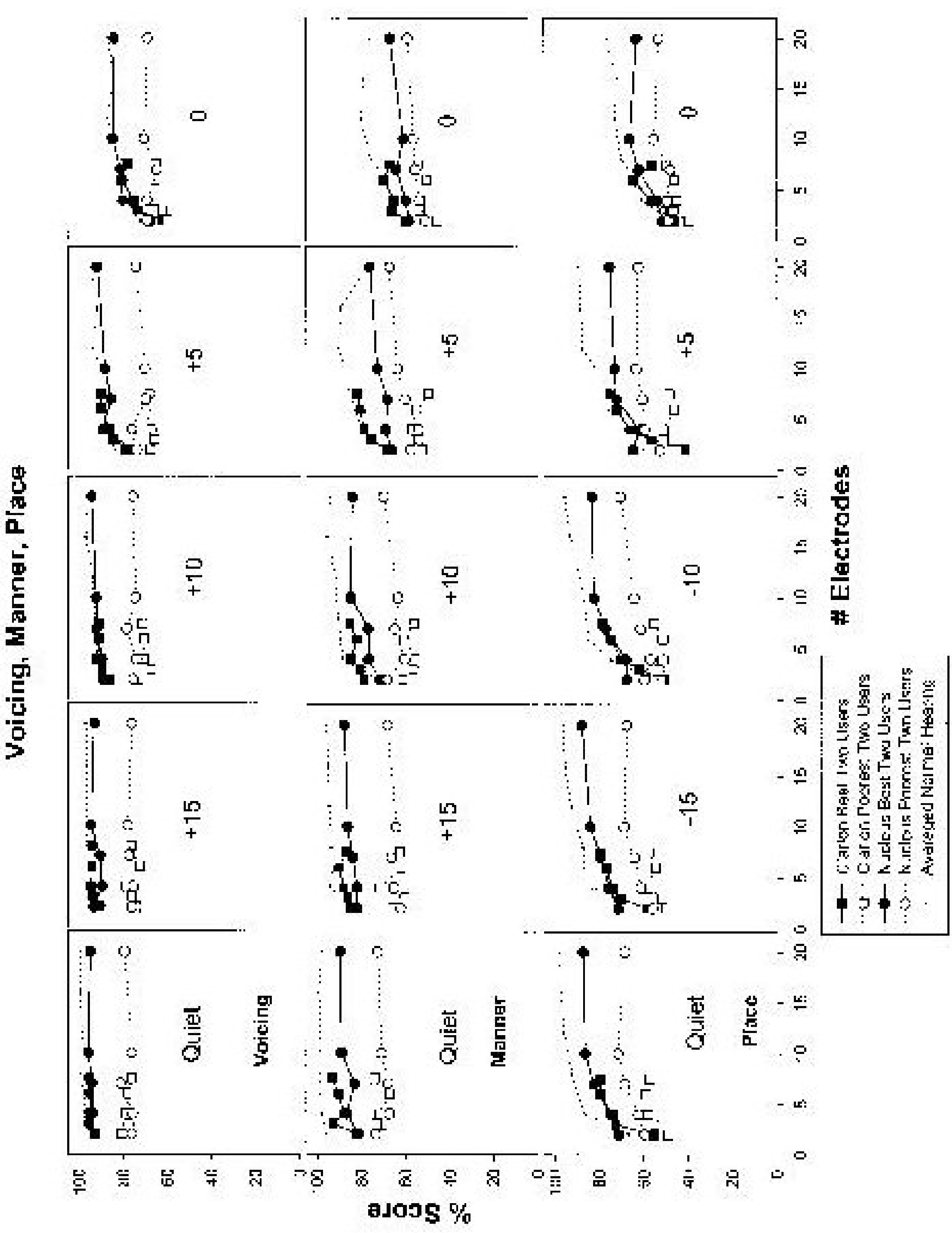
One interesting feature of the results can be observed by comparing the

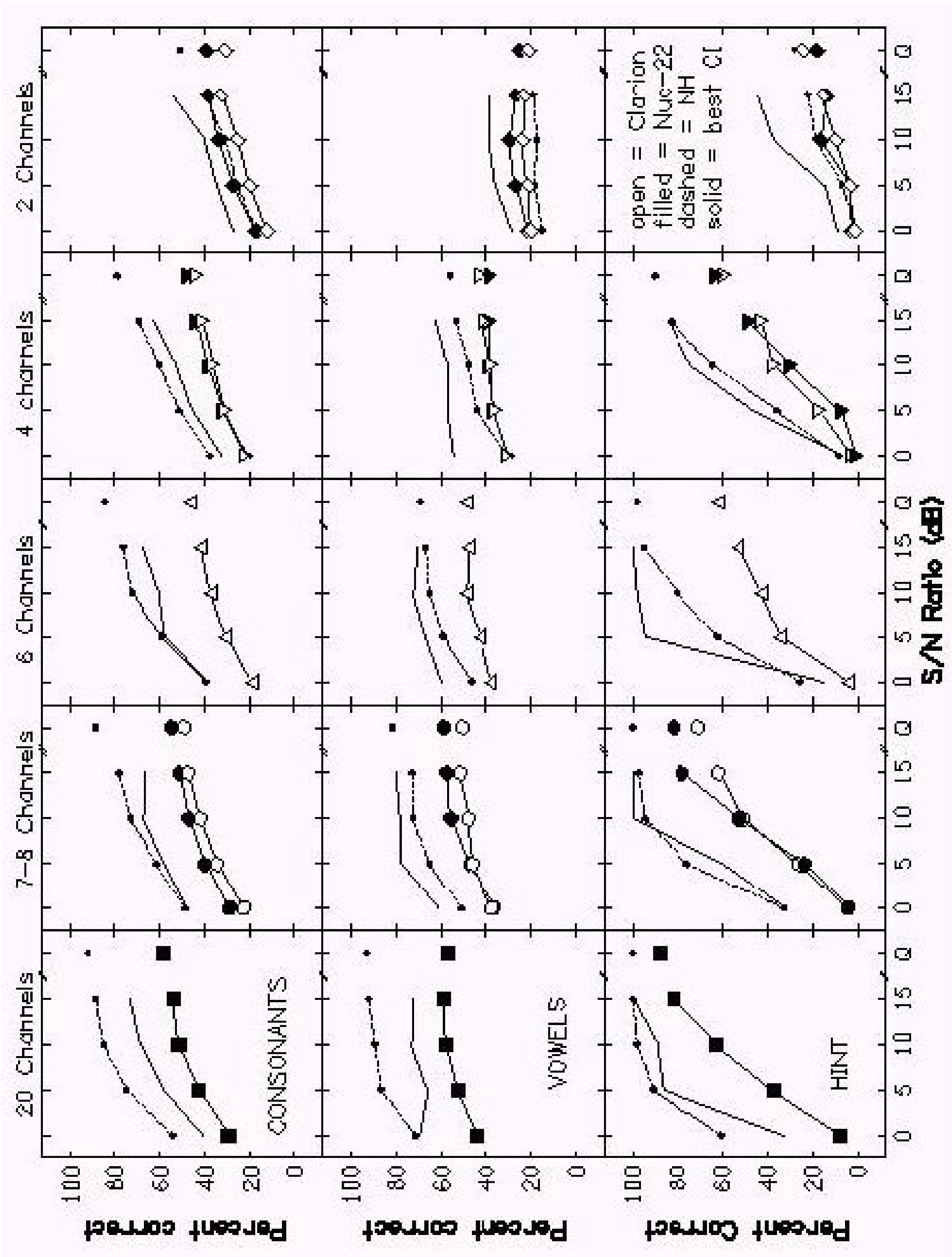
FIGURE 4: HINT Sentences



top and bottom borders of the hatched area in Figures 1 and 2. The better-performing implanted listeners improved as the number of electrodes was increased up to 7 electrodes, while the poorer-performing listeners showed little increase in performance as the number of electrodes was increased above 4. This was particularly true in tests that depended more on spectral cues, such as vowel recognition) and at low S/N levels. This result suggests that implant listeners who do better on speech recognition are able to utilize more spectral channels of information than those who do more poorly on speech recognition. To confirm this observation, consonant confusion matrices were analyzed for the two implant listeners with the best scores and the two implant listeners with the poorest scores.

Consonant recognition was analyzed into the traditional production-based categories of voicing, manner of articulation, and place of articulation. The percent correct for each of these features is presented in Figure 5 for the two best users and the two poorest users of each device and for normal hearing listeners with similar processing. The consonant results, presented in Figure 5, show that the better implant users were able to utilize all categories of cues better than the poorer implant users. At high signal-to-noise ratios, the reception of voicing was not affected by the number of electrodes. However, at poor signal-to-noise ratios the percent correct on voicing increased with the number of electrodes up to four. This may be due to the noise interfering with the perception of the temporal cues for voicing, which should not require much spectral information. Voicing can be conveyed by spectral cues (e.g., the ratio of energy above and below 1500 Hz), but multiple electrodes are required to provide this spectral information. A similar pattern of performance was observed for the reception of manner cues. Percent correct on the place of articulation increased as the number of electrodes increased at all noise levels. In spite of the large differences in electrode design and speech processing strategy, the better listeners with the Nucleus-22 device were receiving similar amount of information on voicing, manner and place as the best listeners with the Clarion device (both CIS and SAS). The pattern of information received for the best implant listeners was similar to that of the normal-hearing listeners with the same number of processing channels.


Voicing, Manner, Place
% Score
# Electrodes
Quiet
+15
+10
+5
0
-15
-10
Voicing
Manner
Place
Clarion Best Two Users
Clarion Poorest Two Users
Nucleus Best Two Users
Nucleus Poorest Two Users
Averaged Normal Hearing



which demonstrates that the limitation is not due to a ceiling effect on performance – the same 7-electrode limit was observed at all noise levels and all performance levels. There appears to be a slight difference in the pattern of performance between the better implant users and the poorer users. The poorer users did not improve as the number of electrodes was increased beyond 3 or 4 on some tests.

Figure 6 plots some of the data from Figures 1-4 in a different format. Here consonant, vowel, and sentence recognition are plotted as a function of the S/N ratio. The panels, from left to right, plot the results as the number of channels is reduced. As in Figures 1-4 the filled symbols plot the data from Nucleus-22 listeners, open symbols plot the average data of 9 Clarion listeners, the small symbols and dashed lines plot the average results from five normal-hearing listeners, and the solid line without symbols plots the best scores for each condition from all 19 implant listeners. For sentence recognition the slopes of the functions decrease as the number of channels decrease. This is also the case for the vowel and consonant functions, although it is less severe than with sentence recognition. Note again that the results from Nucleus-22 and Clarion listeners are quite similar, and the results from the normal-hearing listeners is similar to the best scores from the cochlear implant listeners. In the 2-channel case (right panels) the best scores from implant listeners is actually higher than the average score for the normal-hearing listeners. This may reflect the fact that the implanted listeners have more experience listening to degraded signals than the normal hearing listeners.

Figure 7 presents a 3-dimensional plot of the sentence recognition data from normal-hearing listeners obtained at S/N ratios of 0, 5, 10, 15 dB and in quiet, for 2, 4, 6, 8, 12, 16, and 20 bands. The corner of the figure at low S/N ratios and small number of channels shows the trade-off between the number of channels and sentence recognition in noise. At high S/N ratios, sentence recognition is high for all number of channels greater than 3. At low S/N ratios, however, a reduction in the number of channels is equivalent to reducing the S/N ratio. Even the best performing implant listeners, who are typically using the equivalent of about 8 channels need 5-10 dB better S/N ratio for equivalent performance to normal-hearing listeners using 20 channels. The implication is that if we could increase the number of spectral channels effectively used by implant listeners, they would be able to understand speech much better in noise. Because the sentence recognition function has a slope of 6-10%/db, a 5 dB difference in S/N ratio could produce a 30-50% improvement in sentence recognition. We will derive equations to describe thi surface in Figure 7, which will allow us to specify the exact trade-off between the number of spectral channels and speech recognition in noise.

Consider several hypotheses to explain the differences between implant performance and normal-hearing performance, in particular what factors might limit performance in implant listeners to 7-8 channels.

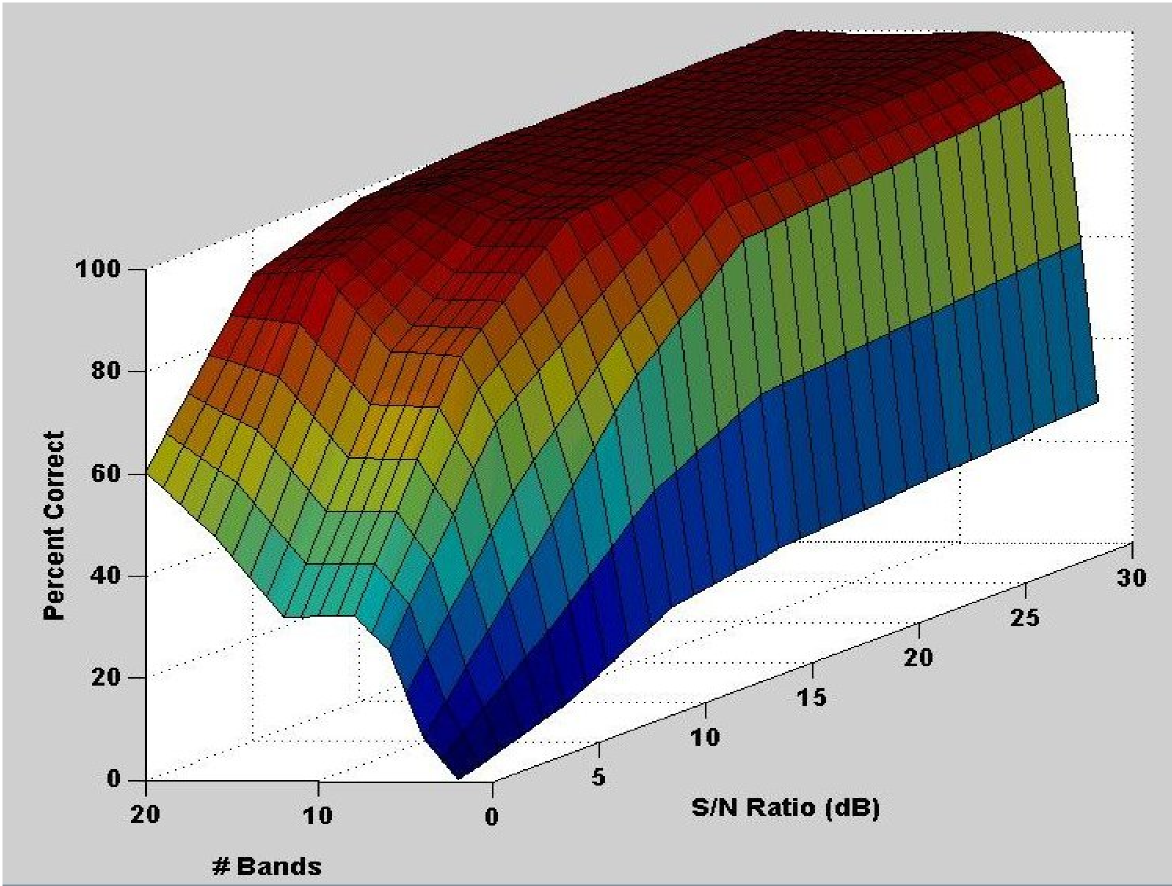


FIGURE 7: Sentence recognition by normal-hearing listeners

Hypothesis 1: Stimulation Rate is the Primary Factor Limiting Performance

Consider the hypothesis that the limitation of implant listeners to 7-8 channels of spectral information is due to the relatively low pulse rate/electrode in the SPEAK processing strategy. If this hypothesis is correct then a SPEAK or CIS system with a faster pulse rate per electrode might show an improvement in performance as the number of electrodes is increased above 7. However, preliminary results with the Nucleus-24 implant system do not show increased overall performance for listeners with the Nucleus-24 system relative to average results with the Nucleus-22 system, regardless of the choice of SPEAK, CIS, or ACE processing strategy (Arndt et al., 1999). The CIS strategies implemented in the Nucleus-24 device allows stimulation of 12 electrodes at stimulation rates of up to 1200 pps/electrode. The ACE strategy, which is a hybrid of SPEAK and CIS strategies allows stimulation of 16 electrodes out of 20 at rates of 900 pps/electrode. These stimulation rates are considerably higher than rates allowed by the Nucleus-22 system. Since the preliminary data indicate no difference between the 22 and 24 systems in the average level of performance, even with the full number of electrodes and higher stimulation rates, it is unlikely that patients with the 24 system are improving in performance beyond 7 electrodes. More data is needed to confirm this observation. If this preliminary result is confirmed it would indicate that stimulation rate is not the primary factor limiting the number of usable channels of spectral information.

Hypothesis 2: Electrode Interaction is the Primary Factor Limiting Performance

If it is the case that cochlear implant listeners are not increasing in performance even when they are presented with high-rate stimulation on 20 electrodes, then a key question in the design of cochlear implants is "What is the factor limiting the number of effective channels conveyed by a cochlear implant?" One candidate is electrode interaction. It may be the case that performance asymptotes at 7-8 channels because additional electrodes are not providing new independent information. There is some evidence that electrode interaction and electrode distinctiveness are correlated with speech recognition performance (Hanekom and Shannon, 1996; Donaldson and Nelson, 2000). Observe the difference between the upper and lower edge of the hatched area in Figures 1 and 2. The upper edge, which represents the best implant scores across all 19 listeners increases in performance up to 7-8 channels. The lower edge, representing the poorest scores across all implant listeners, does not increase substantially from 2 to 20 electrodes. Thus, consistent with the electrode interaction hypothesis, poor implant speech performance is limited to 2-4 effective spectral channels, while good implant speech recognition improves with the number of electrodes, up to 7-8.

If electrode interactions are limiting performance on the top end of implant performance at 7-8 channels, then the poorer users may have an increased amount of electrode interaction that limits their performance even further – not allowing performance to improve beyond the 3-4 channel level, no matter how many electrodes are used. Inspection of the results in Figures 1 and 2 shows that the implant listeners who are poor at speech recognition did not improve as the number of electrodes was increased above 2-4. If this is the case then it is of utmost importance to discover the cause of the electrode interactions and either correct this problem in signal processing or with new electrode designs.

Electrical field interactions could be reduced if electrodes were closer to the excitable neurons. New Contour electrode design by Cochlear Corp (Parker *et al.*, 2000) and Hi-Focus electrode design by Advanced Bionics (Kuzma, 2000) are designed to position the electrodes against the inner wall of the scala tympani (modiolus hugging). The preliminary results in patients with these new electrode designs are very promising, with significant improvements in average word and sentence recognition scores (Osberger, 2000; Staller, 2000). These results could indicate that modiolus-hugging electrodes are improving performance by reducing electrode interactions and so increasing the number of effective independent channels.

Hypothesis 3: Warping in the Spectral-Tonotopic Mapping is the Primary Factor Limiting Performance

Another possible cause of the limitation in the use of all channels is the presence of distortion in the representation of the spectral information. Fu and Shannon (1999a) found that speech recognition was reduced when the spectral information was represented at cochlear locations that were shifted either apically or basally from their normal location. Shannon et al. (1998) found that speech recognition was reduced when the tonotopic distribution of spectral information was warped nonlinearly from its normal mapping. Both studies observed that a warping in the tonotopic distribution could not only result in a reduced number of effective channels of spectral information, but could also reduce the reception of what are thought to be primarily temporal cues in speech, like voicing and manner. Shannon et al. saw significant reductions in voicing, manner, and place information received on consonants when the tonotopic mapping was distorted. A similar pattern was observed in the poorer implant listeners in this study, i.e., their reception of voicing and manner were significantly poorer than in the implant users with better performance. Not only were the poorer users receiving fewer effective channels of spectral information, they were also showing poorer reception of voicing and manner cues in consonants. In normal-hearing listeners nearly 100% of the voicing and manner information in consonants is received for all processors with two or more channels (Shannon et al., 1995). Compare the reception of voicing and manner cues in figure 5 between the normal-hearing listeners and the two groups of implant listeners. The implant listeners with poor performance received less voicing and manner information with 20 electrodes than the better-performing implant listeners received with only two electrodes. Thus, a reduction in the reception of voicing and manner cues indicates more than simply a loss of the number of effective channels of spectral information. Based on the Fu and Shannon (1999a) results on tonotopic shifting and the Shannon et al. (1998) results on tonotopic warping, we suggest that overall poor reception of voicing and manner cues could indicate the presence of a shift or warping in the frequency-tonotopic mapping in those patients.

CONCLUSIONS

Speech recognition was similar for the Clarion and Nucleus-22 cochlear implants both as a function of noise level and as function of number of electrodes. This result suggests that the number of electrodes is the primary determinant of performance - and that the differences between these two devices in speech processor strategy and electrode design do not result in significant differences in performance.

Comparison to normal hearing listeners in the simulation experiment suggests that some cochlear implant listeners can fully utilize the spectral information provided (up to 8 electrodes), but others do not. The relatively small improvement for the poorer implant performers as the number of spectral channels was increased suggests that these individuals are not able to utilize the spectral information provided. The reason or reasons for this inability to process spectral information is not clear, but might be related to distortion in the tonotopic distribution of spectral information.

The lack of improvement in performance for all implant listeners when the number of electrodes is increased above 8 is puzzling. Normal-hearing listeners improve in performance when listening to similar processing. It appears that implant listeners are only able to utilize a maximum of 8 channels of spectral information, no matter how many channels of information are presented. The reason for this limitation is not clear. We speculate that this limitation may be due to electrode interactions, and to possible tonotopic shifts and warping in the frequency-to-place mapping of spectral information.

REFERENCES

Arndt, P., Staller, S., Arcaroli, J., Hines, A., and Ebinger, K. (1999). Within subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant, Cochlear Corporation, Englewood, CO..

Baer, T and Moore, B.C.J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise, *Journal of the Acoustical Society of America* 94, 1229-1241.

Baer, T and Moore, B.C.J. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech, *Journal of the Acoustical Society of America* 95, 2277-2280.

Battmer, R.D., Zilberman, I., Haake, P., and Lenarz, T. (1999). Simultaneous analog stimulation (SAS)-continuous interleaved sampler (CIS) pilot comparison study in Europe, *Ann. Otol. Laryngol.*, 108, 69-73.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (1996). Effects of spectral smearing on phoneme and word recognition, *Journal of the Acoustical Society of America* 100, 1807-1818.

Breeuwer, M. and Plomp, R. (1984). Speechreading supplemented with frequency-selective sound-pressure information, *Journal of the Acoustical Society of America* 76, 686-691.

Brill, S., Gstöttner, W., Helms, J., v.Ilberg, C., Baumgartener, W., Müller, J., and Keifer, J. (1997). Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+, *American Journal of Otology Supplement*, 18, S104-S106.

Clarion by Advanced Bionics. *SCLIN 98 for windows device fitting manual.* Sylmar, California. 1998.

Cochlear Corporation, *Technical Reference Manual*. Englewood, Colorado, 1995. Cochlear Corporation (1995), Audiologist Handbook, Englewood, Colorado.

Donaldson, G.S. and Nelson, D.A. (2000). Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies, *J. Acoust. Soc. Amer.*, 107(3), 1645-1658.

Dorman, M.F., Loizou, P.C. and Rainey, D. (1997). Speech understanding as a function of the number of channels of stimulation for processors using sine-wave and noise-band outputs, *Journal of the Acoustical Society of America* 102, 2403-2411.

Dorman, M.F. and Loizou, P.C. (1997). Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants, *American Journal of Otology*, Supplement 6, 18, S113-S114.

Dorman, M.F. and Loizou, P.C. (1998). Identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations processors with two to nine channels, *Ear and Hearing*, 19(2), 162-166.

Dowell, R.C., Seligman, P.M., Blamey, P.J., and Clark, G.M. (1987). Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise, *Acta Otolaryngologica (Stockholm)* 104, 439-446.

Dubno, J.R. and Dorman, M.F. (1987). Effects of spectral flattening on vowel identification, *Journal of the Acoustical Society of America* 82, 1503-1511.

Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladevosky, M.G., and Parkin, J.L. (1978). Auditory prosthesis research with multiple channel intracochlear stimulation in man, Ann. Otol. Rhinol. Laryngol., 87, Suppl. 53, 1-39.

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, *Journal of Speech and Hearing Research* 40,1201-1215.

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

Fu, Q.-J. and Shannon, R.V. (1999b). Recognition of spectrally degraded speech in noise with nonlinear amplitude mapping, *Proc. of the 1999 IEEE Conf. on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 369-372.

Fu, Q.J, Shannon R.V., and Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing, *Journal of the Acoustical Society of America,* 104: 3586 –3596.

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

Hanekom J.J., and Shannon R.V. (1996). Place pitch discrimination and speech recognition in cochlear implant users, *South African Journal of Communication Disorders*, 43, 27-40.

Hillenbrand J., Getty, L.A., Clark, M.J., and Wheeler K. (1995). Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America* 97, 3099-3111.

Hochberg, I., Boothroyd, A., Weiss, M., Hellman, S. (1992). Effects of noise and noise suppression on speech perception by cochlear implant users, *Ear and Hearing*. 13, 263-271.

Horst, J.W. (1987). Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects, *Journal of the Acoustical Society of America* 82, 874-885.

House Ear Institute and Cochlear Corporation. (1996). *Minimum Speech Test Battery for Adult Cochlear Implant Users CD*.

Kiefer, J., Müller, J., Pfennigdorf, T., Schön, F., Helms, J., von Ilberg, C., Gstöttner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., and Baur, S. (1996). Speech understanding in quiet and in noise with the CIS speech coding strategy (Med-El Combi-40) compared to the Multipeak and Spectral peak strategies (Nucleus),*ORL-Journal for Oto-Rhinology, Laryngology, and its Related Specialties* 58, 127-135.

Kompis, M. and Dillier, N. (1994). Noise reduction for hearing aids: combining directional microphones with an adaptive beamformer, *Journal of the Acoustical Society of America,* 96(3), 1910-1913*.*

Kuzma, J. (2000). Evolution of electrode design, *6th International Cochlear Implant Conference*, Miami, FL, Feb 3-5.

Lawson D.T., Wilson B.S., Zerbi M. & Finley C.C. (1996). "Speech processors for auditory prostheses", Third Quarterly Progress Report, *NIH Contract N01-DC-5-2103***.**

Lawson, D., Wilson, B., & Finley, C. (1993). New processing strategies for multichannel cochlear prostheses. In J.A. Allum, D.J. Allum-Mecklenburg, F.P. Harris, and R. Probst (Eds.), *Natural and Artificial Control of Hearing and Balance*, (pp. 313-321), *Progress in Brain Research*, vol. 97, Amsterdam: Elsevier.

McDermott, H.J., McKay, C. M., Vandali, A. E. & Clark, G. M. (1992). A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (Multipeak) processor, *Acta Otolaryngologica (Stockholm)*, 112, 752-761.

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

Müller-Deiler, J., Schmidt, B.J., and Rudert, H. (1995). Effects of noise on speech discrimination in cochlear implant patients*, Annals of Otology, Rhinology, and Laryngology*. 166, 303-306.

Nejime, Y. and Moore, B.C.J. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise, *Journal of the Acoustical Society of America* 102, 603-615.

Nilsson, M., Soli, S., Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise, *Journal of the Acoustical Society of America*, 95(2), 1085-1099.

Osberger, M.J. (2000). Clinical studies update: Performance with Clarion using advanced electrode designs, *6th International Cochlear Implant Conference*, Miami, FL, Feb 3-5.

Parker, J.L., Treaba, C.G., Cohen, L., Tykocinski, M., Saunders, E., Gibson, P., Darley, I., Dadd, F., Garg, R., and Zhang, L. (2000). Nucleus 24 Contour electrode with stylet, *6th International Cochlear Implant Conference*, Miami, FL, Feb 3-5.

Robert, ME. (1997). *AIPSS-ID: Phoneme Identification Software*. Los Angeles; House Ear Institute.

Seligman, P., and McDermott, H. (1995). Architecture of the Spectra 22 Speech Processor, *Annals of Otology, Rhinology, and Laryngology* 104, Supplement. 166, 139-141.

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.

Skinner, M.W., Clark, G.M., Whitford, L.A., et al. (1994). Evaluation of a new spectral peak coding strategy for the Nucleus-22 channel cochlear implant system, *American Journal of Otology*, 15, Supplement 2, 15-27.

Staller, S.J. (2000). The Nucleus 24 Contour cochlear implant: Preliminary results and expanded criteria, *6th International Cochlear Implant Conference*, Miami, FL, Feb 3-5.

Stelmachowicz, P.G., Jesteadt, W., Gorga, M.P., and Mott, J. (1985). Speech perception ability and psychophysical tuning curves in hearing impaired listeners, *Journal of the Acoustical Society of America* 77, 620-627.

ter Keurs, M., Festen, J.M., and Plomp, R. (1992). Effect of spectral envelope smearing on speech reception I., *Journal of the Acoustical Society of America* 91, 2872-2880.

ter Keurs, M., Festen, J.M., and Plomp, R. (1993). Effect of spectral envelope smearing on speech reception. II, *Journal of the Acoustical Society of America* 93, 1547-1552.

Turner, C.W. and Henn, C.C. (1989). The relation between vowel recognition and measures of frequency resolution, *Journal of Speech and Hearing Research*. 32, 49-58.

Turner, C.W., Chi, S.-L., and Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss, *Journal of.Speech and Hearing Research*, 42(4), 773-784.

Van Hoesel, R.J. and Clark, G.M. (1995a). Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients, *Journal of the Acoustical Society of America* 97, 2498-2503.

Van Hoesel, R.J. and Clark, G.M. (1995b). Manikin and cochlear implant patient test results with a portable adaptive beamforming processor to suppress the effect of noise, *Annals of Otology, Rhinology, and Laryngology* Supplement 166, 144-146.

Van Tasell, D.J. and Yanz, J.L (1987). Speech recognition threshold in noise: effects of hearing loss, frequency response, and speech materials, *Journal of Speech and Hearing Research*, 30, 377-386.

Van Veen, T.M. and Houtgast, T. (1985). Spectral sharpness and vowel dissimilarity, *Journal of the Acoustical Society of America* 77, 628-634.

Weiss, M.R (1993). Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor, *Journal of Rehabilitation, Research, and Development* 30, 117-128.

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.

Plans for the Next Quarter

Experiments – Psychophysics. We will measure electrode interaction with the Clarion device for patients with the original electrode, the original electrode with positioner and the new Hi-Focus electrode with and without positioner. Differences in electrode interaction should indicate if these new designs are meeting the goal of reducing electrode interactions by placing the electrodes closer to the modiolus. Once the Nucleus-24 interface is integrated into laboratory software we will also make electrode interaction measures in patients with the original electrode and with the new "Contour" electrode.

Experiments – Speech Processor Design. (1) Finish data collection on the "holes in hearing" experiment with normal-hearing listeners, (2) continue data collection on the effects of stimulation rate on speech recognition with Clarion and Nucleus-24 SPEAK, CIS and ACE listeners, and (3) initiate collection of speech recognition results from patients with the original and new modiolus-hugging electrode designs.

Publications in this Quarter

Chatterjee, M., Fu, Q.-J., and Shannon, R.V. (2000). Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners, Journal of the Acoustical Society of America, 107(3), 1637-1644.

Fu, Q.-J. and Shannon, R.V. (2000). Effect of stimulation rate on phoneme recognition in cochlear implants, Journal of the Acoustical Society of America, 107(1), 589-597.

Shannon, R.V., Fu, Q.-J, Friesen, L, and Galvin, J. (2000). Present and Future Research on Cochlear Implants, CICI Contact, 13(4), 25-39.

Presentations in This Quarter

Chatterjee, M. and Robert, M.E. (2000). Perception of noisy pulse trains by Nucleus-22 cochlear implant listeners. ARO Midwinter Research Meeting, St. Petersberg, FL, Feb 20-24. (poster)

Fu, Q.-J, Shannon, R.V., and Wang, X. (2000). Loudness growth in cochlear implants: effects of stimulation rate and electrode location, ARO Midwinter Research Meeting, St. Petersberg, FL, Feb 20-24. (poster)

Shannon, R.V. (2000). Present research and future directions in cochlear implant research, Cochlear Implant Workshop, California School of Professional Psychology, Feb 11 Los Angeles, Feb 12 San Diego. (KEYNOTE Speaker)

Shannon, R.V. (2000). Overview of Cochlear Implant Research, Teacher Continuing Education Program, San Bernadino County School System, Rancho Cucamonga, CA, March 11.

Shannon, R.V., Galvin, J., and Baskent, D. (2000). Holes in hearing, ARO Midwinter Research Meeting, St. Petersberg, FL, Feb 20-24. (poster)

Stickney, G. and Shannon, R.V., Opie, J., and Assman, P. (2000). Simultaneous masking as a measure of electrode interaction in cochlear implant listeners, 6th International Cochlear Implant Conference, Miami Beach, FL, Feb 2-6. (poster)

Stickney, G., Shannon, R.V., Opie, J., and Assman, P. (2000). Electrode interaction in multichannel cochlear implants with different electrode designs and positions, ARO Midwinter Research Meeting, St. Petersberg, FL, Feb 20-24. (poster)