

Eleventh Quarterly Progress Report

February 1 through April 30, 1998

NIH Project N01-DC-5-2103

Speech Processors for Auditory Prostheses

Prepared by

Dewey Lawson, Blake Wilson, and Mariangeli Zerbi

Center for Auditory Prosthesis Research

Research Triangle Institute

Research Triangle Park, NC 27709

Contents

[I. Introduction](#)

[II. Design of new speech test materials and comparisons with standard materials](#)

[References](#)

[III. Plans for the Next Quarter](#)

[IV. Acknowledgments](#)

[Appendix 1: Summary of Reporting Activity for this Quarter](#)

I. Introduction

One of the principal objectives of this project is to design, develop, and evaluate speech processors for implantable auditory prostheses. Ideally, the processors will represent the information content of speech in a way that can be perceived by implant patients. Another principal objective is to develop new test materials for the evaluation of speech processors, given the growing number of cochlear implant subjects enjoying levels of performance too high to be sensitively measured by existing tests.

Work in the present quarter included:

- Studies with Ineraid subjects SR15 (week of February 9), SR9 (weeks of February 23 and March 2), and SR16 (week of March 9), principally to evaluate speech reception performance with their portable Innsbruck/RTI processors, following six months or more of daily use for each subject. The studies also included measures of channel interaction using recordings of intracochlear evoked potentials, as in prior studies with other subjects. Intracochlear EPs also were recorded for stimuli used in prior psychophysical studies of pitch scaling for subjects SR15 and SR16 (such recordings already had been made for subject SR9). Additional studies with SR9 and SR15 included (1) measures of stimulus levels corresponding to threshold and most comfortable loudness in the context of multichannel stimulation, (2) comparison of speech processors using mapping functions derived from those measures versus speech processors using mapping functions derived from standard single-channel measures of threshold and MCL, and (3) evaluation of "n-of-m" processors.
- Ongoing studies with Ineraid subject SR2 (usually one morning per week). Studies for this quarter included preliminary investigations of several new processor designs, such as processors using "conditioner" stimuli in

conjunction with the normal outputs of CIS processors, and such as processor using high rate stimuli in various other ways. The studies with SR2 also included measures of intracochlear EPs for modulated and unmodulated pulses presented with various levels of high-rate conditioner pulses.

- Continued studies with subject NU-4, a local patient with standard transcutaneous Nucleus implants on both sides.
- Ongoing development of instrumentation in the evoked potentials laboratory, along the lines indicated in Quarterly Progress Report 9 for this project.
- Presentation of project results in an invited lecture at the University of Iowa (Lawson, February 20) and in an invited poster at the NIH Bioengineering Symposium (Wilson and Pierschalla, February 27 and 28).
- Continued preparation for studies here at RTI with subjects implanted at the University of Iowa with CI24M implants on both sides. Preparation included (1) ongoing development of an interface system, with the cooperation of Cochlear Ltd., for simultaneous laboratory control of two CI24M implants, and (2) ongoing modification of existing speech processor and psychophysical test code for use with the interface system.
- Continued analysis of speech reception, evoked potential, and psychophysical data from prior studies.
- Continued preparation of manuscripts for publication, including one in collaboration with Jay Rubinstein and Paul Abbas, on ways in which spontaneous activity can be reinstated in the auditory nerves of implant patients, and including invited chapters for two books on cochlear implants, one edited by John Niparko and the other edited by Noel Cohen and Susan Waltzman.

In this report we describe development of new speech test materials for use with cochlear implant patients at the high end of the performance spectrum. We also present comparisons of these materials with other, existing materials from tests with implant patients, using within-subject controls.

II. Design of new speech test materials and comparisons with standard materials

Speech test materials are expected to fill two distinctly different roles in our speech processor research: (1) the rapid, sensitive, and reliable assessment of relevant differences in performance among a large number of candidate processors during fitting, optimization, and within-subject comparative studies, and (b) the measurement of levels of performance in terms that will allow comparisons with tests of other patients, with other processors, and at other times and places.

Rapid Comparisons of Processor Performance

For the first of these roles, we have relied on consonant identification tests, initially medial consonant tests using available IOWA videodisc recordings [Tyler *et al.* 1987]. Our earliest use of those materials was in 8 and 16 consonant versions designed for administration in sound alone, sound plus vision, or vision alone conditions. [The original 16 consonants were m, n, f, v, s, z, sh, voiced th, p, b, t, d, k, g, j, and l].

We wrote our own software to administer and analyze these tests, carefully selected exemplars from among the IOWA recordings, and edited the onset and offset points for each one. We first attempted to eliminate extraneous cues -- both visual and audible -- and, failing that, then manipulated the distribution of such defects so as to minimize their impact on test results [Lawson *et al.* 1989]. Our version of the 16 consonant materials included two exemplars each with male and female talkers, all with ah as the carrier vowel.

Each administration of the test involved five presentations of each consonant from the same talker (male or female). Guided by a video display, subjects identified each presentation orally, using a numeric label for each consonant to avoid possible misinterpretation of the subjects' speech. No feedback was provided as to correct or incorrect responses. An experimenter entered each response into the computer which then proceeded to present the next token automatically. The use of oral responses was designed to allow the subjects great flexibility in posture and in distance from the video screen, substantially reducing the motor and visual fatigue involved in prolonged testing. The continuous human communication implicit in the use of oral responses seemed to make the testing experience more pleasant for the subjects and afforded the experimenters additional information regarding such matters as patients' level of confidence, degree of fatigue, changes in test-taking strategy, etc. Finding the oral response approach so effective, we rarely have used any hardware for direct entry of responses by patients. The tokens were presented in sets, so that the *n*th presentation of every consonant occurred before the (*n*+1)th presentation of any. Within each set the order of

presentation was randomized, subject only to the additional requirement that no consonant be immediately repeated (i.e. that the first consonant of one set never be the same as the last consonant of the previous set). Each set was treated as an independent percent correct measurement, allowing us to monitor the statistical uncertainty in a single measurement and in the mean of all measurements as we proceeded. (Data from previous tests under identical conditions were automatically located and included in such computations of the evolving uncertainty.) This organization also allowed us to monitor for learning effects on relatively short time scales as the tests proceeded. We elected to use a limited number of randomizations cyclically, facilitating retrospective searches for systematic variations (e.g., in measured performance as a function of randomization context). Each individual response was recorded in an archival database, along with a specification of the test conditions and any prospective, concurrent, and/or retrospective comments deemed appropriate by the researchers. We found that this 16 consonant test produced rapid and reliable results; typically, two repetitions of the test with each processor for a given condition proved sufficient to assess relative strengths and weaknesses reliably.

Such efficiency, of course, was dependent on obtaining enough errors (consonant confusions) per test to support the necessary statistical accuracy. When one of our research subjects reached performance levels high enough to render the 16 consonant test insensitive, we turned to a version of the same test with 24 consonants. [The additional consonants were ch, zh, unvoiced th, w, y, r, and ng.] We found it much more difficult to minimize the impact of extraneous visual cues within the IOWA videodisc materials in this case. Since research subjects with high enough levels of performance to require a 24 consonant test wouldn't really need evaluation in "with vision" conditions, we reedited the IOWA materials without regard to visual consequences and were able to obtain three exemplars each for male and female talkers and with both ah and ee vowel carriers. The organization and administration of the test remained the same. We were concerned that maintaining five presentations of each consonant per test, resulting in a total of 120 rather than 80 presentations per test, might make the tests tiring or boring enough to reduce the quality or (ultimately, through impact on subject morale) quantity of the data. The first subject to take the 24 consonant version initially did criticize its length. As in the 16 consonant version, a uniform delay had been imposed between entry of the subject's response to one token and presentation of the next -- just enough to cover variations in videodisc search time. That subject was happy to accept such variations in order to shorten the overall duration of the test, and no difficulty has arisen in following that practice with subsequent subjects who have needed 24 consonant tests.

[A similar identification test for eight vowels in a /hVd/ context and with male and female talkers was developed early in our research and continues to be used occasionally. Information transmission scores from such vowel tests are not highly correlated with performance on open set speech tests (Wilson *et al.* 1990).]

Additional features added to our consonant testing software over the years have included (1) the immediate availability of information transmission analyses as confusion matrices were obtained for various processing strategies; (2) provision for the occasional need to repeat the presentation of a token that was completely missed by a subject [When it is necessary to repeat a token a warning notation is automatically entered into the data archive files. This notation then will be displayed automatically whenever the associated test is accessed from the archive.]; (3) an alternate version with token presentation from digital waveform files rather than from laserdisc, and a computer mouse interface for direct response entry by subjects; (4) provisions for superposing digitally recorded tokens over continuously looped digital recordings of various types and amounts of noise, in a way that preserved both noise continuity and reproducibility of noise context for each token from test to test; and (5) alternative displays of confusion matrices for researchers during the tests, based on various types of distinctions (e.g., overall frequency spectrum of the consonant sounds, duration, manner of articulation, place of articulation).

Nineteen copies and updates of our speech testing and analysis software have been supplied to other researchers at eleven institutions at their request. At least two of the institutions have, in turn, supplied the software to clinical centers as part of a testing protocol. Recently we have prepared special 16-consonant test versions utilizing the IOWA videodisc for use with patients whose native languages are Italian and German. The Italian version was developed for use by Dr. Deborah Ballantyne for her patient studies in Rome, while the German version was developed for our own use with a bilateral implant patient from Germany. Selection of consonant sounds and token labeling were, of course, different for each version.

When we first developed the 24 consonant test for use with one exceptional research patient, it allowed us almost immediately to identify a range of CIS processor designs that further improved his speech reception performance to the point that even a 24 consonant test lacked enough sensitivity to discriminate among them. In recent years we have

needed to resort to 24 consonant tests with five other patients, while needing tests including vision with only one.

During the current contract we have addressed the need for still more difficult speech materials for use in rapid assessment of strengths and weaknesses among competing processor implementations. We have done this (1) by using existing tests with materials presented in appropriate types and amounts of competing noise, (2) by evaluating new materials developed by other researchers to serve similar purposes, and (3) by developing new materials and test software of our own.

New materials. Having decided to produce such new sound-alone speech testing materials, we wanted to preserve the strengths of our previously developed techniques and repair as many as possible of the weaknesses we have noted in previously used materials. All of our experience strongly indicated that among the various types of tests that might be considered, consonant identification offered the best combination of sensitivity, relevance, analysis potential, and efficiency of testing for our purposes. [See the discussion of correlations between consonant identification scores and principal open set test scores in Wilson *et al.* 1990.] A major weakness, for our purposes, in the medial consonant tests had been an effect of the initial vowels. High intensity formant transitions of initial vowels into some of the medial consonants provided additional cues as to the identity of the consonants. We decided to remedy that situation by eliminating initial vowels from the test tokens. Our previous consonant materials had been limited to one male and one female talker, to only two vowel contexts, and to only two or three exemplars of each condition. We included in our plans the ability to test, in a carefully balanced but efficient way, across more contexts, more talkers, and with more exemplars. [The structure of our medial consonant tests, with 5 presentations of each consonant, complicates the even distribution of two or three exemplars and talkers across such tests. The number of exemplars, the number of tokens per set, and the number of sets per randomization, for instance, must all be related to ensure balanced use of exemplars over each test (or few consecutive tests) and over each full cycle of randomizations. The best solution we have found is to calculate the exemplar number for each presentation as: $\text{exemplar} = (\text{randomization} + \text{token} + \text{set}) \bmod (\text{total number of exemplars})$. For the 16 consonant test, for instance, this method ensures randomized but balanced presentation of both exemplars over each pair of tests using consecutive randomizations, and for the 24 consonant test of all three exemplars over each three consecutive tests. If the number of randomizations is even for the 16 consonant test and a multiple of three for the 24 consonant version, then exemplar use is even over the long term as well.] Finally, while ensuring enough variation among exemplars to make the tests appropriately difficult, we have exercised as much care as possible to minimize the number and seriousness of extraneous cues (defects) within the recorded tokens.

We have recorded materials with five distinct talkers. Two were adult males with significantly different voice pitches, two adult females with different voice pitches, and one a child. This will allow our tests effectively to span the full range of frequency contexts found in human speech.

We used isolated CV syllables as tokens for our new consonant identification tests. As noted above, this will eliminate high intensity transitional cues from initial vowels, making the test more difficult and providing what is arguably a more relevant context for testing consonant recognition abilities vis a vis connected speech. The same advantages would be true of CVC syllables, but we elected to avoid the use of final consonants for two reasons. First, the presence of final consonants would provide more possibilities for the occurrence of extraneous cues across exemplars. The task of finely balancing the tests was simplified by not having to screen for such unintended cues or distractions. Second, we were concerned about the possible biasing effects of mixing words and nonsense syllables in the same test [Boothroyd and Nittrouer, 1988]. Restricting a consonant identification test either to all words or to all nonsense syllables would have made it extremely difficult, if not impossible, to cover a useful range of consonant sounds within a reasonably short test. It was our judgment that a mixture of the two was likely to be less of a distraction and have less of a biasing effect in a CV rather than CVC context.

One of the 24 consonants included in our previous medial consonant tests, ng, was not suitable for use in the initial position. Each of the remaining 23 was paired with three vowels to make up our full set of CV tokens. The vowels we chose are ah, ee, and oo. They were selected under the dual criteria of wanting to minimize the length of the tests while representing virtually the full frequency range of the first three vowel formants.

The following table lists the 23 consonants, along with codes indicating the manner and place of articulation of each. In the three columns on the right are the CV syllables in English word context. Where the CV syllable itself is interpretable as an English word (here including letter names, slang, and foreign words often used in English) it is shown in upper case. Otherwise, an English word beginning with the CV syllable is shown in lower case. Five entries

are blank, indicating that we know of no common English usage of the corresponding CV syllable in initial position.

consonant	manner	place	ah context	ee context	oo context
P	pu	L	PA	PEA	POOH
B	pv	L	BAH	BEE	BOO
T	pu	A	tot	TEA	TO
D	pv	A	dot	DEE	DO
K	pu	V	cot	KEY	COO
G	pv	V	got	GHI	GOO
F	fu	LD	FA	FEE	food
V	fv	LD	volley	VEE	VOUS
uTH	fu	D		think	Thule
vTH	fv	D		THEE	
S	fu	A	sot	SEE	SUE
Z	fv	A	Zagreb	ZEE	ZOO
SH	fu	P	SHAH	SHE	SHOE
ZH	fv	P	Zsa-Zsa		
CH	fu	P	chop	CHI	CHEW
J	fv	P	jot	GEE	JEW
H	fu	G	HA	HE	WHO
M	n	L	MA	ME	MOO
N	n	A	NAH	KNEE	NU
W	s	L	watt	WE	WOO
Y	s	A	yacht	YE	YOU
L	l	A	LA	LEE	LOO
R	l	A	RAH	read	RUE

Manner of articulation key:

p=plosive, f=fricative, n=nasal, s=semivowel, l=liquid

u=unvoiced, v=voiced

Place of articulation key:

L=labial, LD=labiodental, D=dental, A=alveolar, P=palatal, V=velar, G=glottal

Notice that, under this rather inclusive definition of "English word" over 80% of our CV syllables ending in ee and oo can be heard as words, compared to only about half as many of the ones ending in ah. If we restrict our definition to words in common everyday use the ah list is seen as predominantly nonsense syllable, in even greater contrast to the situation for the other two lists.

The 16-bit digital recordings were made directly, with a minimum of intervening analog circuitry and at a sampling rate of 44.1 kHz. All the original recordings have been preserved, in case qualification and calibration tests indicate a need for adjustments in the testing corpus. All the final recordings in the testing corpus are edited to share a common delay between beginning of playback and onset of audible sound and a common duration of recording (to facilitate digital mixing with a looped continuous noise recording).

The testing corpus recordings exist as WAVE chunks in standard RIFF file format on CD-ROM and on hard disk, along with Windows-based software for test administration and analysis.

Both qualification and calibration tests have been undertaken for the new consonant identification materials in vowel ah context with one of the male talkers. The major qualification tests have included evaluations of test-to-test reliability across randomizations and assessment of degree of difficulty and sensitivity to processor differences. Difficulty and sensitivity assessments have been conducted with both 16 and 23 consonant test versions.

As an example of the testing done on our new CV materials, we present data from a two-week visit by patient SR3 devoted primarily to such testing. During that visit, SR3 was tested with a variety of processors chosen to provide an appropriate range of performance. Figure 1 and Figure 2 compare male talker percent correct scores for 16 and 23 consonant sets, respectively, in CV context with percent correct 24 medial consonant scores (also male talker) obtained in the same conditions. (Many of the same conditions also were tested with HINT sentence materials [Nillson *et al.*, 1994] at +10 dB S/N and with the female talker version of the 24 medial consonant test. Examples from those cross calibration results will be included below.)

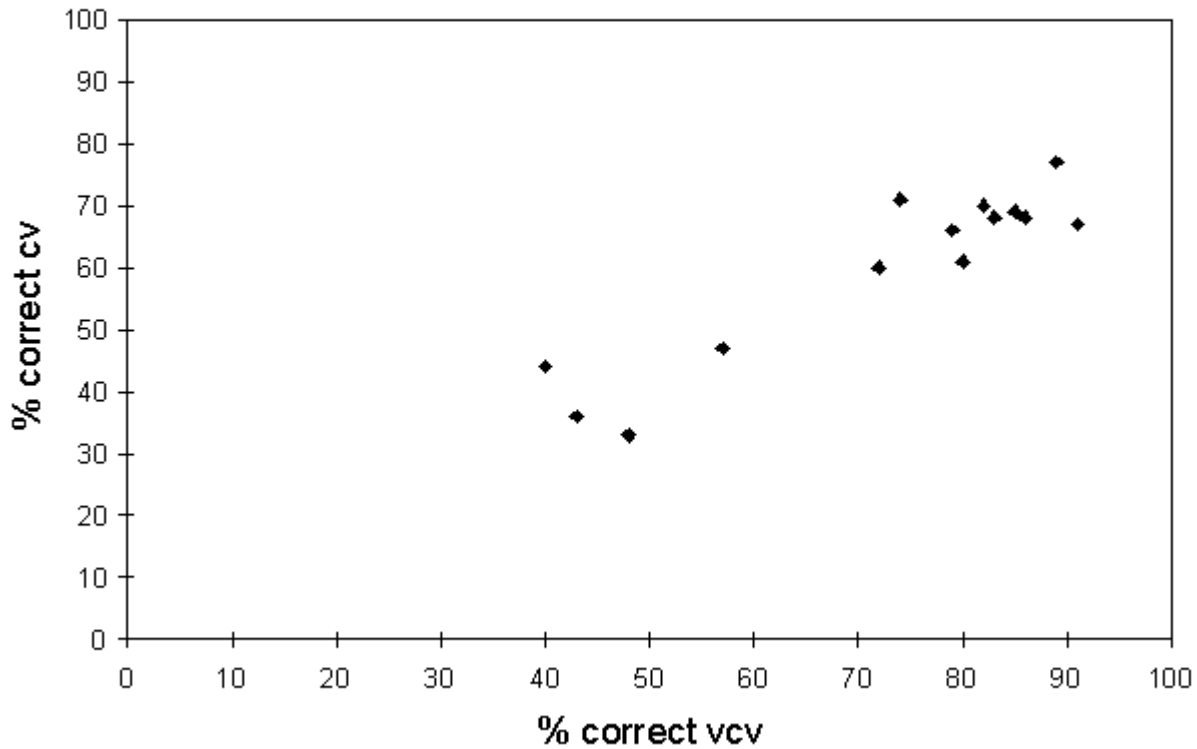


Figure 1. Percent correct identification: 16 consonants in CV context vs. 24 in VCV context.

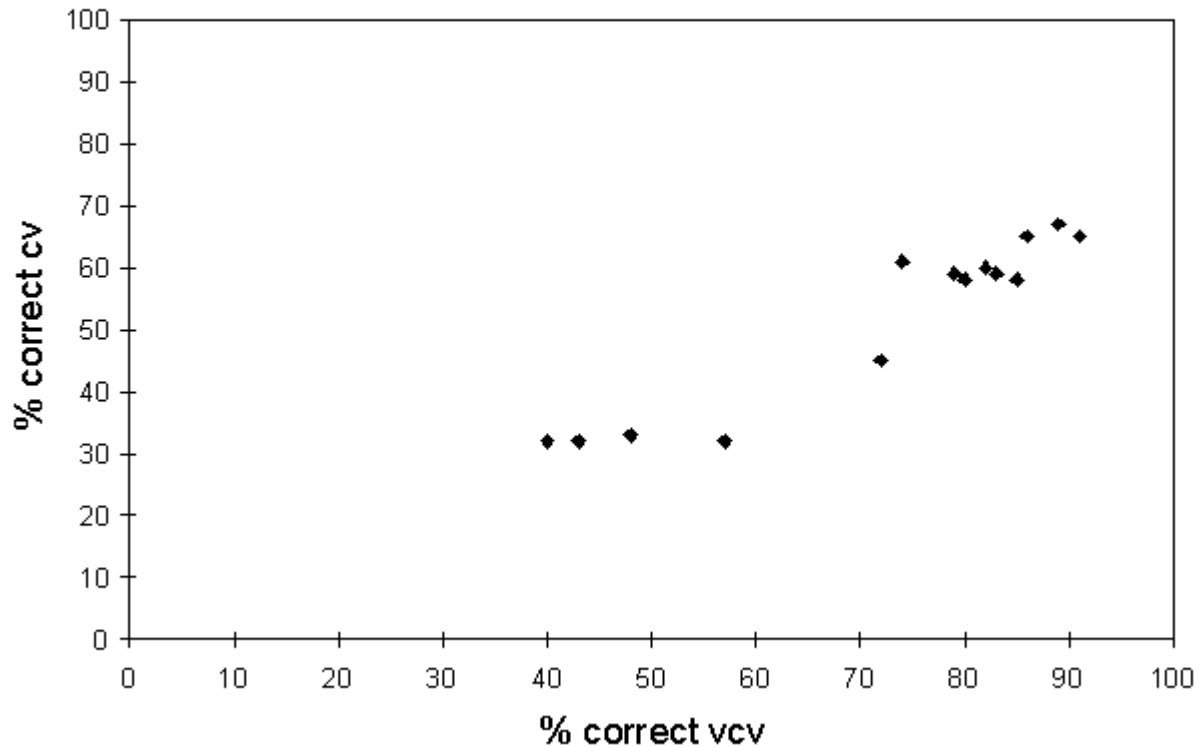


Figure 2. Percent correct identification: 23 consonants in CV context vs. 24 in VCV context

These figures indicate that with these materials the 16 consonant CV test is in fact more difficult than the 24 medial consonant one, and that the 23 consonant CV version is somewhat more difficult still. A concern with the latter data (Figure 2) is an apparent lack of sensitivity in the 23 consonant CV version over the range of 40 to 60 % in 24 medial consonant scores.

While, as we shall discuss below, there is generally a strong correlation between percent correct consonant identification scores and the corresponding overall information transmission scores, there are occasions on which the two may yield different indications of relative processor performance. Most often, this happens when a patient has had little experience with a new processor and is able consistently to discriminate differences that she or he cannot yet correctly label. This visit by SR3, which involved only the briefest exposure to each of a set of processors chosen specifically to produce a wide range in performance, produced ideal conditions for such situations.

In Figure 3 and Figure 4 we compare overall information transmission scores for the same consonant tests represented in the previous two figures. The principal conclusions are the same as before, although the range of difficulty among the three tests here appears even smaller.

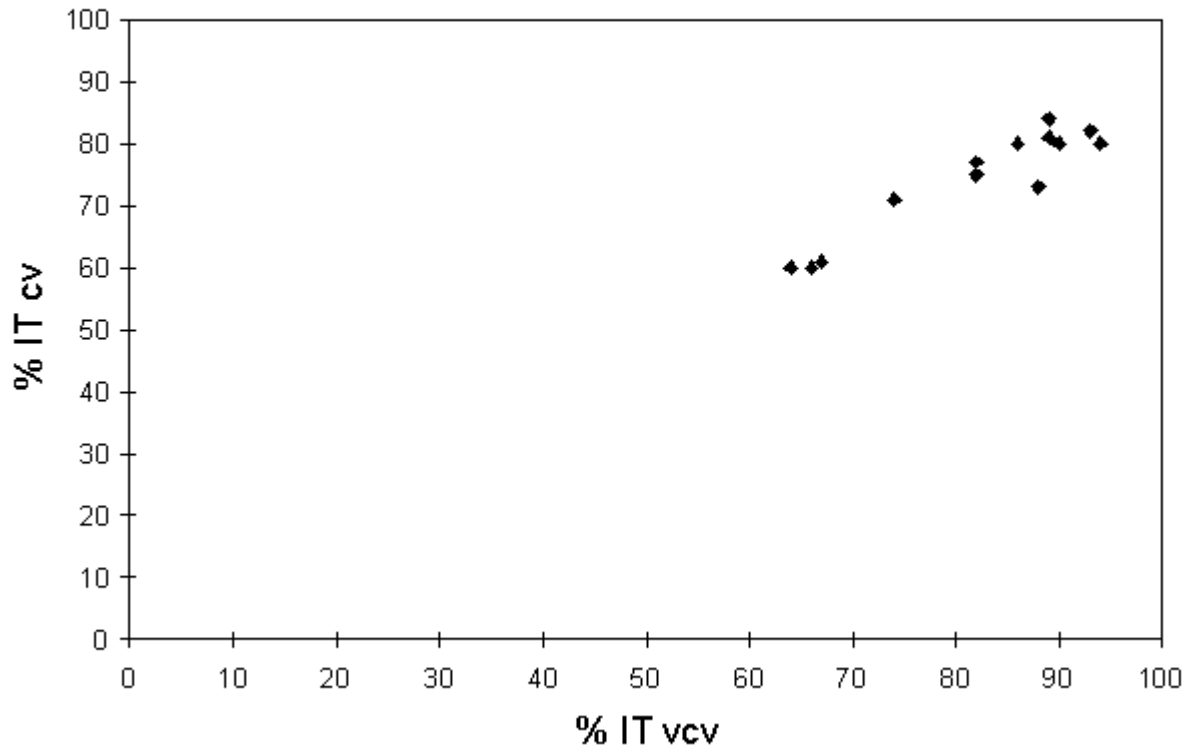


Figure 3. Percent overall information transmission: 16 consonants in CV context vs. 24 in CVC context.

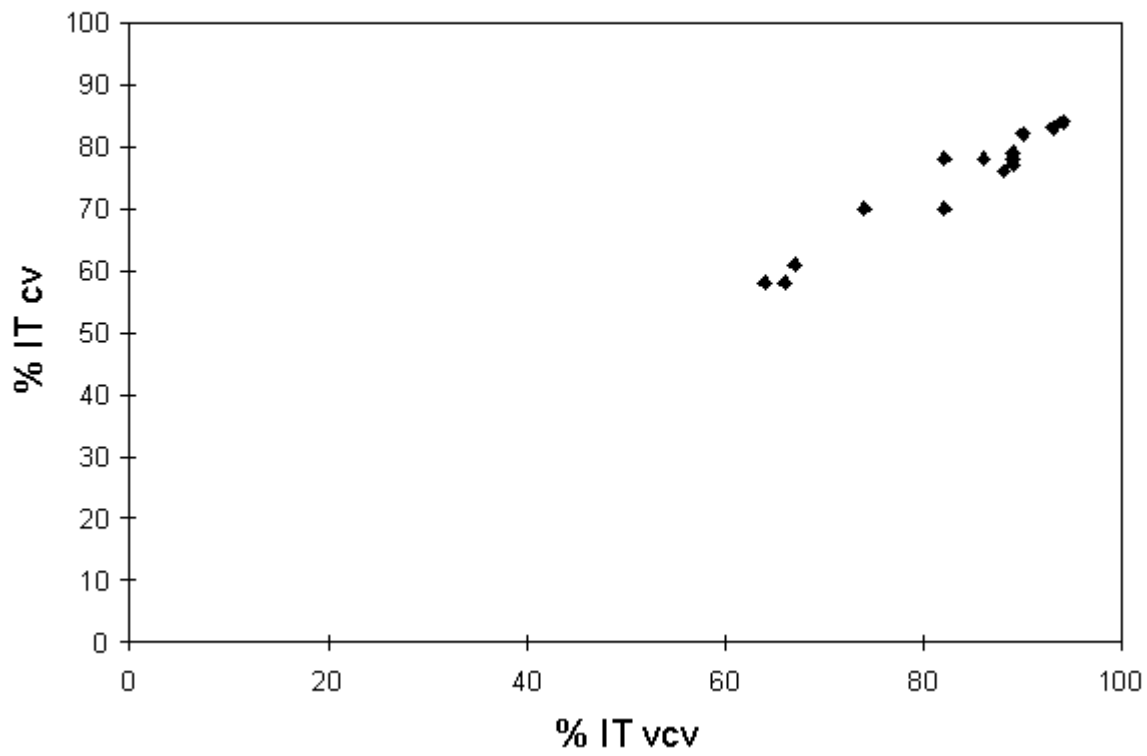


Figure 4. Percent information transmission: 23 consonants in CV context vs. 24 in VCV.

A notable feature of this analysis, however, is that it demonstrates that results from the tests with the four lowest scores are in fact consistent with the others.

Documenting Performance Levels

The second principal role of speech testing materials in our speech processor research is the measurement of levels of performance in terms that will allow comparisons across patients, implanted devices, and research groups. Tests for this purpose are not subject to two major constraints on the tests we use for fitting and optimizing processor designs: these tests do not have to be so brief because they are administered much less often, and they do not have to remain valid and sensitive when administered hundreds of times to the same subject. It is especially important, on the other hand, that these tests accurately reflect the ability to comprehend speech in real world situations in the everyday lives of the subjects [Neuman *et al.*, 1994b]. As with the rapid assessment tools, it is crucial that for future work test materials and methods be identified that will support comparisons of speech comprehension across an increasingly wide range as the best cochlear implant results continue to improve.

We shall begin our discussion with an examination of the tests we have used extensively for documenting subject performance levels with various processors.

Materials used extensively in our research. In our initial studies comparing CA and CIS processors in 11 subjects, we found that the combination of four open set tests from the Minimal Auditory Capabilities (MAC) battery [Owens *et al.*, 1985] nicely spanned a wide range of speech reception performance and provided sensitivity to relatively small differences in performance across that entire range. The four tests are recognition of 25 two-syllable words (Spondees), recognition of 100 key words in the Central Institute for the Deaf (CID) sentences of everyday speech, recognition of the final word in 50 high context sentences from the Speech Perception in Noise (SPIN) test (presented without noise in our studies), and recognition of 50 one-syllable words from Northwestern University Auditory Test 6 (NU-6). All these tests are presented from standard analog recordings of a male talker, sound alone. In the case of the NU-6 words we use only tapes 2, 3, 5, and 6 from the Cochlear Corporation set of materials. The simplest among these tests (Spondee recognition and CID sentences) provided sensitivity for the subjects at the low end of our performance range but, because of ceiling effects, were not at all sensitive to differences among the top performers. The most difficult tests in the group (NU-6 words and SPIN sentence final words), on the other hand, provided sensitivity for subjects at the upper end of the performance scale but, because of floor effects, were insensitive to differences among the poorest performers. Each subject's performance fell within the sensitive range for at least two of these four tests.

In addition, performance of the 11 subjects (and subsequent research subjects in our laboratory) is highly correlated across these four open set tests, allowing them as a group to provide a consistent and continuous scale across the entire range of their sensitivities. Correlation coefficients among our results for the four tests, all significant at $p < .001$, are as follows:

	Spondee	CID	SPIN
CID	.96		
SPIN	.92	.87	
NU-6	.92	.88	.95

While we found this group of MAC open set tests very useful and powerful in the 11 subject study and in much of our subsequent work, there are at least three ways in which these materials fall short of our ideal for documenting performance levels: (1) Each of the four sets of test materials is spoken by a single adult male voice; (2) Only one of the four tests (SPIN) is designed and calibrated for administration in noise; and (3) We are, happily, encountering more and more subjects whose performance with some processors is too high for even the NU-6 test to provide the needed measurement sensitivity.

Our use of medial consonant and CV syllable identification tests as tools for rapid and repeated assessment of relative performance among candidate processors was discussed at some length above. We also have found such tests to be a

very reliable and sensitive instrument for documenting performance levels. Overall information transmission scores for identification of such materials in a sound-alone condition are highly correlated with each of the four open set tests just discussed. Each of the following correlation coefficients is significant at $p < .001$:

	Spondee	CID	SPIN	NU-6
Consonant Overall IT	.93	.91	.92	.92

We routinely have included both male and female adult voices in such consonant identification tests, and also have presented them at various signal to noise ratios with respect to continuous multitalker speech babble and CCITT noise.

In the course of our collaborative study with Duke University Medical Center and Cochlear Corporation we have assessed some additional open set tests and calibrated them against our extensive body of data for the tests discussed above. Each of the first five subjects in the collaborative study has undergone testing to document levels of performance with at least four different processing strategies. At least two of those strategies were tested at the end of each of three visits to our laboratory, while the other two were tested at the end of at least two of the three visits. In every case the tests included recognition of the 200 or so words in 24 CUNY sentences -- both in quiet and at a +10 dB signal to noise ratio -- and recognition of 50 CNC monosyllabic words. All three of these tests used Cochlear Corporation analog recordings. For assessment and calibration purposes we also presented the NU-6 test in every case.

We shall display some examples of cross-calibrations among those materials.

It is important to note that, in all our cross-calibration tests of other instruments against the NU6 monosyllabic word test, we have used only three of the five Cochlear Corporation tapes, numbers 3, 5, and 6. We have found that the lists of 50 NU6 words on each of those three tapes have excellent test-retest reliability and sensitivity. (Tapes 1 and 2 of the original Cochlear Corporation series were recorded under noticeably different conditions than the other three, while the NU6 test on the original tape 4 used three words twice each in the same test, and also repeated all three practice items within the test.) No NU6 test was used twice with the same patient during the same visit to our laboratory. Since the NU6 word lists were with male talker, we have chosen male voice consonant materials for comparisons.

As noted above, overall information transmission scores based on our medial consonant identification tests generally correlate quite closely with the corresponding raw percent correct scores. This is illustrated in Figure 5 and Figure 6. Exceptions to this generally high correlation, also as noted above, typically involve the ability to distinguish tokens from unfamiliar processors without yet being able to label them correctly. Such exceptions might be expected to occur most often at relatively low levels of absolute performance, and would correspond to vertical displacements of points on these plots, to lower percent correct values.

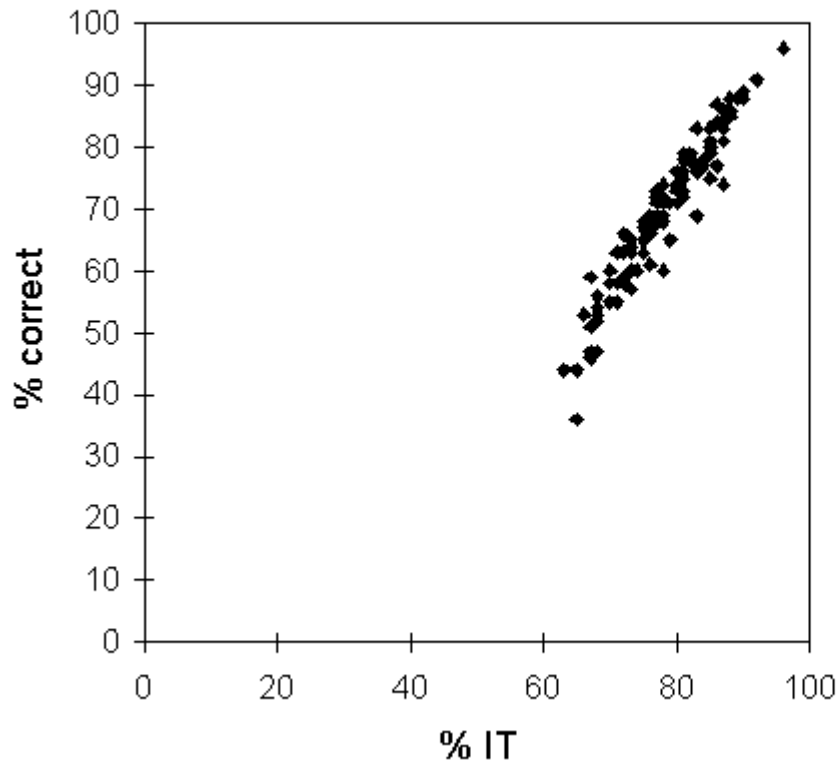


Figure 5. Percent correct vs. overall information transmission scores: identification of 16 medial consonants, male talker, patients NP3-NP5.

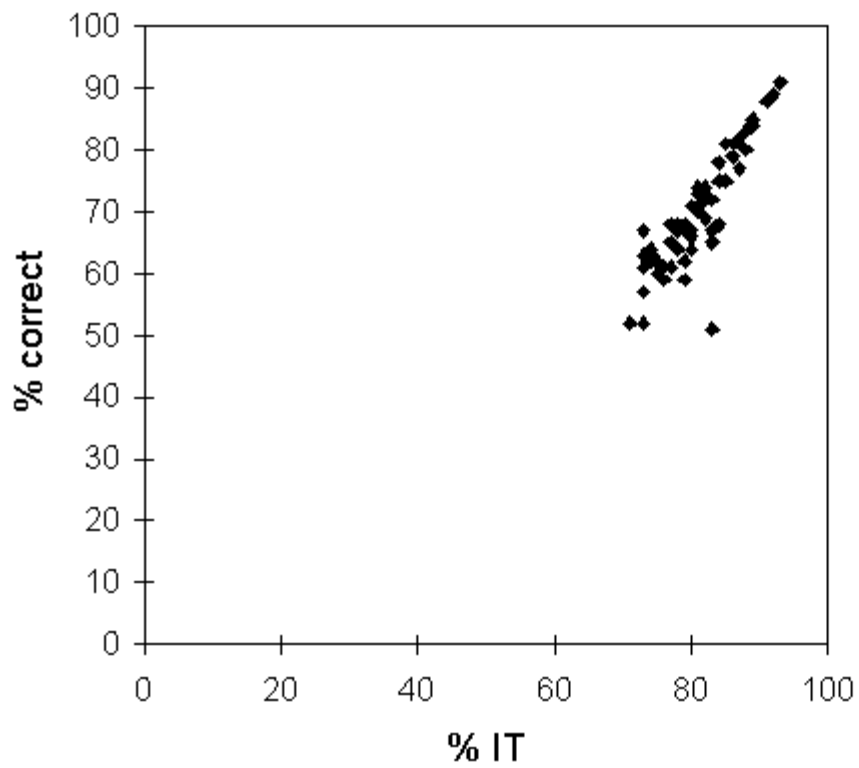


Figure 6. Percent correct vs. overall information transmission scores: identification of 24 medial consonants, male talker, patients NP1 and NP2.

In Figure 7 and Figure 8 we compare 16 medial consonant identification data to NU6 monosyllabic word identification scores for the same conditions in the first five patients of the Nucleus percutaneous series. Clearly, there can be ceiling effects for 16 consonant identification data at open set performance levels as low as 30% in terms of NU6 word identification scores.

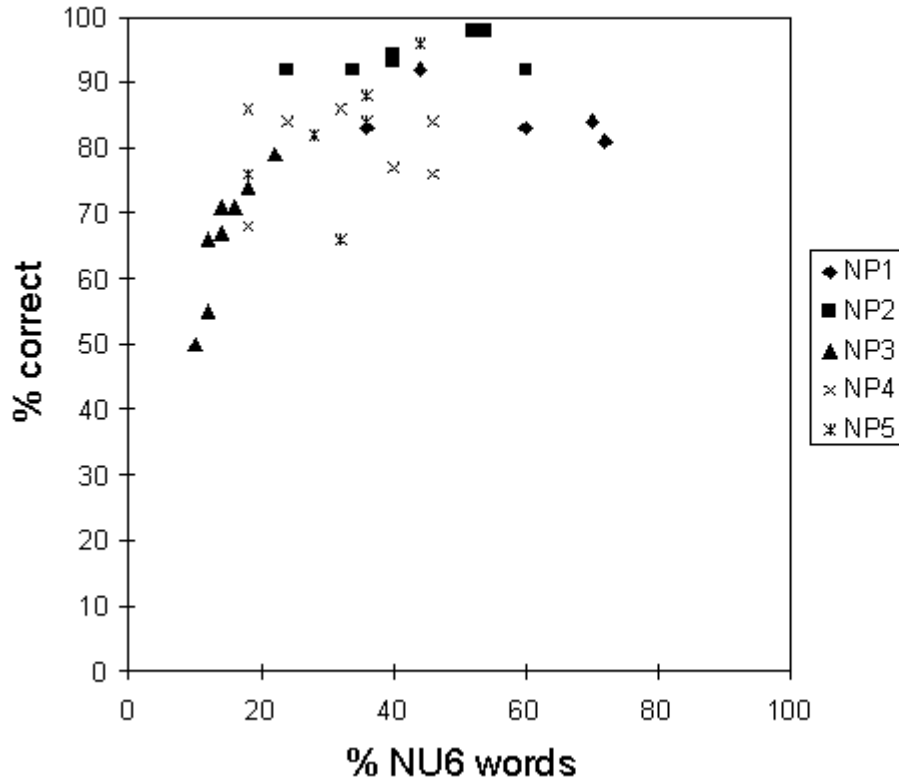


Figure 7. Identification of 16 medial consonants as a function of NU6 monosyllabic word scores: male talker, patients NP1 – NP5.

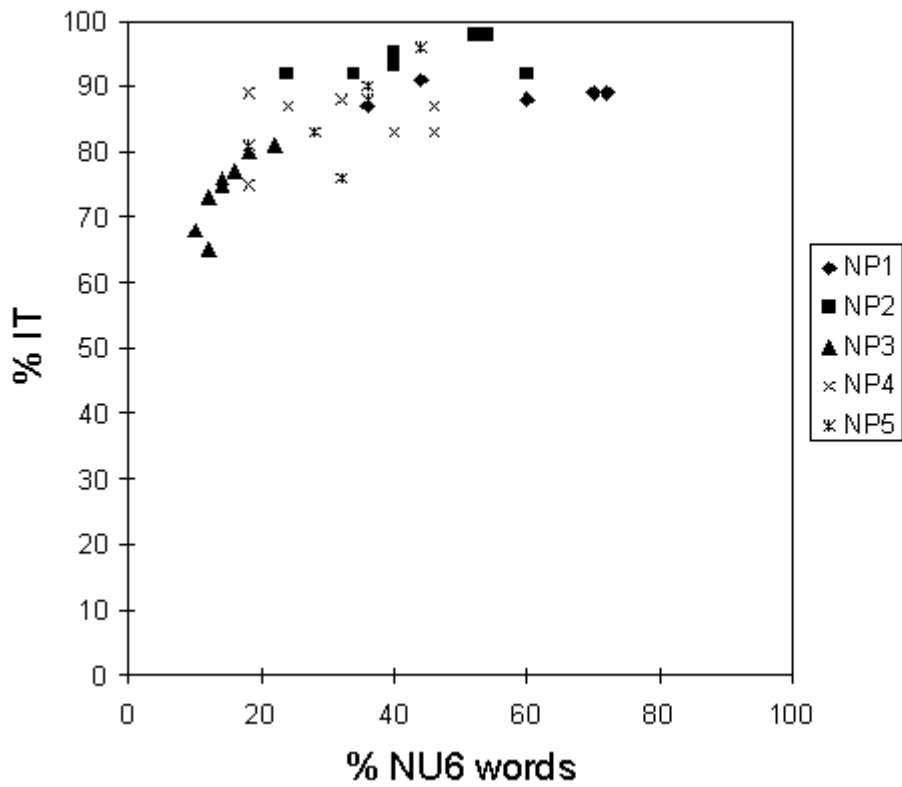


Figure 8. Overall information transmission scores, based on identification of 16 medial consonants, as a function of NU6 monosyllabic word scores: male talker, patients NP1 – NP5.

In Figure 9 are shown data of the same sort for identification of 24 medial consonants by two of the same patients.

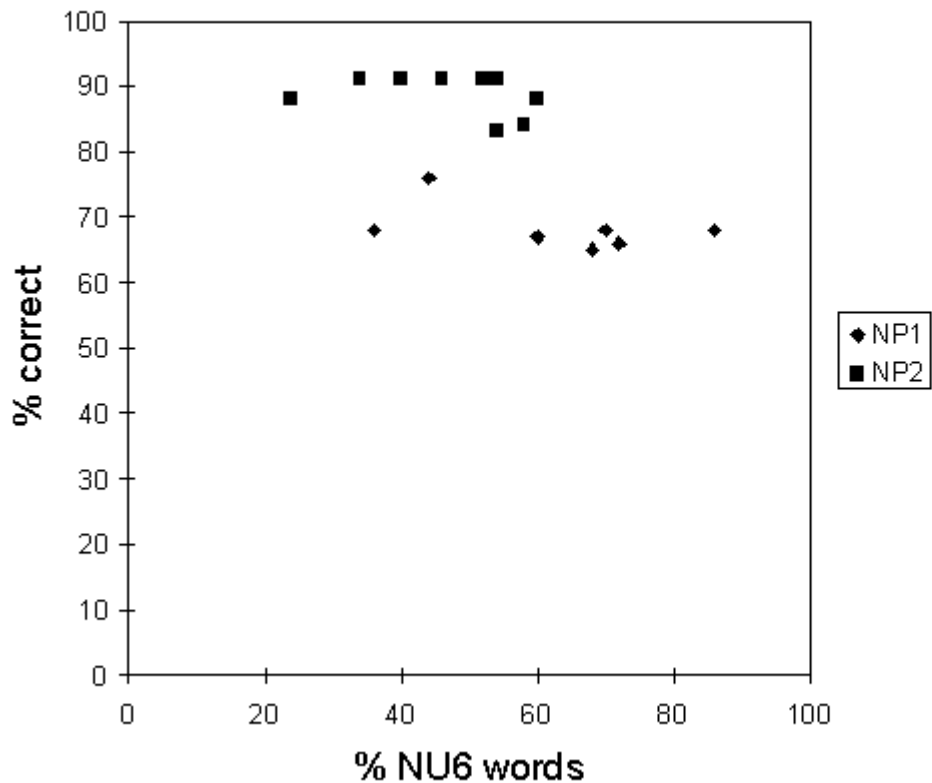


Figure 9. Identification of 24 medial consonants as a function of NU6 monosyllabic word scores: male talker, patients NP1 and NP2.

These data illustrate dramatically that certain patients can display different characteristic relationships between consonant identification and open set scores. Also evident here is that sometimes consonant identification tests may not distinguish among processors that have substantial differences in open set performance.

[It should be mentioned that these data were obtained in an unusual context, in that these two patients were being asked to switch frequently between taking 16 and 24 consonant tests. The study protocol required the former, while these two patients' levels of performance indicated use of the latter. Our impression at the time was that both subjects were kept off-balance by the repeated disappearance and reappearance of the additional 8 tokens and available responses.]

Similar data for two Ineraid patients – SR3 and SR10 – are shown in Figure 10. Here the picture is one of sensitivity comparable to that of the 16 medial consonant version, but with ceiling effects absent at levels of open set performance corresponding to NU6 scores less than 60 to 70%.

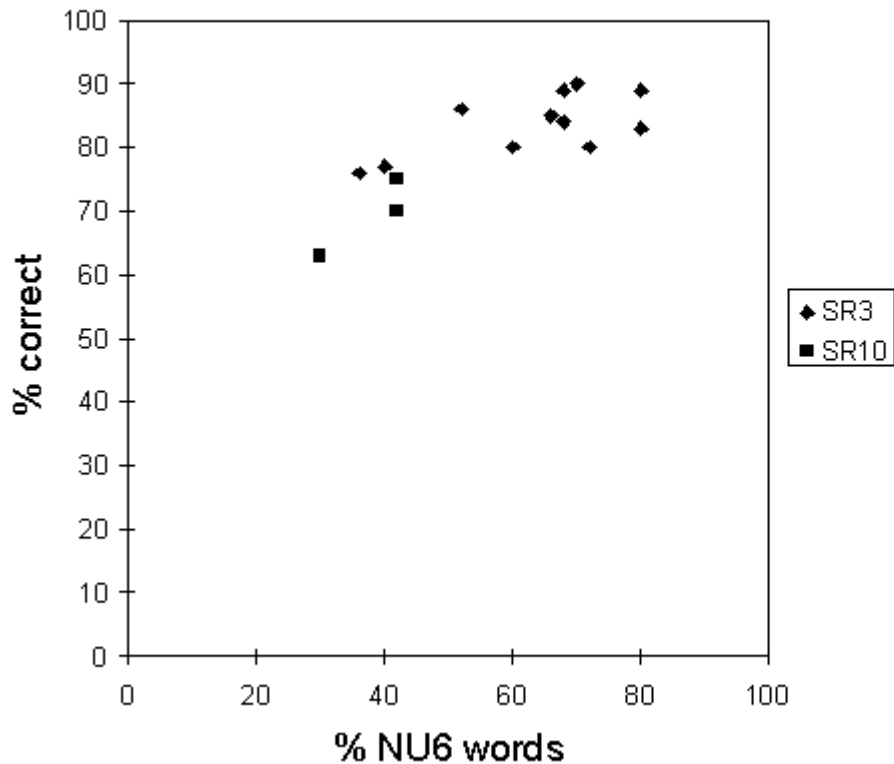


Figure 10. Identification of 24 medial consonants as a function of NU6 monosyllabic word scores: male talker, patients SR3 and SR10.

We will continue to collect cross-calibration data among 24 medial consonant, 23 CV syllable, and NU6 tests with appropriate patients as opportunities arise.

Another illustration of substantial differences in open set performance that may not be revealed by consonant identification testing is included as Figure 11. Here the test being compared to NU6 monosyllabic word identification is one using sentences from the HINT [Hearing in Noise Test, Nillon *et al.*, 1994]. Those materials were designed for use in an adaptive test whose result is a S/N ratio in dB corresponding to a particular performance level. In that context, whether or not all the key words of the previous sentence were correctly identified determines the S/N ratio for the next sentence presented. We and other research groups have used these sentence lists at constant S/N ratios with cochlear implant patients, and scored them in percent correct word identification terms.

While we have more recently been using S/N ratios of +3 dB and +6 dB, the data of Figure 11 are for +10 dB. Clearly, this test using HINT sentences achieved greater sensitivity to processor differences than NU6 administered in quiet. Each point in this figure corresponds to the administration of two HINT sentence lists, comprising from 103 to 107 words. We regard the use of two such lists as an absolute minimum for obtaining a reliable result for comparison purposes. This special use of the HINT materials is inherently more sensitive to sentence-to-sentence variations than the adaptive procedure for which they were created, and we have observed substantial test-to-test variation in scores. One afternoon's scores recently, for example, included pairs of scores in the same conditions of 63 and 70%, 78 and 86%, 66 and 62%, but also a pair of 70 and 58% and a sequence of three tests in the same condition of 82, 53, and 65%. We plan to continue to use these materials, but with the awareness that significant time will be required to obtain each accurate measurement, as well as consumption of a significant fraction of the available materials, which are even more memorable than NU6 word lists.

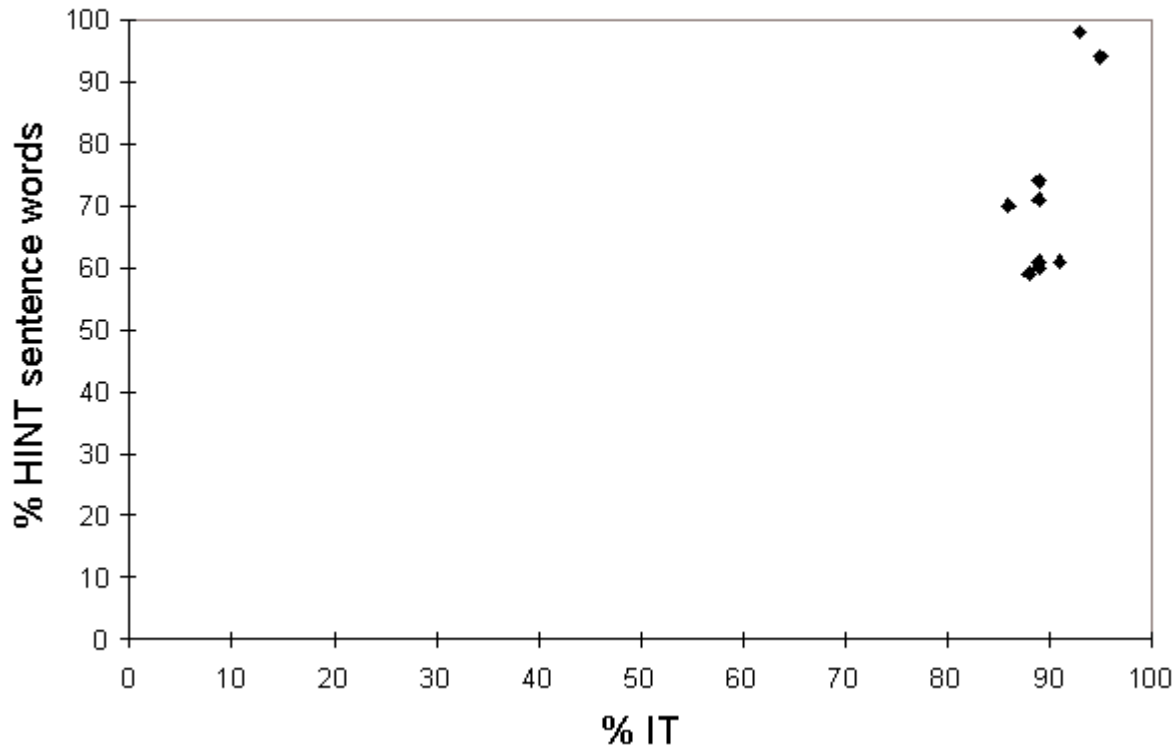


Figure 11. Identification of words in HINT sentences presented at +10 dB S/N with respect to speech spectrum noise, vs. overall information transmission scores based on identification of 24 medial consonants in the same conditions: male talker, patient SR3. Each point corresponds to the administration of two HINT sentence lists, including 103 - 107 words.

The protocol for our Nucleus percutaneous studies included use of the CNC lists of monosyllabic words. Scores from those tests are compared with NU6 results for the same conditions in Figure 12. While the CNC materials have the advantage of offering a much greater number of different lists without repetition, and of being somewhat more difficult than NU6 on the average, we have found the former to have poorer test-retest consistency.

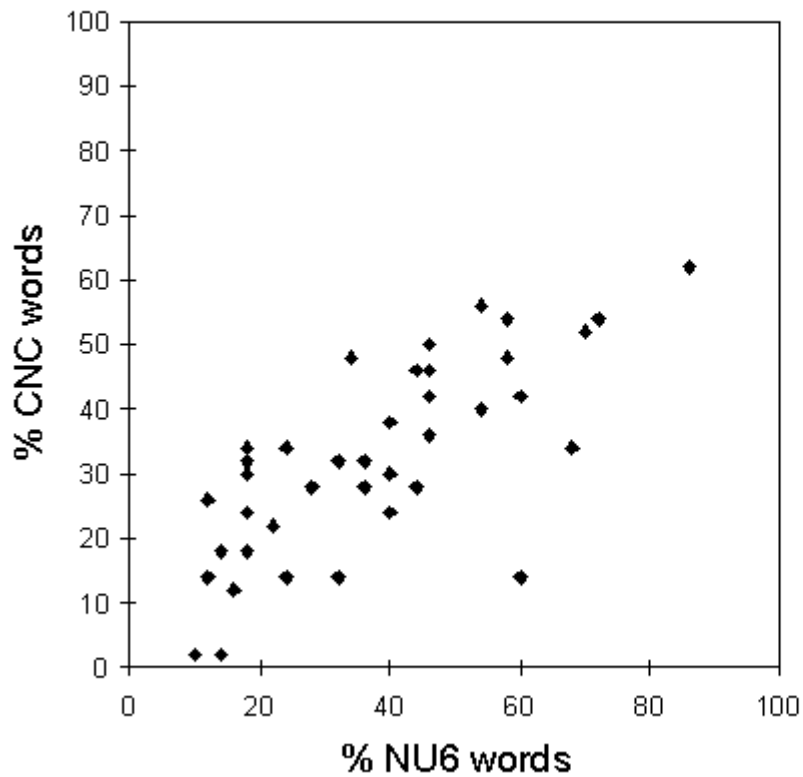


Figure 12. Identification of CNC monosyllabic words vs. identification of NU6 monosyllabic words for the same conditions.

Our cross-calibration studies have been very useful in assessing the relative merits of different tests and materials. We shall cite a few examples, displaying results from several other tests on our standard NU6 word identification scale.

Figure 13 shows such a comparison for the CUNY sentences in quiet. Clearly this test shows an abrupt transition at a level of performance corresponding to a NU6 score of about 20% and shows little sensitivity elsewhere.

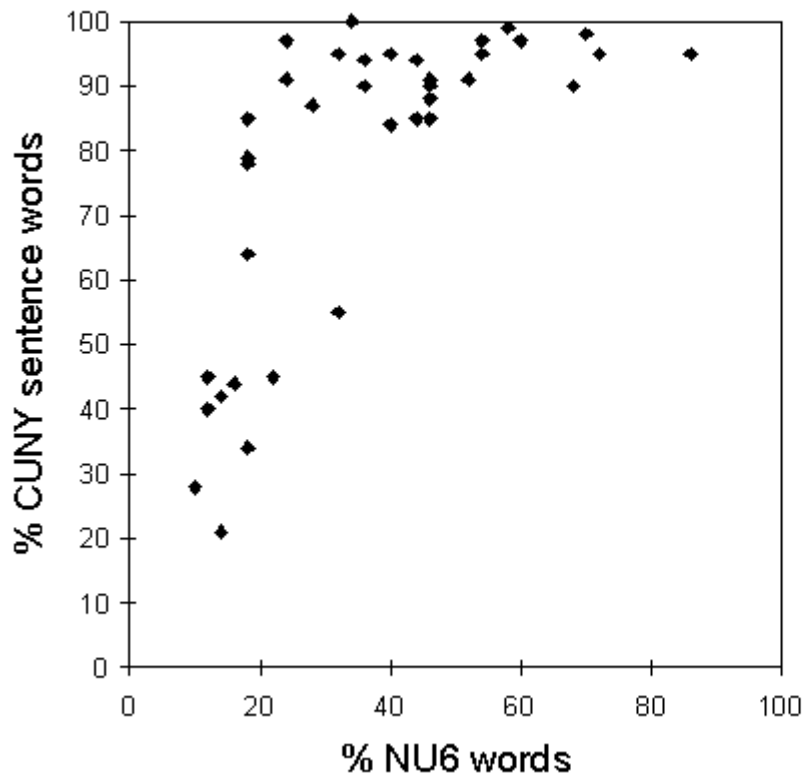


Figure 13. Identification of words from CUNY sentences in quiet, as a function of NU6 monosyllabic word scores: patients NP1 – NP5.

The CID sentences, also administered in quiet in earlier testing in our laboratory, show a somewhat more gradual transition, with substantial sensitivity for performance levels up to 40% or so on the NU6 scale. This is shown in Figure 14.

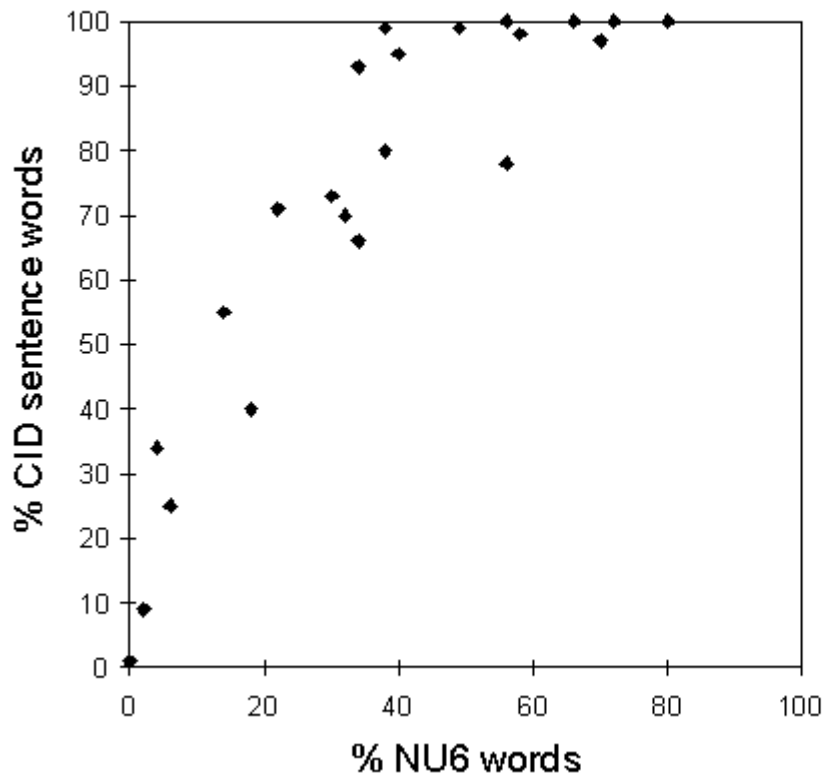


Figure 14. Identification of words from CID sentences in quiet, as a function of NU6 monosyllabic word scores: patients SR1 – SR11.

When the CUNY sentences were administered at +10 dB S/N using a prepared tape provided by Cochlear Corporation, the results shown in Figure 15 were obtained. Again there appears to be an abrupt transition at a performance level corresponding to a NU6 score of about 20%, with a lot of scatter among higher performance data.

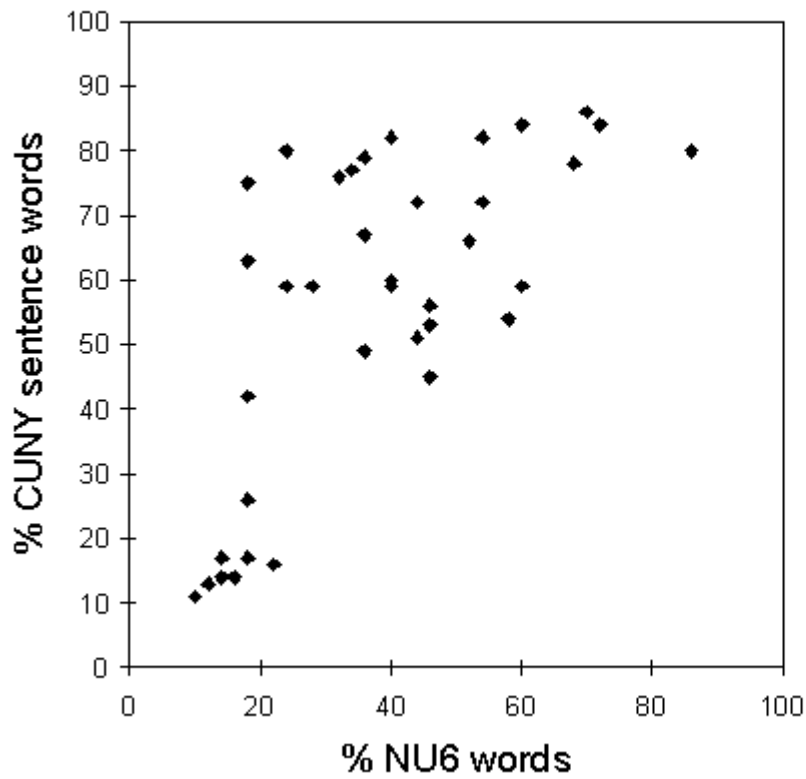


Figure 15. Identification of words from CUNY sentences presented at +10 dB S/N with respect to multitalker babble, as a function of NU6 monosyllabic word scores: patients NP1 – NP5.

In previous work in our lab, the SPIN sentences presented without noise provided more useful sensitivity up to a NU6 score of 70% or so, as shown in Figure 16.

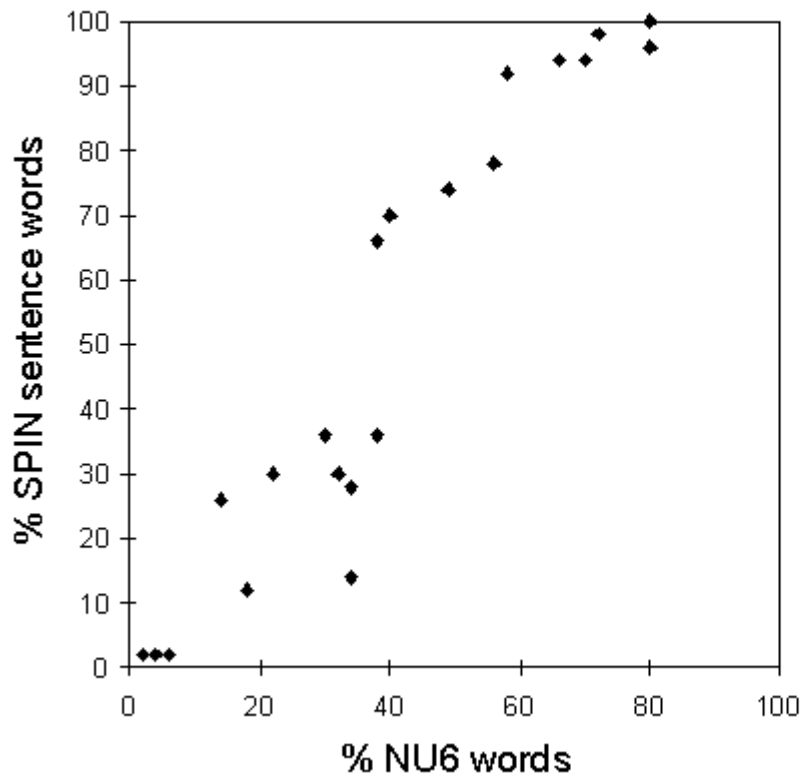


Figure 16. Identification of words from SPIN sentences in quiet, as a function of NU6 monosyllabic word scores: patients SR1 – SR11.

Two final figures are provided for comparison with previous ones, to allow consideration of the possible impact of using NU6 rather than CNC as a reference standard. Figure 17 plots results for the CUNY sentences in quiet with respect to CNC monosyllabic word identification, for comparison with Figure 13. Figure 18 provides a similar comparison for Figure 15, plotting results for CUNY sentences at a S/N ration of +10 dB.

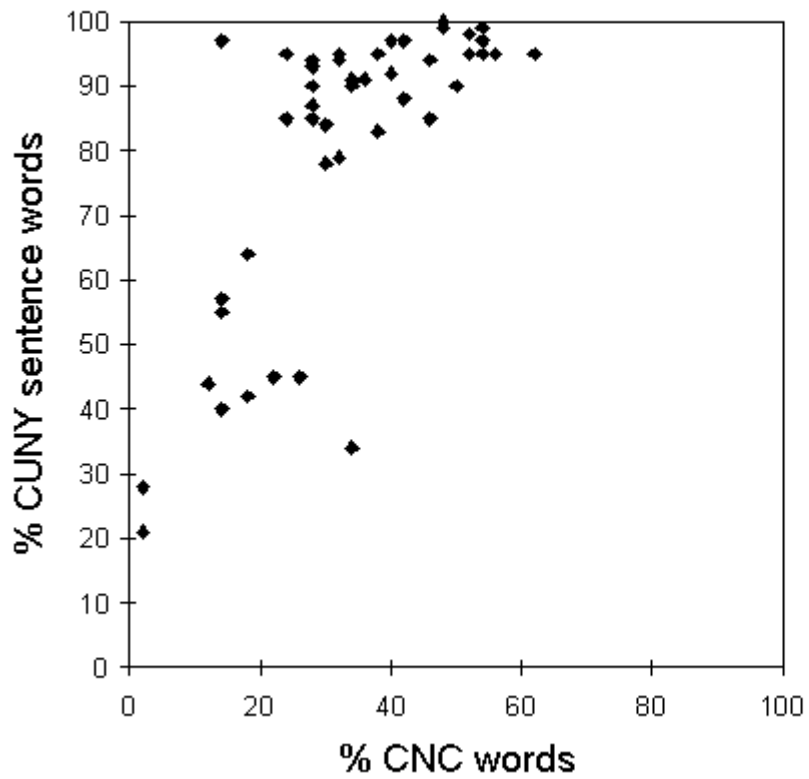


Figure 17. Identification of words from CUNY sentences in quiet, as a function of CNC monosyllabic word scores: patients NP1 – NP5.

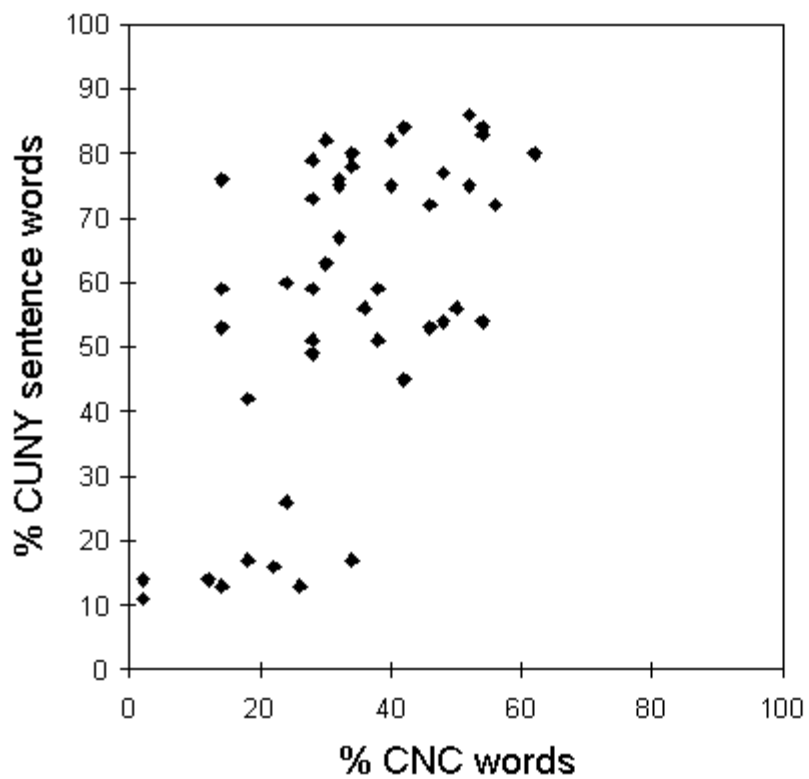


Figure 18. Identification of words from CUNY sentences presented at +10 dB S/N with respect to multitalker babble: patients NP1 – NP5.

Special German language materials. As discussed in more detail in other QPRs, we now have considerable experience with the use of the German language HSM sentence materials [Westra CD 15] presented at various S/N ratios with respect to CCITT speech spectrum noise. Those materials also were processed for us by Sig Soli and Michael Nillson of the House Ear Institute for use with hardware and software developed jointly by HEI and Starkey Laboratories, Inc., for binaural testing. Briefly, both speech and noise underwent signal processing to apply head-related transfer functions (HRTFs) appropriate to various angles of sound incidence at a listener's ears. The special processed recordings were presented using the HEI/Starkey hardware and software designed for presentation of the English HINT materials discussed above. Our experience indicates similar patterns of list-to-list variation in word identification scores for the German and English sentences. In association with the same studies of a German patient (ME2), we gained experience with the use of a set of German monosyllabic words – the Freiburger words [Westra CD1]-- both in quiet and at fixed S/N ratios with respect to CCITT noise.

Other tests employed in our laboratory. We have on occasion used the Speech Perception in Noise (SPIN) sentence materials in their intended form, at a +8dB signal to noise ratio with respect to continuous multitalker speech babble. We now have the facilities to administer such tests at any convenient S/N with respect to babble, speech spectrum noise, or any other form of competing noise.

We also have conducted testing using the analog recordings of the NU-6 word lists mixed with an analog recording of the same multitalker speech babble used in the SPIN tests. The results were encouraging, but we were concerned by the variability in noise context from one administration of the test to the next (*i.e.* lack of synchronization of the test tokens with the noise recording).

In addition to the open set tests discussed above, we also have used certain closed set subtests of the MAC battery extensively. However, as the highest absolute performance levels of our subjects have tended to increase over the years, we have found fewer instances in which such segmental tests as the initial consonant, final consonant, and vowel MAC subtests were difficult enough to provide the necessary sensitivity to performance differences. We are quite familiar also with the everyday sounds subtest of the MAC battery, but have not found such tests to be of any help in improving processors. Other MAC subtests, such as the same/different and four-choice spondee tests, question/statement, noise/voice, and accented word tests are so simple as to be insensitive for an increasingly large majority of adult postlingually deafened users of auditory prostheses.

We have evaluated a considerable number of other available speech test materials and techniques in our continuing effort to improve the sensitivity and reliability and extend the range of such testing of cochlear and brainstem implant subjects. Among these were two instruments developed by Arthur Boothroyd and colleagues at CUNY -- the AB word lists and the Speech Pattern Contrasts (SPAC) tests [Boothroyd, 1987].

We analyzed each of 15 lists of ten AB words for occurrences of initial and final consonants and then grouped them in sets of three, yielding five 30-word groups. Four of those five groups included all the final consonants, with the remaining group lacking two. Two of the groups lacked single (different) initial consonants, two other groups lacked two (different) initial consonants each, with the final group (the same one lacking two final consonants) lacking three initial consonants. Thus four of the five groups were especially well matched for our purposes. Within each group were from one to three pairs of rhyming words, potentially memorable enough to compromise the test's results in repeated use with the same subject. We wrote our own software to administer tests based on these groups of AB words, using the laserdisc recordings and onset and offset addresses for each word as supplied by CUNY. Our software provided a cycle of six different randomizations for each of the groups, and allowed choice of any of the four talkers available on the CUNY laserdisc. A researcher entered the subject's oral responses via keyboard and a side-by-side list of presented words and responses was later printed out for computing phoneme scores and correcting word scores for homonyms. In a week of extensive tests with one high performance subject, involving 12 different processor designs, we conducted repeated AB word list tests with male and female talkers and compared the results with those from NU-6 monosyllabic word tests with a male talker. We found that, in this form and for our purposes, the AB word lists provided roughly the same difficulty range as the NU-6 test and produced word and phoneme scores with somewhat more test-retest variability.

We found the SPACII test particularly elegant in its design and evaluated it with several of our subjects -- users of three different clinical devices representing a wide range of implant performance -- in the hope that it might prove even more useful than our medial consonant identification tests as a guide to processor optimization. We wrote our own software to administer and score the test, using the CUNY laserdisc recordings. For subjects with relatively high performance levels, the test was not difficult enough to produce the errors needed for analysis. For subjects with poorer levels of performance, differences between processors had to be relatively large in order to produce reliable differences in SPACII scores.

We have conducted vowel identification tests with synthetic vowel recordings supplied by Michael Dorman. While we regard such tests as valuable, the higher correlations between consonant scores and open set word and sentence performance than is the case for vowel scores has led us to continue to concentrate on testing materials that emphasize consonants.

Definitely among the speech test materials possessing the greater difficulty we seek are the Harvard/IEEE sentences. Those materials were very useful in studies with one of our highest performance subjects. Many of the sentences are highly memorable, however, making them unsuitable for repeated use with the same subject. [We used only sets of sentences supplied by William Rabinowitz as never having been presented to our common research subject, who delights in recalling some of these sentences, even years after hearing them once.] In addition to the fact that these sentences should not be reused, there is some concern that the difficulty of these materials may vary significantly depending on a subject's education and socio-economic status.

We once evaluated an additional assortment of speech testing materials for appropriate difficulty with a single subject using a single processor providing a high level of performance. The materials included stimulus variability tests supplied on analog tape recordings by colleagues at Indiana University (IU) [Sommers *et al.* 1992], and selected tests from the Department of Veterans Affairs (VA) audio compact discs developed for assessment of auditory function [RH Wilson, 1993]. A table of percentage correct scores is included as an indication of each test's potential sensitivity for subjects currently among those enjoying the highest levels of cochlear implant performance.

IU Rate Tests:		
Medium Rate: 52%		
Mixed Rates	Slow: 58%	Medium 46% Fast: 46%
IU Voice Tests:		
Single Voice	Easy: 46%	Hard: 40%
Multiple Voices	Easy: 39%	Hard: 20%
IU Harvard Sentences		
Single Talker 87%		
Multiple Talkers 71%		
VA Processed NU-6	Word	Phoneme
Low-Pass 1500 Hz	28%	48%
High-Pass 2 kHz	22	52
45% Compression	48	71
45% Compr + 0.3s Reverb	40	65
65% Compression	40	65
65% Compr + 0.3s Reverb	16	46
VA NU-6 Female Talker	78	91
VA Maryland CNC Male Talker	76	91
[<i>cf.</i> NU-6 Malke Talker, CC tape	94	97]
VA CID W-22 Male Talker	70%	

VA Synthetic Sentence ID	64% word
[cf. CUNY everyday sentences	99%]
[cf. SPIN low predictability sentences	72%]
.....	

Evaluation of the IU tests of the effects of intratest variations in speech rate and talker was complicated by errors in the scoring sheets at that time. The scores shown above include all corrections noted on the sheets supplied to us and take into account a number of additional errors we found. Both mixed rate and multiple talker conditions produced decrements in our subject's scores and merit further attention and development of test materials.

Among the VA materials a variety of signal processing manipulations were seen to increase the difficulty of the NU-6 monosyllabic word recognition test. For our purposes, of course, the low-pass and high-pass filtering and compression operations raise issues of relevance to the everyday hearing performance we seek to gauge. The presence of a 0.3 second reverberation time, on the other hand, is quite relevant to everyday listening tasks. The VA compact discs include reverberation only in conjunction with two different levels of compression, however, and our results indicate that the effect of 0.3s reverberation on NU-6 scores is a strong function of the amount of that compression. We expect that the unprocessed NU-6 lists with a female talker on the VA CD will prove very useful in our work, but will not alone provide sensitivity over the expanded difficulty range we require.

Similarly, these comparison results indicate that the Maryland CNC lists with a male talker may provide a useful adjunct to the limited number of male talker NU-6 lists we have been using. [A contemporaneous score for the same subject and same processor for the NU-6 test using the Cochlear Corporation analog tape recordings is included in the table for comparison.]

Also included in the table are two additional tests from the VA discs and two from other sources included in the same assessment of relative difficulty.

New Materials. We are making male and female talker recordings of the individual words in the NU-6 lists for presentation in various randomizations under computer control, including presentations at fixed synchronization points with respect to continuous digital noise loops. This should allow us to extend the range of a proven measuring instrument without decreasing its sensitivity or test-to-test reliability through noise context variations. Only minimal modification would be required to existing software for NU-6 word testing in continuous noise using digital recordings.

We also expect that, just as has been the case with the medial consonant identification tests, the new CV syllable consonant identification tests in quiet and in reproducible continuous noise contexts will prove valuable in documenting performance levels as well.

References

- Boothroyd A (1987) Perception of speech pattern contrasts via cochlear implants and limited hearing. *Ann. Otol. Rhinol. Laryngol.* 96, Suppl. 128: 58-62.
- Boothroyd A, Nittrouer S (1988) Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am.* 84: 101-114.
- Lawson DT, Wilson BS, Finley CC (1989) Speech processors for auditory prostheses. *Fourteenth Quarterly Progress Report, NIH Project N01-NS-5-2396*. Bethesda, MD, National Institutes of Health, Neural Prosthesis Program.
- Neuman AC, Levitt H, Dillon H, Rubin-Spitz J (1994b) Evaluation of speech materials for hearing aid assessment. Preprint, submitted to *J. Acoust. Soc. Am.*
- Nilsson M, Soli SD, Sullivan JA (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.

Owens E, Kessler DK, Raggio M, Schubert ED (1985) Analysis and revision of the Minimal Auditory Capabilities (MAC) battery. *Ear Hear.* 6: 280-287.

Tyler RS, Preece JP, Lowder MW (1987) *The Iowa Audiovisual Speech Perception Laser Videodisc*. Iowa City, IA, University of Iowa Hospitals and Clinics, Department of Otolaryngology -- Head and Neck Surgery.

Sommers MS, Nygaard LC, Pisoni DB (1992) The effects of speaking rate and amplitude variability on perceptual identifications. *J. Acoust. Soc. Am.* 91: 2340ff.

Wilson BS, Lawson DT, Finley CC (1990) Speech processors for auditory prostheses. *Fourth Quarterly Progress Report, NIH Project N01-DC-9-2401*. Bethesda, MD, National Institutes of Health, Neural Prosthesis Program.

Wilson RH (1993) Development and use of auditory compact discs in auditory evaluation. *J. Rehab. Res. Devel.* 30: 342-351.

III. Plans for the Next Quarter

Our plans for the next quarter include the following:

- Ongoing studies with Ineraid subject SR2 (usually one morning per week). Studies anticipated for the next quarter include evaluation of a series of single-channel, *continuous sampling* processors, with systematic manipulations in pulse rate and in the cutoff frequency of the lowpass filter in the envelope detector, over wide ranges. One purpose of these studies is to evaluate effects of rate of stimulation in the absence of electrode interactions. (Interactions can be exacerbated by increases in rate for multichannel processors, and we want to provide a control for that effect.) Additional studies may include (as time permits) continued investigation of several new processor designs, such as designs using "conditioner" stimuli in conjunction with the normal outputs of CIS processors.
- Studies with Ineraid subjects SR3 (weeks of May 11 and May 18) and SR9 (weeks of July 20 and July 27), including electrophysiological, psychophysical and speech reception measures. Both of these subjects have participated in many prior studies in our laboratory. SR3 enjoys high levels of speech reception performance with her implant, whereas SR9 is in the low-performance category. Studies with these subjects will extend the prior studies.
- Continued studies with subject NU-4, a local patient with standard transcutaneous Nucleus implants on both sides.
- Resumption of studies with one or more subjects implanted with the Clarion device. The studies will include comparisons among (1) the Clarion implementation of a CIS-like strategy in conjunction with monopolar electrodes, (2) the CIS-like strategy in conjunction with "enhanced bipolar" electrodes, (3) the "simultaneous analog strategy" (SAS) of the Clarion device in conjunction with monopolar electrodes, and (4) the SAS in conjunction with enhanced bipolar electrodes. All processors will use the same number of channels. The processors will be implemented and applied using our laboratory speech processor system and fitting procedures.
- Presentation of a keynote lecture at the 4th *European Symposium on Paediatric Cochlear Implantation*, to be held in 's-Hertogenbosch, The Netherlands, June 14-17.
- Discussion of collaborative projects with the Innsbruck Team (*e.g.*, our joint studies with recipients of bilateral implants), during a trip by Wilson to Innsbruck following the 's-Hertogenbosch conference.
- Continued analysis of speech reception, evoked potential, and psychophysical data from prior studies.
- Continued preparation of manuscripts for publication.

IV. Acknowledgments

We thank subjects NU-4, SR2, SR9, SR15 and SR16 for their participation in the studies of this quarter.

Appendix 1. Summary of Reporting Activity for this Quarter

Reporting activity for the last quarter, covering the period of February 1 through April 30, 1998, included the

following:

Publication

Wilson BS, Rebscher S, Zeng F-G, Shannon RV, Loeb GE, Lawson DT, Zerbi M. Design for an inexpensive but effective cochlear implant. *Otolaryngol Head Neck Surg* 118: 235-241, 1998.

Invited Presentations

Lawson DT: Measures of thresholds in the context of multichannel stimulation and application of those measures in the fitting of speech processors for cochlear implants. University of Iowa, Department of Otolaryngology, Head & Neck Surgery, Iowa City, IA, February 20, 1998.

Wilson BS, Pierschalla M: Development of cochlear prostheses. Invited poster, NIH Bioengineering Symposium, *Building the Future of Biology and Medicine*, Bethesda, MD, Feb. 27 and 28, 1998.