Eighth Quarterly Progress Report

October 1, 1997, through December 31, 1997

Speech Processors for Auditory Prostheses

NIH Contract N01-DC-6-2100

submitted by

Donald K. Eddington
William R. Rabinowitz
Joseph Tierney
Victor Noel
Margaret Whearty

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, MA

1.0 Introduction

Work performed with the support of this contract is directed at the design, development, and evaluation of speech processors for use with auditory prostheses implanted in deaf humans. Major research efforts are proceeding in four areas: (1) developing and maintaining a laboratory based, software controlled, real time, speech processing facility where processor/stimulator algorithms for monaural and binaural eight-channel implants can be implemented/tested and a wide range of psychophysical measurements can be made, (2) using the laboratory facility to refine the sound processing algorithms used in the current commercial and laboratory processors, (3) using the laboratory facility to explore new sound processing algorithms for implanted subjects, and (4) designing and fabricating programmable, wearable speech processors/stimulators and using these systems to: (a) field test processor algorithms developed and tested in the laboratory, (b) evaluate the effects of learning using longitudinal evaluations of speech reception, and (c) compare asymptotic performance of different speech processors across subjects.

A good deal of this contract period was spent preparing an application in response to the Request for Proposals for the next three-year period of this contract area. In this report we present material describing work directed at the acoustic simulation of speech processors used in cochlear implant systems. We expect the results from this line of work to: (1) suggest refinements of present sound processing systems and (2) aid in the exploration of new sound processing algorithms.

2.0 Acoustic Simulations, Background

Recently there has been considerable interest in developing acoustic simulations of speech processors used in cochlear implant systems (e.g., Dorman et al., 1997; Shannon et al., 1995). The input acoustic signal is processed initially as in an implant speech processor (typically, band-pass filtering and envelope detection), but instead of delivering the frequency-specific information from each channel to a different electrode, the channel inputs modulate tonotopically ordered acoustic carriers. The carriers can be either sinusoids or bands of noise. The modulated carriers are summed to form a composite signal which is presented acoustically. The intent of these simulations is to enable normal-hearing listeners to experience an approximation of what an implant subject hears. If successful in reaching this goal, normal listeners could gain insight into the capabilities and limitations of a given processing system and, in addition, these systems could be used to explore the consequences of a range of processing manipulations.

We utilized acoustic simulations to explore the effects of different numbers of CIS (Wilson et al., 1991) processing channels in quiet and in noise. The simulations were implemented using a real-time signal processing system based on a Motorola 96000 floating-point DSP with a sample rate of 12 kHz. A block diagram of the simulation is shown in Figure 1. In our experiments: (1) the Preemphasis/AGC was bypassed because it is not used by our pool of implantees; (2) the Nonlinear Mapping was disabled because the small dynamic range associated with electric stimulation is not an issue for normal hearing subjects, (3) the band envelopes modulated tones (or noise-bands) at their respective filter's geometric center frequency (this presents the band energy at the correct cochleotopic place), and (3) the modulated tones were summed for presentation with a speaker or headphone. The number of processor channels varied from 2 to 16 with the carrier frequencies set as shown by the filled circles in Figure 2.

In one set of measurements, a single normal-hearing listener was tested on our 24-initial consonant identification test. Speech-spectrum noise (from the HINT corpus(Soli and Nilsson, 1994)) was mixed with the test speech at different speech-to-noise ratios (SNRs), at the input to the simulation system, so that both the speech and the noise were processed via the simulation. Measurements were made using simulations with 3, 6, and 12 channels, and with the simulation bypassed -- an unprocessed condition.

The same 24-initial consonant test was also administered to three of our better implant users. Two used a CIS Clarion system (8 channels; 833 pps/electrode) and one, an Ineraid subject, used the MIT version of CIS processing (6 channels; 2000 pps/electrode) implemented on a Geneva wearable sound processor. These subjects were selected on the basis of achieving scores near 80% on the NU-6 monosyllabic word recognition test. Because each of these subjects had worn their CIS sound processor for at least three months at the time of testing, one might be concerned about comparing the implant results with those from the acoustic simulations where the normal-hearing subject did not experience months of continuous listening. While one should be mindful of this difference, we do not believe it is a serious problem because the subject spent many hours training on these materials (more than 4000 presentations with feedback) to establish a level of asymptotic performance.

3.0 Acoustic Simulations: Speech Reception Results

The results plotted in Figure 3 show an orderly decrease in performance as the number of channels and the SNR decrease. In quiet (SNR = ∞), direct listening (Unproc) produced 96% correct, the 12-channel simulation (AS-12) result was about 11 percentage points lower, and each subsequent halving of channels (to AS-6 and AS-3) decreased the performance another 11% percentage points. As SNR decreases, performance falls at varying rates. For intermediate scores, each decrement in the system from unprocessed to AS-3 requires an increase of about 8 dB in SNR to maintain a given level in performance.

The open symbols of Figure 3 represent the mean scores for the three high-performing cochlear implant users described above. Note that in both quiet and in noise, the performance of the implantees is essentially equivalent to that of the normal listener using a 6-channel simulation. This is also evident in Figure 4 where these consonant results are presented in the form of percent information transfer for three consonant features. Notice in the panel for 6-channel processing that the distribution of information transfer across the consonant features for the implantees is very similar to that for the normal-hearing subjects for the noise conditions examined. This is consistent with the view that the implantees and normal hearing listener are extracting similar information from their respective processors.

Figure 4 shows that in quiet (SNR = ∞), the main effect of reducing the number of channels is a large reduction in place information. Since the representation of spectral shape varies directly with the number of processing channels, this effect on place scores is to be expected. In contrast, scores for manner and voicing are affected relatively little, even for the three-channel (AS-3) condition. For decreasing SNR, scores on all features decrease. Manner is generally affected least, place is affected most, and voicing is intermediate. As mentioned above, the average scores from the three high-performing implantees show confusion patterns that are very similar to those of the normal listener using a 6-channel simulation.

In another set of measurements, two groups of four normal-hearing listeners were tested on sentence reception in noise using the HINT test. One group was tested in the unprocessed condition and with simulations using 12, 6, and 3 channels. The second group was tested with simulations using 16, 8, 4, and 2 channels, and at one SNR for unprocessed listening.[1] The results (Figure 5) show that performance falls with decreasing channel number and decreasing SNR. The results are quite orderly within each group (Group-1: open symbols, dashed lines; Group-2: filled symbols, solid lines). Some "crossovers" occur between the groups, e.g., Group-2 with 8 channels did better than Group-1 with 12 channels. This may reflect a real difference between the groups and/or variability intrinsic to the test materials.

The three high-performing implantees were also tested on the HINT materials. Figure 6 shows their results along with those from the normals using simulations with a corresponding number of channels (6 and 8). Again, both sets of scores are quite similar. These results also show that for the best performing implantees (with 6 to 8 channels) to attain the same word scores as normals listening without processing (i.e., listening naturally), the implantees require about a 6 dB higher SNR. In some ways, this deficit is remarkably small. It is much less than the deficits obtained with earlier multichannel implants (4-channel Ineraid and feature-based Nucleus systems) and it is similar to that obtained by some listeners with moderate-to-severe hearing impairments using hearing aids. Nevertheless, doubling the effective number of channels (up to 12 or 16) could halve this deficit to about 2 – 3 dB.

4.0 Acoustic Simulations: Possible Implications

The main picture is clear: with everyday-like sentences, only a few (perceptually independent) channels are needed to allow high performance in quiet (Dorman et al., 1989; Dorman et al., 1997; Fishman et al., 1997; Holmes et al., 1987; Lawson et al., 1993; Lawson

[1] Two subject groups were required in order to avoid repeating a sentence list during the testing of any individual subject (the limited number of HINT sentence lists was not sufficient for the number of channel/SNR conditions).

et al., 1996; Shannon et al., 1995). However, to allow performance in noise that approaches normal listening, a much higher number of channels is required.

The equivalence in performance between the high-performing implantees and normal-hearing subjects listening to 6- and 8-channel simulations on both sentence and consonant tests implies that these implant subjects are extracting all of the information relevant to speech reception that is available from the implant. This result is impressive and provides a challenge to developing improved systems. It suggests that further gains in performance can be obtained only by increasing the number of CIS processor channels/electrodes that can be utilized by subjects or by using existing electrodes to deliver additional information beyond that represented by the standard band-envelope signals.

Figure Captions

Figure 1. Block diagram of system used to simulate a CIS processor for normal-hearing listeners. In many CIS systems, preemphasis and/or AGC conditions the input signal. Because the CIS system our implantees use does not include this type of conditioning, it was not used in the simulation. The number of channels (band-pass filters) was a parameter that was varied. The bandwidths for the various conditions are shown in Figure 2. The envelopes were computed using full-wave rectification and a low-pass filter with cut-off frequency of 400 Hz. The nonlinear map used in CIS processors was disabled for these simulations. For each channel, a tone at the channel's geometric center frequency (see Figure 2) was modulated by its respective envelope. These AM signals were summed and presented to normal-hearing listeners by speaker or headphone.

Figure 2. Cut-off frequencies and "center frequencies" for the individual sets of channels. The numbers at the right specify the number of channels for each channel set. The vertical lines represent the cutoff frequencies of the band-pass filters used to implement each set of channels. The filled circles show the frequency of the sinusoidal carrier associated with each channel (Tone 1, ... Tone *n* in Figure 1).

Figure 3. Speech reception results for a single, normal-hearing listener (filled symbols; dashed lines) listening to: unprocessed 24-initial consonants, and 12, 6, and 3 channel simulations of implant speech processors whose inputs are the 24-initial consonants. Average results for three high-performing implant subjects using their normal sound processor for the same inputs are also shown (open symbols, solid line). Data are shown for a range of additive noise conditions.

Figure 4. A summary of consonant confusions, in terms of the percentage information transmission for the features: voicing, manner, and place. Data are presented for the various conditions of unprocessed, 12, 6, and 3 channel simulations, and for a range of SNRs at which the 24-consonants are presented. Results for the cochlear implant users (CI) are presented in

the same panel as the 6-channel (AS-6) results for the normal-hearing subject listening to the acoustic simulation.

Figure 5. Results from two independent groups of four listeners tested on sentence reception in noise using the HINT test. One group was tested in the unprocessed condition as well as with simulations using 12, 6, and 3 channels. The second group was tested with 16, 8, and 2 channel simulations, and at one SNR for the unprocessed condition.

Figure 6. Individual data for the sentences in noise test (HINT) for the same three high-performing implanted subjects (open symbols, solid lines). Average data from normal subjects without processing (solid symbols, solid line) and normal subjects using the 6 and 8 channel simulations (solid symbols, dashed lines).

Bibliography

Dorman MF, Dankowski K, McCandless G, Smith LM (1989): Consonant recognition as a function of the number of channels of stimulation by patients who use the Symbion cochlear implant. Ear and Hearing 10:288-291.

Dorman MF, Loizou PC, Rainey D (1997): Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. 102:2403-2411.

Fishman KE, Shannon RV, Slattery WH (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 (in press).

Holmes A, Kemker FJ, Merwin G (1987): The effects of varying the number of cochlear implant electrodes on speech perception. American Journal of Otology 8:240-246.

Lawson D, Wilson B, Finley C (1993): New Processing Strategies for Multichannel Cochlear Prostheses. In Allum JA, Allum-Mecklenburg DJ, Harris FP, Probst R (eds): "Natural and Artificial Control of Hearing and Balance." Amsterdam: Elsevier, pp 313-321.

Lawson DT, Wilson BS, Zerbi M, Finley CC (1996): Speech processors for auditory prostheses, Third Quarterly Progress Report. NIH Contract N01-DC-5-2103.

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

Soli SD, Nilsson M (1994): Assessment of communication handicap with the HINT. Hearing Instruments 45:12-16.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991): Better speech recognition with cochlear implants. Nature 352:236-238.

Acoustic Simulations of CIS Processing

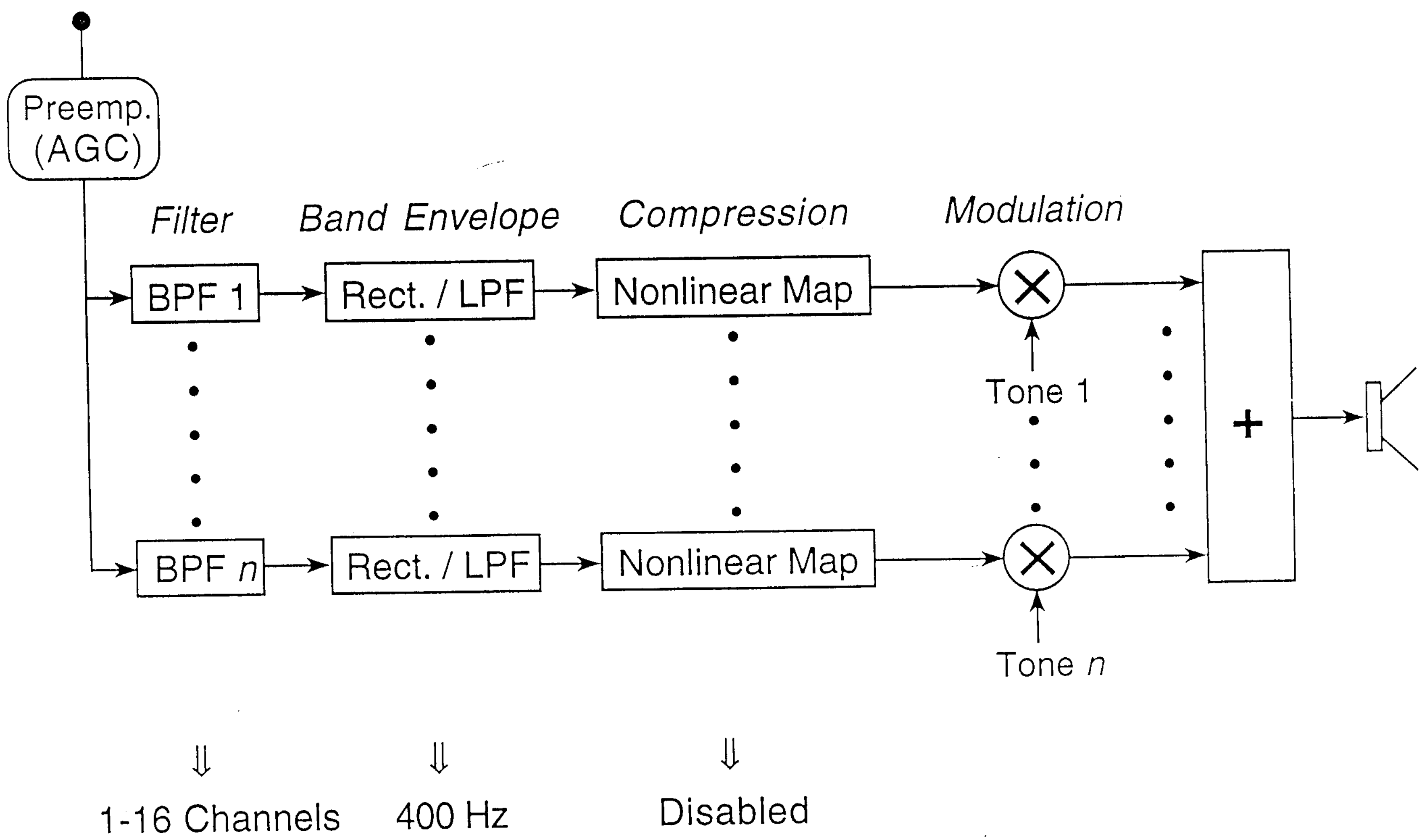
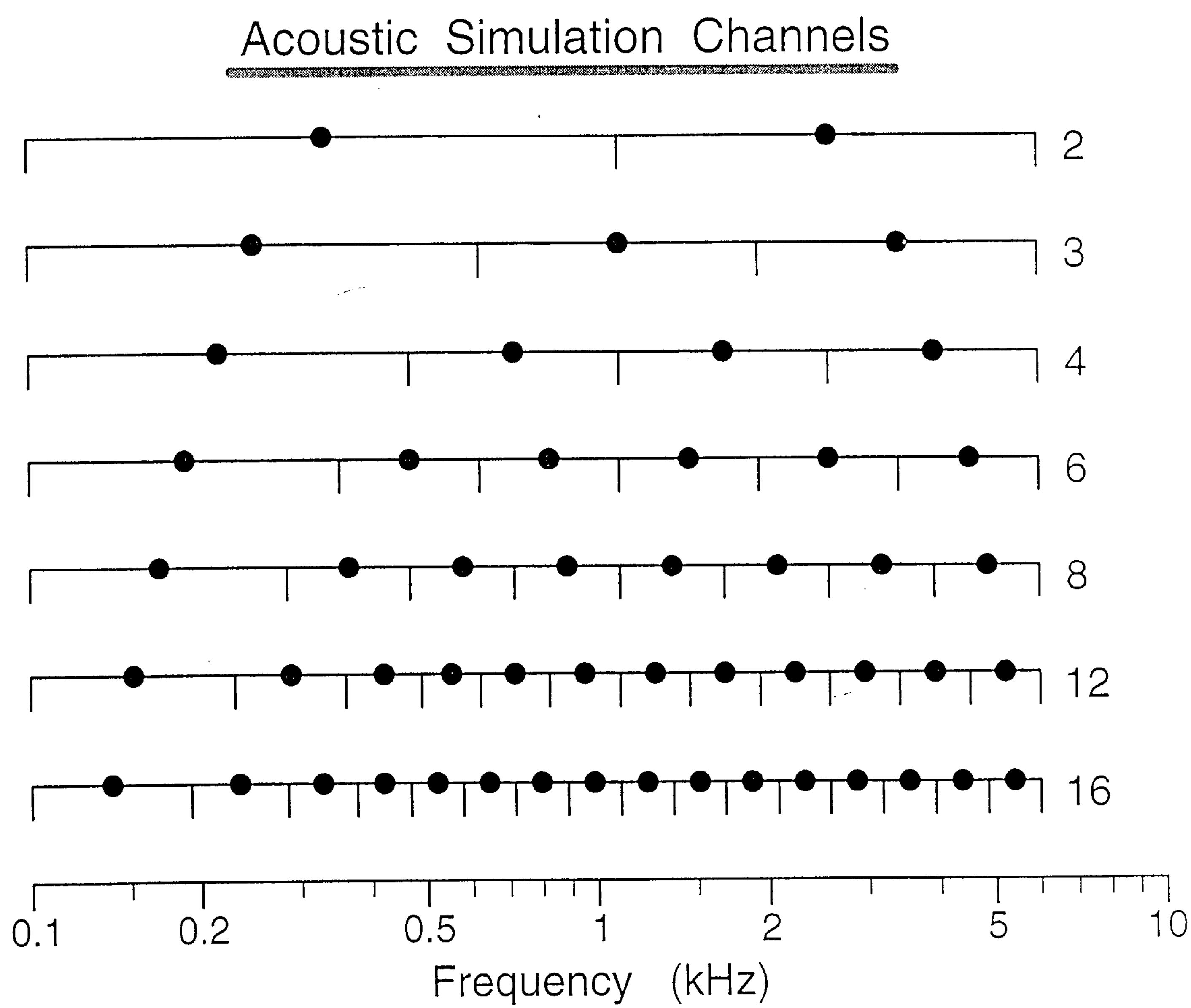


Figure 1



Acoustic Simulation Channels
2
3
4
6
8
12
16
0.1
0.2
0.5
1
2
5
10
Frequency (kHz)

Figure 2

Figure 3

24 Initial Consonants in Noise

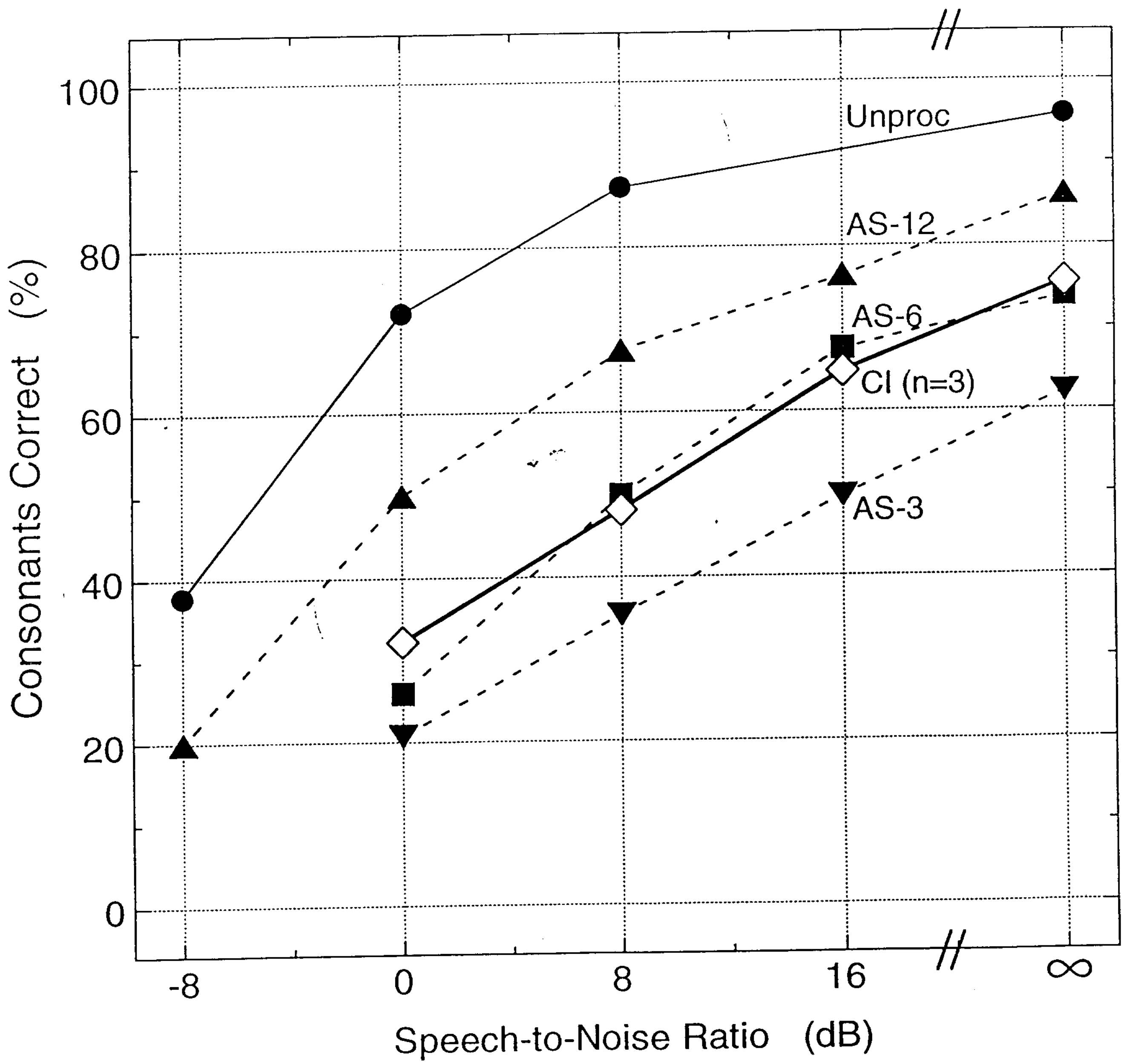


Figure 4

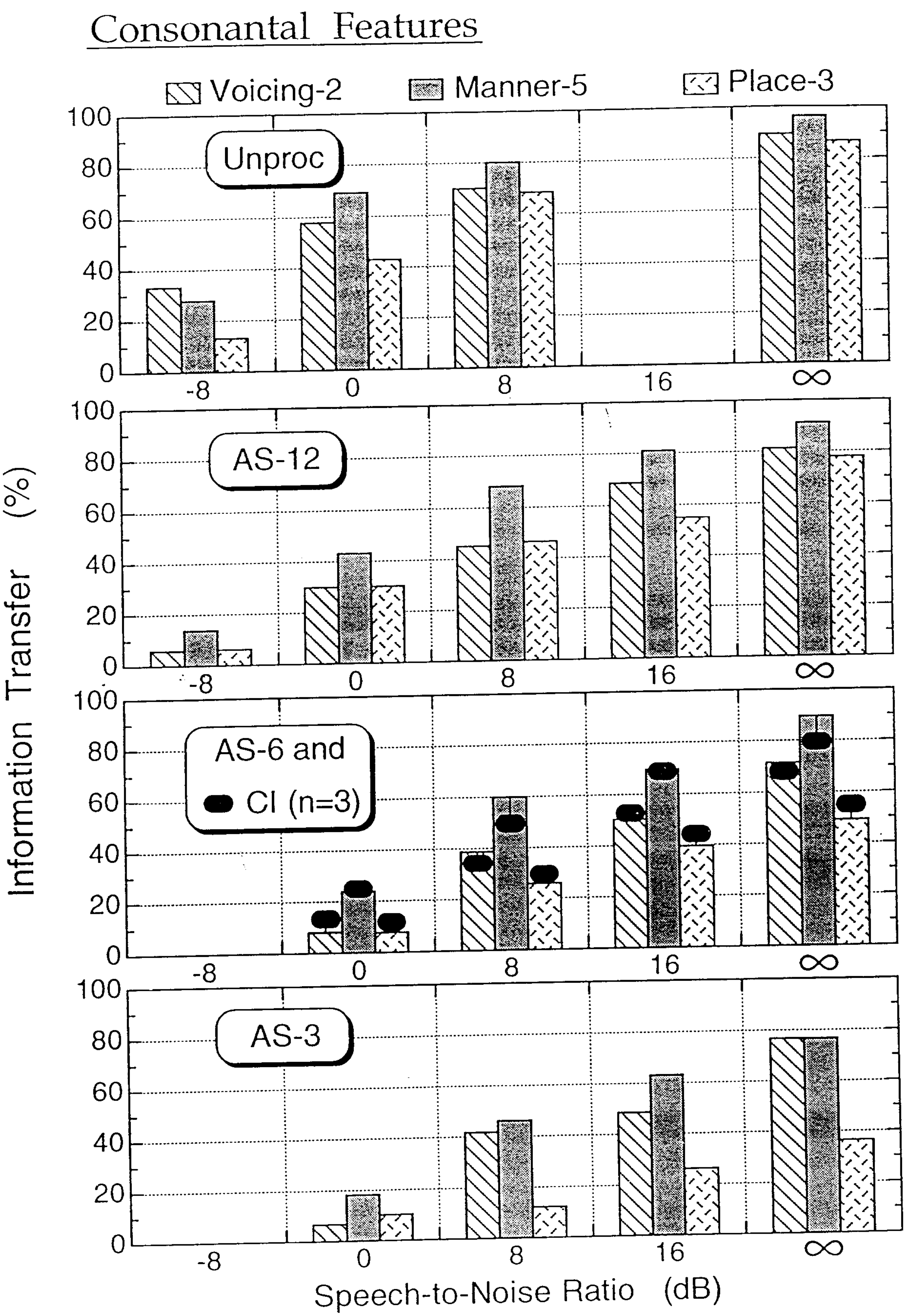

Consonantal Features
Voicing-2
Manner-5
Place-3
Unproc
AS-12
AS-6 and
CI (n=3)
AS-3
Information Transfer (%)
100
80
60
40
20
0
-8
0
8
16
∞
Speech-to-Noise Ratio (dB)

Figure 5

Acoustic Simulations of CIS

- Sentences in Noise (HINT)
- Normal-Hearing Listeners
- Two Groups, n = 4/Group, Mean +/-Std. Error

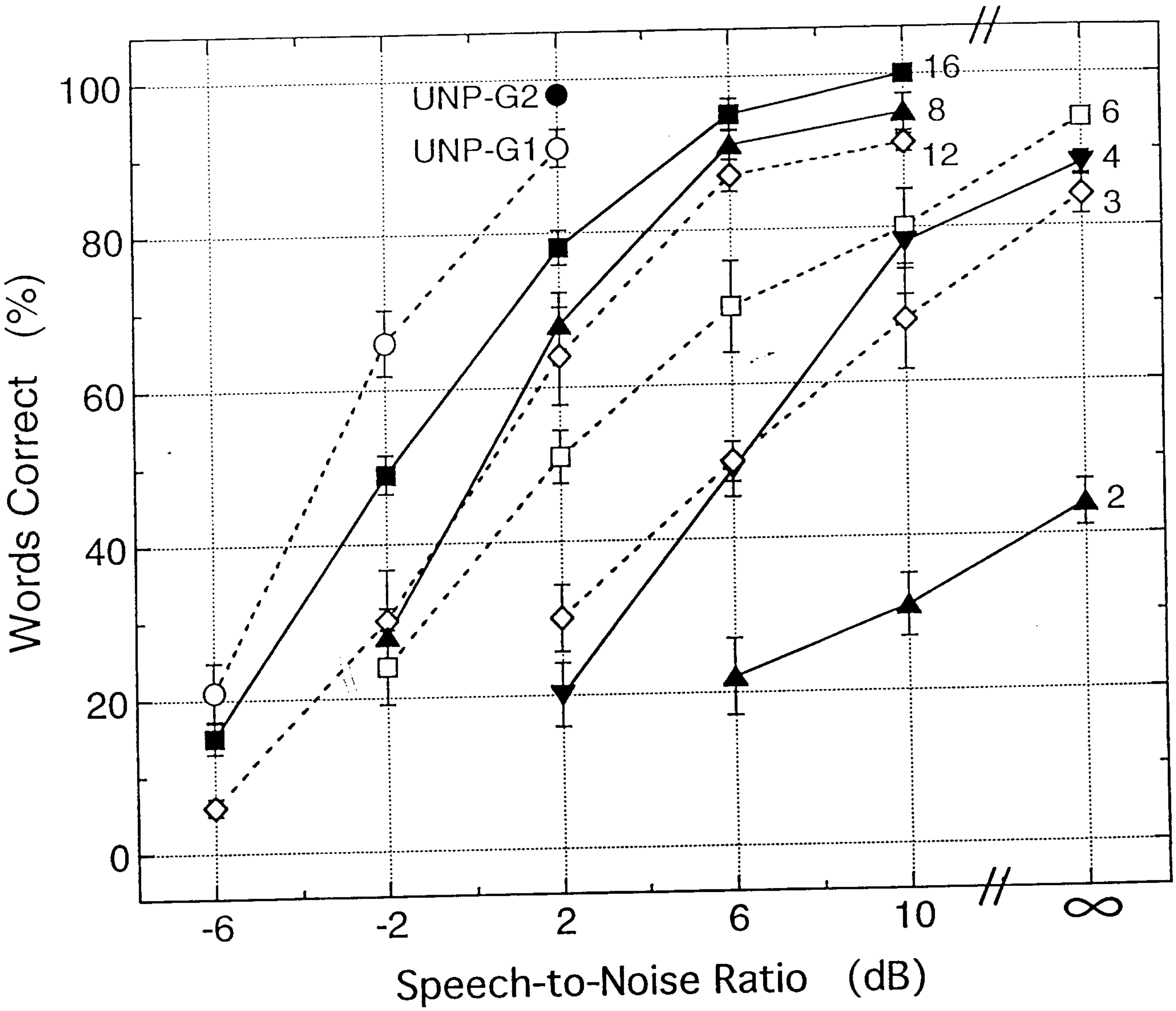
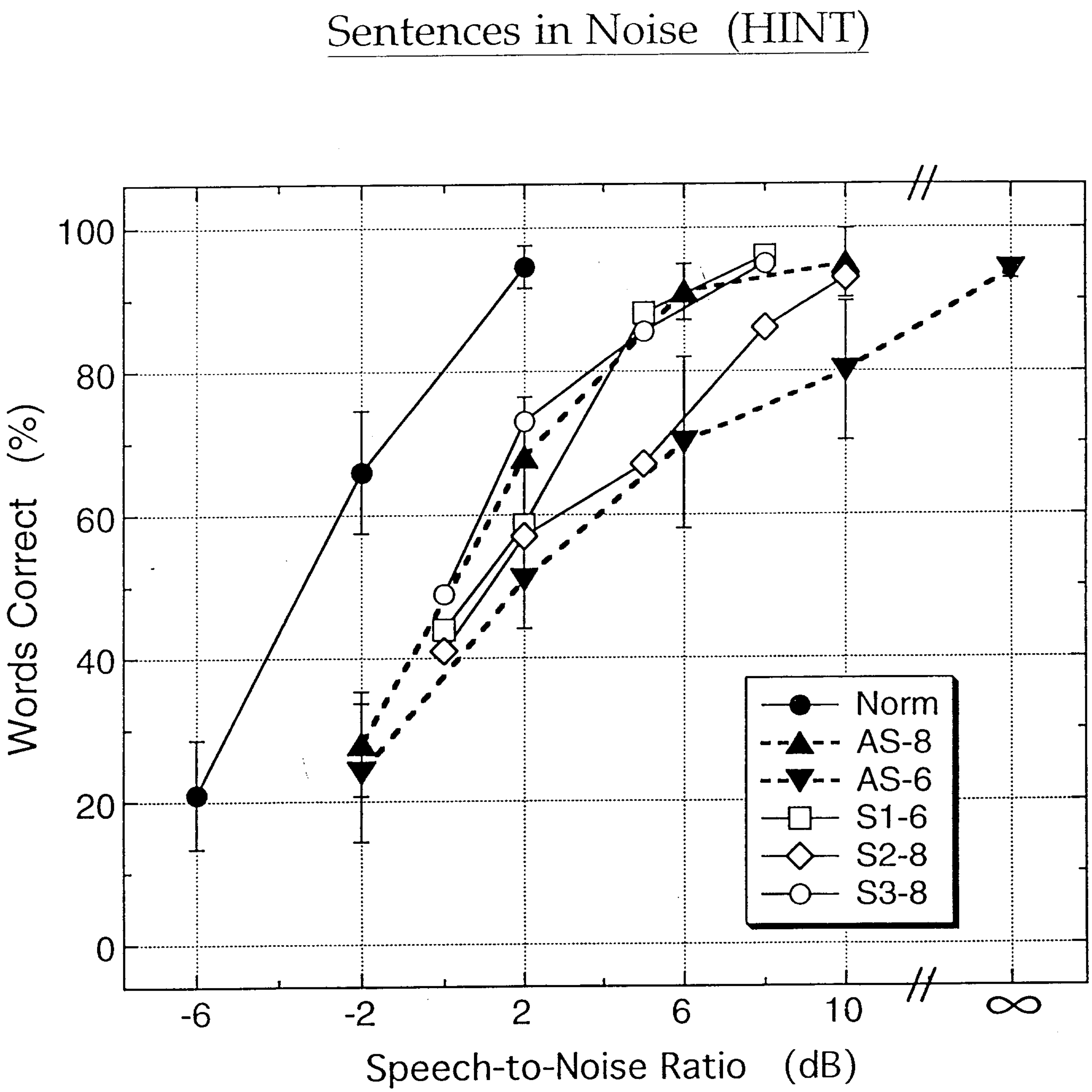


Figure 6



Sentences in Noise (HINT)
Words Correct (%)
100
80
60
40
20
0
-6
-2
2
6
10
∞
Speech-to-Noise Ratio (dB)
Norm
AS-8
AS-6
S1-6
S2-8
S3-8