Tenth Quarterly Progress Report

April 1, 1998, through June 30, 1998

Speech Processors for Auditory Prostheses

NIH Contract N01-DC-6-2100

submitted by

Donald K. Eddington
Joseph Tierney
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.

In this Quarterly Progress Report (QPR), we describe two limitations that restrict the performance obtained with today's CIS sound processors: the limited number of channels and the limited within-band information. The within-band information that is discarded by CIS processors serves to motivate a simple processor implementation that, theoretically, restores a measure of the discarded information (along with additional distortion products). The implementation and fitting of this processor in a small group of Ineraid subjects is described and preliminary results are compared with a standard CIS system.

2.0 Limitations of Today's CIS Processors

In continuous interleaved sampling (CIS) processors (Wilson, Finley et al. 1991) the envelopes of the band-pass filtered outputs are computed and used to modulate trains of biphasic pulses that are interleaved in time across channels. The results from several laboratories have shown that, in general, Ineraid implantees (monopolar electrode configuration) score better on speech reception tests with CIS processing than with CA processing (e.g., Wilson, Finley et al. 1991; Pelizzone, Boex-Spano et al. 1995; Eddington, Tierney et al. 1996; Dorman and Loizou 1997).

While improvements (like CIS and SPEAK (Seligman and McDermott 1995)) in the design of sound processors for cochlear implants have improved the performance of patients over the past ten years, even today's best-performing users suffer a significant deficit in speech reception (see Figure 3 and associated text below). We find it useful to assess the limitations of today's CIS processors from two perspectives: (1) number of channels (e.g., the ability to provide information about an utterance's spectral shape through patterns of spatial excitation) and (2) within-channel, temporal information.

2.1 Relatively Small Number of Channels

In a previous QPR (Eddington, Rabinowitz et al. 1997), we presented speech-reception data from acoustic simulations suggesting that a significant increment of the number of "relatively independent" channels in today's CIS processors would likely result in better speech reception. For the convenience of the reader, these data are included here as we discuss some of the limitations imposed by the relatively small number of channels available to today's cochlear implant users.

We measured the speech-reception capability of normal-hearing subjects listening to speech processed by a system designed to simulate CIS processing (see Figure 1). In these experiments, the simulation was implemented by using a standard CIS processor and: (1) turning off the Preemphasis/AGC (not used by our implantees) and Nonlinear Mapping (the small dynamic range of electric stimulation is not an issue for normal hearing subjects), (2) the band envelopes modulated tones (or noise-bands) at their respective filter's center frequency (this presents the band energy at the correct cochleotopic place), and (3) summing the modulated tones (or noise-bands) for presentation with a speaker or headphone. The number of processor channels was varied from 2 to 16 with the carrier frequencies set as shown by the filled circles in Figure 2.

The filled symbols of Figure 3 show consonant scores for a normal-hearing subject listening to processors of varying channel number in a number of noise conditions. That the test is difficult can be appreciated by the fact that even in quiet (SNR $= \infty$), the normal-hearing subject does not score 100% when listening in the unprocessed condition. In general, as the SNR decreases, the scores decrease for all processors. Similarly, as the number of processing channels decreases, the scores decrease for any SNR.

The open symbols of Figure 3 represent the mean scores for three of our best cochlear implant users. Two were Clarion users (8 channels) and one an Ineraid implantee using the 6-channel, MIT version of CIS processing implemented on a Geneva wearable sound processor. 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 have months of continuous listening experience. 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.

Notice that the implantee scores are virtually identical to the normal-hearing listener using a 6-channel simulator. 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.

Similarity of performance between implantees and the normal-hearing subjects when listening with processors of like numbers of channels is also seen in Figure 5. Here, the mean scores of four, normal-hearing listeners tested for their ability to recognize HINT sentences presented at several SNRs are plotted (filled symbols) for three processing conditions: unprocessed, 8-channel processor and 6-channel processor. The open symbols represent the scores of our three best cochlear implant users. Note that the scores for the implant users fall between the 6- and 8-channel simulation scores of the normal hearing subjects.

The data presented in Figures 3, 4 and 5 indicate that the normal-hearing subjects and the high-performing implantees receive similar information when they both listen through 6- or 8-channel processors of like architecture. Thus, it is very likely that the common processing structure is responsible for limiting the information transferred by these 6- and 8-channel processors to the normal-hearing and implanted listeners. To the extent the normal-hearing subjects are able to make use of all the information available in the output of these processors, improved performance will only be achieved for the best implant users by increasing the information the processors provide and presenting it to the brain in a form that is usable. One hopes that increasing the information in this way will benefit users across the whole range of performance.

The simulation data suggest that increasing the number of "independent" channels/electrodes may increase the information in the output of the CIS processing systems tested. The data in Figure 3 show that in quiet, even the best performing cochlear implant subjects score below normal-hearing subjects listening to the12-channel processor (76% vs. 86%) or to the unprocessed speech (76% vs. 96%). In noise, the room for improvement is even greater. The simulation scores for normal-hearing subjects improve for all SNRs as the number of channels increase from 3 to12. To the extent the simulation results can be related to the implantees, they suggest that increasing the number of channels in CIS processors may increase the overall scores in the implant users. Unfortunately increasing the number of channels does not simply mean increasing the number of analysis bands in the sound processing system. The information in each channel must also be translated to a format the brain can utilize. At a minimum, this means providing a set electrodes capable of producing the same number of (relatively independent) channels of neural information as there are analysis channels in the sound processor.

2.2 Within-Channel Information Limited

Another method for increasing the information delivered by a CIS processor is to increase the amount of information within the stimulus signal of each processing channel or, said another way, decrease the information discarded by each channel. Consider, for example, the signal output of a band-pass filter (bandpass(t)) that can be represented as the product of an envelope signal (env(t)) and a phase-modulated carrier ($\cos[\phi(t)]$) so that

$$\text{bandpass}(t) = \text{env}(t) * \cos[\phi(t)].$$

If a CIS processing channel accurately extracts the band envelope, env(t) is the portion of the band-pass filter output signal that, after being level-mapped, modulates the pulse train to form a stimulus signal. The phase-modulated carrier signal (also referred to as the excitation signal or the fine temporal structure of the band-pass filtered output signal) represents the information discarded by the CIS processing channel.

One important question is how much information is lost when an N-channel, CIS processor discards the excitation signals $\cos[\phi(t)]_k$, for $1 \leq k \geq N$? If the information in the excitation signals is large relative to the information in the collection of envelopes, considerable performance gains might be realized by restoring some of the information discarded in the excitation signals.

Insight into this issue can be gained from a study where normal-hearing subjects listened to the output of a bank of band-pass filters and speech reception was measured as a function of the number of band-pass filters and the envelope bandwidth (Drullman, Festen et al. 1994). In Figure 6 (a reproduction of Figure 3 from the Drullman, et al. Paper), speech reception is plotted as a function of the bandwidth of the band envelope signals with the band-pass filter bandwidth (i.e., number of channels for the fixed frequency range of 100-6400 Hz) as a parameter. Note that the bank of one-octave filters includes six channels and corresponds most closely to the 6- and 8-channel CIS processors and CIS simulations used to collect portions of the data shown in Figures 3-5.

In the case of an "LP cutoff frequency" of zero, there are no envelope variations, and the only information transmitted to the listener is the sum of the excitation signals from each band-pass channel. Listening to the six-filter system that delivered only the excitation signals, Figure 6 shows that normal-hearing subjects scored over 80% correct on a thirteen-sentence test. This demonstrates that considerable information is included in the 6- or 8-channel excitation signals (i.e., in the fine-time structure of the band-pass filters' output signals) that are discarded in CIS sound processors.

As the filters increase in number (i.e., the bandwidth of the band-pass filters decreases), the excitation signal carries less information. In the case of the 24-channel system (1/4 octave filter bandwidths) with an "LP cutoff frequency" of zero (excitation-only system), normal hearing listeners score only a few percent correct on the sentence test.

Another way to compare the spectral information in a band-pass filter's output signal to the information in the envelope signal is to consider their respective waveforms and spectra for a specific example. The time waveforms displayed in Figure 7 represent a steady-state segment from the vowel portion of a synthesized /da/ that has been processed in four different ways. In each case the /da/ has been processed by a band-pass filter with

–3 dB cutoff frequencies of approximately 360 and 640 Hz (this corresponds to Channel 2 of a 6-channel processor of overall bandwidth 200-6500 Hz). Moving from the bottom panel to the top panel the waveforms are: the output of the band-pass filter (BPFo); the half-wave-rectified output of the band-pass filter (HWR); the full-wave-rectified output of the band-pass filter (FWR), and the quadrature envelope of the band-pass filter output (env(t)). Figure 8 displays the corresponding log-magnitude spectra in dB versus frequency in Hz for each of the waveforms of Figure 7.

Notice that the spectral information for the output waveform of the band-pass filter between 400 and 700 Hz is well represented by the spectrum of the half-wave-rectified (HWR) case. Below 400 Hz, the HWR spectrum also includes spectral detail associated with the envelope (env(t), top panel) and above 700 Hz the energy from the harmonic distortion of the rectification process is evident.

In the FWR case, the env(t) spectrum is better represented than when half-wave rectification is used. Like the HWR case, the distortion products related to the rectification process are represented above 700 Hz. Note that the spectral details associated with the information carried in the output waveform of the band-pass filter is degraded by the combination of the env(t) and harmonic-distortion components in the range from 300 to 800 Hz. This means there is little information describing the original waveform's fine structure (excitation).

Since today's CIS processor implementations tend to use either a quadrature or FWR operation to generate the envelope of the band-pass filter's output waveform (often with low-pass filtering at 400 Hz or lower), it is clear that the details of the waveform's fine structure are discarded. One possible method for increasing the information content within a channel would be to restore some or all of the information contained in the excitation signal in a way that the brain can interpret. In the following section we describe preliminary results from the first in a series of experiments where we investigate methods of restoring some portion of the excitation signal to the individual channels of CIS processing.

3.0 Hybrid (2HWR/4QUAD) CIS Processor

Figure 8 suggests that simply using HWR without low-pass filtering would restore the excitation information (plus the HWR distortion products) to the channel's waveform. This section describes a hybrid CIS processing system that combines two such HWR channels (channels 1 and 2) with four regular CIS processing channels (channels 3-6). This hybrid system is used to test whether subjects are able to extract more information when the excitation signal (and the HWR distortion products) is included with the env(t) in the modulator signal for the two lowest-frequency channels. Future reports will describe the results of other methods we are/will be testing to increase the information in 6- and 8-channel processing systems by restoring the excitation signal to the individual processing channels in ways that result in neural responses the brain is able to interpret.

3.1 Implementation of the Hybrid Processor

We modified the two lowest channels of a CIS processor configuration as shown in Figure 9. The top block diagram shows a standard configuration that employs quadrature detection to compute the envelope of the band-pass filter output (env(t)). The env(t) is then "compressed" by a level-mapping function and used to amplitude modulate a biphasic pulse train (~ 4000 Hz). The modified implementation for channels 1 and 2 (shown in the lower block diagram of Figure 9) uses the same band-pass filters as our standard CIS processor, but the modulator signal is produced with HWR without low-pass filtering.

Since the two HWR channels carry information with bandwidths extending to approximately 1400 Hz, we increased the carrier rate of these channels to approximately 8 kHz. The carrier modulated by the level-mapped env(t) in channels 3 through 6 was approximately 4 kHz. The order of interleaving for the carrier pulses is shown in the bottom panel of Figure 9. This interleaved pulsatile sequence provides the non-overlapping pulse trains associated with CIS processors, while providing such pulse trains at twice the update rate for channels one and two.

3.2 Psychophysical Procedures Used for Fitting the Hybrid Processor

For each electrode, psychophysical measures of threshold (THR) and the most comfortable stimulus level (MCL), were made using a 300 ms duration, unmodulated segment of the carrier used in that electrode's channel (e.g., the repetition rate for electrodes 3-6 was approximately 4000 pps and for electrodes 1 and 2 approximately 8000 pps).

The hybrid processor was designed by first implementing a standard, 6-channel CIS processor (see Eddington, Delhorne et al. 1995; Eddington, Tierney et al. 1996) using 4000 pps carriers for channels 3-6, 8000 pps carriers for channels 1 and 2, and quadrature-derived envelopes across all channels. After verifying that the standard system performed as expected (e.g. speech reception equal to the subject's wearable CIS system), the two HWR channels were implemented. Each of the two HWR channels were then balanced for loudness with their corresponding standard quadrature channel by adjusting the I_{max} parameter of each HWR channel's level-mapping function to give the same loudness while listening to the Iowa, 16-consonant review list (Tyler, Preece et al. 1983). After fitting five subjects using this method, we find I_{max} for the HWR channels is increased by a factor of 1.15-1.20 over a standard CIS implementation.

3.3 Comparison Data for CIS and 2HWR/4QUAD Hybrid Processors

The data of Table I present consonant and vowel identification scores for the standard CIS and the 2HWR/4QUAD hybrid sound processors. The consonant scores represent percent correct identification using the Iowa, 16-consonant test (/aCa/, e.g., "asha," "apa") and the Iowa, 8-vowel test (/hVd/, e.g., "heed", "had"). Each of the consonant scores represents 5-20 presentations of the 16-consonant set and each vowel score is derived from 6-12 presentations of the 8-vowel set.

Table I.
Consonant and Vowel Recognition for the Standard and Hybrid CIS Sound Processors
[percent correct (standard deviation)]

Subject	16-Consonant Test		8-Vowel Test	
	Standard CIS	Hybrid CIS	Standard CIS	Hybrid CIS
A01	69% (4.3)	73% (3.0)	81% (6.0)	92% (0)
S02	72% (3.8)	77% (2.1)		
S11	63% (3.2)	54% (2.2)	69% (3.6)	67% (2.4)
S18	51% (3.2)	56% (2.6)	67% (2.4)	65% (1.7)
S27	61% (3.4)	59% (2.9)	46% (0)	61% (4.5)

These data do not show large, consistent differences in speech reception between the two processors. In the case of consonant recognition, the only significant difference (.01 level) between the standard and hybrid systems is the lower performance exhibited by subject S11 with the hybrid system. Two subjects, A01 and S27, posted vowel scores that showed a significant improvement for the hybrid system.

4.0 Future Work

If the improvements in vowel scores for these two subjects hold up with continued testing, it will be interesting to determine the nature of the information these subjects are using to make their judgements and to investigate reasons the other subjects do not experience these improvements. We also plan to test whether those subjects like A01 and S27 (where some portion of the information content of the excitation signal may have been restored) will demonstrate better noise immunity with the hybrid system than with the standard.

The hybrid system described in this QPR is only a first step in a series of experiments we are conducting to restore each channel's excitation signal in a way that will convey more information to the brain than is accomplished by today's standard CIS processors. These experiments will include hybrids of analog and CIS processors and processors that employ high-rate carriers designed to increase the stochastic nature of the elicited neural responses.

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. 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).

Figure 6. Reproduction of Drullman, et.al. Figure 3 (Drullman, Festen et al. 1994).

Figure 7. The time waveforms (arbitrary amplitude vs. sample number) derived from a steady-state segment from the vowel portion of a synthesized /da/ that has been processed in four different ways. In each case the /da/ has been processed by a band-pass filter with –3 dB cutoff frequencies of approximately 360 and 640 Hz (this corresponds to

Channel 2 of a 6-channel processor of overall bandwidth 200-6500 Hz). Moving from the bottom panel to the top panel the waveforms are: the output of the band-pass filter; the half-wave-rectified output of the band-pass filter (HWR); the full-wave-rectified output of the band-pass filter (FWR), and the quadrature envelope of the band-pass filter output (env(t)). The sample interval associated with each waveform 62.5 μs.

Figure 8. Amplitude spectra of the waveforms shown in Figure 7.

Figure 9. Block diagrams and carrier-timing structure for the two styles of processing used to implement a hybrid CIS processor. The top block diagram represents the processing used by the four highest-frequency channels (channels 3-6). A quadrature method is used to compute an envelope that is "compressed" using a level-mapping function and then serves to amplitude modulate a biphasic pulse train (16 μs/phase, cathodic first, 3907 pps). The lower block diagram represents the processing used by the two lowest-frequency channels (1 and 2) and shows the modulator being derived by half-wave rectification (HWR) and level mapping. The carrier is a biphasic pulse train like that used for channels three through six with a repetition rate of 7813 pps. The bottom panel shows the ordering of the interleaved pulses across the channels. This interleaved pulsatile sequence provides the non-overlapping pulse trains associated with CIS processors, while providing twice the update rate for channels one and two. For these lower channels, the waveform's zero crossings are defined within +/−64 μs. Obviously the pulsatile carrier rate for channels one and two can be increased if this jitter is shown to degrade the information presented to implanted electrodes.

Bibliography

Dorman, M. F. and P. C. Loizou (1997). "Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with Continuous Interleaved Sampling (CIS) processors." Ear and Hearing **18**: 147-155.

Drullman, R., J. M. Festen, et al. (1994). "Effect of temporal envelope smearing on speech reception." Journal of the Acoustical Society of America **95**(2): 1053-1064.

Eddington, D. K., L. Delhorne, et al. (1995). Speech Processors for Auditory Prostheses, Final Report. Cambridge, MIT Research Laboratory of Electronics.

Eddington, D. K., W. R. Rabinowitz, et al. (1997). Speech Processors for Auditory Prostheses, Eighth Quarterly Progress Report. Cambridge, MIT, Research Laboratory of Electronics.

Eddington, D. K., J. Tierney, et al. (1996). Speech Processors for Auditory Prostheses, First Quarterly Progress Report. Cambridge, MIT Research Laboratory of Electronics.

Pelizzone, M., C. Boex-Spano, et al. (1995). "First field trials with a portable CIS processor for the Ineraid multichannel cochlear implant." Acta Otolaryngol **115**(Stockholm): 622-628.

Seligman, P. and H. McDermott (1995). "Architecture of the Spectra 22 speech processor." Ann. Otol. Rhinol. Laryngol. **104**(Suppl. 166): 139-141.

Tyler, R. S., J. P. Preece, et al. (1983). The Iowa Cochlear Implant Tests. Dept of Otoloaryngology Head Neck Surgery, University of Iowa Hospital Clinic.

Wilson, B. S., C. C. Finley, et al. (1991). "Better speech recognition with cochlear implants." Nature **352**: 236-238.

Acoustic Simulations of CIS Processing

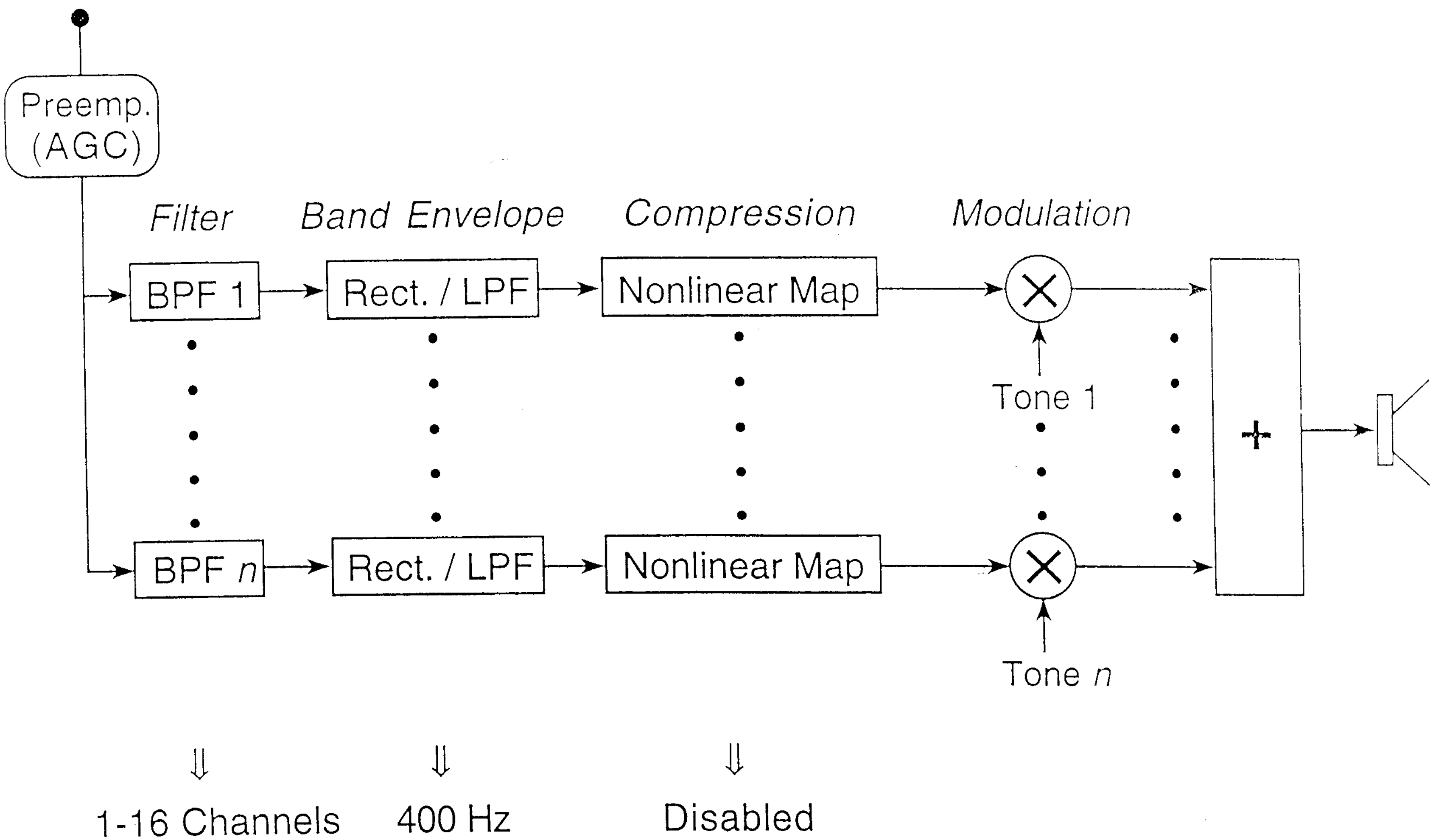


Figure 1

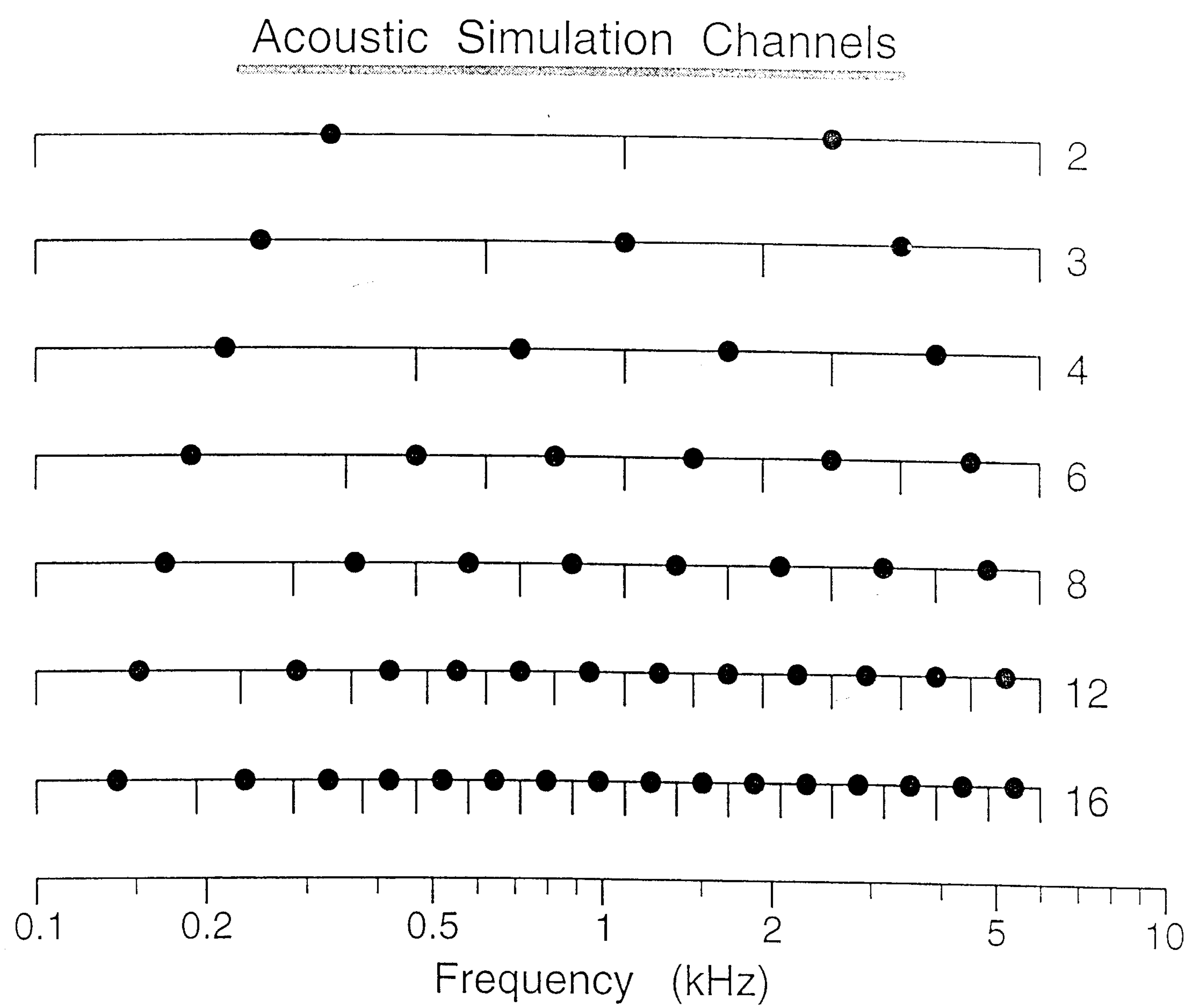

Acoustic Simulation Channels
2
3
4
6
8
12
16
0.1
0.2
0.5
1
2
5
10
Frequency (kHz)

Figure 2

Figure 3

24 Initial Consonants in Noise

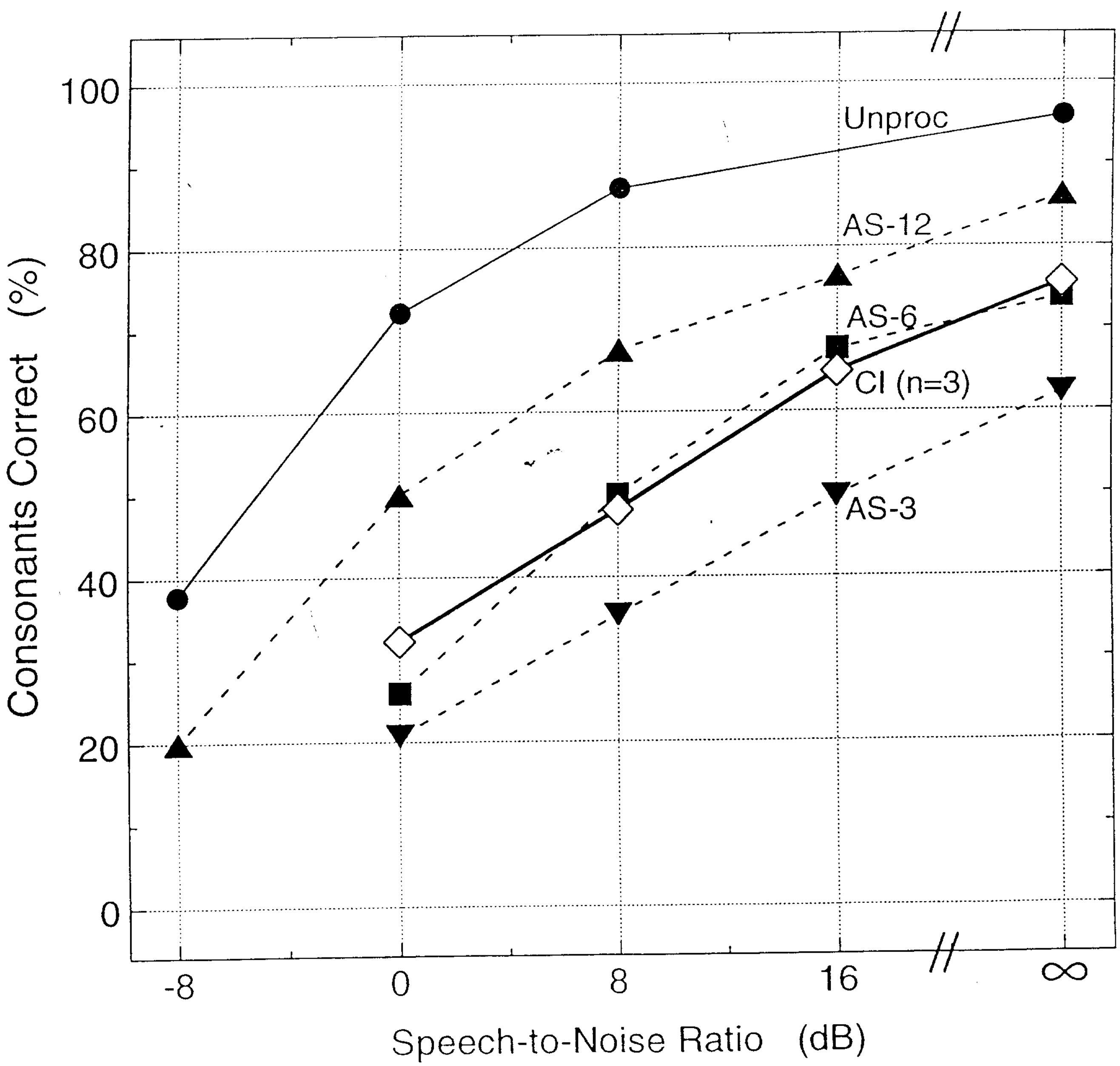
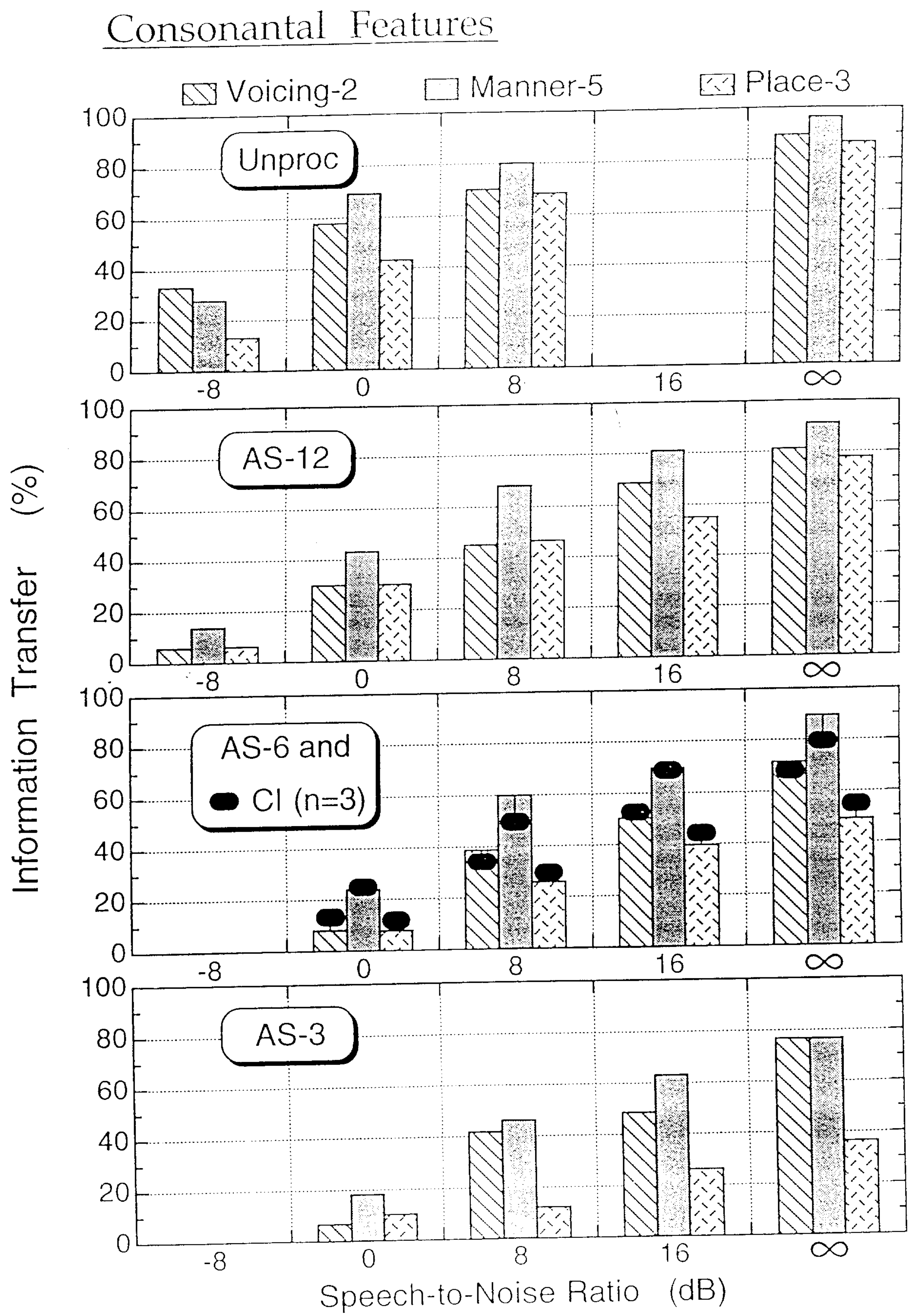


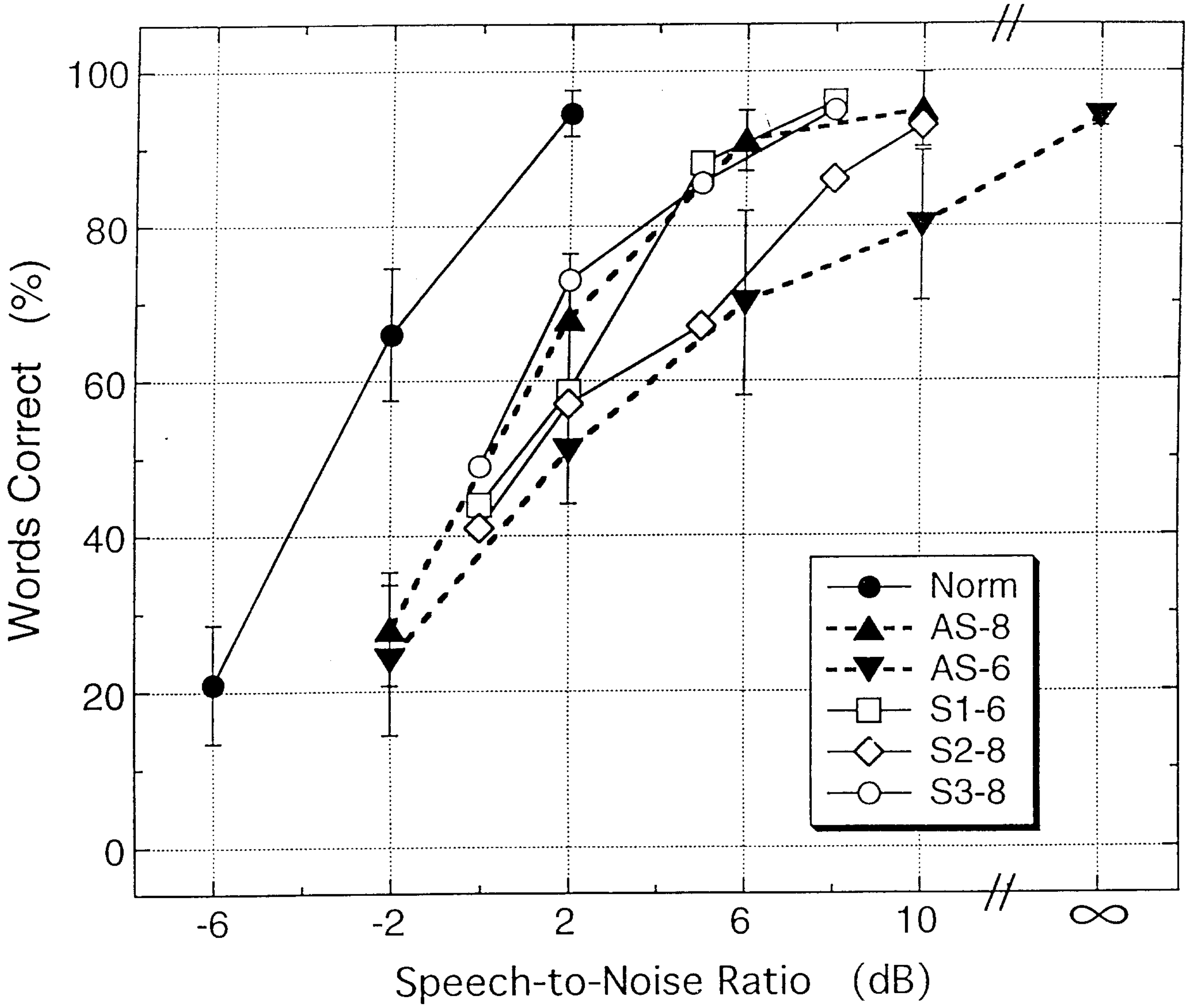
Figure 4

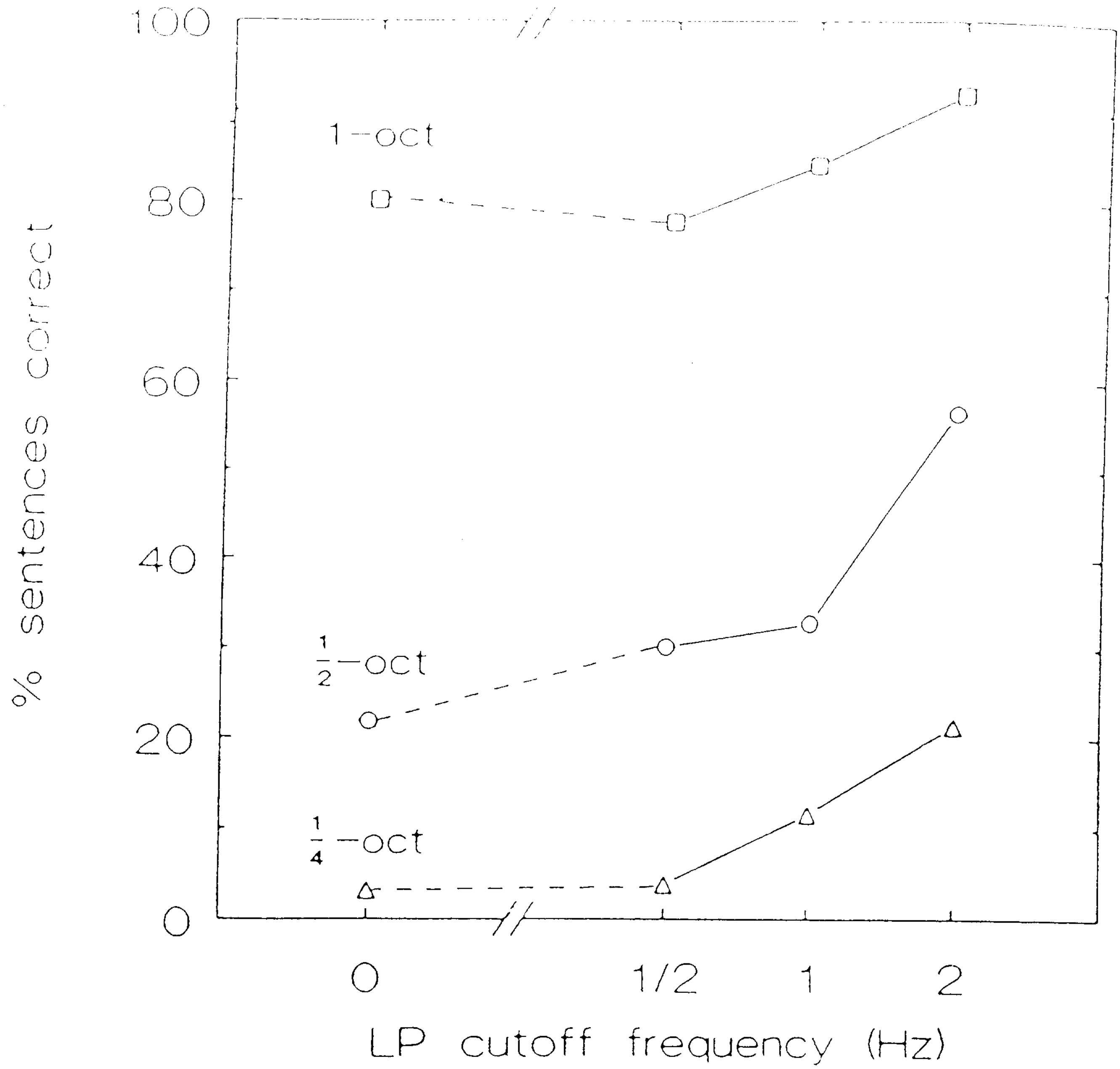


Consonantal Features
Voicing-2
Manner-5
Place-3
Unproc
AS-12
AS-6 and
CI (n=3)
AS-3
Information Transfer (%)
100
80
60
40
20
0
-8
0
8
16
∞
Speech-to-Noise Ratio (dB)

Figure 5

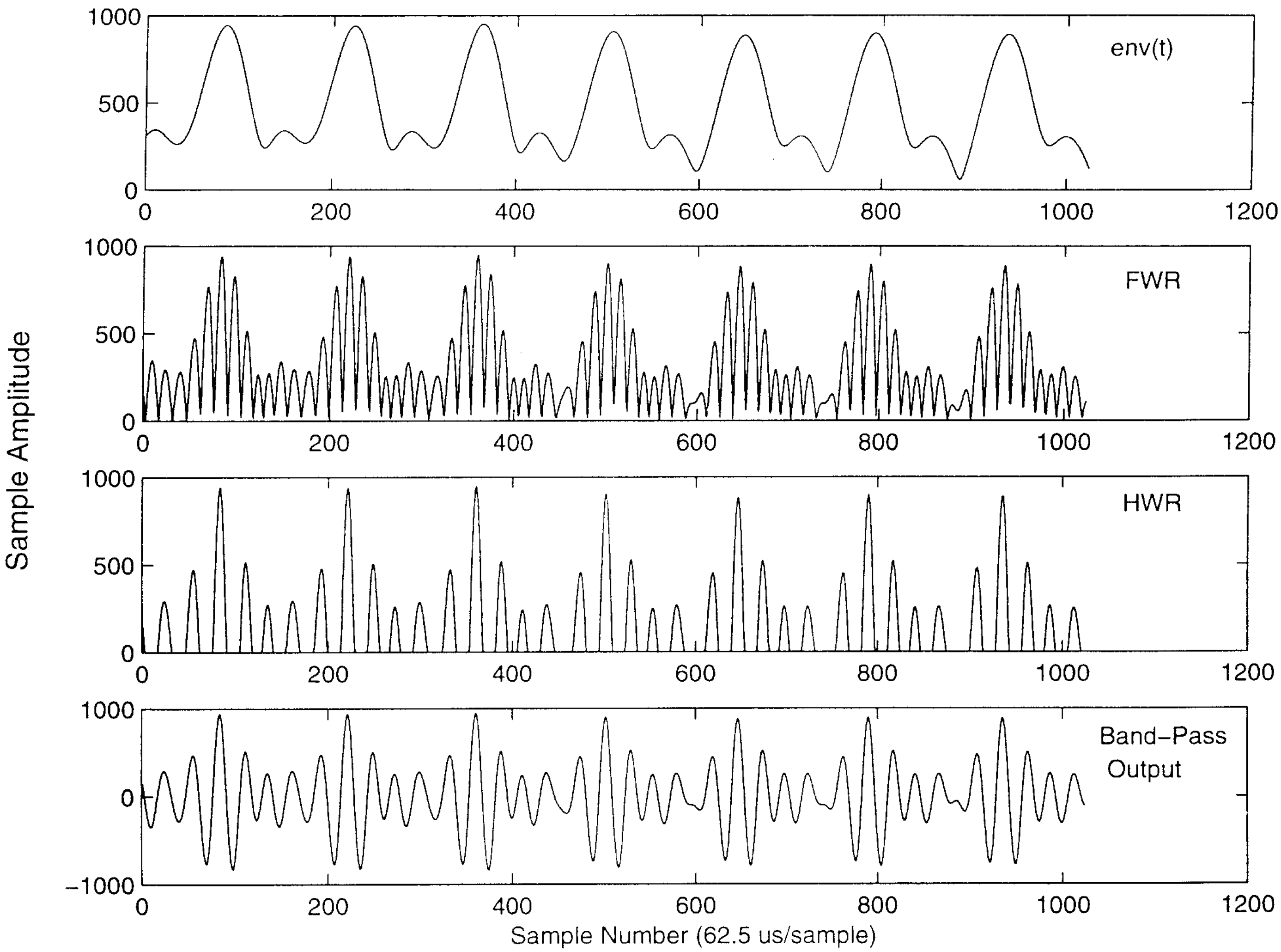
Sentences in Noise (HINT)





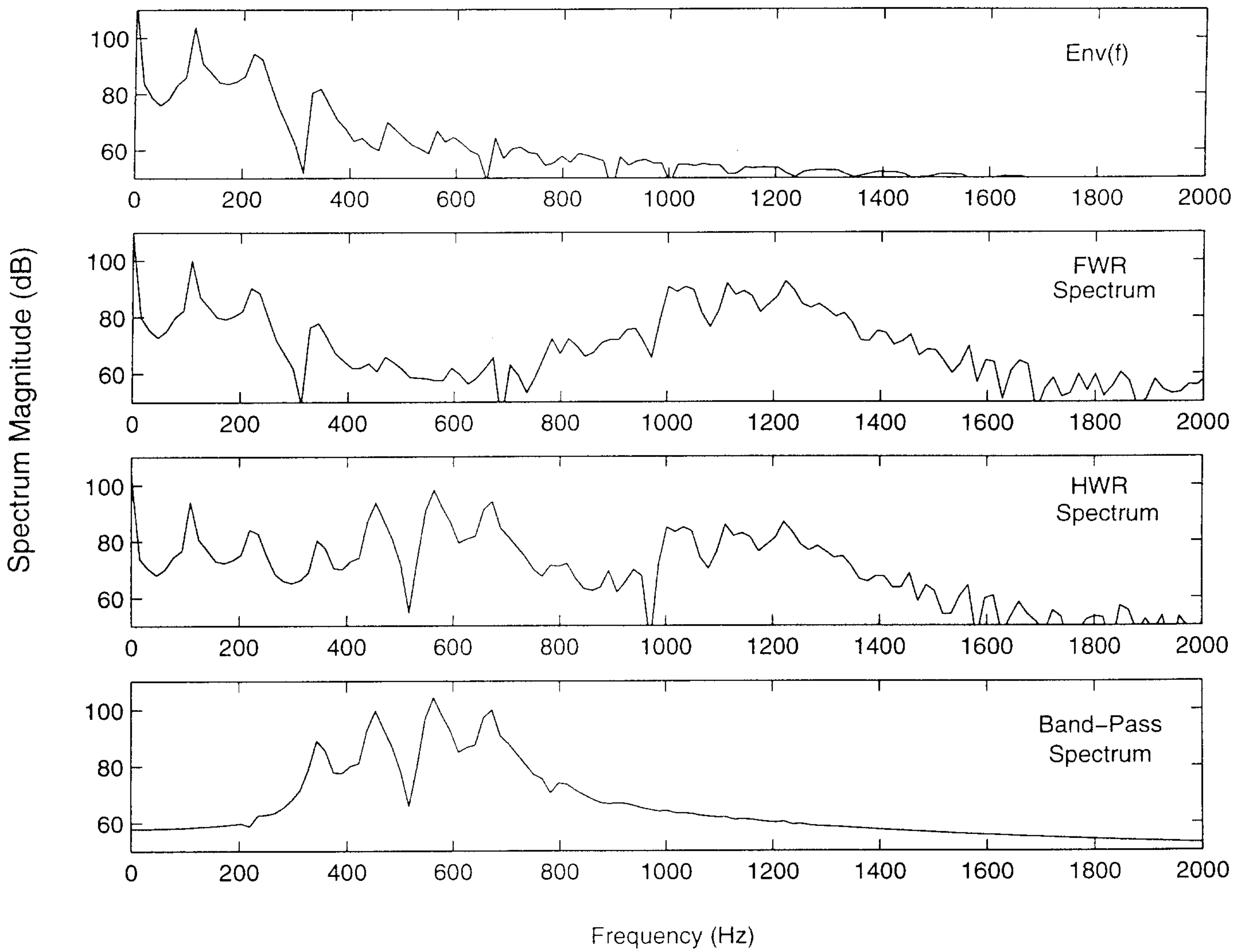
IG. 6 Mean score of sentences in quiet as a function of cutoff frequency, ith processing bandwidth as parameter. (AFTER DRULLMAN ET AL 1994)

Figure 7



env(t)
FWR
HWR
Band-Pass
Output
Sample Amplitude
Sample Number (62.5 us/sample)
1000
500
0
-1000
0
200
400
600
800
1000
1200

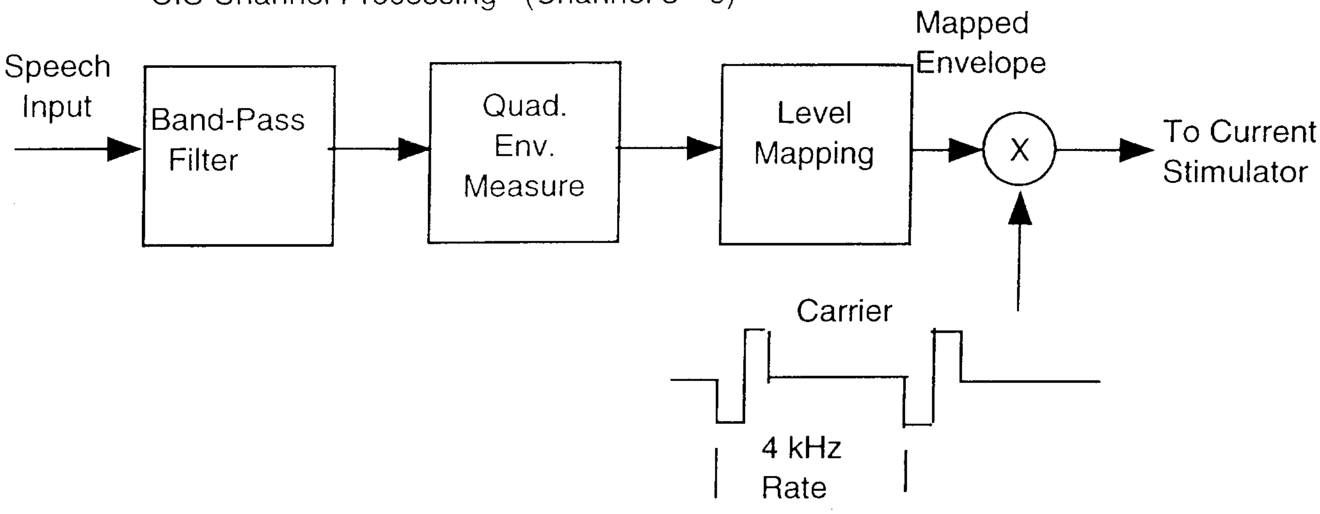
Figure 8



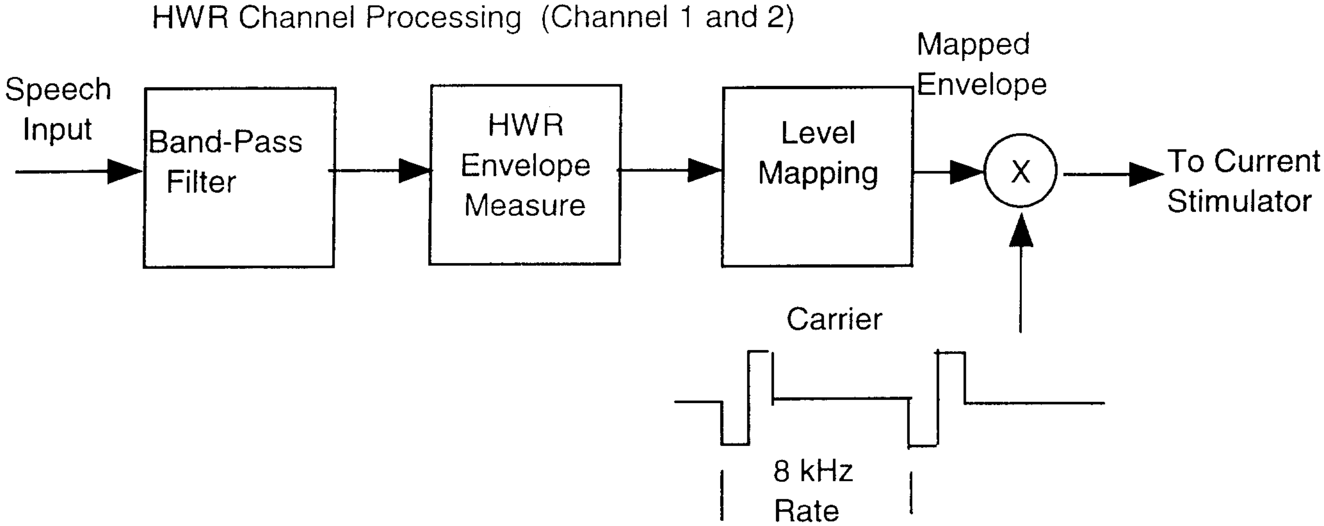
Spectrum Magnitude (dB)
Env(f)
FWR Spectrum
HWR Spectrum
Band-Pass Spectrum
100
80
60
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency (Hz)

Figure 9

CIS Channel Processing (Channel 3 - 6)



HWR Channel Processing (Channel 1 and 2)



Pulsatile Timing Sequence (Channels 1 and 2 @ 8 kHz, 3-6 @ 4 kHz)