

# Improving vowel recognition in noise using the CIS strategy

Philip C. Loizou and Xiaohong Liu

Applied Science Department  
University of Arkansas at Little Rock  
Little Rock, AR 72204-1099  
E-mail: loizou@ualr.edu

<http://giles.ualr.edu/asd/cimplants/>

Research supported by NIDCD.



# Abstract

- The present study focuses on the development of signal processing techniques for improving speech intelligibility in noisy environments. The proposed strategy improves speech intelligibility by enhancing spectral contrast among channel outputs thereby increasing the effective spectral dynamic range. Channel amplitudes are computed according to the Continuous Interleaved Sampling (CIS) strategy, i.e., through bandpass filtering and envelope detection, and then “enhanced” to increase the spectral contrast. The enhanced channel spectrum has sharper peaks and deeper valleys. To test the effectiveness of the proposed method we processed 11 vowels in speech-shaped noise at +5 dB S/N through a simulation of the CIS strategy and the enhanced CIS strategy, and presented them to normal-hearing listeners for identification. Preliminary results indicated that the enhanced CIS strategy yielded a 12% improvement in vowel recognition over the CIS strategy.

# Background

- The effect of noise on speech understanding is far more detrimental for cochlear implant patients (e.g., Loizou *et al.* 1997) than for normal-hearing listeners.
- Noise reduction strategies or signal processing techniques which are robust to noise are needed. Such techniques could be stored in the patient's processor as an additional program since current commercial devices (e.g., Med-El, Clarion) can store multiple programs. One of those programs could be used exclusively, for example, in noisy situations.
- The noise reduction strategies need to be computationally easy to implement, if they are to be used in portable processors.
- The present study presents some preliminary results on such a strategy.

# Effect of noise on CIS strategy

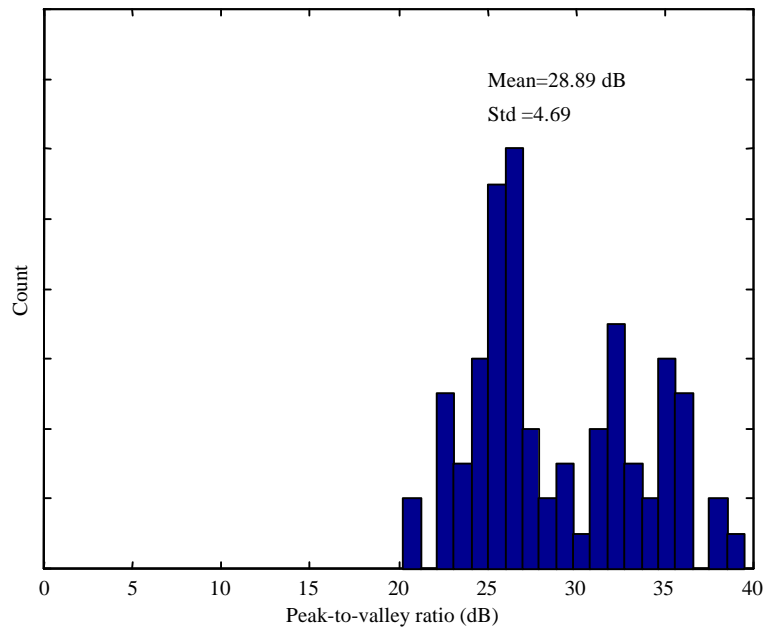
- To investigate the effect of noise on the channel output spectrum, we compared vowels in quiet with vowels in noise processed through the CIS strategy. The channel amplitudes were computed by bandpass filtering the spectrum into 6 channels, and performing envelope detection on the filtered channels.
- Examining the channel spectra of vowels processed in speech-shaped noise (+5 dB/SN) and vowels processed in quiet, we noticed that the noise reduced the spectral dynamic range, i.e., it flattened the channel spectra.
- For most vowels, the location of the spectral peak and the location of the valley was preserved, however the peak-to-valley ratio was greatly reduced.

# Effect of noise on CIS channel output spectra

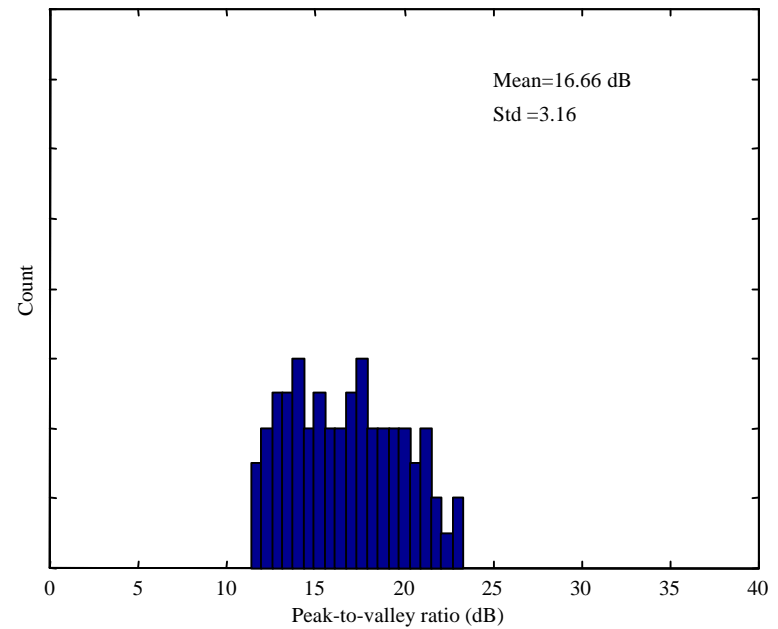
- The spectral peak-to-valley ratio, which essentially defines the spectral contrast, is very important for vowel recognition.
- The reduced spectral contrast, and consequently the reduced spectral dynamic range, makes discrimination of spectral peaks (which contain important formant frequency information) and spectral valleys very difficult. That in turn reduces speech intelligibility.
- The fact that the locations of the peaks and the valleys are preserved but the spectral contrast is reduced suggests a method of enhancing the spectral contrast by sharpening the peaks and deepening the valleys of the channel spectra.
- The proposed method is based on such a principle.

# Histograms of spectral peak-to-valley ratios of the vowel /i/

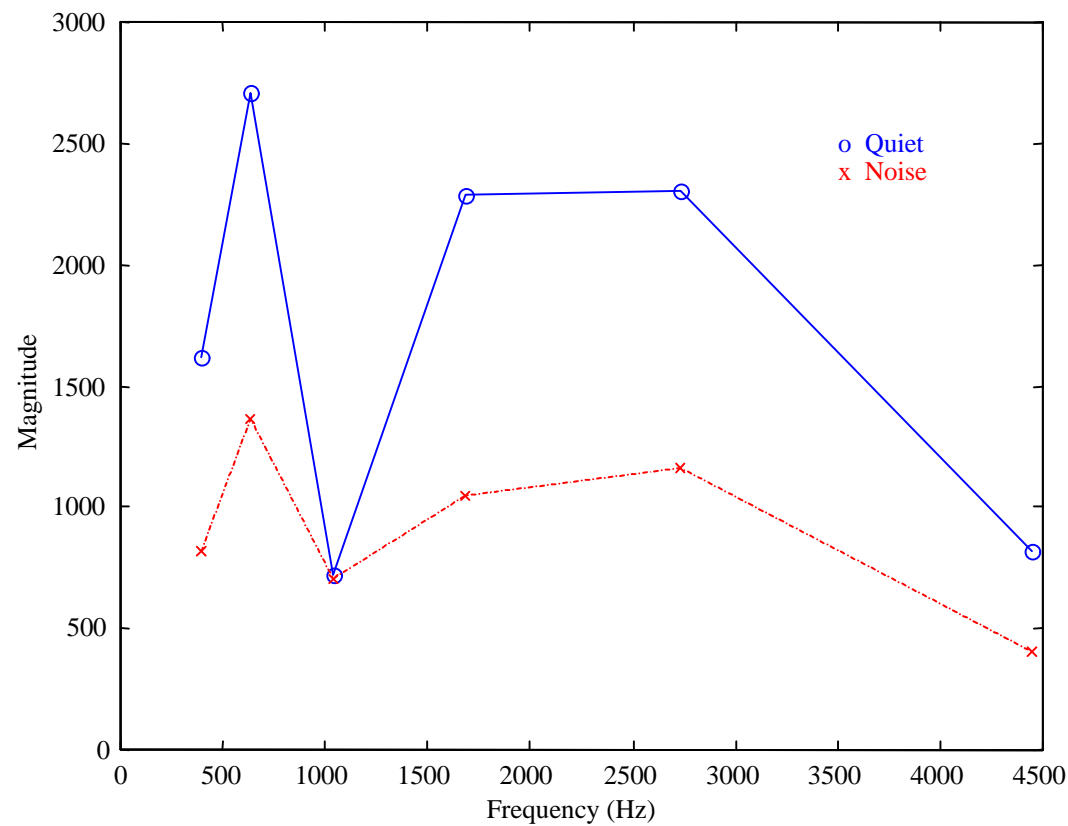
In Quiet



In +5 dB S/N noise



# Effect of noise on channel spectrum of the vowel /ε/



# Methods

- **Subjects**

Four graduate students from the Applied Science Department, UALR, served as subjects. All subjects were native speakers of American English and had normal hearing.

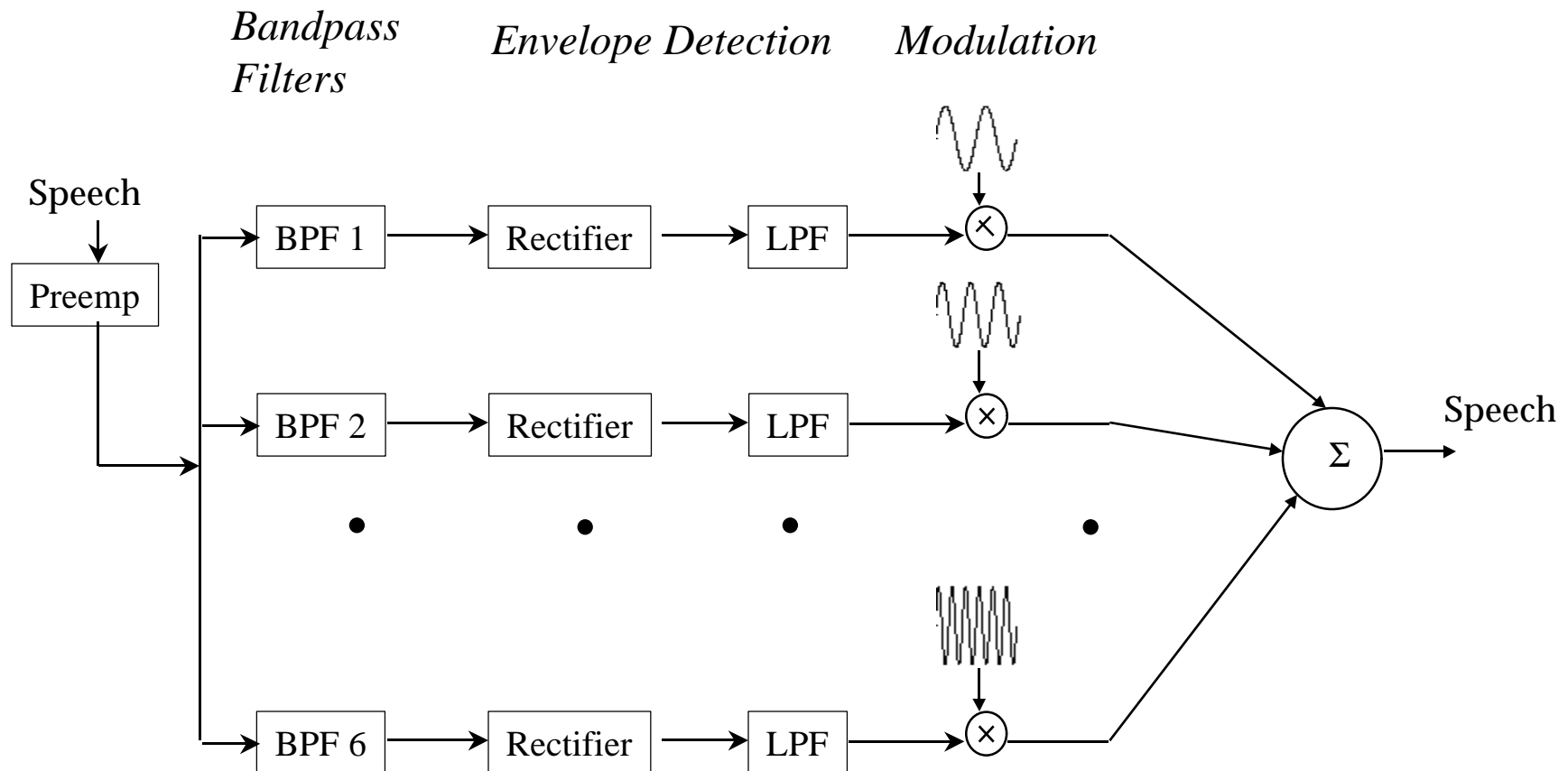
- **Speech Material**

11 vowels in /hVd/ context taken from the Hillenbrand *et al.* (1995) study were used. The vowels were contained in the words: “heed, hid, hoed, hayed, hod, had, head, hood, who’d, hud, heard”. Each vowel was produced by a different male speaker.

# Signal Processing (CIS)

- The vowels were processed through a simulation of the CIS strategy (Wilson *et al.*, 1991).
- Signal was first pre-emphasized, and then applied to a bank of six (logarithmically-spaced) bandpass filters. The bandpass filter outputs were full-wave rectified and low-pass filtered with a cutoff of 400 Hz. Speech was synthesized as a sum of six sinusoids with amplitudes equal to the rms energy of the envelopes (computed every 4 msecs), and frequencies equal to the center frequencies of the bandpass filters.
- For more information about the acoustic simulation see Dorman *et al.* (1997) or Loizou (1998).

# Signal processing



# Proposed noise-reduction strategy (Enhanced CIS)

- Channel amplitudes are computed in a manner similar to the CIS strategy, and then “enhanced” to increase the spectral contrast. The enhancement is performed as follows:
- Let  $A_i$  ( $i=1,2,\dots,6$ ) be the channel amplitudes computed according to the CIS strategy (prior to compression).
- Let  $A_{\max}$  and  $A_{\min}$  be the largest and smallest channel amplitudes computed in each cycle.
- The enhanced channel amplitudes  $B_i$ , are then computed as:

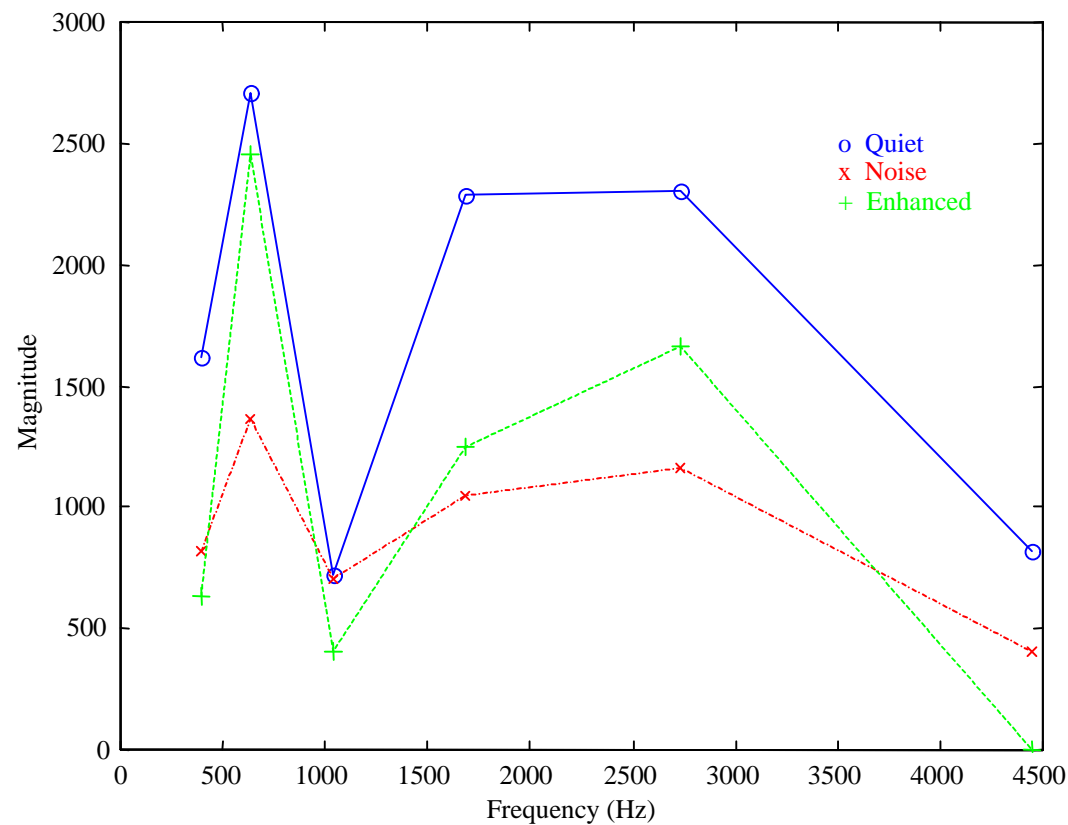
$$B_i = \left[ \frac{A_i - A_{\min}}{A_{\max} - A_{\min}} \right] c A_i \quad i = 1, 2, \dots, 6$$

where  $c > 1$  is a constant, determined experimentally.

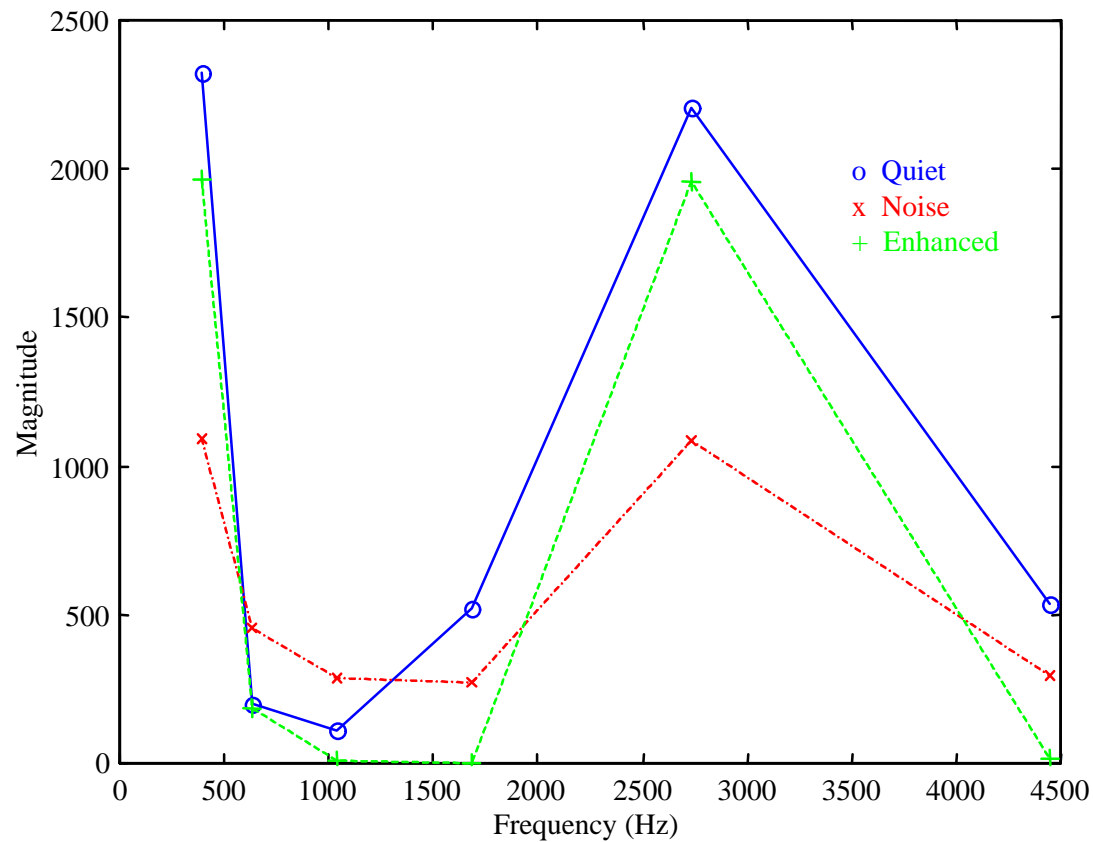
# Enhanced CIS strategy

- Note that the weighting term in brackets takes values in the range of 0 to 1.
- If  $A_i$  has a value close to  $A_{\max}$  (i.e., it is near a spectral peak), then it receives a large weight (number close to 1)  $\gg A_i$  value is enhanced
- If  $A_i$  has a value close to  $A_{\min}$  (i.e., it is near a spectral valley), then it receives a small weight (number close to 0)  $\gg A_i$  value is reduced
- Note that in the actual implementation, if the enhanced channel amplitude  $B_i$  is greater than the corresponding most-comfortable level (MCL) of the  $i$ -th electrode, then it needs to be saturated to MCL. Likewise,  $B_i$  are floored to the threshold values if they are smaller than the threshold value of the  $i$ -th electrode.

# Example of enhanced channel spectrum of the vowel /ε/



# Example of enhanced channel spectrum of the vowel /i/



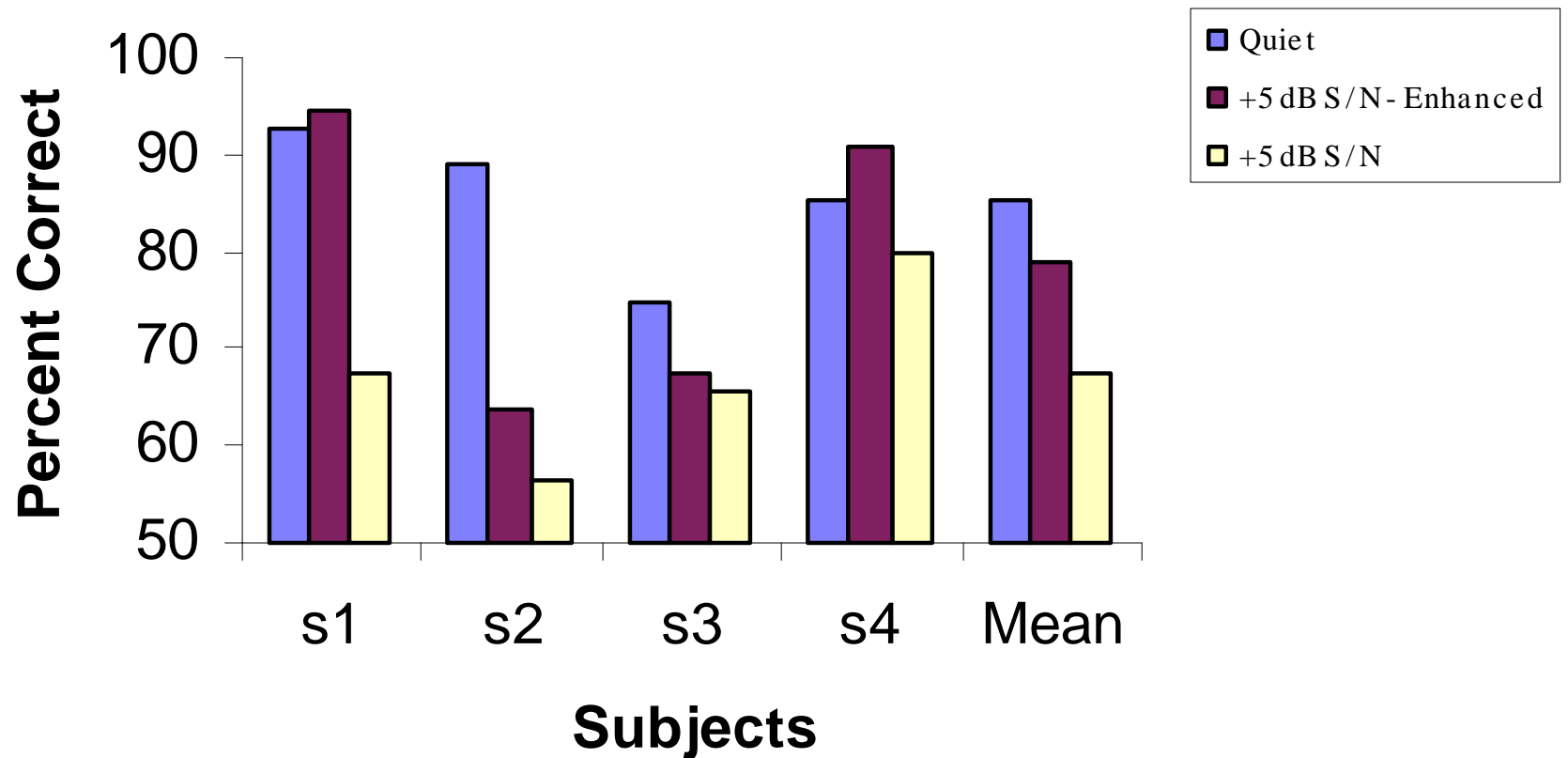
# Procedure

- The experiment was performed on a PC equipped with a sound card. The subjects listened to the processed vowels via closed ear-cushion headphones at a comfortable level set by the subject. A simple graphical interface was used that allowed the subjects to select the vowels they heard.
- The vowels were processed in three conditions:
  - in quiet, using a simulation of a 6-channel CIS strategy.
  - in +5 dB S/N using a simulation of a 6-channel CIS strategy.
  - in +5 dB S/N using a simulation of the enhanced 6-channel CIS strategy.
- The processed vowels were randomized and presented to the listeners 5 times for identification.
- Before testing, the subjects were given a practice session with examples of vowels processed through the same conditions.

# Results

- The mean percent score for vowels processed through the CIS strategy in quiet was 85%
- The mean percent score for vowels in +5 dB S/N processed through the CIS strategy was 67%
- The mean percent score for vowels in +5 dB S/N processed through the enhanced CIS strategy was 79%

# Vowel recognition in quiet and in +5 dB S/N noise



# Conclusions

- A noise reduction strategy was proposed based on the hypothesis that noise reduces the spectral dynamic range of speech processed through the CIS strategy.
- Our results showed that by enhancing the spectral contrast, we can improve recognition of vowels in noise. The proposed noise reduction strategy yielded an improvement of 12% in vowel recognition over the (simulated) CIS strategy.
- Although the patient's electrode dynamic range is limited, future speech processing strategies need to maintain a large spectral contrast between channel amplitudes, especially in noisy conditions.

# Acknowledgements

- This research was supported by a grant from NIDCD.

To download a copy of this poster, visit our web site:  
<http://giles.ualr.edu/asd/cimplants/>

# References

- [1] Dorman, M. Loizou, P. and Rainey, D. (1997). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sinewave and noise-band processors," *J. Acoust. Soc. Am*, 101, 2403-2411.
- [2] J. Hillenbrand, Getty, L, et al. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am*, 97, 3099-3111.
- [3] P. Loizou, S. Graham, et al. (1997). "Comparing the performance of the SPEAK strategy (Spectra 22) and the CIS strategy (Med-El) in quiet and in noise," *1997 Conference on Implantable Auditory Prostheses*, Asilomar, Monterey, CA.
- [4] Loizou, P. (1998). "Mimicking the human ear: An overview of signal processing strategies for cochlear prosthesis," *IEEE Signal Processing Magazine*, 15(5), 101-130.
- [5] Wilson, B., Finley, C., et al. (1991). "Improved speech recognition with cochlear implants," *Nature*, 352, 236-238.