Minimum spectral contrast for vowel identification in CIS-type processors

Oguz Poroy and Philip C. Loizou

Department of Electrical Engineering & Callier Center
University of Texas at Dallas
Richardson, TX
E-mail: loizou@utdallas.edu

http://www.utdallas.edu/~loizou/cimplants/

Research supported by R01 DC03421 from NIDCD.

Introduction

- In the CIS strategy, the channel amplitudes are computed by bandpass filtering the speech signal through a small number of channels, and estimating the envelopes of the filtered waveforms.
- Due to the limited electrode dynamic range, the channel amplitudes are mapped, using a logarithmic-type function, to a smaller dynamic range.
- This mapping not only reduces the amplitude dynamic range, but also reduces the **spectral contrast** - the difference in amplitudes between the channel peaks and the channel troughs.

Introduction

- Background noise may also reduce the spectral contrast (Loizou and Liu, 1998).
- Since the location and relative amplitudes of the peaks cues vowel identity, the differences in channel amplitude between peaks and troughs need to be maintained to some degree for vowel identification.
- It is therefore of interest to know how small of an amplitude difference between the channel peaks and troughs is needed for vowel identification, particularly when vowels are processed through a small number of channels (e.g., as in cochlear implants).

Objectives

- *To determine the minimal spectral contrast needed for vowel identification particularly when vowels are processed through a finite number of channels.*
  - To address this objective, 8 vowels in /hVd/ context were processed through a finite number of channels in a manner similar to the CIS strategy, and reconstructed with varying degrees of spectral contrast ranging from 1-20 dB.

Objectives

- *To determine whether the minimal spectral contrast needed for vowel identification is dependent on the spectral resolution (number of channels). It is hypothesized that a larger spectral contrast is needed for vowels processed through a small number (4) of channels.*
  - To address this objective, the vowels were processed through 4, 6, 8 and 12 channels, and reconstructed with varying degrees of spectral contrast.

Methods

- **Subjects**

  Nine 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**
  - 8 monopthong vowels in /hVd/ context from the Hillenbrand *et al.* (1995) database.
  - The vowels were contained in the words: "heed, hid, hod, had, head, hood, who'd, hud".
  - The vowels were produced by a male speaker (F0=115 Hz).

Signal Processing

- The vowels were processed through a simulation of the CIS strategy (Dorman *et al.*, 1997; Loizou, 1998).
- Signal was first pre-emphasized, and then applied to a bank of *n* (n=4, 6, 8, 12) bandpass filters. The bandpass filter outputs were full-wave rectified and low-pass filtered with a cutoff frequency of 400 Hz.
- Channel amplitudes were estimated by computing the rms energy of the envelopes every 4 msecs.
- The channel amplitudes were then modified to obtain the required peak-to-trough ratio (1, 2, 4, 6, 8, 10, 15 and 20 dB) using the algorithm shown in the next slide.
- Speech was synthesized as a sum of *n* sinusoids with amplitudes equal to the modified amplitudes and frequencies equal to the center frequencies of the bandpass filters.

Algorithm for modifying spectral contrast

- Identify spectral amplitude peak, Ap, and spectral valley, Av
- Define Av1=Ap-ndB (new spectral valley amplitude)

  where ndB = desired peak-to-valley ratio in dB
- Define

$$Ac = \frac{A - Av}{ndB}$$

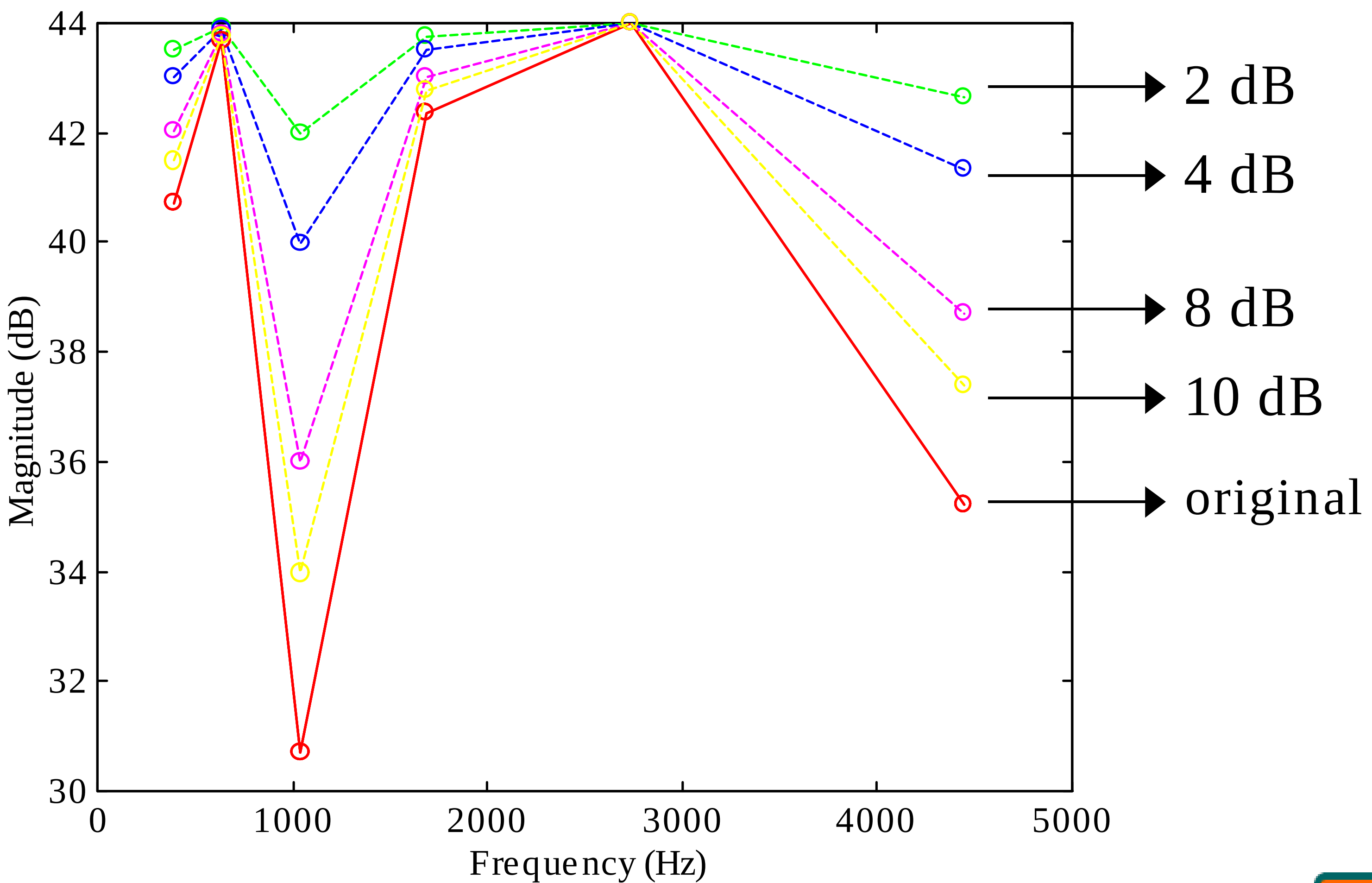
  where **A** is the original channel amplitude vector in dB
- Anew = ndB*Ac + Av1 - Modified channel amplitude vector with ndB peak-to-valley ratio

UTD

Algorithm for modifying spectral contrast

- Note that the algorithm does not simply change the spectral valley amplitude to get the required peak-to-valley ratio. That would distort the shape of the spectrum.
- The proposed algorithm preserves the shape of the channel amplitude spectrum and also maintains the relationship between adjacent channel amplitudes.

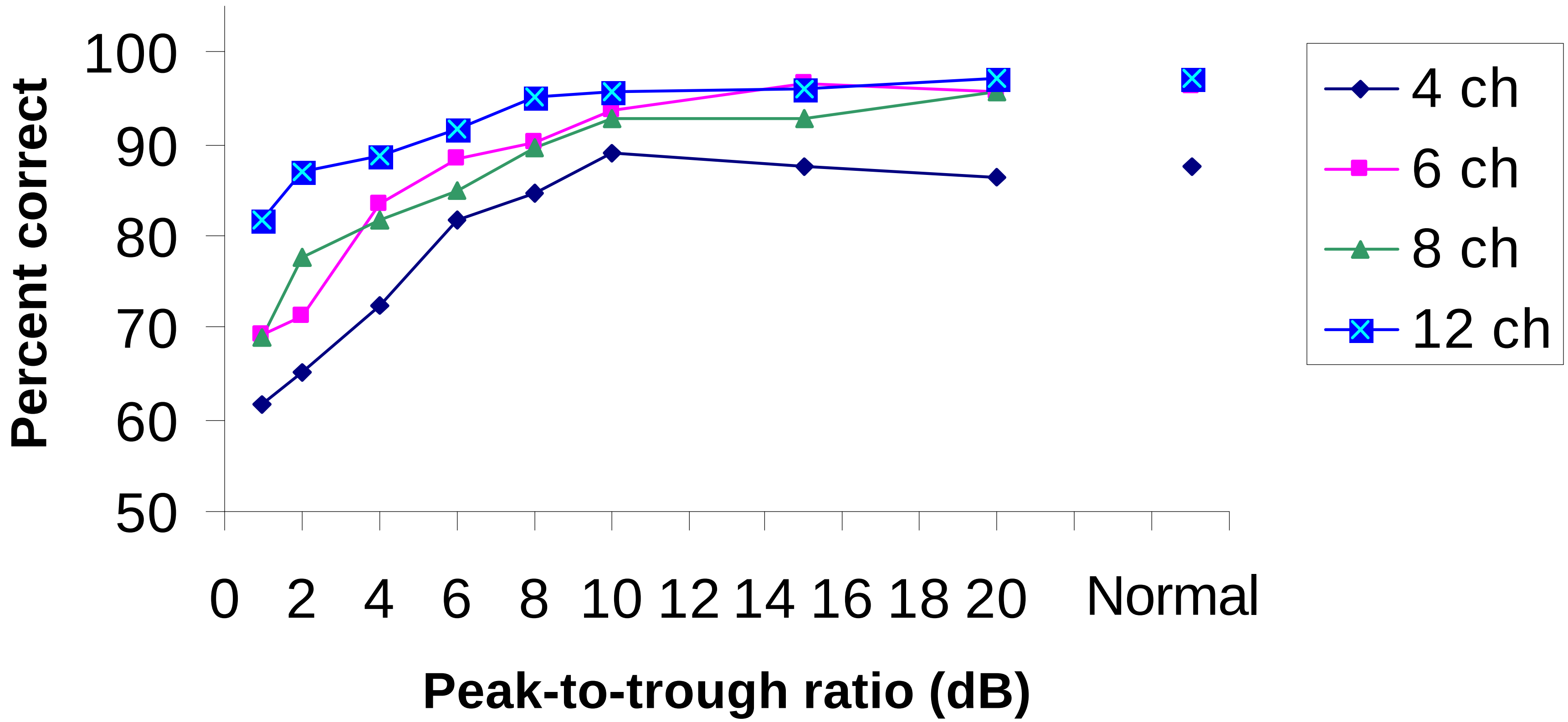
Channel spectra of /ε/ with various peak-to-trough ratios (dB)



Procedure

- The vowels were processed through 4, 6, 8 and 12 channels, and in each channel condition, they were processed into 8 different peak-to-trough levels (1, 2, 4, 6, 8, 10, 15, 20 dB) plus one condition where the original peak-to-trough level was maintained.
- For each channel condition, the peak-to-trough levels and the vowels were all randomized and presented to the normal-hearing listeners 5 times for identification.
- The channel conditions were counterbalanced across subjects to avoid any order effects.
- Before testing, the subjects were given a practice session with examples of vowels processed through the same number of channels.

Results



Percent correct
100
90
80
70
60
50
0 2 4 6 8 10 12 14 16 18 20 Normal
Peak-to-trough ratio (dB)
4 ch
6 ch
8 ch
12 ch

Statistical Analysis

- A two-way (number of channels and spectral contrast) repeated measures ANOVA showed:
  - A significant effect of number of channels, $F(3,24) = 8.7$, $p < 0.0001$
  - A significant effect of spectral contrast, $F(8,64) = 67.3$, $p < 0.0001$
  - A significant interaction between number of channels and spectral contrast, $F(24,192) = 3.7$, $p < 0.0001$

Post-hoc analysis

- 4 channels
  - Performance asymptotes at 10 dB
- 6 channels
  - Performance asymptotes at 10 dB
- 8 channels
  - Performance asymptotes at 8 dB
- 12 channels
  - Performance asymptotes at 8 dB

Discussion

- High vowel identification scores (82%) were obtained even with a 1-dB peak-to-trough difference for the 12-channel vowels.
- In contrast, a 6 dB peak-to-trough difference was needed to achieve the same level of vowel identification for the 4-channel vowels, and a 4 dB difference was needed for the 6- and 8-channel vowels.
- These results suggest an inverse relationship between the importance of spectral resolution (number of channels) and spectral contrast.

Discussion

- The results obtained with 12 channels are consistent with those obtained by Leek *et al*. (1987) who showed that normal-hearing listeners only need a 2-dB peak-to-trough difference to identify four vowel-like complex sounds. [Hearing-impaired listeners needed a 7-dB peak-to-trough difference.]
- Unlike the synthetic vowels used by Leek *et al*, the vowels used in the present study were more natural and contained dynamic spectral information. We believe that this explains the higher identification scores obtained in this study with the 1-dB 12-channel vowels, 82% correct (present study) vs. 60% correct (Leek *et al*. study).

Conclusions

- The minimum spectral contrast needed to recognize vowels with high accuracy (>80%) depends on the spectral resolution (number of channels).
  - A small (1 to 2 dB) spectral peak-to-trough difference is needed for vowels processed through a large number (12) of channels.
  - A relatively larger (6 dB) spectral peak-to-trough difference is needed for vowels processed through a small number (4) of channels.
- Independent of the number of channels, asymptotic performance was achieved when the peak-to-trough difference was greater than 8-10 dB.
- Vowel identification could be improved, for implant patients receiving a small number (4) of channels, if some post-processing of the CIS channel amplitudes is done to enhance the spectral contrast.

Acknowledgements

- This research was supported by Grant No. 1 R01 DC03421 from NIDCD.

To download a copy of this poster, visit our web site:
http://www.utdallas.edu/~loizou/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] Leek, M., Dorman, M. and Summerfield, Q. (1987). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.*, **81(1),** 148-154.

[4] Loizou, P. and Liu, X. (1998). "Improving vowel recognition in noise using the CIS strategy," *29th Annual Neural Prosthesis Workshop*, NIH, Bethesda, MD.

[5] Loizou, P. (1998). "Mimicking the human ear: An overview of signal processing strategies for cochlear prosthesis," *IEEE Signal Processing Magazine*, **15**(5), 101-130.

[6] Wilson, B., Finley, C., et al. (1991). "Improved speech recognition with cochlear implants," *Nature*, 352, 236-238.