BIOMEDICAL SIGNAL PROCESSING - BME 511 – FALL 2020

MATLAB Project: Simulations of Cochlear Implant Signal Processing

Assigned: October 13th, 2020
Due: October 25th, 2020

Submit: 1) ZIP with all MATLAB code (shared among your group); 2) PDF of write-up with all figures to demonstrate results (figures can be the same among groups member, but all words should be independently written).

INTRODUCTION:

Sensorineural deafness can be caused either by cochlear damage or by damage within the auditory nerve or the neurons of the central auditory system. If the damage to the auditory system is peripheral (inner ear) and severe, then a cochlear implant (CI) can be used. CI devices have been successful in restoring hearing to profoundly deaf patients through electrical stimulation of the auditory nerve with an array of fine electrodes (typically fewer than 24) inserted into the scala tympani of the cochlea. Almost all current CI's have adopted speech processing strategies that focus on extracting and representing the slowly varying amplitude modulation (AM) cues from a limited number of frequency bands (Loizou, 1998). This approach has been justified by the finding that speech perception is possible with only a few (3-4) channels providing only AM information (Shannon et al. 1995).

While AM encoding has allowed CI users to achieve good speech recognition in quiet, their performance in noise is severely compromised. While the steady-state speech information is faithfully represented using AM cues, the emotion and prosodic information that is encoded in the time varying formants and pitch is not reliably transmitted to the CI listener. For similar reasons, most CI listeners cannot appreciate music very well. Smith et al. (2002), with their clever auditory chimaeric stimuli, suggested that encoding frequency modulation (FM) in current CI speech processors may help the CI listeners to obtain an appreciation for music and better localization of sounds. While these research findings are very exciting, some technical issues have been raised in the interpretation of Smith et al's perceptual results (Zeng et al., 2004).

In this project, you will explore these issues. The project is divided into three sections. In the first section, you will evaluate the current CI speech processing algorithm with only AM cues. In the second section, you will implement the auditory chimeras to understand the significance of encoding FM into current CI signal-processing algorithms. In the third section, you will learn about one of the technical issues raised in interpreting the results using auditory chimeras and discuss its potential implications for the future of CI research and development.

Note: This "research" project represents a transition from the standard problem sets you have been doing to the independent research project you will do in the last ~4 weeks of the course. This project goes into more depth and is more open ended than the problem sets. Thus, it is advised to take full advantage of the time allotted to do the journal paper reading and MATLAB coding to fully engage with this material. The readings from journal articles are provided so that you can get reasonably up to speed in this research area and dig into it a bit to explore some of the issues that are surrounding a current area of active debate in the field of cochlear-implant and hearing-science research. There are not necessarily right or wrong answers for some parts of this project.

The point is to explore these signal-processing issues in a research-like manner. Note that some sections (e.g., Section B. parts d and e) ask for research-like interpretations related to the field. Thus, a significant amount of the grading will be based on your ability to discuss the research implications of this work, not just getting the MATLAB code to work. This emphasis is reflected in the point values for each section, which are provided at the end. Note, even if you can't get some part of the code to work, you can (and should) still answer the research interpretation questions by assuming a certain outcome and providing a logical interpretation of what the implications would be if you had observed such an outcome.

PART A: Introduction to cochlear-implant signal processing. Replicate the study by Shannon et. al. (1995) to show that only a few channels are needed in a cochlear-implant simulation to obtain relatively good speech perception in quiet, and that perception improves as you add more channels. Specifically, replicate their Fig. 2C.

a) Design filter-bank reconstructions with number of frequency bands (Nb) equal to 1, 2, 4, 8, and 16 (spread across the same total frequency range, e.g., 0 to 4 kHz). Use the same basic filter parameters as in the Shannon et al. study and use a noise carrier in each channel. *[Note: you may benefit in Part A from the chimera code provided for Part B.]*
b) Use various envelope extractions
    a. Rectification/low-pass filter (16 Hz)
    b. Rectification/low-pass filter (160 Hz)
    c. Hilbert transform (see Smith et al. 2002 study below).
c) Plot several example spectrograms of the outputs to show your filterbank is working correctly (see Fig. 1 in Shannon et al.). [NOTE: MATLAB's *spectrogram* function allows control of various parameters related to time/frequency resolution, e.g., *spectrogram(X,WINDOW,NOVERLAP,NFFT,Fs),* which may need to be varied to obtain good looking spectrograms. <u>See supplemental audio lecture and code</u>].
d) Setup a simple perceptual study to replicate the basic findings in Fig. 2C. Twenty sentences are provided in the file SPINsents.ZIP, which are from the SPIN-R sentence database. You don't necessarily need to use all of these stimuli; they are just provided for your convenience/flexibility in setting up a simple demonstration of the basic results in Fig. 2C. Use each lab-group member as a subject in a quantitative (but very simple) perceptual task, i.e., collect quantitative data to demonstrate this effect, but do so in a simple way (i.e., it is OK if this is not a rigorous perceptual design, but it should be quantitative). Show figures to demonstrate your results and comment on their relation to the Shannon results.

PART B: Exploring the potential for frequency modulation information to improve cochlear-implant signal processing. Replicate the study by Smith et al. (2002) to show that envelope information is most useful for speech perception and fine-time information is most useful for music perception. Specifically, replicate Figs. 2 (speech-speech) and 3 (music-music).

a) Take advantage of the chimera-generation MATLAB code provided from the Chimera website (http://research.meei.harvard.edu/chimera/). Be sure to go through the code to understand how each step in Fig. 1 of Smith et al. is implemented in MATLAB. *Since the code is provided, describe the implementation details for 1) the Hilbert transform and 2) the spectrally matched noise. These are clever implementations that are worth taking note of.*
b) Setup a simple perceptual study to replicate the basic findings in Figs. 2 and 3 (Smith et al., 2002) for speech-speech and music-music chimeras. As above, use quantitative data collected

from you and your lab partners. A simple task of "did you hear S1 or S2 (M1 or M2)?" with fixed stimulus pairs should suffice to demonstrate the basic findings.

c) Repeat the chimera synthesis using a rectifier/low-pass filter (16 and 160 Hz) mechanism to extract the envelope, rather than the Hilbert envelope (still use the Hilbert-derived fine-time information). Does this change in envelope extraction affect the perception of chimeric stimuli in a significant way? If so, why do you think this happens?

d) Provide a qualitative description of the findings in this study, and how they support the basic conclusions about speech and music made by Smith et al. Discuss the significance of the transitions observed in Figs. 2 and 3 and how they relate to the basic conclusions made by Smith et al.

e) Provide a qualitative description of the implications of this study for cochlear-implant processing in terms of how speech and music are currently encoded in cochlear implants (Loizou, 1998), and what improvements should be made. Provide a brief discussion of the prospects for making these improvements in CI electrical-stimulation strategies.

PART C: Considering a technical artifact with chimera synthesis and its potential implications

We know that the chimera synthesis algorithm used by Smith et al. (2002) is creating an "acoustic" stimulus with the envelope of one sound (S1) and the fine-time structure of a second sound (S2). However, the brain does not have access to the "acoustic" stimulus directly for perception. Rather, the brain only has access to the output of the cochlea (auditory nerve, AN), which consists of roughly 30,000 narrowly tuned AN fibers with characteristic frequencies (CFs) ranging from 20 Hz -20 kHz. So, one question raised by Zeng et al. (2004) is whether the chimera synthesis algorithm is successfully creating AN responses with the envelope from S1 and the fine-time structure from S2 for all conditions (number of bands) examined in the Smith et al. (2002) study. This question suggests the possibility that the transitions between envelope and fine-time salience seen in Smith et al's Figs 2 and 3 may result from artifacts.

In this part of the project, you will examine two studies (one perceptual: Gilbert and Lorenzi, 2006; one physiological: Heinz and Swaminathan, 2009) that have quantitatively evaluated in a more rigorous manner whether one of the potential artifacts discussed by Zeng et al. (2004) is significant for interpreting the Smith et al. (2002) study. These are more rigorous analyses than the feasibility demonstration done by Zeng et al. (2004) and so it is possible that you will find that some of Zeng et al's claims are not correct, or more likely that they are correct, but that they may not be as significant as claimed (or maybe they are). Thus, this section is getting more into real research and considering ways to evaluate these issues quantitatively.

The first technical issue raised by Zeng et al. (2004) concerns the situation where the chimera filters are much broader than the AN filters (i.e., for a small number of chimera bands). Because auditory filters are narrow relative to the broad chimera bands, it is not necessarily true that the AN output will have the same envelope properties as the "acoustic" chimeras. This may provide an explanation for the transition between envelope and fine-structure salience for speech observed in Smith et al's Fig. 2.

a) Discuss the perceptual and physiological approaches (in general terms) used by Gilbert and Lorenzi (2006) and Heinz and Swaminathan (2009), respectively, to explore the technical issue raised by Zeng et al. (2004) in Section II related to "envelopes recovered by cochlear filtering." Discuss the signal processing issues involved in this issue and the metrics used to quantitatively evaluate this phenomena in each study. NOTE: you do not need to understand every detail of each study (e.g., rectifier distortion removal), but rather focus on the key elements/approaches related to the potential artifact being discussed here in Part C. Based on these studies, do you think this issue can account for the reversal in envelope/fine-time salience for speech for 2 bands observed in Smith et al's Fig. 2, as claimed by Zeng et al? Explain qualitatively why this issue arises.

b) Discuss how this artifact, if true, affects the interpretation of the results in Fig. 2 of Smith et al. and how this affects the basic conclusions they made about the relative salience of envelope and fine-time information for speech and music perception. Does this artifact affect the implications of Smith et al.'s study for CI processing?

References: (further reading provided in ZIP file for those interested)

Gilbert, G., and Lorenzi, C. (2006). "The ability of listeners to use recovered envelope cues from speech fine structure," J Acoust Soc Am, 119, 2438–2444.

Heinz, M. G., and Swaminathan, J. (2009). "Quantifying envelope and fine-structure coding in auditory-nerve responses to chimaeric speech," J. Assoc. Res. Otolaryngol., 10, 407–423.

Loizou, P.C., 1998. Mimicking the human ear. Ieee Signal Processing Magazine, 15, 101-130.

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

Smith, Z.M., Delgutte, B., and Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature, 416, 87-90.

Zeng, F.G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.Y., and Chen, H., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues. J. Acoust. Soc. Am., 116, 1351-1354.

<u>Point Distribution (out of 100 total)</u>

<u>Part A (35 points total)</u>

a) 10
b) 10
c) 5
d) 10

<u>Part B (35 points total)</u>

a) 5
b) 10
c) 5
d) 5
e) 10

<u>Part C (30 points total)</u>

a) 15
b) 15