NCSI (Spring 2019)

Project I

Submission: Due Date 6th March 5 pm by email (sharba@ece.iitkgp.ernet.in). It must be 1 *pdf* file and 1 *m-file* per person with the following:
a) A cogent write up with figures appended at the end, as asked below; the writing should not exceed 2 pages – please refer to figures by figure numbers.
b) A single MATLAB m-file code that should generate all the figures
PLEASE DO NOT SEND ZIP FILES

PART A

Responses to Tones (Rate representation)

1) Use the AN model, as informed in class, to generate the following: Use two high spontaneous rate auditory nerve fibers (HSR ANF) with Best Frequency (BF) of 500 Hz and 4 kHz and obtain their tuning curves (response rates as a function of frequency) at 10 different intensities: -10 dB SPL to 80 dB SPL (in steps of 10 dB). Use tone frequencies of 62.5 Hz to 32 kHz (a total of 9 octaves) with 8 frequencies in each octave (1/8th octave frequency difference). That is the tone frequencies will be 62.5*2.^[0:1/8:9] Hz. Use a duration of 200 ms for each tone and modulate the tones with onset and offset ramps of 10 ms. Use 50 repetitions of each tone and obtain the average rates. Plot all the tuning curves of each ANF in one figure (Figure 1 and Figure 2). Obtain the rate vs intensity function for BF tone of each ANF at the 10 intensities above and plot them (Figure 3). What are the observations?

Responses to Speech (Rate representation)

2) Now have a bank of ANFs starting with BF 62.5 Hz (or check what lowest frequency is allowed) up to 8 kHz (a total of 7 octaves) with 8 ANFs in each octave spaced 1/8th octaves apart (like frequencies presented in Part 1; ANF BFs 62.5*2.^[0:1/8:7] Hz; a total of 57 different BFs of ANFs).

Fixing sound level: Use a steady state portion of the speech sound wavfile provided (‘ah’ part of b**“a”**sketball). Use the *audioread* (or *wavread*) function to read it into MATLAB. Separate the “ah” out from your speech signal waveform. Use the root mean square value of the segment to calculate its dB SPL level [re 20*10^(-6)]. Use this steady state sound level and multiply the entire speech signal with appropriate factors to input in the ANFs (bank) for 3 different sound levels. Determine the 3 sound level as follows: Use the steady state portion and modify it with onset and offset ramps as in Part 1 and find the rate responses to the vowel “ah” of a 500 Hz BF ANF at -20 to 80 dB SPL in 10 dB steps, plot (Figure 4) the rate intensity function (comment by comparing it with the BF tone rate intensity function). Choose 3 sound levels one near (but above) threshold, one in the dynamic range and one in the saturation level close to the end of the dynamic range. After having determined the 3 sound levels determine the spike trains (50 repetitions each) of each ANF in the bank (57 fibers) to the entire speech signal at the 3 sound levels.

Plot (Figure 5) the spectrogram of the speech signal with appropriate window size (25.6 ms *hanning* windows maybe used with overlap of successive windows by 50%, that is, a resolution of 12.8 ms). Now compare the spectrogram with the following: Represent the responses determined above from each ANF as an average rate (number of spikes per unit time) as a function of time. Use windows of 4 ms, 8 ms, 16 ms, 32, ms 64 ms and 128 ms (with overlap between successive windows by 50%, Figure 6A-F, 6 different window sizes). Plot the rate in an image in color with one axis as time (centre of each successive window) and the other axis as BF of the ANFs: It is akin to a spectrogram, only that now you have rate response instead of energy and BF instead of frequency. Also in comparing the spectrogram with the above images do not forget that the spectrogram has a linear frequency axis whereas the ANF BFs are on a logarithmic axis.

Responses to Speech (Fine timescale representation)
3) Use the PSTHs from 50 repeats of the stimulus in every ANF, using a window size 0.01 ms or 10 microseconds. Consider the 12.8 ms long successive windows (50% overlap in successive windows) and get the Fourier Transform (use the *fft* function) of the PSTH. This is an indirect way of looking at phase locking, that too relative amounts of locking to many different frequencies can be observed simultaneously. Find the frequency to which a fiber locks the most, that is, find the peak in the Fourier transform and it's corresponding frequency which is the dominant frequency. Get the dominant frequency in each successive window. Mark the frequency and time location on top of the spectrogram (say with an asterisk). Do not use all the BFs of ANFs for this purpose. Use only BFs ½ octaves apart and from 125 Hz up to 2 kHz. So there will be total of 9 fibers, use 9 colors of asterisks and overlay them on the spectrogram at appropriate frequencies (dominant frequency) and time (Figure 7). Comment on your observations. Comment on comparisons with the rate response plots of 2.

The next part maybe slightly difficult for non-EE/ECE/CS students. However, please seek help from the TAs or me or others in class to do it. ***Do not worry about grading – try to do it and understand the concepts***.

PART B
Read the paper: 1) https://www.ncbi.nlm.nih.gov/pubmed/7569981 available at: http://www.utdallas.edu/~assmann/hcs6367/shannon_zeng_kamath_wygonski_ekelid95.pdfShannon et al 1995, Speech recognition with primarily temporal cues, Science. 1995 Oct 13;270(5234):303-4.
And a further paper: 2) https://www.ncbi.nlm.nih.gov/pubmed/11882898 available at https://www.ee.columbia.edu/~dpwe/e6820/papers/SmithDO02-chimaeric.pdf
Smith et al 2002, Chimaeric sounds reveal dichotomoies in auditory perception, Nature. 2002 Mar 7;416(6876):87-90.

The goal of the next part is to implement a part of the paper (1) Shannon et al 1995 and try to gain understanding of coding principles used in speech perception higher in the auditory pathway based on ANF response properties. Use the methods described in the paper (see Note 7) to modify the provided speech signal. Have 5 cases: 1 band, 2 bands, 4 bands, 8 bands and 16 bands – the filters should be logarithmically spaced and should span 90 Hz to 5.76 kHz (6 octaves). Thus for 1 band case the lower and higher cut-offs should be 90 Hz and 5.76 kHz, and for the 2 band case the 2 filters should be 90 Hz to 720 Hz and 720 Hz to 5.76 kHz. Use fourth order bandpass Butterworth (MATLAB function *butter*) filters instead of elliptic IIR and no need for the pre-emphasis filter. For the filtering operation use the *filtfilt* function. For extraction of envelope use the Hilbert transform (https://in.mathworks.com/help/signal/ug/envelope-extraction-using-the-analytic-signal.html) and then low pass filter (again use *butter* for the low pass filter). Create the new sounds and have someone who has not heard the sentence tell you what they hear (give comments). No need to do elaborate statistics with multiple speech sounds as in the paper. Get a qualitative idea whether intelligibility increases or not and by how many bands is the sound clearly understood.

Next use the 1 band sound and 4 band sounds and repeat what you did in Part A2 and A3. Now comment on the observations and how it relates to your friend's (listener in previous paragraph) qualitative assessment of the speech sound.