HW 3: Information Theory Problem Set II
42-699B/86-595 Neural Data Analysis
Due @ noon on Tues 10/9/12.

In this homework set, you will apply the techniques you learned in class to determine how neurons in the inferior colliculus convey information about two sound localization cues: interaural level differences (ILD) and spectral notches (SN). Specifically, you will investigate how one (representative) neuron encodes the two cues in spike count and in spike timing.

In the data you will analyze, anesthetized cats listened to sounds presented over headphones. These sounds were processed to have a particular ILD and SN cue superimposed on them. In total, 5 different ILDs were used and 5 different SNs were used, for a total of 25 different possible ILD/SN combinations. Over the course of this experiment, each stimulus was presented 102 times. You will compute the mutual information carried by the spike trains about either ILD or SN, MI(R;ILD) and MI(R;SN), where R is either the spike count of the neuron (question 1), or R is an estimate of ILD or SN derived using the spike distance metric method.

Download the file HW3_MIdata.mat from blackboard and load it up in Matlab. It contains one matrix, snild_dat, that has a size of nspikes x 4. (nspikes is the total number of spikes recorded from this neuron during the experiment.) The first column of this matrix gives the identity of the ILD cue used (a number ranging from 1 to 5) when this spike was recorded. The second column gives the identity of the SN cue used (also a number ranging from 1 to 5) when this spike was recorded. The third column gives the trial number (which repetition of this stimulus the spike was recorded on). It ranges from 1 to 102. Finally, the fourth column gives the time of the recorded spike (in seconds), relative to stimulus onset.

As an example, the command *trial_1_inds=find(snild_dat(:,1)==2 & snild_dat(:,2)==3 & snild_dat(:,3)==1);* will give you all the indices of the spikes recorded during the first presentation of the sound with ILD=2 and SN=3. The command *snild_dat(trial_1_inds,4)* will display all of those spike times.

First, you will compute the information carried in this neuron’s spike count about both ILD and SN cues.

1. **(40 pts.) Matlab analysis: Information carried by spike count.**

    1a. (5 pts) Plot the mean and std of the spike count as a function of ILD (from 1 to 5).

    1b. (5 pts) Plot the mean and std of the spike count as a function of SN (from 1 to 5).

    1c. (5 pts) In one plot, make 2 histograms: p(r|ild=3) and p(r|ild=5). That is, have r on the x-axis, and p(r|ild) on y-axis as the height of the bar. (Note, if you concatenate the two column vectors p(r|ild=3) and p(r|ild=5) into one nx2 matrix, and use the bar command in Matlab, they will show up in different colors in a single plot.)

1d. (5 pts) Repeat 1c for p(r|sn=3) and p(r|sn=5).

1e. (5 pts) From 1c and 1d, would you expect spike count to carry more information about ILD or SN? Why?

1f. (5 pts) Compute H(ILD), H(SN), and H(R). What's the maximum value of MI(R;ILD)? MI(R;SN)?

1g. (5 pts) Compute H(R|ILD) and H(R|SN).

1h. (5 pts) Compute MI(R;ILD) and MI(R;SN).

Now you will compute the information carried in spike timing about ILD and SN cues. You will do this using the spike-distance metric method discussed in class. The file HW3_spkd.m available on blackboard will calculate the V&P distance between two given spike trains. It takes as input the times of the first spike train, the times of the second spike train, and a cost parameter. Download the file and play with it until you know how to use it.

Computing the distance between every pair of recorded spike trains included in our data set would take too long without recoding HW3_spkd.m to be much more efficient. So, for question 2 you are going to use only the first 5 repetitions of each stimulus. The following commands should be used to make a reduced data set snild_dat2 for the next question:

```
inds=find(snild_dat(:,3)<=5);
snild_dat2=snild_dat(inds,:);
```

2. (55 pts.) Matlab analysis: Information carried by spike timing.

Note: 2a, 2b, and 2c should be plotted in one figure. Use the commands after 2c to make it look nice.

2a. (2 pts) Suppose you have the following two spike trains, with spike times of:

spike_train_1 = [0, .1, .2, .3] seconds, and spike_train_2 = [0, .2, .3] seconds.

Use HW3_spkd.m to compute the distance between these two spike trains for the following costs: [0,1,3,10,30,100,300,1000,3000,10000]. Plot the distance as a function of cost.

2b. (2 pts) Repeat 2a for these two spike trains: spike_train_1 = [0,.1,.2,.3] seconds, and spike_train_2 = [.01,.11,.21,.31] seconds.

2c. (2 pts) Repeat 2b for these spike trains: spike_train_1 = [0,.1,.2,.3] seconds, and spike_train_2 = [.05,.15,.25,.35] seconds.

Note: the following commands will give you a nice plot for 2a, 2b, and 2c:

```
figure;
set(gcf,'Name','2a,b,c');
plot([.1,costs(2:end)],dist_2a,'k'); hold on;
plot([.1,costs(2:end)],dist_2b,'b');
plot([.1,costs(2:end)],dist_2c,'r'); hold off;
xlabel('Cost [s^{-1}]');
ylabel('Distance');
title('2a,b,c: Distance v Cost');
legend({'2a: D([0,.1,.2,.3],[0,.2,.3])',...
   '2b: D([0,.1,.2,.3],[.01,.11,.21,.31])',...
   '2c: D([0,.1,.2,.3],[.05,.15,.25,.35])'},'Location','NorthWest');
set(gca,'XScale','Log');
set(gca,'XTick',[.1,1,10,100,1000,10000]);
set(gca,'XTickLabel',{'0','1','10','100','1000','10000'});
ylim([-.05,8.5]);
```

Use HW3_spkd.m to compute the pairwise distances between all spike trains in the snild_dat2 data set at the following costs: [0,1,3,10,30,100,300,1000,3000,10000]. It will be needed to answer the questions below.

2d. (8 pts) Plot a histogram of all of the distances between the spike train for ILD=1, SN=1, and trial=1, and every other spike train that had SN=1 in snild_dat2, for cost=1000. What is the mean distance to other spike trains with SN=1?

2e. (8 pts) Plot a histogram of all of the distances between the spike train for ILD=1, SN=1, and trial=1, and every other spike train that had SN=5 in snild_dat2, for cost=1000. What is the mean distance to other spike trains with SN=5? For a cost of 1000, is this spike train closer to the SN=5 group or the SN=1 group?

2f. (8 pts) For each spike train, compute the ILD and SN group that it is "closest to", on average, for a particular cost. Use this to build a set of confusion matrices, confmat_ild, such that confmat_ild{q}(i,j) gives the number of times a spike train from the ILD=i group was closest to the ILD=j group, under cost q. Display confmat_ild for q=1000. (Hint: don't include the "self-distance" when computing which group the spike train belongs to!) Should there be a lot of information about ILD at this cost?

2g. (8 pts) Repeat 2f for SN instead of ILD. Should there be a lot of information about SN at this cost?

2h. (8 pts) Using the confusion matrices, compute (and plot) MI(ILD_estimate;ILD) and MI(SN_estimate;SN) as a function of cost.

2i. (9 pts) Write a short summary describing how SN and ILD cues are represented in this neuron.

3. (5 pts.) About how much time did you spend on each question of this problem set? Which (sub) problem taught you the most, and which taught you the least?