Information in the brain

We will look at two applications of information theory to the study of the brain; the first, dealt with here is calculating the information content of spike trains. This is a example of data analysis for interpreting experimental data in neuroscience. The second, infomax, is an interpretation of information processing in the brain, which may have something to tell us about how to compute with data; this will be in the next chapter.

The information content of spike trains

Neurons communicate using action potentials, discrete voltage spikes that travel from neuron to neuron. This is the main way information propagates through the brain. The action potentials, often also called spikes, a stereotypical in amplitude, so the information that carry is represented in their timing. Spike trains are very noisy, repeated responses to the same stimulus tend to be very different; for example, in Fig. 1 there are raster plots from neurons in the auditory area of the zebra finch brain; a raster plot is a plot with a little dash for each spike. The individual trials run horizontally. As you can see there is considerable variation from trial to trial. This variability is unsurprising, the brain is made up of multiple interlocking ad hoc networks so there is likely to be considerable background noise. The question is then, how much information do spike trains carry.

An attempt to answer this question is found in **?**. They consider the response of the large movement sensitive 'H1' neurons at the back of the brain of blowfly. They recorded from these neurons for some time while the fly was shown a moving stimulus. They then discretized the spike trains, binning them into small time bins, as in Fig. 2. If the time bins are chosen small enough, this is a sequence of zeros and ones. The sequence of zeros and ones can then be split into a sequence of words

$$010001000000100 \rightarrow 01000, 10000, 00100 \tag{1}$$
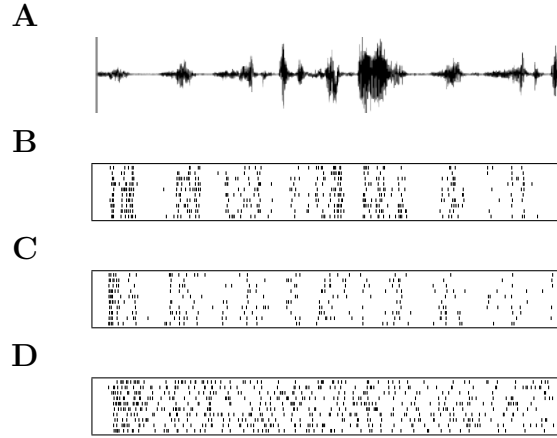
A

B

C

D

Figure 1: The response of cells in the auditory fore brain of zebra finch to a conspecific song stimulus. **A** shows the sonogram for the song; **B**, **C** and **D** show rasterplots for three responses, each one second long. For each cell there are ten trials, stacked vertically, with each horizontal line corresponding to a single trial. Each small dash is a single spike.

There words can be considered a set of outcomes for a random variable:

$$
\begin{aligned}
\mathbf{w}_0 &= (0,0,0,0,0) \\
\mathbf{w}_1 &= (0,0,0,0,1) \\
\mathbf{w}_2 &= (0,0,0,1,0) \\
\mathbf{w}_3 &= (0,0,0,1,1) \\
&\quad \ldots
\end{aligned}
\tag{2}
$$

Now the idea is to calculate the mutual input between the $\mathbf{w}_i$s and the stimulus. To this end they recorded two cases, in the first they varied the input for some time, allowing them to calculate $H(W)$; in the second they repeated the same stimulus sequence many times, allowing them to calculate $H(W|S)$; S represents the stimulus. In practice this means that they played the same movement sequence many times. This movement sequence is split up into 30 ms sections, each section counts as one stimulus, say s represents on of these stimuli. It is possible to calculate $p_{S|W}(W|S = s)$ by counting

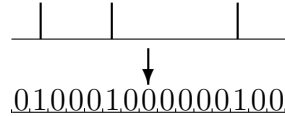Figure 2: Discretization: the spike train is changed into a series of zeros and ones by binning then using a small time step.

the occurance of each word during the repititions of s. From this $H(W|S = s)$ is calculated. Now, under the assumption that each stimulus is equally likely, this is averaged over the stimuli to give $H(W|S)$. A similar procedure using the non-repeating stimulus give $H(W)$ and from the difference comes $I(W;S)$.

One question is how to decide how small to make the discretization time and how long to make the words. It is argued that the blowfly responds very quickly to attempts to swat it, so the information coming from H1 is being interpreted by other brain areas over a time scale of 30 ms. The discretization size is $\delta t = 3$ ms. In **?** they report 78 ± 5 bits per second, which, given the firing rate, equates to 1.8 ± 0.1 bits per spike.

Obviously one interesting question is how the mutual information varies with δt, the discretization size; if it increases as δt decreases, this indicates timings carry information at smaller time scales. The problem is that the smaller δt, the more words there are; with $\delta t = 1$ ms there are 2^{30}, about a billion different words. Another approach is to preturb the spike train, for example by jittering the spike times. Looking at how this changes the mutual information tells us how information is coded in the spike trains.

The difficulty with this approach to the information content of spike trains is the huge number of words: with 30 ms words and 3 ms letters there are $2^{10} = 1024$ words. If six seconds of data are used for the repeating stimulus, that is 100 different stimuli, then even for a three hour recording there are 1800 trails for each stimulus, not a huge amount for estimating 1024 probabilities. Thus means that it would take a lot of data to reliably estimate the probabilities needed to calculate the mutual information. Of course, in the case of blowfly considerable amounts of data are available and the quick reaction time also makes the restriction to 30 ms words reasonable,

but the size of data set does present a challenge. There are clever ways to improve this situation; already in **?** they look at how the estimate varies with the amount of data and then extrapolate. In **?** the estimation of probability from counting is improved in a way that improves the estimate for mutual information. I also look at this problem in **??**