Neural response time integration subserves perceptual decisions - K-F Wong and X-J Wang's reduced model

Chris Ayers and Narayanan Krishnamurthy

December 15, 2008

Abstract

A neural network describing the dynamics of lateral interparietal area (LIP) neurons in decision making was reduced from eleven differential equations to two. The reduced model was used to explain psychophysical data recorded during a two choice discrimination task. Investigation focused on effects of stimulus strength on decision accuracy, decision latency, and the conditions under which the model could further be reduced.

1 Introduction

'How do neurons in a decision circuit integrate time-varying signals, in favor of or against alternative choice options?' Neurophysiological studies reveal numerous cortical areas perform decision making. For example, one's ability to choose a target amongst distractors has been shown to be mediated by the frontal eye field (FEF).

Volitional attention to parts of a visual scene operates on salience maps in LIP through a similar process.

X-J Wang proposed a spiking neural network model to address the decision making process [5]. Although this model was successful, it was later simplified to better support analysis [6]. In section 2 we describe the reduction methodology. In 2007 Wong et al. [7] use the reduced model to explain the observed response of LIP neurons when the random dot stimulus was overlaid with a brief background movement. The ability of their reduced model to reproduce the experimental results of Huk and Shadlen [3] validates the flexibility and correctness of their model.

1.1 Background

Movement detection and perception was first studied by Britten et al. [2] in the medial temporal (MT) area. Subjects were presented with a display of randomly moving dots, as shown in figure 1 (a). For each trial c'% of the random dots were selected to move coherently. Subjects were trained to indicate indicate the motion direction with either a button press, or an eye movement.

Shadlen and Newsome [9] used the same coherent random dot motion technique to record responses from LIP. LIP neurons receive input from MT neurons and exhibit a slow ramping activity that is correlated with the formation of perceptual decisions about movement direction. The difficulty of these decisions were varied from trial to trial by changing c'. At lower motion coherence the subject's response time was longer, and the ramping activity of an LIP neuron was slower. Yet for all levels of c' and for all reaction times the firing rates reached the same level at the time of decision. This phenomenon is the ramp to threshold decision making process of neurons. See results figure 3, which replicates the LIP neuron's response

in Shadlen and Newsome's experiments. Figure 2 replicates their psychometric data from neurometric latency in response to stimulus.

2 Model Reduction

Wong et al. reduced the full scale spiking model to a two variable model using the mean field theory technique proposed first by Wilson and Cowan [10]. Models developed by Abbott and Chance [1] showed that individual excitatory synaptic inputs can act as either a driver or modulator depending on how they are linked with inhibition. This approach to modeling relies on the central limit theorem, which posits that the net input to a neuron from N neurons, as $(N \to \infty)$, can be treated as a Gaussian random process. The mean activity of a homogenous network of N neurons can then be represented by the activity of a single unit.

The three populations of neurons involved in the decision making circuit are:

- *direction selective neurons*, which encode complimentary signals $V+$ and $V-$ as discovered by Romo, and are further used to model frequency discriminability by Manchens et al. [4].

- *inhibitory interneurons*, which play a crucial role in maintaining timing that subserves oscillations, implicated in working memory

- *non-selective neurons*, which are indifferent to the stimulus of interest, whose main role is to provide noisy background activity.

In addition to the above simplifications it was proposed that behavioral response times were determined by the large time constant of NMDA receptors. This is supported by AMPA and GABA-A activity (fast excitation and fast inhibition) which attains steady state values quickly relative to NMDA. Figure 1 (b) illustrates the

approach to arriving at two mutually inhibitory populations of neurons with recurrent excitation and external inputs.

3 Methods

3.1 Algorithm

As discussed, the full spiking model was reduced to two ODEs. The simplified model describes two populations, each selective for a movement direction. The dynamics of these regions are modeled as rate driven gates, dominated by the NMDA decay time.

$$\frac{dS_i}{dt} = -\frac{S_i}{\tau_{NMDA}} + (1 - S_i)\gamma f(I_i), \quad i = 1, 2 \tag{1}$$

$$\tau_{NMDA} = 60 \text{ ms}, \quad \gamma = 0.641$$

The parameter γ is a unitless conversion from Hz to open gate probability.

The function $f(I_i)$ is an input-output relationship derived from the full model. This simplification was motivated by a desire to remove the self-consistency checks involved in calculating current and firing rate. The firing rate was solved as function of the two currents within the operating region and fit to the current-frequency relationship developed Abbott and Chance [1] using the parameters a, b & d.

$$r_i = f(I_i) = \frac{aI_i - b}{1 - \exp[-d(aI_i - b)]} \tag{2}$$

$$a = 270 \text{ Hz/ nA}, \quad b = 108 \text{ Hz}, \quad d = 0.154 \text{ s}$$

The system is driven by synaptic input from MT. For all tests the stimulus was in the receptive field of the first population. Hence, I_1 is an increasing and I_2 is a decreasing

function of c':

$$\begin{aligned} I_1 &= J_{A,ext}\mu_0\left(1 + f\frac{c'}{100\%}\right) \text{ nA} \\ I_2 &= J_{A,ext}\mu_0\left(1 - f\frac{c'}{100\%}\right) \text{ nA} \\ J_{A,ext} = .5e^{-3} \text{ nA} \cdot \text{ Hz}^{-1}, &\quad \mu_0 = 30 \text{ Hz}, \quad f = .45 \end{aligned} \tag{3}$$

Each population is recurrently coupled in addition to receiving inhibitory input from the opposing population:

$$\begin{aligned} I_A &= J_{N,11}S_1 - J_{N,12}S_2 + I_1 + I_{noise,1} \\ I_B &= J_{N,22}S_2 - J_{N,21}S_1 + I_2 + I_{noise,2} \\ J_{N,11} = J_{N,22} = 0.3725 \text{ nA}, &\quad J_{N,12} = J_{N,21} = 0.1137 \text{ nA} \end{aligned} \tag{4}$$

This simple coupling scheme can represent both decision and memory, as will be shown. Finally, noise has been reintroduced into the system:

$$\begin{aligned} \tau_{noise} = \frac{dI_{noise,i}}{dt} &= -(I_{noise,i} - I_0) + \eta_i\sqrt{\tau_{noise}}\sigma_{noise} \\ \sigma_{noise} = 0.009, &\quad I_0 = 0.3297 \text{ nA}, \quad \tau_{noise} = 2 \text{ ms} \end{aligned} \tag{5}$$

All equations are simulated in Matlab R2008a using Euler integration with a .1 ms time step.

3.2 Caveats

Although this project seeks to reproduce the findings of [6], the model itself was adapted from a later incarnation[7]. It was found that the original curve parameters

of the I/O function did not yield the expected results. To achieve an approximate conversion the currents I_1 & I_2 were reduced.

The model as described is to be simulated on the seconds, not milliseconds, timescale. That having been said, for clarity all units of time hereafter are presented in milliseconds.

4 Results and Discussion

4.0.1 Model Validation

The goal of the reduced model is to efficiently and accurately reproduce the behavior of the full spiking neural network. One method of validation discussed is to recreate the psychometric curves of the full model (Fig. 2). This is accomplished by simulating over many trials with different noise realizations and calculating the percentage of correct outcomes. It was shown that the behavior of the reduced model matches well the curve of the original, and data collected from LIP.[8]

A similar approach was used to compare the mean decision times. The decision reached by the system was determined by whichever population exceeded the threshold of 15 Hz first (Fig. 3). As coherence increased the decision accuracy increased and decision time decreased.

4.0.2 Phase Space Analysis

The primary advantage of producing such a simplified model was its ready adaptation to phase plane analysis. Several important characteristics of the system can be easily described by the interaction of nullclines.

Without a stimulus the intersection of the two nullclines created five fixed points, three of which were stable (Fig. 4). The “central” attractor represented the steady-

state of the system before/between stimuli. The other two stable points could be used to represent working memory. Figure 7 (a) is the response of the two populations to memory guided decisions. It shows the cells maintain their activity after the stimulus is turned off.

When a stimulus was applied the nullclines moved from the origin, with their interaction determined by coherence strength, c' (Fig. 5). At low levels of coherence the three central fixed points merged into a single saddle node, which drove the evolution of the system. The dynamics of the saddle node, combined with the added noise forced the system to one of the competing populations. This is the method by which decisions are made.

At high levels of coherence the basin for the winning population grew larger, and furthermore was biased towards the origin (Fig. 6). As a result the ease with which the system leaves the stable manifold for the correct population increases, until the correct decision is made in almost every trial. Above 70% coherence the attractor for the opposing population will be annihilated, after which only the correct decision can be made.

The time to a correct decision was shown to be inversely proportional to movement coherence. The reason for this becomes clear when viewing the system in phase space. At low motion coherence the basins are roughly equal and the system will progress down the stable manifold to the central saddle node. Initially the gradient of the saddle will dominate motion, but eventually the noise will succeed in driving the system into one of the basins. Once leaving the stable manifold the decision can be reversed, but there is a point at which the gradient of the winning attractor can not be negated by the noise.

Motion coherence increases the natural "desire" for the system to move towards a preferred state preceding a decision. The curvature of the stable manifold increases

and the probability of the system progressing along the manifold decreases. Moving from the current state requires both a longer path and time as the dynamics near the saddle node is governed by small gradients. As the system nears the saddle the decision is furthered delayed until the noise forces a choice. For this reason error trials take longer with increasing coherence because the state must pass close to the saddle node to make it into any basin.

4.0.3 Working memory and stimulus driven neural activity in different epochs preceding a decision

Perceptual decisions occur during epochs when a dominant population ramps past a threshold. The temporal evolution of neural activity in memory guided decision making and its dependence on stimulus strength is shown in figures 7 (a) and (b). Cells maintain their activity after stimulus withdrawal. The epoch with largest dependence on motion strength and greatest separation of the two neural responses happens at (0.5-1s), which is the late phase of stimulus onset. There is little dependence of neural response to motion strength in the late memory period, which is intuitive as the driving stimulus is no longer present.

5 Conclusion

We have been able to replicate the results of X-J Wang's 2006 paper. The results from the reduced model were validated and used to interpret experimental observations using phase plane analysis. The model was used to interpret the effects of stimulus strength on decision accuracy, decision latency. The model is able to maintain a decision over time without input stimuli, analogous to working memory in memory guided decisions.

References

[1] Abbott, L.F. and Chance F.S., *Drivers and modulators from push-pull and balanced synaptic input*, Progress in Brain Research 149, 147-155 , 2005.

[2] Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA, *A relationship between behavioral choice and the visual responses of neurons in macaque MT*, Visual Neuroscience 13:87100, 1996.

[3] Huk AC, Shadlen MN, *A neural integrator underlying perceptual decision-making in macaque parietal cortex*, Journal of Neuroscience, 25:10420 10436, 2005.

[4] Christian K. Machens, Ranulfo Romo, Carlos D. Brody, *Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination* , Science, Vol. 307, pages: 1121 - 1124 , 2005.

[5] Xiao-Jing Wang, *Probabilistic Decision Making by Slow Reverberation in Cortical Circuits*, Neuron, Vol. 36 , pages: 955 - 968, 2002.

[6] Kong-Fatt Wong and Xiao-Jing Wang, *A Recurrent Network Mechanism of Time Integration in Perceptual Decisions*, Journal of Neuroscience, Vol. 26 No.4 , pages: 1314 - 1328, 2006.

[7] Kong-Fatt Wong , Alexander C. Huk , Michael N. Shadlen and Xiao-Jing Wang, *Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making* , Frontiers in Computational Neuroscience, Vol. 1 No.6 , doi:10.3389, 2007.

[8] Roitman JD, Shadlen MN, *Response of neurons in the lateral interparietal area during a combined visual discrimination reaction time task*, Journal of Neuroscience 22:9475-9489.

[9] Shadlen MN, Newsome WT, *Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey*, Journal of Neurophysiology, 86:1916 1936, 2001.

[10] Wilson H, Cowan J, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophysical Journal J 12:124, 1972.

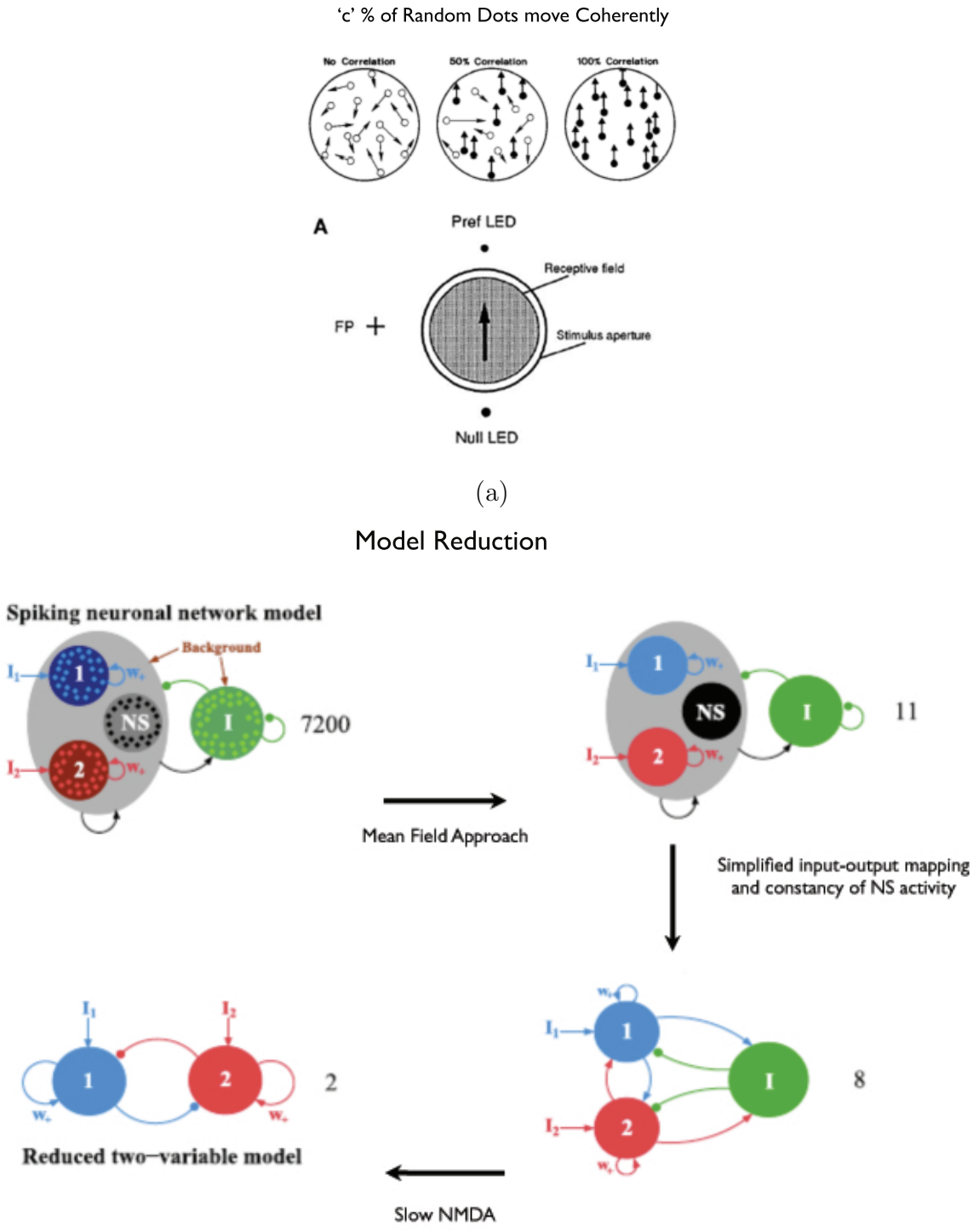


Figure 1: (a) c'% of random dots move coherently (b) model reduction flow chart - see section 2 for details

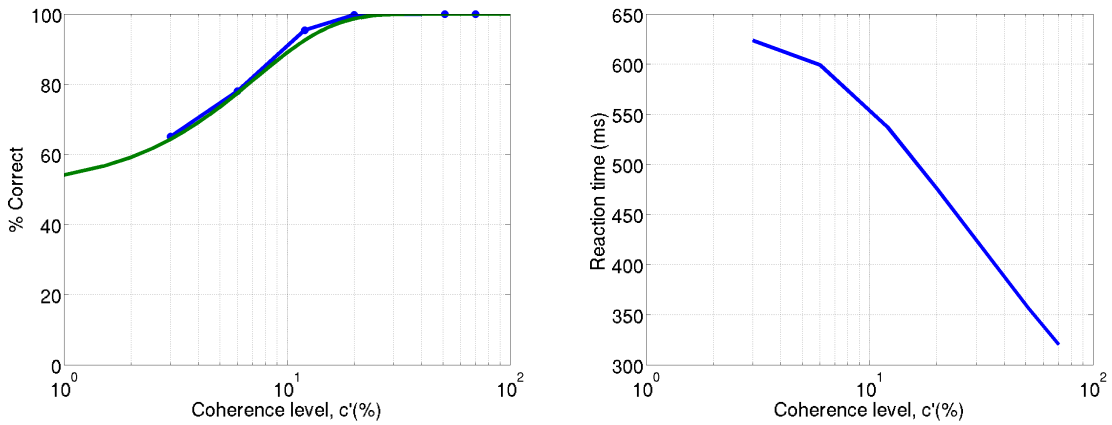


Figure 2: Psychometric curves of simplified model. Left, decision accuracy as a function of coherence strength. Generated data in blue, original curve fit shown in green. Right, Reaction time as a function of coherence strength. Only successful trials are shown here

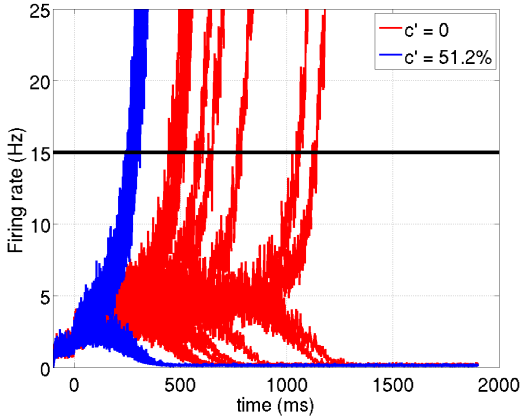


Figure 3: Firing rates as a function of coherence. The decision threshold of 15 Hz is shown here. Note the difference in reaction time between coherence levels

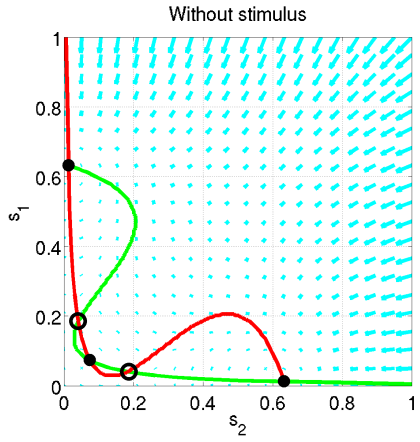


Figure 4: Phase space representation of model. Model prior to application of stimulus. Stable fixed points show as filled circles, unstable as open circles. $dS_2/dt = 0$ shown in red, $dS_1/dt = 0$ shown in green.

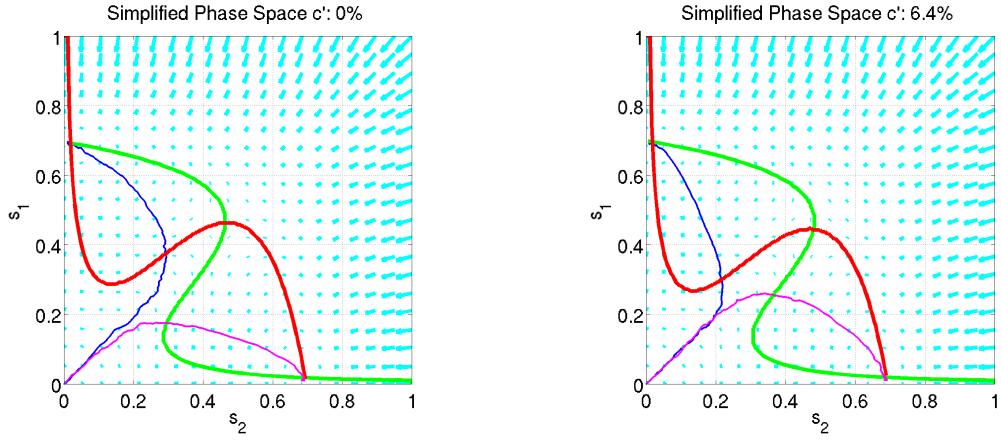


Figure 5: Phase space at low coherence. Left, Model after application of $c = 0'\%$ stimulus. Success in blue, Error trial in magenta. Right, Even at low stimulus values the model predicts the majority of decisions will be correct. Note the longer path of the error trial.

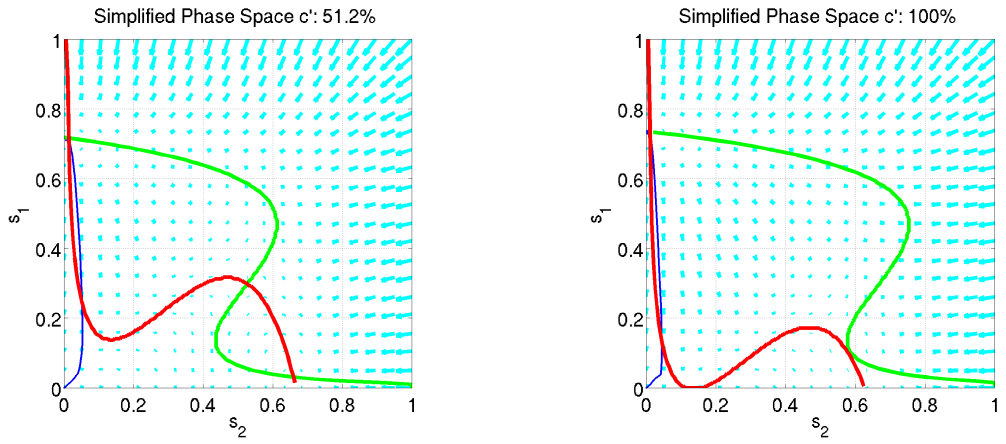
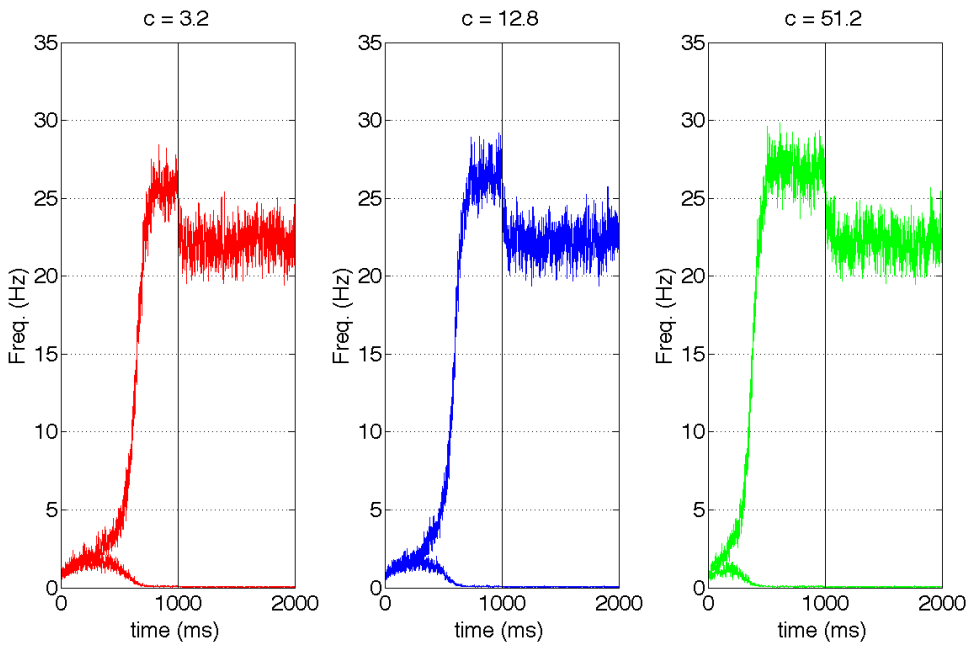
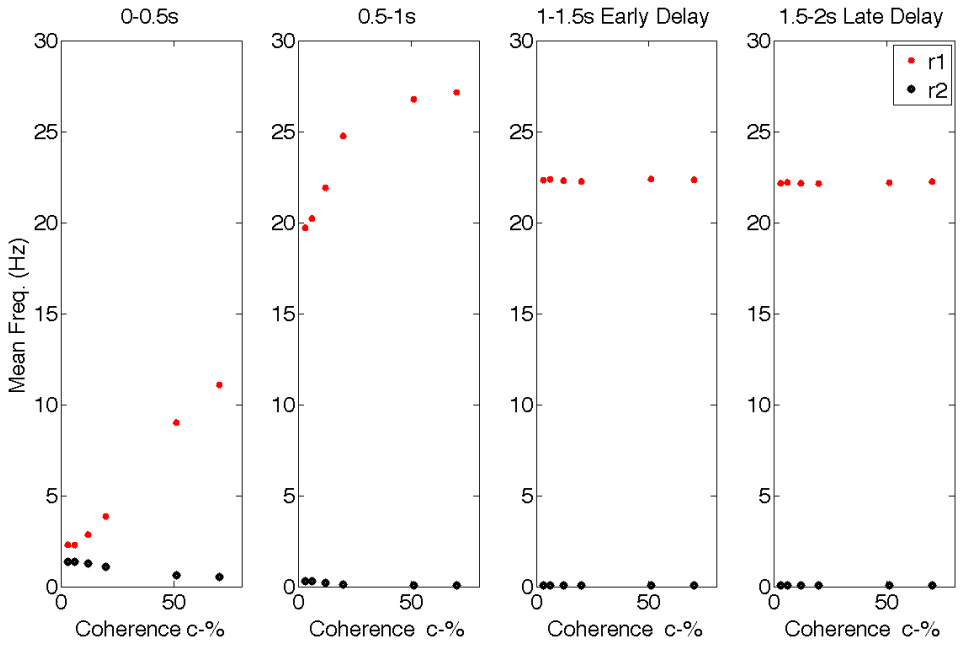


Figure 6: Phase space at high coherence. Left, At 51.2% coherence the correct decision is almost always made. Right, eventually the attractor for the opposing population is annihilated.



(a)



(b)

Figure 7: (a) Delayed response samples for different coherence levels. Neural populations maintain its activity after an initial drop in the rate, when their input stimulus is turned off (1s). Maintenance of activity in the absence of input stimuli is posited as the mechanism for working memory. The response for stronger motion coherence stimulus, reaches the threshold more rapidly than a weaker motion coherence stimulus. (b) Dependence of neural activity on motion strength in different epochs, averaged over ten trials. The epoch with largest dependence on motion strength and greatest separation of the two neural responses happens at (0.5-1s), which is the late phase of stimulus onset. There is little dependence of neural response, to motion strength in the late memory period, which is intuitive as the driving stimulus is no longer present.