Development of an Integrated Model of Primary Visual Cortex

by

Benjamin Selby

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Network-level models of visual processing have potentially important insights for applications such as computer vision and robotics. Primary visual cortex is a key stage of visual processing with involvement in many circuits proposed for these applications including motion tracking, object recognition, and control of eye movements. However, no model of V1 to date has captured the complete set of observed behaviour in a large-scale model. Linear kernel methods with threshold and divisive non-linearities can reproduce classical receptive field behaviour, but not the full range of non-classical behaviours. The stabilized supralinear network (SSN) provides a simple scheme of lateral interactions that produce a wealth of observed V1 behaviour not previously captured with linear kernel methods. However, the SSN is restricted in stimulus selectivity and is not pixel-computable, limiting its utility for real-world applications. Integrating a linear kernel model with the SSN resulted in a model that is pixel-computable and produces a wide range of classical and non-classical behaviour. With further development this network model will be usable in visual processing circuits.

The SSN was also expanded to use binocular stimuli. Using an optimization procedure, SSN parameters were found that produce interocular transfer of suppression in excitatory units, but not inhibitory ones. The lack of interocular transfer in inhibitory units may indicate that an alternate inhibition-stabilization scheme is more biophysically realistic.

Mammalian visual perception is enabled not only by neural processing but also by precise eye movements, which allow for efficient scanning of the environment. This thesis describes the requirements for a robot that can orient cameras with the same dynamics as macaque monkey eyes as well as a camera system that reproduces macaque visual acuity.

Acknowledgements

One thing I have learned through graduate school is the value of support and the many places it can be found. I guess Jon Bon Jovi was right - "No man is an island".

I first must give thanks to Bryan Tripp, my supervisor, who gave me the opportunity to pursue this work (and let me complete it from afar after I moved to Calgary). Bryan is a mentor who leads by example, and I am very grateful for his numerous insights, patience, and generosity with his time, as well as for introducing me to the world of neuroscience research.

My family has provided invaluable support and encouragement throughout. Thanks to my parents, Keya and Steve, and my sister Elly, for always being supportive and available even from a distance. Thanks also to Carol and John Beckel, who have helped make me feel at home in Calgary.

Thanks to my friends in Waterloo, Toronto, Calgary, and elsewhere for keeping me sane after too many days locked in an office. Special thanks are due to Karl Price and Brian Luptak for valuable technical discussions particularly regarding chapter 4.

Thanks to my labmates Omid Rezai and Salman Khan for their companionship and countless coffee breaks at the CnD. Special thanks are due to Omid for his constant encouragement and many long Skype sessions debugging code.

Thanks to Chris Eliasmith and John Zelek for reading drafts of this thesis.

Last but not least, I must give thanks to Molly Beckel. Without Molly's ever-present love and support this thesis would not have been possible.

Dedication

To Molly, my best pal

Table of Contents

List of Tables

List of Figures

Introduction

Motivation and goals

The major goal of this thesis is to develop a foundation for study of the mammalian visual system and how it can be usefully applied to problems in computer vision and robotics. This requires both biologically reasonable models of visual processing, as well as robotic systems which can emulate the movements that enable mammalian perception. Primary visual cortex, or V1, is the largest visual area in the mammalian brain and also the first cortical stage of visual processing. It is involved in many suggested circuits with diverse functions including object recognition, motion tracking, and eye-movement control.

In this work a model of V1 is developed which integrates two approaches: receptive field and network modelling. The resulting model produces a wealth of behaviour observed in V1 including diverse stimulus selectivity and several more subtle or "non-classical" behaviours. In addition a set of specifications is presented for the design of a "monkey head" robot which can orient two cameras independently with dynamics similar to those of a macaque monkey. The cameras used in this robot were selected to reproduce to non-uniform visual acuity of the macaque eyes. This specification was provided to a contract and at the time of this writing the detailed design is in its final revisions.

The long-term goal of this project is to produce biologically-plausible networks that can produce animal-like performance for various visual tasks. By conducting a network-modelling study of visual processing, further understanding of neural processes is developed and can likely be applied to other brain areas. The work presented here is simply a starting point for achieving this goal - there is much more left to do than has been accomplished here. The model and robot developed here will provide a platform for more work towards this goal to be done.

Thesis organization

Chapter 1 provides a brief overview of the neuroanatomy and physiology relevant to this model of V1.

Chapter 2 describes my earlier attempt at producing a *receptive field* model of V1, which abstracts neural processing using linear kernels and basic nonlinearities, similar to some earlier models. This model is successful in producing various stimulus selectivities, but fails to produce some non-classical behaviour.

Chapter 3 describes a network model called the stabilized supralinear network (or SSN),

developed by Kenneth D. Miller's lab. This model produces a variety of non-classical behaviours not seen in the receptive field model, but does not have the same level of stimulus selectivity and is not pixel-computable. By integrating the receptive field model developed in chapter 3 into the SSN, an *integrated model* of V1 is produced which is pixel computable and produces a wide range of behaviour.

Chapter 4 describes the development process of the monkey head robot including dynamic requirements and visual requirements, as well as camera and lens selection. Two design alternatives are presented and compared. The detailed design is currently being completed by a contractor.

Finally, chapter 5 describes some immediate next steps to be taken with the model, including integration with the finalized and fabricated robot.

Chapter 1

V1 Anatomy and Physiology

How do we see? Today there is no definitive answer, though not for lack of searching for one. The potential impact of finding an answer to this question is enormous, carrying implications for neuroscience, medicine, and intelligence research to name a few. One reason to study vision is to develop a framework for understanding the brain - processes which produce vision may be similar to processes elsewhere in the brain. Today, many researchers study neurological vision in order to improve the performance of computer vision, which currently as a field cannot produce performance comparable to that of humans in tasks including object recognition, motion tracking, and navigation (Kruger et al., 2013).

Primary visual cortex, or V1, has long been an area of interest for neuroscientists and vision researchers. Its large size and location have made it relatively easy to study, and many principles developed and discovered by studying V1 may be applied to other areas of the brain. In this chapter the anatomy and physiology of mammalian vision are briefly reviewed, with emphasis on primary visual cortex.

1.1 Neural processing

The neuron is the basic cell of the brain. There exist about 100 types of neurons found across species and in different areas of the brain. Neurons have been studied in great detail, most of which is out of the scope of this work (see Schwartz, Barres, and Goldman (2013) for a comprehensive review). What is critical is that neurons are the substrate of information processing in the brain, transmitting electrochemical impulses in selective response to stimuli. Figure 1.1 depicts a neuron from the hippocampus.

Neurons are electrochemically excitable - that is, they can be made to respond to electrical and chemical stimuli. They receive stimuli through their dendrites, which can receive impulses from many other neurons. They respond to such stimuli electrochemically - through the transmission of current through the axon, culminating in the release of neurotransmitter. This event is called an *action potential* or *spike* (Schwartz, Barres, and Goldman, 2013). Depending on the type of neurotransmitter released at the termination of the action potential, other neurons can be excited (increasing likelihood of producing an action potential in the post-synaptic neuron) or inhibited (decreasing that likelihood). Neurons have a *spike threshold* - when net excitation to a neuron exceeds this threshold, that neuron will spike.

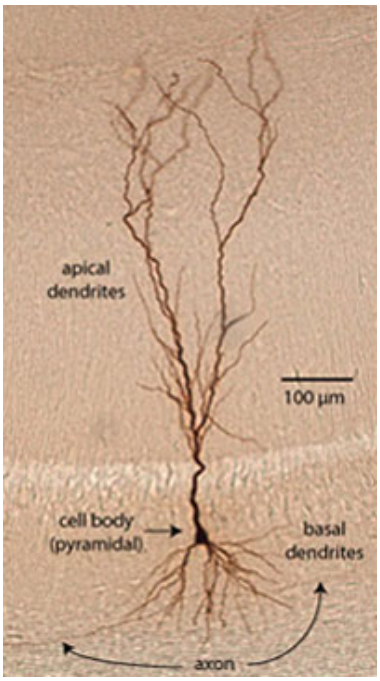

Figure 1.1: A pyramidal neuron. Both the axon and dendrites branch extensively, so this neuron receives input from many others and outputs to several others. Image from Spruston (2009).

Spikes are the all-or-nothing response of a neuron, without them no information is transmitted from one neuron to another. Spikes propagate unidirectionally down the axon from the cell body to the synapse - information is only be transmitted from the pre-synaptic cell to the post-synaptic one. Neurons can produce several hundred spikes per second (Schwartz, Barres, and Goldman, 2013).

While there are many subtleties in neuronal communication, many aspects of behaviour can be understood by studying the spike rate of neurons (Ermentrout, 1998; Dayan and Abbott, 2001). A neuron's spike rate depends on the driving stimulus. The stimulus which produces a maximal firing rate is called the *preferred stimulus*. Neurons in different areas prefer different types of stimuli - some may respond to a certain smell or a sound, while others may respond to pictures of Jennifer Aniston (Quiroga et al., 2005). A neuron's preference for certain types of stimuli (called its *selectivity* or *tuning*) is determined by the connections to that neuron.

1.2 Visual processing in the mammalian brain

Historically the brain is divided in areas based on distinct structure and function. Areas are implicated in serving specific functional roles and do so through specialized structure. Brain areas vary in terms of neuron types, density, connections between neurons, and in many other ways.

The defining feature of the mammalian brain is the *cerebral cortex*, the wrinkled outer layer of the brain. The cerebral cortex (or simply *cortex*) is associated with higher-level

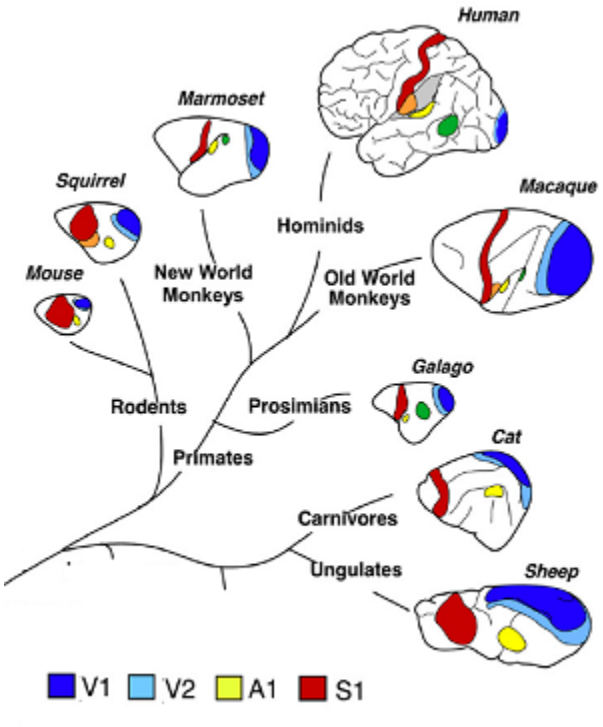

Figure 1.2: Variation of location of functional cortical regions amongst 8 species with their taxonomic relationships. Figure from Carandini (2012).

processing observed in mammals including motor skills, visual perception, and language and consciousness in humans. Cortex is specialized into functionally distinct regions across its surface. The location and size of a functional region varies across species. Figure 1.2 shows several functional regions and their locations in different species.

Cortex is made up of 6 major layers, enumerated from outermost (layer 1) to innermost (layer 6). The layers have similar connectivities throughout the brain, and columns of cortex tend to be excited together, indicating that the cortical column functioning as a whole enables cortical processing (Mountcastle, 1978; Mountcastle, 1997). For instance, layer 4 is the primary recipient of feedforward connections from lower brain areas, and layer 6 also receives some feedforward input. Layers 2 and 3 receive input from layer 4, and send their output to higher brain areas. Similar computations may be performed across cortex despite functional differences, which has prompted researchers to search for evidence of canonical cortical computations.

Visual processing is both parallel and constructive. It is parallel in that the entire visual scene is processed by a brain area simultaneously. Information is transmitted (via spikes) from one area to the next, and through neural selectivity this information is processed, and so perception is constructed. Figure 1.3 illustrates the flow of visual information in the brain. Light enters the eye and stimulates photoreceptor cells, which stimulate responses in *retinal ganglion* cells. The axons of these neurons then form the optic nerve, which carries visual information to the *lateral geniculate nuclei* (LGN), where contrast in the visual information is sharpened. The neurons in LGN then stimulate neurons in layer 4 of

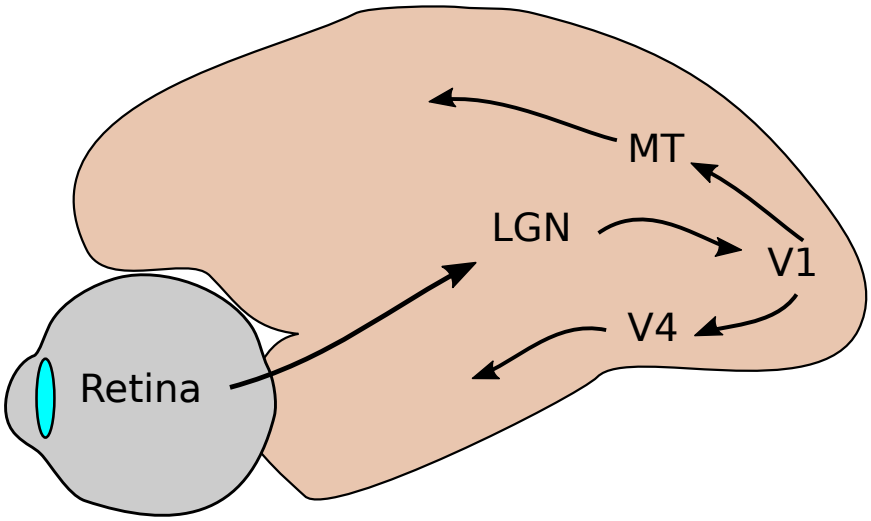

Figure 1.3: Flow of visual information in the macaque brain. The LGN are in fact two nuclei in the thalamus, below the cortical surface. V1, V4, and MT are major cortical areas. The arrows indicate feedforward connections, however cortical areas receive an even larger degree of feedback connections from higher areas.

V1, where additional selectivity is introduced. V1 feeds forward to several other brain areas, each of which processes the information differently. For instance, area MT extracts motion information, while area V4 extracts shape and colour information (Gilbert, 2013b). These "higher" areas also project feedback connections, and indeed these feedback connections outnumber feedforward connections to V1 one by a factor of 10 (Gilbert, 2013b).

1.3 Primary visual cortex

Primary visual cortex, or V1, is the first cortical stage of visual processing, receiving feed-forward input from the lateral geniculate nuclei (LGN) in the thalamus. It is located in both hemispheres of the posterior occipital lobe. In the macaque monkey, V1 accounts for 13% of the total cortical surface area (Van Essen, 2004). V1 sends feed-forward connections to other cortical areas including V2, V3, MT, MST, and FEF, as well as several subcortical regions (Carandini, 2012). The right hemisphere of V1 processes the left visual field while the left hemisphere processes the right visual field. Like all of the neocortex, macaque V1 can be divided into 6 primary layers with several additional sublayers. In the macaque connectivity between these layers is quite specific and distinct (Casagrande and Xu, 2004). Parallel pathways extend from the retina through LGN to V1 carrying distinct visual information integrated at higher levels. These pathways carry into specific layers of V1 for separate processing. Callaway (1998) provides a comprehensive review of V1 circuitry including laminar connections and cell types.

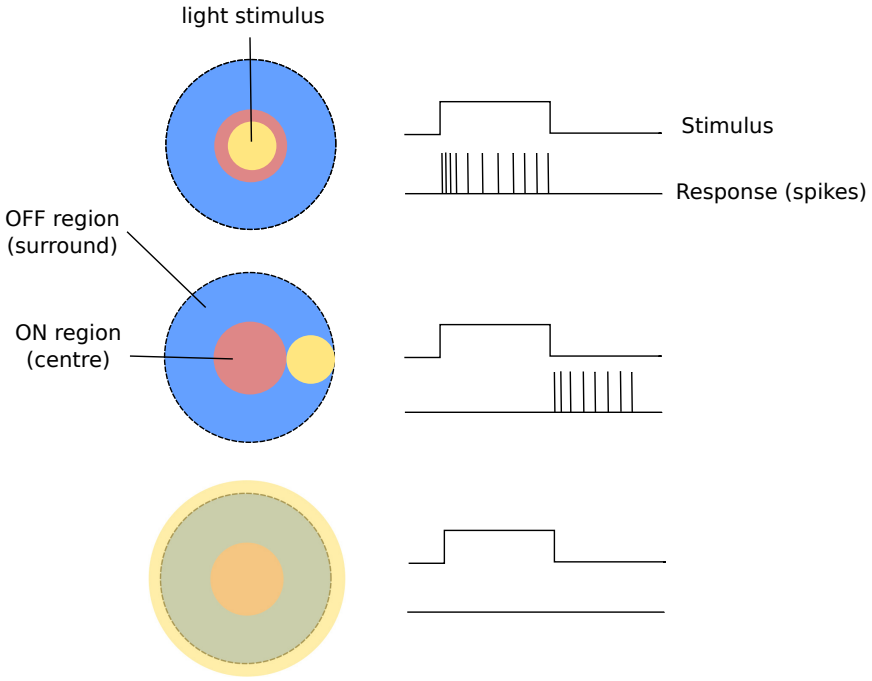

Figure 1.4: Sample receptive field for relay neuron in the lateral geniculate nucleus (LGN). When a light stimulus is applied to the ON-centre, the neuron responds by spiking. Similarly, when a light stimulus is removed from the OFF-surround region, the neuron responds. When both regions are stimulated simultaneously the neuron does not fire. Adapted from Gilbert (2013b).

1.3.1 Receptive fields

Visual neurons have a property called a *receptive field*, referring to the region of the visual field in which stimuli can elicit a response from that neuron (Gilbert, 2013b). The definition of a receptive field includes not only general position in the visual field, but also the excitatory and inhibitory regions which produce stimulus selectivity (Kuffler, 1953). Relay cells in LGN typically have two co-centric, circular receptive field regions: an ON-region and an OFF-region, called a centre-surround receptive field. ON-regions are excited when stimulated by light while OFF-regions are excited when a light stimulus is removed, and inhibited while stimulated by light. Figure 1.4 illustrates this receptive field arrangement and behaviour.

In V1, the most apparent trait of neurons is their selectivity for orientation of contrast in their receptive field (Hubel and Wiesel, 1959; Hubel and Wiesel, 1968) - contrast bars of one orientation produce spikes while those of orthogonal orientation produce none. This is a result of several V1 relay cells (whose receptive field structure is similar to that of LGN relay cells) with partially overlapping centre-surround receptive fields forming synapses onto individual V1 neurons, effectively elongating the ON- and OFF- regions into oriented bars. Figure 1.5 illustrates this elongation through combination of presynaptic input.

V1 is a *retinotopic map*, meaning the neurons of V1 are arranged as a mapping of the visual field - adjacent neurons respond to adjacent areas of the visual field. Neurons corresponding to the centre of the visual field have the smallest receptive field size, and the

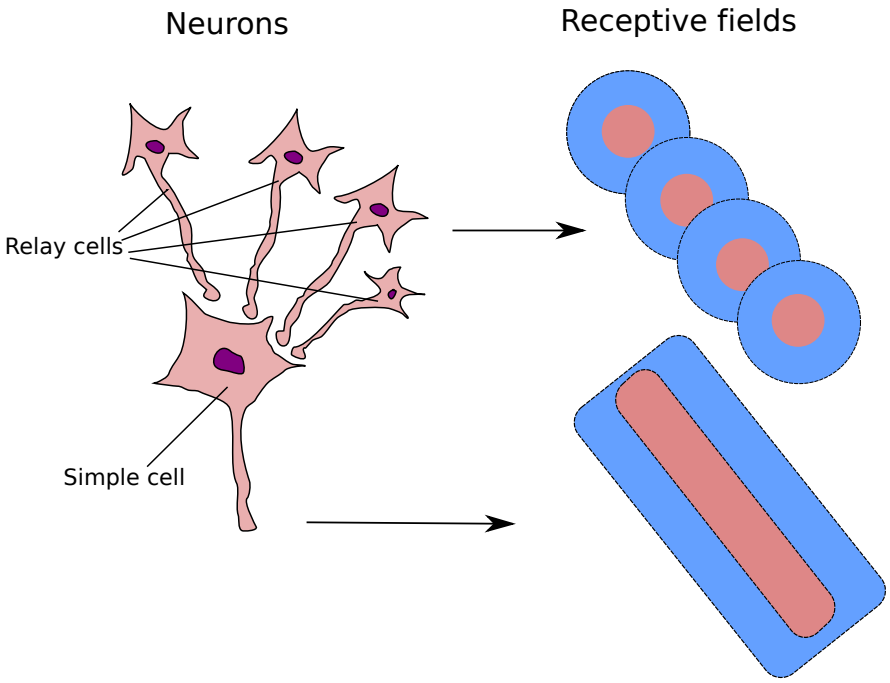

Figure 1.5: Several centre-surround relay cells (in V1 layer 4Cβ) form synapses onto a V1 simple cell. These relay cells have overlapping receptive fields which result in the elongated ON region flanked by OFF regions, producing the orientation and phase-selective receptive field typical of a simple cell. Adapted from Gilbert (2013a).

receptive field size increases monotonically with distance from the centre (Gilbert, 2013b). Receptive field size ranges from 0.1° in the centre of the visual field to 10° in the periphery (Gilbert, 2013b).

The receptive field is not an intrinsic property of an individual neuron. Instead, the receptive field is an emergent property of the combination of the connections to that neuron, and the processing performed by those earlier neurons which connect to it. Understanding these processes and computations which produce the apparent receptive field is the goal of theoretical modelling. Receptive fields are explored further in the subsequent chapter.

1.3.2 Simple and complex cells

V1 neurons are typically separated into two categories: simple and complex cells. This stems from the work of Hubel and Wiesel (1959), who first described these neurons in studies of cat primary visual cortex. Like relay cells in earlier stages, simple cells have distinct ON-regions and OFF-regions. Unlike simple cells, complex cells have overlapping ON and OFF regions, and are therefore phase-invariant to oriented contrast in their receptive fields. Complex cells typically have larger receptive fields than simple cells (Hubel and Wiesel, 1968). There is also third type of cell sometimes referred to as "hypercomplex" or "end-inhibition" cells. These cells have receptive fields of ON-regions terminated by OFF-regions, making them sensitive to line endings and corners. The distribution of simple and complex cells varies depending on cortical layer, with input layer 4 having more simple cells and layers 2 and 3

containing more complex cells (Hubel and Wiesel, 1968).

Since their qualitative identification by Hubel and Wiesel, more quantitative approaches have been used to distinguish simple and complex cells (Skottun et al., 1991).

1.3.3 Selectivity maps

Not all V1 neurons are sensitive to the same type of stimuli. Indeed, V1 shows several different functional maps of stimulus selectivity across its surface. These maps are a result of selective connectivity from earlier areas (Nauhaus and Nielsen, 2014), though their functional value is still unknown (Wilson and Bednar, 2015). There are selectivity maps of orientation, spatial frequency, disparity, and others, all superposed over the V1 area. These maps are periodic, repeating many times over the cortical surface. This periodicity is likely useful for pooling of responses in higher areas (Nauhaus and Nielsen, 2014). Not all maps are present in all species, for instance, rodents do not possess a structured orientation preference map (Carandini, 2012).

1.3.4 Functional roles of V1

Given the orientation selectivity of V1 and observed contour sharpening (Kapadia et al., 1995; Lamme, 2004), it has been suggested that V1 is the site of contour integration, sharpening contours which are used for higher-level processing in other areas. Since V1 is very large and highly connected to several other areas, it likely acts as a fairly raw representation of the visual field which is separately processed by other areas. It has been proposed that V1 acts as a saliency map (or precursor) for other brain areas such as the superior colliculus, which controls eye movements (Li, 2002).

Chapter 2

Receptive Field Model of V1

This chapter describes the development of a receptive field model of V1 based on several sources of physiological data. The model is initially restricted to use macaque monkey data exclusively. The goal is to produce a single model neuron whose parameters can be changed to produce the diverse set of responses observed in V1, and then to integrate that model into a network which can be used in visual cortex circuits.

2.1 Modelling approach

There are two primary distinctions to make in the description of this model. The first is the level of biophysical detail included. It is assumed here that fine biophysical details such as membrane conductance, neurotransmitter types, action potential shapes, and otherwise are relatively unimportant compared to the rate and timing of neuron action potentials in producing a model which is useful in visual processing circuits. This simplifies the development of the model because while V1 behaviours have been characterized via extracellular recording (recording firing rates of individual neurons), some of these details for other properties may be unexplored or ambiguous. To this end the model will be composed of point neurons whose outputs are scalar firing rates.

The second distinction is that the model developed in this chapter is a *receptive field* (or phenomenological) model. A receptive field model reproduces what a neuron does, but not how it does it. This is in contrast to *mechanistic* models, which attempt to explain how the receptive field properties of the neurons arise. The behaviour of a given neuron is influenced by all earlier neural processing (via feedforward connections) and also by later processing (via feedback connections). By probing at one point and presenting some visual stimulus, the influence of these early and late stage processes is also recorded. An ideal receptive field model therefore encapsulates all the processing the brain performs at a given point. In the case of V1, a receptive field model needs to encapsulate behaviour from the retina, LGN, and V1, as well as any processing which arises as a result of feedback from higher areas.

Because RF models forgo a large degree of biophysical realism, they can be made more computationally efficient by reducing the number of processing stages required to produce behaviour - receptive field models do not need to model several layers of synapses as would a fully mechanistic model. By applying the following approach to several visual brain areas,

one could conceivably construct circuits that "see" in using brain-like computations and potentially producing brain-like performance.

2.1.1 State of the art

Models of the early visual system including V1 have evolved greatly since Hubel and Wiesel's work in cat striate cortex. No model to date has yet encapsulated all of the observed behaviour of V1. Originally, V1 models were based on linear receptive fields alone, and then passed through some spiking non-linearity.

Movshon, Thompson, and Tolhurst (1978) showed that much of the spatiotemporal behaviour of V1 simple cells could be explained through linear summation of the stimulus in the cell's receptive field, including orientation, spatial, and temporal frequency.

Jones and Palmer (1987) evaluated the use of Gabor functions as linear kernels and found them to be an effective model of cat striate cortex receptive fields.

Heeger (1992) proposed a model in which simple cells compute a weighted linear summation of their input, thresholds the sum, and then is normalized by surrounding units with overlapping receptive fields. This model shows contrast normalization and also produces differences in response to moving and stationary stimuli. This "divisive normalization" stage has since been used in several models of neural processing in several areas (see Carandini and Heeger (2012) for a review). Indeed, the term *normalization* has been applied to several nonlinearities observed in V1, including for contrast saturation, surround suppression, and cross-orientation inhibition. This model was later extended Somers et al. (1998) developed a simple mechanistic model which produced a wide variety of non-classical effects including surround suppression, contour facilitation, and contrast normalization. This model used both excitatory and inhibitory interactions to examine how neurons integrate inputs outside their classical receptive fields. A compelling case for processing by small groups of neurons is made, but assumes potentially unrealistic asymmetry between excitatory and inhibitory properties. Also, the model is not pixel-computable and therefore needs to be adapted for use with real images.

Li (1998) proposed a model of contour integration using exclusively data from V1. This was later extended into a saliency map model, proposing that V1's primary function is to indicate the most salient or potentially important parts of the visual scene (Li, 2002). Like Somers' mechanistic model, these are models which focus on neural interactions resulting in the observed behaviour, and do not deal with real images.

Spratling (2010) developed a model of V1 based on the notion of "predictive coding (PC)", that is minimizing the error between an internal (neural) representation and the observed sensory inputs. The PC model uses Gabor-like receptive fields and is restricted to stationary greyscale images. It is unclear how suitable the PC model is for large-scale applications.

2.2 Simple cell model architecture

A model of the simple cell receptive field was developed based on the above literature. Figure 2.1 shows the basic processing performed by this model.

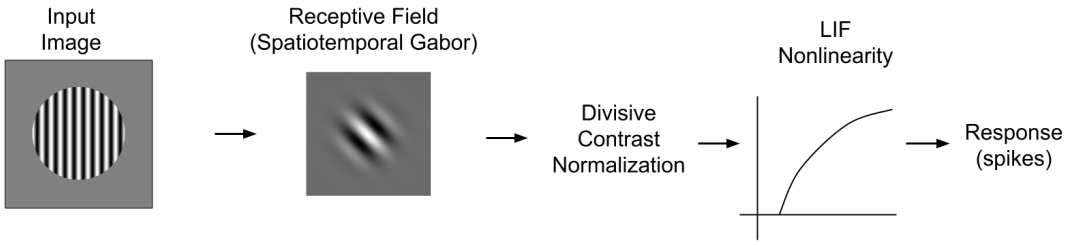

Figure 2.1: Simplified schematic of the receptive field model.

The basic receptive field for all three models is a 2D Gabor filter as used in many other models of V1. A Gabor function is the product of a Gaussian *envelope* and a sinusoidal *carrier*. The equation for a Gabor function is as follows:

$$\begin{aligned} g(x,y) &= \tfrac{1}{2\pi\sigma_x\sigma_y}\exp\left(\tfrac{-x'^2}{2\sigma_x^2} - \tfrac{y'^2}{2\sigma_y^2}\right)\cos\left(2\pi k x' - \phi\right) \\ x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned} \tag{2.1}$$

where σ_x and σ_y are the standard deviations of the Gaussian envelope, θ is the carrier orientation, k is the carrier spatial frequency, and ϕ is the carrier phase shift. The 2D Gabor filter is then element-wise multiplied by the input image and the result is summed:

$$J_{stim} = \sum_{i=-N/2}^{N/2}\sum_{j=-N/2}^{N/2} g(i,j)h(i,j) \tag{2.2}$$

where J_{stim} is the stimulus-driven current, N is the number of pixels (to a side), $g(x,y)$ is the Gabor function, and $h(x,y)$ is the input stimulus. The input drive current J_{stim} is then passed into a leaky-integrate-and-fire (LIF) neuron model as presented in equation 2.3 below:

$$\begin{aligned} a &= G[\alpha J_{stim} + J_{bias}] \\ G[J] &= \tfrac{1}{\tau_{ref} - \tau_{RC}\ln\left(1-1/J\right)}, \\ J &< 1 = 0 \end{aligned} \tag{2.3}$$

where a is the activation or firing rate of the neuron in spikes per second, α is a current gain, J_{bias} is a current bias, τ_{RC} is an RC-circuit time constant, and τ_{ref} is a refractory period time constant. While this LIF non-linearity is not a highly-realistic simulation of neuron spiking, it has been argued that it does offer the best trade-off between biophysical realism and computational efficiency (Eliasmith and Anderson, 2004). The LIF stage adds a thresholding of the input current which has been shown to be crucial for producing reasonable responses to non-preferred stimuli (Carandini et al., 2005). A non-linear least-squares curve fitting (via Matlab's `lsqcurvefit` function) was used to fit the model parameters to various sets of data showing different receptive field properties. In order to do so, visual stimuli were produced as specified in the experiment that produced the data (typically

sinusoidal gratings). The curve fitting was performed for each set of data separately in order to confirm that the model format could produce selectivity for each stimulus property.

For all the following results, a minimum of 8 parameters were searched over: the 5 Gabor filter parameters (σ_x, σ_y, θ, k, and ϕ), and 3 LIF parameters: gain α, current bias J_{bias}, and an RC factor which allowed for scaling of the τ_{RC} by a factor of 1-10. Eliasmith and Anderson (2004) showed that increasing τ_{RC} in a LIF neuron makes the neuron respond more linearly.

In order to accommodate time-dynamic stimuli (for modelling of temporally-sensitive selectivies such as temporal frequency and direction of motion), the 2D Gabor filter was extended to 3 dimensions with some duration T and a time discretization dt (both measured in seconds):

$$g(x,y,t) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(\frac{-x'(t)^2}{2\sigma_x^2} - \frac{y'(t)^2}{2\sigma_y^2}\right)\cos\left(2\pi k x'(t) - \phi\right) \tag{2.4}$$

where

$$x'(t) = x\cos\theta - y\sin\theta + s(t)\cos\theta y'(t) = x\sin\theta + y\cos\theta + s(t)\sin\theta \tag{2.5}$$

and

$$s(t) = 2\pi f t dt \tag{2.6}$$

where f corresponds to the preferred temporal frequency. This time-dimension extension is similar to the energy model of Mante and Carandini (2005), but the filter operates in the time domain instead of the frequency domain.

The following results show that this simple model is able to reproduce a variety of behaviour observed in both simple and complex V1 cells, including orientation selectivity, spatial, and temporal frequency selectivity. In the following experiments an arbitrary filter size of 25×25 pixels was selected. This receptive field size can later be adjusted for population models of V1, but requires finding new model parameters via optimization.

2.2.1 Orientation selectivity

Orientation selectivity in V1 neurons was first identified by Hubel and Wiesel (1959) in cats, who later found similar tuning in macaque monkeys (Hubel and Wiesel, 1968). DeValois, Yund, and Hepler (1982) present orientation tuning curves for several hundred macaque V1 cells. Orientation tuning information was attained by presenting both moving and stationary oriented gratings across the visual field while recording spike rates from individual V1 neurons. A large variation in orientation tuning was found, with cells varying in tuning bandwidth (the range of orientations which produce a non-zero response), maximum response rate, and direction selectivity (for drifting gratings). V1 cells show large variation of orientation bandwidth, with some cells responding to as narrow as 10 degrees while others show no orientation preference at all. Figure 2.2 shows the RF model's response to stationary gratings of various orientations. Model parameters were fitted to results considered "close to the average for [the] whole cell population" (DeValois, Yund, and Hepler, 1982), with a bandwidth of about 90 degrees. The orientation selectivity of the model is a product of

the orientation of the sinusoidal carrier portion of the Gabor filter, therefore changing the parameter θ alone produces a neuron with identical tuning properties to the one presented in figure 2.2, but with different orientation preference. Figure 2.3 shows some of the gratings used to probe the model response, as well as the Gabor filter found by the opimization procedure.

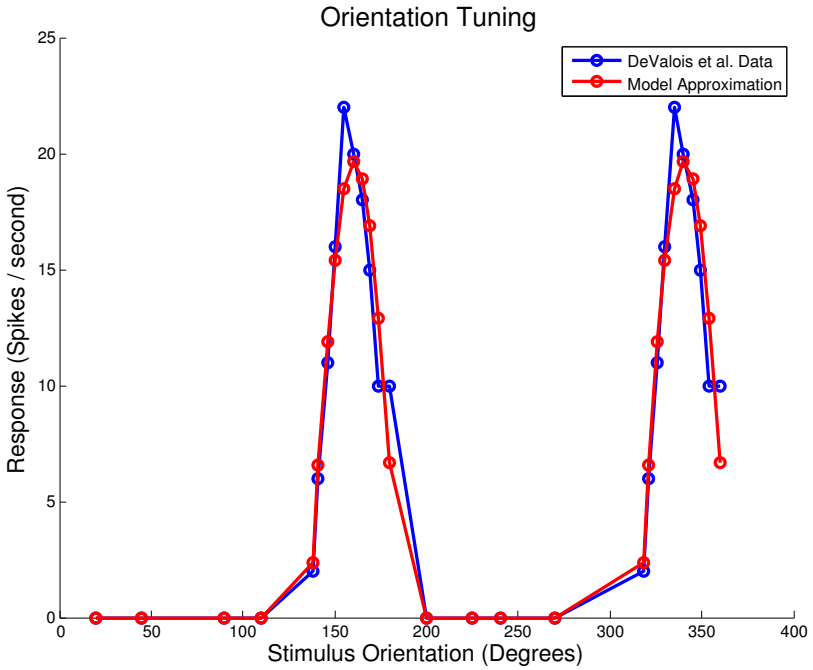

Figure 2.2: Receptive field model results based on orientation tuning data from DeValois, Yund, and Hepler (1982).

2.2.2 Spatial frequency selectivity

V1 neuron firing varies with the spatial frequency of stimuli. This is demonstrable with sinusoidal gratings of variable frequency. Foster et al. (1985) presented various spatial frequency tunings evident in macaque V1. Gratings of fixed orientation and varying spatial frequency were presented and the neuron responses were recorded. They found a range of spatial frequency tuning, however 91% of neurons recorded showed bandpass behaviour. Figure 2.4 shows that our receptive field model effectively produces this bandpass selectivity for spatial frequency. This is a result of the carrier frequency of the Gabor filter stage, thus modifying this parameter (k in eq. 2.1) changes the spatial frequency selectivity of the model neuron.

2.2.3 Temporal frequency selectivity

Foster et al. (1985) also examined tuning of macaque V1 cells to stimulus temporal frequency, and found that 68% of neurons studied showed low-pass temporal frequency response, wherein no optimal frequency could be identified. Neurons were probed by first

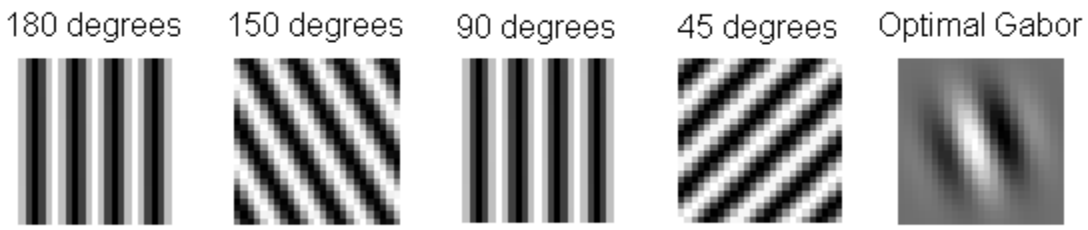

Figure 2.3: Various oriented gratings used to probe the model for orientation selectivity. The rightmost figure shows the Gabor filter which produced the best fit (the preferred orientation was around 155 degrees in the data).

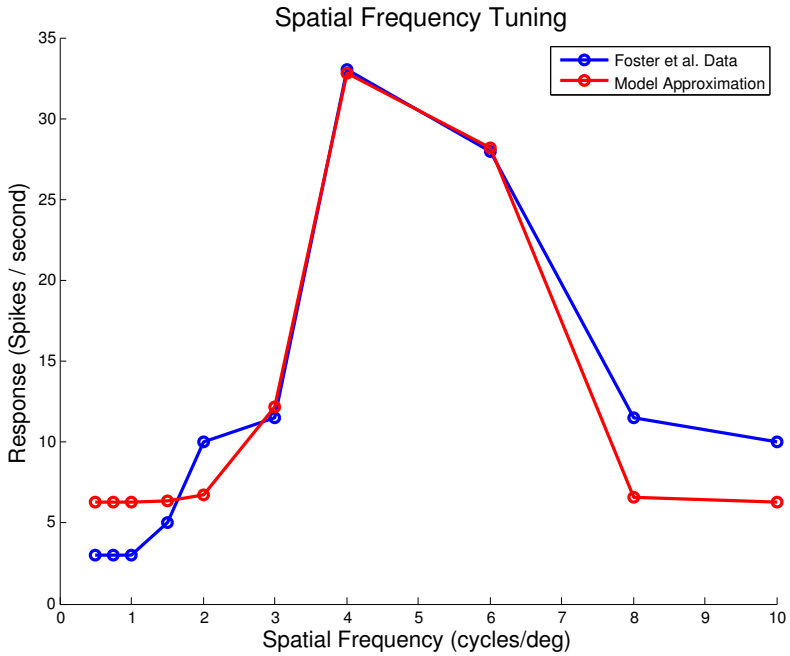

Figure 2.4: RF model spatial frequency response. The model provides a good fit, but fails to capture subtle differences in the observed high and low-frequency responses. Data from Foster et al. (1985).

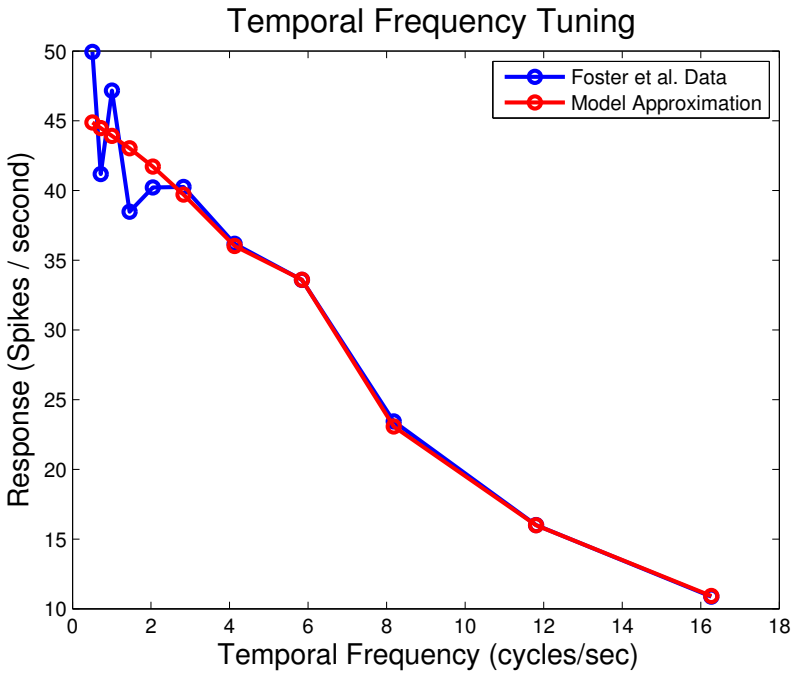

Figure 2.5: RF model low-pass temporal frequency response. The model provides a good fit for drifting gratings with a low-pass response, which is evident in a majority of V1 neurons. Data from Foster et al. (1985).

finding the stimulus orientation and spatial frequency which produced maximal response. Then drifting gratings of various speeds were presented to the neuron and the spike responses were recorded. The remaining 32% of neurons were found to be bandpass, with responses falling below 50% of peak response amplitude at frequencies above and below the preferred frequency. Mean bandwidth for both simple and complex cells was 2.9 octaves, with mean optimum temporal frequency of 3.7 Hz.

Figure 2.5 shows the model response with parameters fit to the response of a low-pass neuron. The neuron shows no preference, with response attenuation increasing with increasing frequency. According to Foster et al. (1985), this is the most common response of neurons in macaque V1 to temporal frequency. A total stimulus length of 1 second was used, with a discretization of $dt = 1/2f_{max}$, where f_{max} is the maximum grating frequency to be tested. The model was optimized over the 5 Gabor filter parameters, 3 LIF parameters, and the temporal frequency selectivity f (see equation 2.4).

Figure 2.6 shows the model response to temporal frequency as tuned to data presented in Foster et al. (1985) for a band-pass selective neuron. Temporal frequency selectivity arises from the phase shifting of the Gabor filter over time and parameterizing the drift rate. Thus the maximum response from the Gabor stage of the model will arise when the drift rate of the Gabor aligns with the drifting grating.

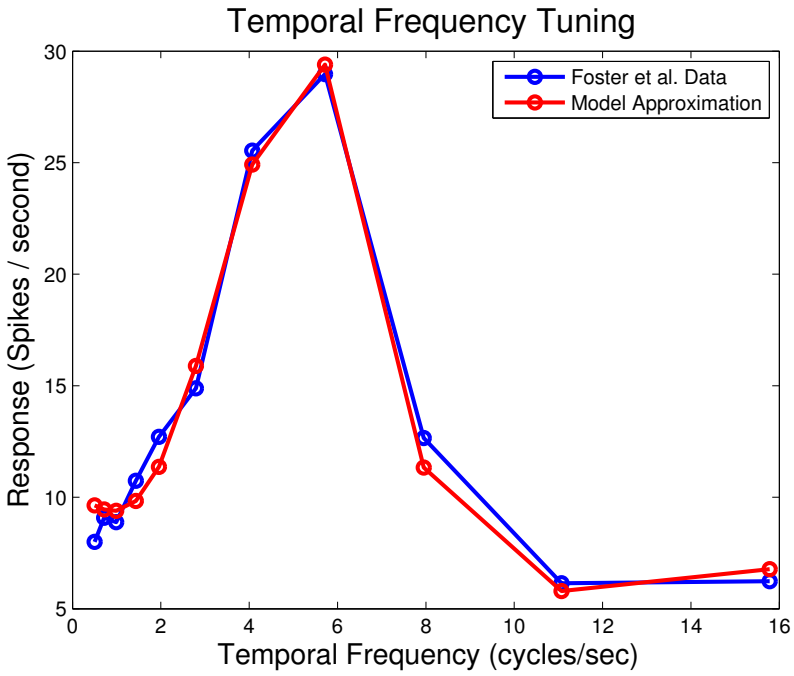

Figure 2.6: RF model band-pass temporal frequency response. The model provides a good fit for drifting gratings with a bandpass response. Data from Foster et al. (1985).

2.2.4 Other receptive fields

V1 cells have been found to be sensitive to several other stimulus properties including colour and direction of movement.

Colour

V1 neurons are selective for colour. Horwitz and Hass (2012) demonstrate that the colour-selectivity of V1 neurons is represented in one of three cardinal directions: black-white, red-green, and blue-yellow. Neurons were recorded for a variety of colour stimuli and it was noted that several stimuli produce the same response, yielding an "isoresponse surface" in three-dimensional colour-space. A number of different surfaces are identified, the major and minor axes of which often align with these three colour cardinal directions.

Colour-sensitive neurons make up a small portion of V1 neurons and they tend to lack sensitivity to other stimulus properties such as spatial and temporal frequency, so they can be treated as a separate class of cell (Livingstone and Hubel, 1984). Because of this, colour sensitivity was not included in the receptive field model.

Direction of movement

Some (20-50%) V1 neurons are sensitive to the direction in which a stimulus moves (De-Valois, Yund, and Hepler, 1982; Foster et al., 1985). Most cells are bi-directional, but there is a significant population of unidirectional cells. The temporal frequency selectivity implemented here is direction-dependent. While the model was not designed for direction

invariance specifically, it could likely be produced from temporal frequency selectivity by a weighted sum of 2 Gabor filters shifted in opposite directions in the time domain, or by implementing a frequency-domain filter as in Mante and Carandini (2005).

2.3 Non-classical receptive fields

The receptive field properties discussed thus far have been confined to the "classical" receptive field of the neuron, that is, the region of the visual field in which stimuli can elicit a response from that neuron. However it has long been observed that the response of a neuron to that stimuli can be modulated by stimuli outside of that receptive field. This surrounding region is sometimes called the *non-classical* receptive field: stimuli in this region do not produce a response when presented alone, but do change neuron response to stimuli within the classical receptive field (Lamme, 2004). This modulation can be suppressive (lessening the neuron's response) or facilitative (strengthening it). Effects of the non-classical receptive field are sometimes called contextual modulation, or "non-linear behaviour", because they cannot be reproduced by linear models (Carandini et al., 2005). These effects include attentional modulation, surround suppression, contrast saturation, contrast adaptation, contour integration, and correlated variability.

2.3.1 Contrast saturation

Albrecht and Hamilton (1982) showed that in monkeys and cats, most cells in V1 respond non-linearly to stimulus contrast. Neurons were probed with oriented gratings of various contrasts. Generally, the response of V1 neurons grows linearly within the first 10-20% of increasing contrast, and then the response saturates to some asypmtotic maximum firing rate. Of 247 cells recorded, 70% showed a contrast response function best fit by a "hyperbolic ratio" function, which is as follows:

$$R_{sat} = R_{max}\frac{c^n}{c_{50}^n + c^n} \tag{2.7}$$

where R_{max} is the asymptotic maximum firing rate, c is the stimulus contrast, c_{50} is the semi-saturation constant (contrast producing 50% of the maximum response), and n is some exponent typically ranging between 1 and 5 (Albrecht and Hamilton, 1982). To incorporate this into the RF model, a saturating stage based on the hyperbolic ratio function was added before the LIF nonlinearity stage. The normalization takes the following form:

$$J_{sat} = \alpha\frac{J_{gabor}^n}{c_{50}^n + J_{gabor}^n}, J_{gabor} < 0 = 0 \tag{2.8}$$

where J_{sat} is the saturated current, J_{gabor} is the output current from the Gabor filter (as calculated in 2.2). Contrast was defined as the difference between the maximum and minimum pixel intensities in the stimulus, with an intensity range of 0-1 (where 1 corresponds to white and 0 to black). J_{sat} is then input to the LIF non-linearity as in eq. 2.3.

The addition of the contrast saturation stage enables the model to reproduce saturating contrast response. Figure 2.7 shows the model response when fit to data presented in Albrecht

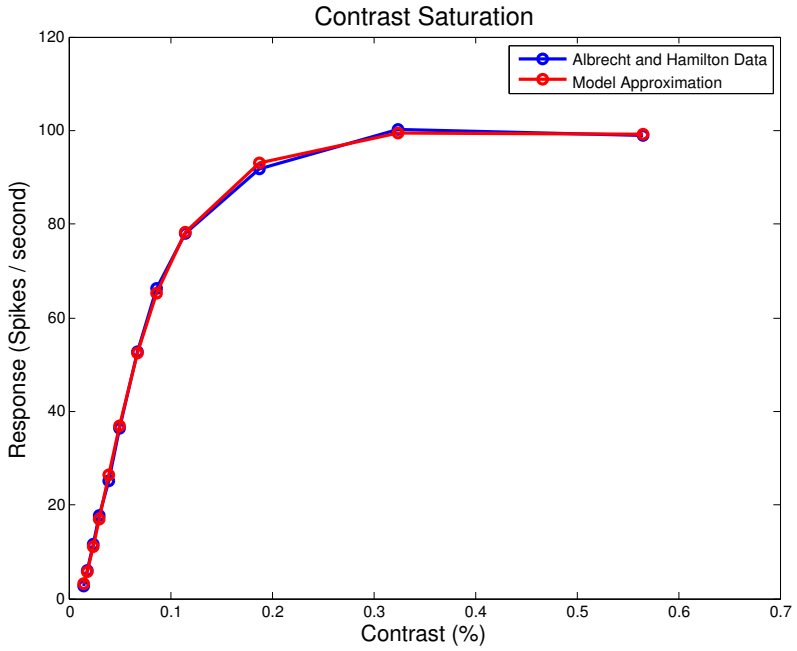

Figure 2.7: RF model contrast saturation response. With the addition of a contrast saturation stage, the RF model response saturates with increasing grating contrast.

and Hamilton (1982). Figure 2.8 shows the model response without the saturation stage. This stage results in a divisive normalization of the LIF input current. When the input current saturates in this way, so must the output of the LIF nonlinearity. Figure 2.9 shows some of the stimuli used to probe the model's contrast response. In this case the parameter optimization space consisted of the 5 Gabor parameters and 3 LIF parameters, as well as the c_{50} and n parameters used in equation 2.8.

2.3.2 Contour facilitation

Another prominent example of contextual modulation in V1 is contour facilitation, or the tendency of V1 neurons encoding similarly oriented lines to facilitate each other and inhibit others with orthogonal orientations. Kapadia et al. (1995) examined several different examples of such contour facilitation using separated line segments in various arrangements, including the effects of coaxial offset, lateral offset, and orientation offset.

In an investigation of contour integration in humans, Field, Hayes, and Hess (1993) proposed an "association field", a grouping of Gabor patches corresponding to the arrangement of line segments that produce contour saliency in human psychophysical experiments. Such perceptual "pop-out" may have its basis in V1 where contour facilitation has been observed.

Based on this speculation, the RF model Gabor-filter stage was expanded by adding a surround region RF which modulates the classical RF. This surround region was modelled by adding two additional Gabor filters which had variable separation from the classical RF and also variable weight. The drive current to the neuron is then:

$$J_{total} = J_{classical} + w_1 * J_{gabor1} + w_2 * J_{gabor2} \tag{2.9}$$

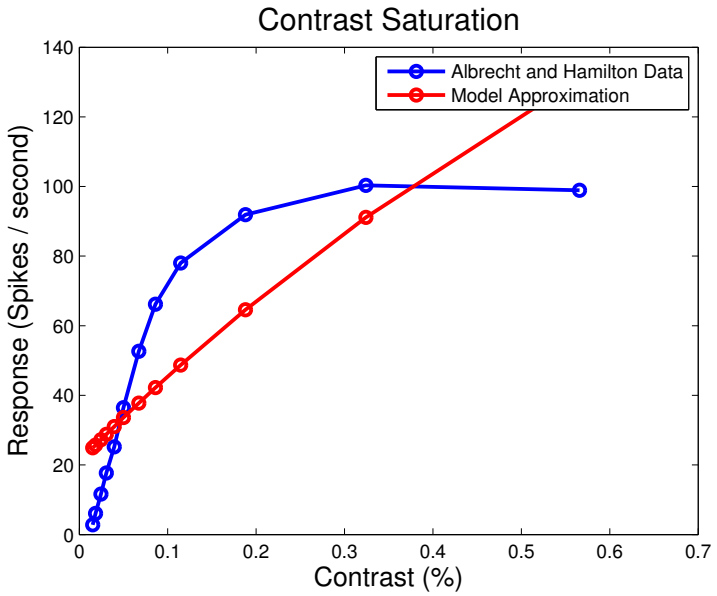

Figure 2.8: RF model contrast response without a saturation stage. Because the output current of the Gabor filter linearly scales with contrast the model does not produce contrast saturation.

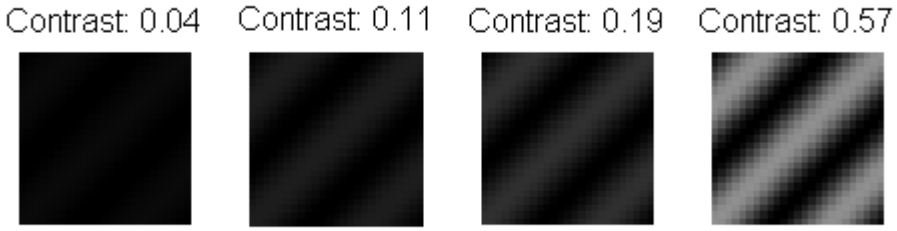

Figure 2.9: Grating stimuli of various contrasts used to probe the model's contrast response. Input to the model is normalized from 0-1 in greyscale, with grating contrast measured as the difference between the maximum and minimum pixel intensities in the stimulus.

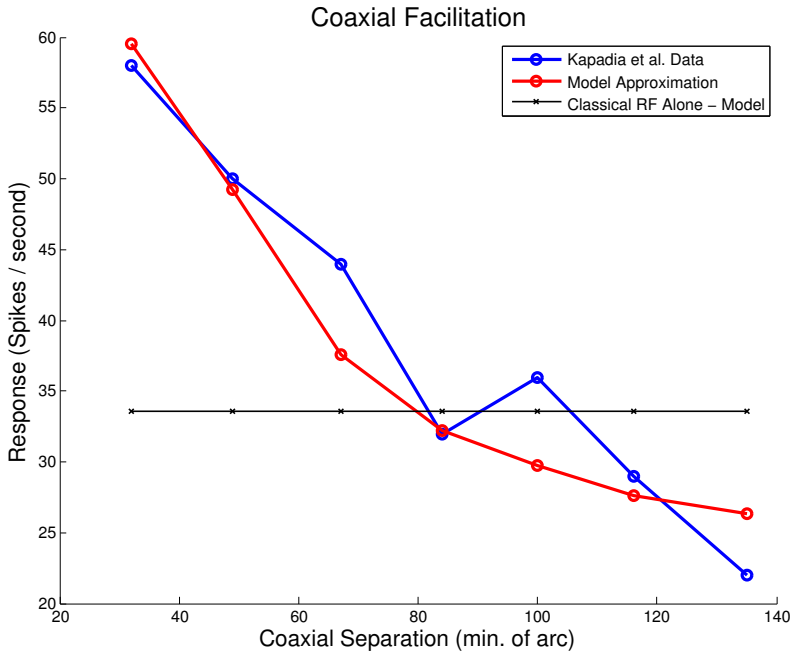

Figure 2.10: RF model coaxial facilitation and observed data. The observed data actually shows a slight increase in the response at a separation of 100 arc-min., which the model is unable to reproduce. The black line indicates the models response to the segment in the CRF alone.

where J_{gabor1} and J_{gabor2} are determined by the same element-wise multiplication, but elsewhere in the stimulus image, and w_1 and w_2 are the weights of these fields. These weights have a range of ± 1. Separation of the Gabor filters was also parameterized. Thus two additional parameters per additional Gabor filter are added to the optimization parameter space: the weight of that filter and its position.

As a proof of concept, the model was expanded only to produce coaxial facilitation, wherein two similarly oriented line segments are aligned and the separation between them increased (see figure 2.11). Model results using this expanded receptive field are shown in figure 2.10. One interesting observation in the data is that there is a second peak in the co-axial tuning at around 100 arc-min. of separation. This was observed in several neurons (Kapadia et al., 1995).

2.3.3 Surround suppression

Surround suppression is an important property of the non-classical receptive field, referring to the observation that the region surrounding the CRF, when adequately stimulated, will suppress the neuron's response (Allman, Miezin, and McGuinness, 1985; DeAngelis, Freeman, and Ohzawa, 1994; Webb et al., 2005). This is important because neurons are rarely stimulated alone in natural settings, and in fact respond differently when regions outside their receptive fields are also stimulated. Therefore surround suppression and contextual influences in general will be essential for producing a model that is useful with natural images.

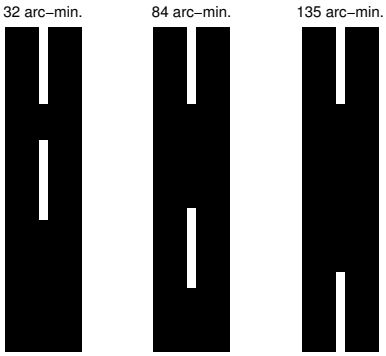

Figure 2.11: Examples of model stimuli for probing co-axial facilitation.

Surround suppression is in direct opposition to contour integration which is also observed in V1. In fact, weak-contrast CRF stimuli can be facilitated by surround stimuli while the same surround is suppressive for high contrast CRF stimuli (Sengpiel, Sen, and Blakemore, 1997). Surround suppression is also spatiotemporally tuned, sensitive to the orientation, temporal, and spatial frequencies of the surround stimuli. This large set of observations proves to be difficult to capture with a receptive field model.

2.4 Limitations of the Receptive Field Model

While the basic RF model presented here performs well in reproducing the breadth of behaviour of the V1 simle cell classical receptive field, it became clear that the same approach was not easily extensible to produce non-classical RF behaviour. The number of parameters associated with each neuron quickly grows, making optimization over the parameters difficult. In addition, V1 shows a number of seemingly paradoxical effects, such as contrast-dependence of facilitation and surround suppression. Two non-linearities are already present in the model (contrast saturation and LIF firing), and to produce additional non-classical behaviour would likely require further non-linear processing. The addition of these non-linear processes makes it difficult to predict how the model will behave to untested inputs. Because of this, it can be concluded that in order to produce the non-classical receptive field behaviour a model including interactions between neurons may perform better in a large-scale context. By further studying the mechanisms which produce V1 behaviour, a model which captures a wider set of behaviour in a more concise and efficient manner may be developed.

Chapter 3

Integrated Model of V1

The receptive field model of the previous chapter produces a range of central receptive field behaviour, but is unable to account for contextual modulation of behaviour such as surround suppression. Examining the neurological mechanisms that produce V1 behaviour may allow for simplification of the model and provide insight into how to the model can best be extended to broaden its set of reproducible behaviour. Mechanistic models are tools for researchers to attempt to explain the emergence of various behaviours. These models provide a framework for thinking about processing in neurons, reconciling observed data with purported functions of neurons and neural circuits. However, it is not feasible to model all the mechanisms present in V1 behaviour - there are simply too many subtleties and unknown properties, and the resulting model would not be useful for real-time applications. One important and well-studied mechanism is lateral interactions - interactions between neurons within a cortical layer. The previous chapter presented a purely feedforward model without any such interactions. In this chapter, we examine a recent and promising model - the SSN - that explains a wide variety of V1 contextual behaviour using a particular lateral interaction scheme, complementary to the receptive field model developed in the previous chapter. The SSN is adapted here to use the receptive field model as input, resulting in a pixel-computable *integrated* model. Also in this chapter the SSN is extended to take binocular input and examined for interocular transfer of suppression, a phenomenon observed in V1.

3.1 The Stabilized Supralinear Network

The stabilized supralinear network, or SSN (Rubin, Van Hooser, and Miller, 2015), provides a simple mechanism which accounts for a wide variety of non-linear behaviour without requiring fine tuning of model parameters, and includes intracortical connections. It is based on four major circuit properties:

- Supralinear (power law) neuron output. All units in the SSN output a power law function of their input. This is supported by the findings of Priebe et al. (2004), who found that V1 neurons have a power-law relationship between average membrane potential and firing rate.

- Excitatory interactions between neurons. Neurons in the SSN are sparsely connected

to each other and excite each other locally, resulting in a given neuron's excitatory drive increasing when neighbouring neurons are excited.

- Inhibitory feedback. Combining the supralinear I/O of the neurons with intracortical excitation results in a rapid rise in spike rate until the neurons all saturate. The SSN proposes that this instability is counteracted by simultaneous excitation of inhibitory cells, which then inhibit the activity of both the excitatory and inhibitory units to a point of network stability.

- Specific spatial properties of connections between neurons. In the SSN, long-range connections are primarily excitatory, and prefer to connect to inhibitory units. Inhibitory units connect to both excitatory and inhibitory units over shorter (more local) distances. This connectivity is critical to some of the SSN's behaviour.

The SSN is developed to account for two cortical computations which are apparent in several areas of cortex. The first of these is surround suppression, wherein stimuli outside a neuron's CRF suppress its response to stimuli within the CRF. This has been observed in cats and monkeys. Surround stimuli can facilitate low-contrast centre stimuli.

The second phenomena the SSN attempts to account for is sublinear response summation or what the authors call "normalization" (which they note has been applied to many other phenomena - see Carandini and Heeger (2012) for a review). Responses to two stimuli shown simultaneously (such as two superposed gratings of different orientations) tend to be closer to the average of the two stimuli, not the sum.

These 2 phenomena have been observed in many cortical areas including V1, V2, V4, MT, MST, and FEF. Thus they can potentially be considered canonical computations of cortex. The model presented by Rubin, Van Hooser, and Miller (2015) focuses on reproducing V1 behaviour. The SSN also makes a number of predictions which were then verified in ferret V1, providing support for the model as a cortical computation.

3.1.1 Model structure

The SSN is based on the idea that supralinear excitation is stabilized by feedback inhibition (Ahmadian, Rubin, and Miller, 2013). Before excitation of the network becomes unstable, connections between units (called *recurrent* connections because they are within the same layer) increase inhibition to both the E and I units, driving the network to a point of stability. This arises from the relationships between the connection weights W_{ab}.

Rubin, Van Hooser, and Miller (2015) focus on modelling V1 behaviour, presenting two simplified models which are then combined into a more complete V1 model. A number of assumptions are made. The SSN uses point neurons whose outputs are scalar firing rates. They use two types of neuron units, excitatory (E) and inhibitory (I), in equal proportion, spaced evenly over a grid. The interactions are of a single layer (layer 2/3), modelling horizontal connections without interlaminar interactons. The units are treated as groups of complex cells which are only selective for stimulus orientation. Stimulus input to the network is a mathematical function of orientation and retinotopic position - the SSN is not pixel-computable.

The model equations are as follows: Let x be a 2D vector representing the retinotopic position in cortex of an E/I pair of units. c is the stimulus strength, corresponding to contrast with arbitrary scale. W_{ab} is the connection weight from a unit of type b to type a, where a and b can be either excitatory (E) or inhibitory (I). $h(x)$ is the input stimulus image (scale of 0 to 1), defined over all cortical positions in the network's grid. Input to the model is as follows:

$$I_E(x) = ch(x) + \sum_{x'}(W_{EE}(x,x')r_E(x') + W_{EI}r_I(x')) \tag{3.1}$$

$$I_I(x) = ch(x) + \sum_{x'}(W_{IE}(x,x')r_E(x') + W_{II}r_I(x')) \tag{3.2}$$

where the sum over x' ranges over all other unit positions. $I_E(x)$ is the excitatory input and $I_I(x)$ is the inhibitory input to the units at position x. $r_E(x)$ and $r_I(x)$ are firing rates of E and I units at position x.

The steady-state firing of a neuron for fixed input I is a power law:

$$r_E^{SS}(x) = k([I_E(x)]_+)^n \tag{3.3}$$

$$r_I^{SS}(x) = k([I_I(x)]_+)^n \tag{3.4}$$

where k is some constant gain, n is an exponent greater than 1, and inputs $I(x)$ are thresholded such that $I(x) < 0 = 0$.

At time t, each unit approaches the steady-state value with first-order dynamics:

$$\tau_E \frac{dr_E(x)}{dt} = -r_E(x) + r_E^{SS}(x) \tag{3.5}$$

$$\tau_I \frac{dr_I(x)}{dt} = -r_I(x) + r_I^{SS}(x) \tag{3.6}$$

Equations 3.5 and 3.6 are then solved using Euler's method with a specified time discretization dt and total steps T.

Connection weights W_{ab} describe the connections between E and I units at each position x. These weights are sparse and computed probabilistically as the product of 2 Gaussians - one for position and one for orientation preference:

$$P(W_{ab}(x,x') \neq 0) = \kappa_b G_{\sigma_{ab}}(x,x') G_{\sigma_{ori}}(\theta(x),\theta(x')) \tag{3.7}$$

where $G_\sigma(x,y) = e^{\frac{-|x-y|^2}{2\sigma^2}}$ (a one-dimensional Gaussian describing likelihood of connection based on orientation preference) and κ is a scale factor (greater for I connections than E connections). Thus a neuron is more likely to connect to nearby neurons, and also more likely to connect to neurons with similar orientation preference. A 75×75 interval grid was used. The orientation map used was based on a method presented by Kaschube et al. (2010). This is depicted in figure 3.1. For non-zero weights, the weights were sampled from normal distributions:

Parameter	Value
Δx	$16^\circ/75$
J_{EE}	0.1
J_{IE}	0.38
J_{EI}	0.089
J_{II}	0.096
κ_E	0.1
κ_I	0.5
n_E*	2.2
n_I*	2.0
$k*$	0.012
τ_E*	20 ms
τ_I*	10 ms
σ_{EE}	$8\Delta x$
σ_{IE}	$12\Delta x$
σ_{EI}	$4\Delta x$
σ_{II}	$4\Delta x$
σ_{ori}	45°

Table 3.1: Parameters used for the 2D SSN model results presented in Rubin, Van Hooser, and Miller (2015) (supplemental materials section). Parameters marked with $*$ are stochastic, sampled from $N(x, 0.05x^2)$ where x is the value indicated in the table.

$$W_{ab}(x, x') = N(J_{ab}, (0.25J_{ab})^2) \tag{3.8}$$

where J_{ab} is the mean strength of connection weight from units of type b to units of type a. Negative weights are set to zero. Using these equations as well as the supplemental material of Rubin, Van Hooser, and Miller (2015), the major results of the SSN were reproduced. Parameter values (found in the supplementary material) are summarized in table 3.1.

3.1.2 Key functional properties of the SSN

The SSN produces a wide variety of behaviour which has been observed in V1, as well as some new behaviours. Here a selection of these behaviours are described, as well as how the SSN accounts for them.

Contrast dependence of network activity

With increasing stimulus strength of a grating the SSN transitions from being primarily eternally driven to primarily network driven (figure 3.2). The network is inhibition-dominated, thus maintaining stability. At sufficient contrast ($c \approx 25$ in this instance) the external input is largely cancelled resulting in slow growth of network activity.

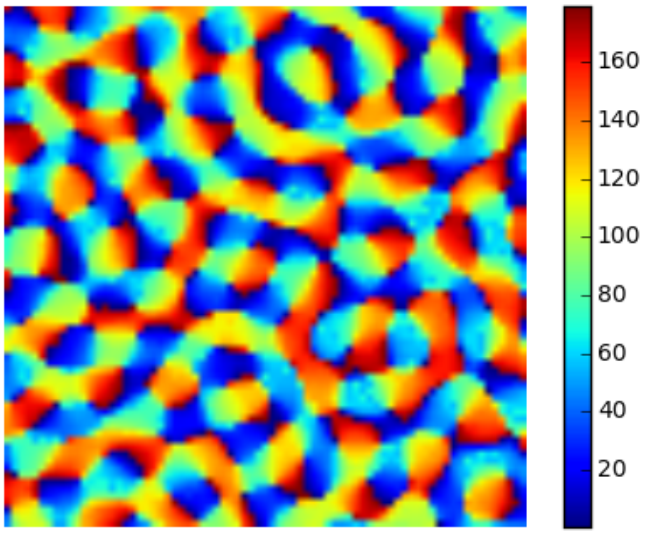

Figure 3.1: 75×75 orientation map used in the original SSN model. This is randomly generated using the method presented by Kaschube et al. (2010).

Surround suppression

Surround suppression has long been considered a product of lateral inhibition in the cortex - neurons within the same layer inhibit each other over long distances (1 or more millimetres) of cortex (Priebe and Ferster, 2008; Carandini and Heeger, 2012; Somers et al., 1998). When a stimulus extends outside the classical receptive field of a neuron, it starts to stimulate surround neurons that have inhibitory connections to the central neuron, reducing its response. This proposition predicts that neurons being surround-suppressed receive increased inhibitory drive, however in surround suppression experiments both excitatory and inhibitory drive is found to decrease (Ozeki et al., 2009). This is called a "paradoxical effect" and was previously observed by Tsodyks et al. (1997).

The SSN is an inhibition-stabilized network in which there are stable and unstable states of network activity. The network moves towards states of stability. To understand how an inhibition-stabilized network functions, consider a stable state of the network (where all E and I units are at stead-state firing rates). If the I input is held constant, then any increase in E will cause the network to recurrently excite itself to instability, while any decrease in E will cause the E to fall to zero (I input would dominate). The details of how these balancing dynamics arise are developed by Ozeki et al. (2009) and further developed by Ahmadian, Rubin, and Miller (2013).

In the SSN, excitatory projections span the longest range, and they preferentially connect to I units (see σ_{IE} and σ_{EE} in table 3.1). Because of this, when a surround stimulus excites surrounding neurons, the excitatory projections increase the inhibitory drive to the central neurons, causing their activity to decrease. Since I units also receive local excitation, the

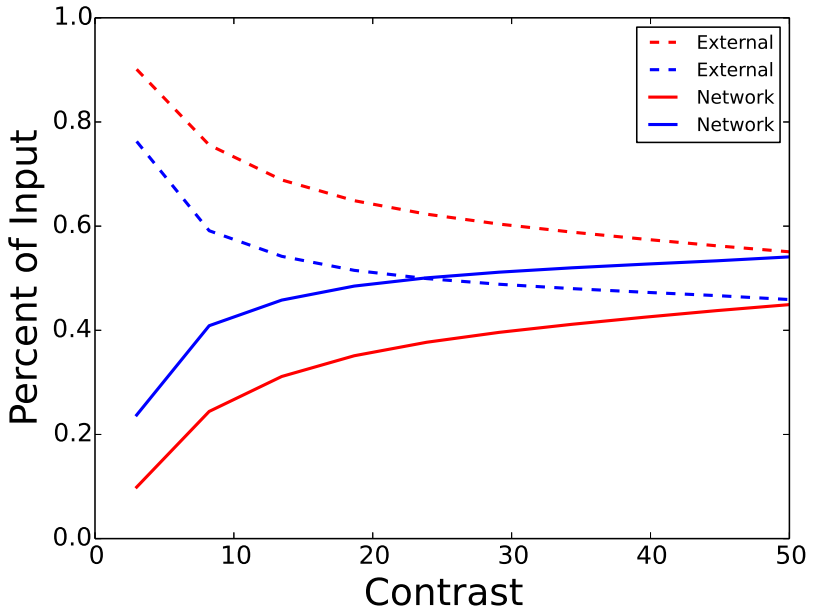

Figure 3.2: SSN transition with stimulus strength from primarily externally-driven to primarily network driven. As contrast increases, the average recurrent drive increases and begins to dominate external (feedforward) drive from input stimuli. Red lines are proportion of total E input to a unit and blue lines are proportions of total I to a unit. This figure was produced as the average of activity for 40 randomly selected E/I pairs. "Contrast" values correspond to arbitrary c values in equations 3.1 and 3.2. Note that external input is identical for E and I populations, so larger proportions of I network activity indicate inhibition dominance.

activity of the I units also decreases until the network reaches a new steady-state.

Periodic size tuning

As a stimulus (typically an oriented, drifting grating) increases in size, it has been observed that most V1 neurons show a similar response. The response rate grows to some peak, after which the response steadily decreases with increasing stimulus size. This is called size tuning, and the stimulus size which produces the peak response is called the *summation field size*. Previously, size tuning has been coarsely sampled (Anderson et al., 2001), however the SSN predicts that V1 neurons show periodicity in their size tuning curves at higher contrasts. Intuitively, as the stimulus' size increases, the centre neurons should be increasingly surround suppressed. However the SSN predicts that as a stimulus reaches a certain size, the neurons which were suppressing the central neurons are themselves surround suppressed, resulting in an increase in activity of the central neurons. This results in a periodic component in the size tuning curves of V1 neurons. This had not been previously explored but was verified in ferret V1 (Rubin, Van Hooser, and Miller, 2015).

Orientation tuning of suppression

Surround suppression is sensitive to the orientation of the surround stimulus (Cavanaugh, Bair, and Movshon, 2002; Webb et al., 2005). Surround stimuli matching the preferred orientation of the the neuron being observed are more suppressive than surrounds of non-preferred orientations. The SSN accounts for this because the units preferentially connect to units with similar orientation preferences.

Contrast dependence of suppression

Surround suppression varies with centre-stimulus contrast. When the centre stimulus is weak, some surround stimuli are found to be less suppressive or even facilitative to central neurons (Sengpiel, Sen, and Blakemore, 1997). However, when the contrast of the centre stimuli increases the surround stimulus becomes increasingly suppressive. This arises in the SSN because at low contrast, the network behaviour is not dominated by network inhibition (external input dominates - see figure 3.2) and so for smaller surround stimuli the centre can be excited. As contrast increases the network becomes increasingly inhibition dominated so the activity of the centre is suppressed.

3.2 RF-Integrated Model

The SSN as presented by Rubin, Van Hooser, and Miller (2015) is not pixel computable - the input stimulus $h(x)$ is a 2D function which describes an input to each grid position based on its orientation preference. The SSN units are assumed to be complex (phase-invariant) cells which are only selective for orientation. In order to use the SSN in a pixel-computable large-scale model of V1, I modified the receptive field model developed in the previous chapter to produce the input to the SSN, using a one-to-one connection of receptive field units to SSN units. Figure 3.3 illustrates the integrated model structure.

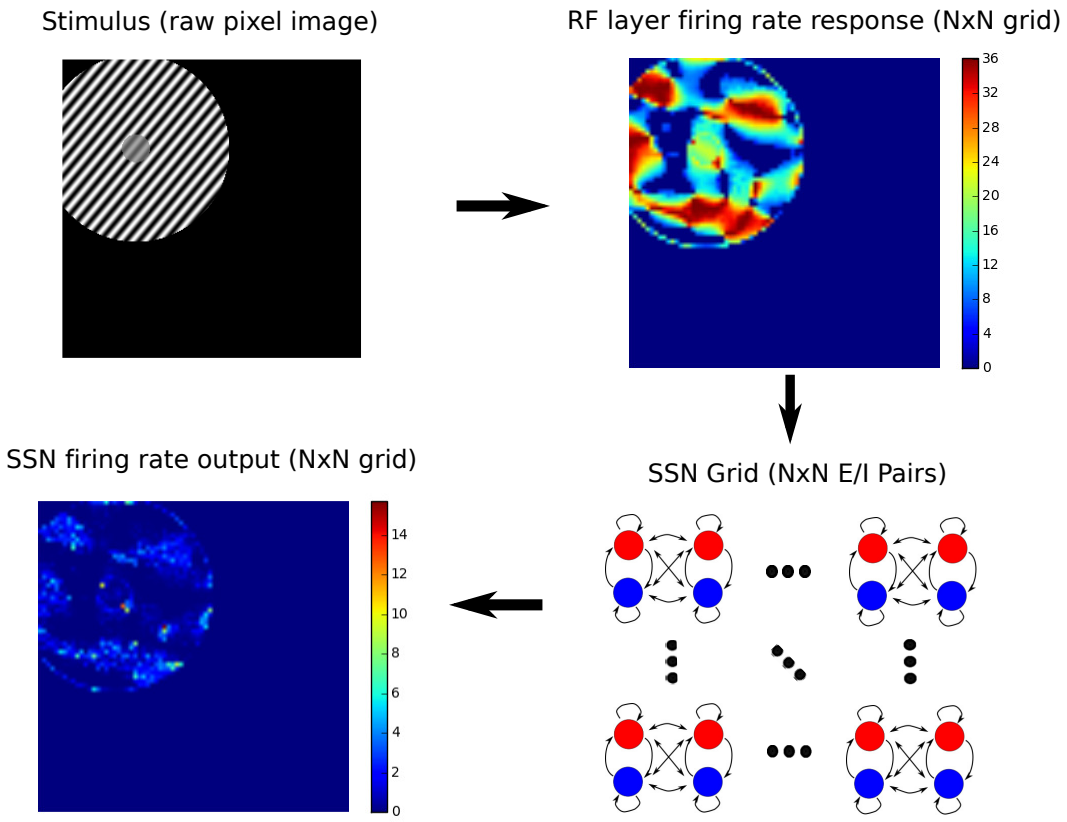

Figure 3.3: Schematic of the integrated model architecture. A stimulus (in grayscale pixel format) is input to a $N \times N$ grid of receptive-field units, whose firing rates are then passed to a $N \times N$ network of supralinear E/I pairs which produce firing rates in response. This integration of a receptive field layer as input to the SSN makes the SSN pixel-computable while preserving much of the original behaviour.

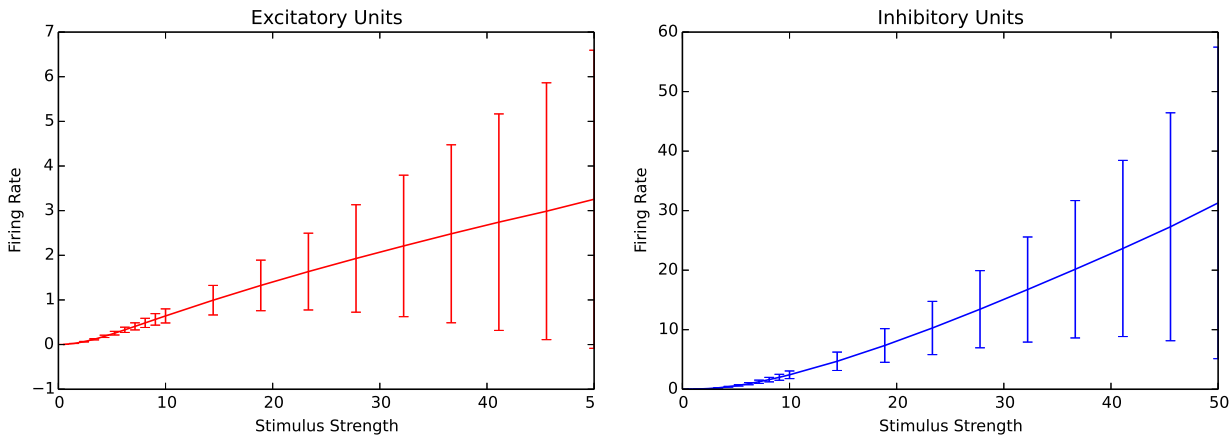

Figure 3.4: Mean contrast response functions of 40 SSN units to full-field gratings. Whiskers indicate standard deviation. The mean E unit response increases linearly while the mean I unit response remains supralinear with increasing contrast.

3.2.1 Contrast response function of the SSN

The SSN requires supralinear unit response functions in order to produce contrast-dependent changes in behaviour including summation field size shrinkage, contrast-dependent surround suppression, and other effects (Rubin, Van Hooser, and Miller, 2015). This is based on the observation that cortical neuron spike-rate is related to membrane potential by a power law with powers between 2 and 5 (Priebe et al., 2004). However the receptive field model in the previous chapter has a saturating contrast response function, as observed in cats and monkeys (Albrecht and Hamilton, 1982). Figure 3.4 shows that the SSN does not account for contrast saturation in the individual responses of the neurons. The stimulus strength c in equations 3.1 and 3.2 corresponds to contrast on an arbitrary scale which linearly scales the input shape $h(x)$ - it is assumed in the SSN that V1 neurons will typically operate in an unsaturated space. Because the model needs to be adapted for use with pixel-images, a contrast saturation function must be developed which maintains SSN behaviour and also reflects the saturation observed in V1 recordings. The SSN input image $h(x)$ can be thought of as feedforward input from layer 4 neurons, which are also known to saturate. Assuming unity gain feedforward input to layer 2/3, the function $h(x)$ should therefore saturate with increasing image contrast. If the slope of the contrast response function is sufficiently high in the early levels of contrast before saturation, SSN contrast-behaviour should be maintained with a saturating input function.

3.2.2 Phase invariance

The receptive field model presented earlier modelled simple cell behaviour exclusively, while the SSN employs exclusively complex cells. This is reasonable for a model of layer 2/3, where cells are primarily complex (Hubel and Wiesel, 1968). In their development of a motion energy model, Adelson and Bergen (1985) used linear filters in quadrature pairs to produce phase-invariant responses. To do this, the input current J is calculated as follows:

$$J_{stim} = \sqrt{J_{cos}^2 + J_{sin}^2} \tag{3.9}$$

$$J_{cos} = \sum_{i=-N/2}^{N/2} \sum_{j=-N/2}^{N/2} g_{cos}(i,j)h(i,j) \tag{3.10}$$

$$J_{sin} = \sum_{i=-N/2}^{N/2} \sum_{j=-N/2}^{N/2} g_{sin}(i,j)h(i,j) \tag{3.11}$$

where $g_{cos}(i,j)$ is a 2D Gabor with cosine carrier, $g_{sin}(i,j)$ is a 2D Gabor with sine carrier, and $h(i,j)$ is a luminance image.

A 2D orientation-contrast curve was constructed to fit RF model parameters. The orientation component is Gaussian in order to match the orientation tuning of the SSN units. A least-squares curve-fitting was used to fit the parameters. The updated receptive field model behaviour is shown in figure 3.5. In order to provide the best fit across a range contrasts and orientations the model's contrast response function does not saturate, but is sublinear which is a behaviour observed in some neurons (Albrecht and Hamilton, 1982). A hyperbolic-saturating response which also maintained sharpness of orientation tuning was not realizable with the implemented model.

3.2.3 Integrated model results

In order to ensure that the use of RF-model input maintains key SSN behaviours, a number of the experiments presented by Rubin, Van Hooser, and Miller (2015) are recreated.

Network activity transition

The integrated model shows the same transition from external to network drive dominance as seen in the original SSN (figure 3.6). The transition occurs at contrasts of about 18%. This transition indicates that SSN behaviour of inhibition stabilization can be maintained with the saturating input function. The network proportion grows more slowly than in the original SSN because the contrast function is sublinear in the integrated model (see figure 3.2 for comparison).

Size tuning and periodicity

As in the original SSN, the integrated model produces size tuning that has a periodic component for the I units. Figure 3.7 shows mean size tuning for units that showed strong surround suppression among 40 E/I pairs. Units were probed for their size tuning by stimulation with preferred gratings of increasing size. As in Rubin, Van Hooser, and Miller (2015), only units with surround suppression index (SSI) > 0.25 were used to produce the size tuning plots (40 E units, 11 I units). The I units show a secondary peak in response as a result of suppression of I units which were suppressing the centre units, causing the centre units to increase their firing rates. SSI was calculated as follows:

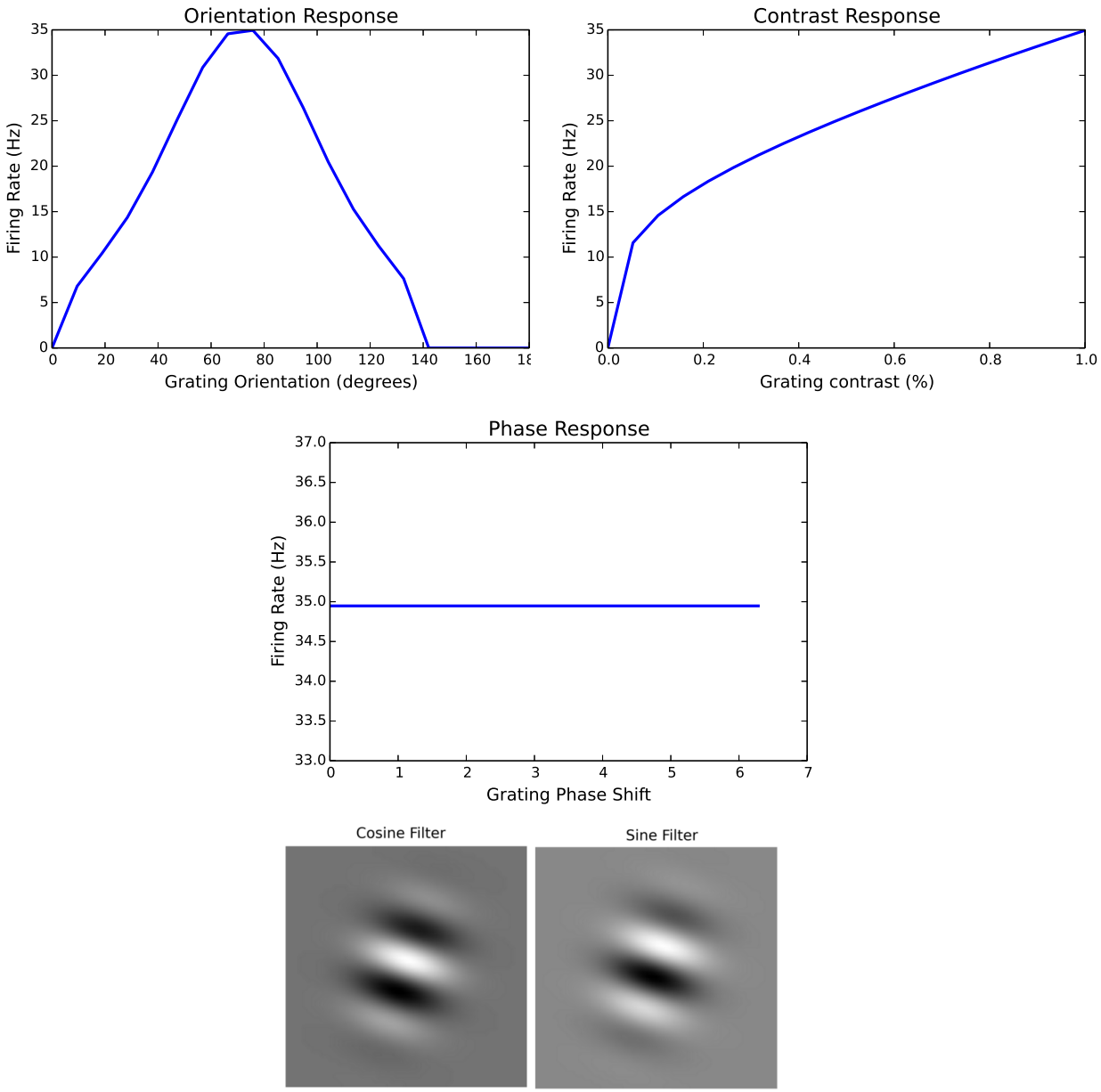

Figure 3.5: Receptive-field model output updated for usage with the SSN. *Top left:* Orientation response at 100% stimulus contrast. The orientation curve used for fitting was Gaussian to match the bandwidth shape of the original SSN. *Top right:* Contrast response for stimuli of preferred orientation and spatial frequency. The best fit found by the optimization was a sublinear one. *Middle:* The phase response to gratings of preferred orientation and spatial frequency at 100% contrast. The quadrature pairing of the Gabor-RF produces a phase-invariant response. *Bottom:* The quadrature pair of Gabor filters which produced these responses.

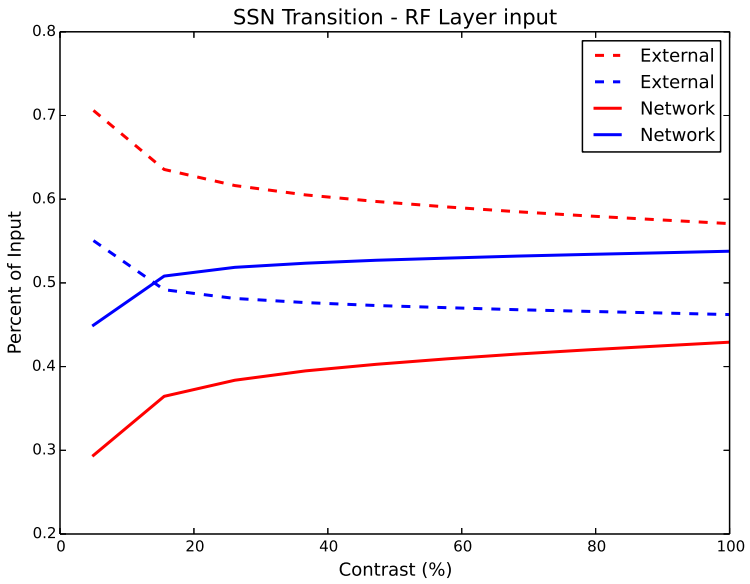

Figure 3.6: Proportion of network activity and external drive versus contrast, using the RF model as input. As contrast increases, the proportion of total drive to the units is increasing from recurrent network activity, not from the external stimulus $ch(x)$. Network drive quickly dominates for both E and I units, stabilizing the power-law output of the excitatory units.

$$SSI = \frac{r_{max} - r_{full}}{r_{max}} \tag{3.12}$$

where r_{max} is the peak firing rate to stimuli less than 10° in size, and r_{full} is the response to 16° stimuli.

Contrast sensitivity of surround suppression

SSN units were probed for surround suppression dependence on centre stimulus strength and surround size. Figure 3.8 shows an example of the stimuli used for these experiments.

Figure 3.9 shows the model response for surround suppression of a single unit with increasing contrast of the centre stimulus. Various surround sizes were tested. In this case, no facilitation is observed for any surround stimuli.

Figure 3.10 shows another E unit which does show slight facilitation for one surround stimuli. This unit does not show complete transition from facilitation to suppression with increasing contrast, as has been observed in the physiology and produced with the SSN. Of 50 units examined for their contrast-sensitivity in suppression, this unit was the only one to show slight facilitation. It is not clear what portion of units showed facilitation in the original SSN.

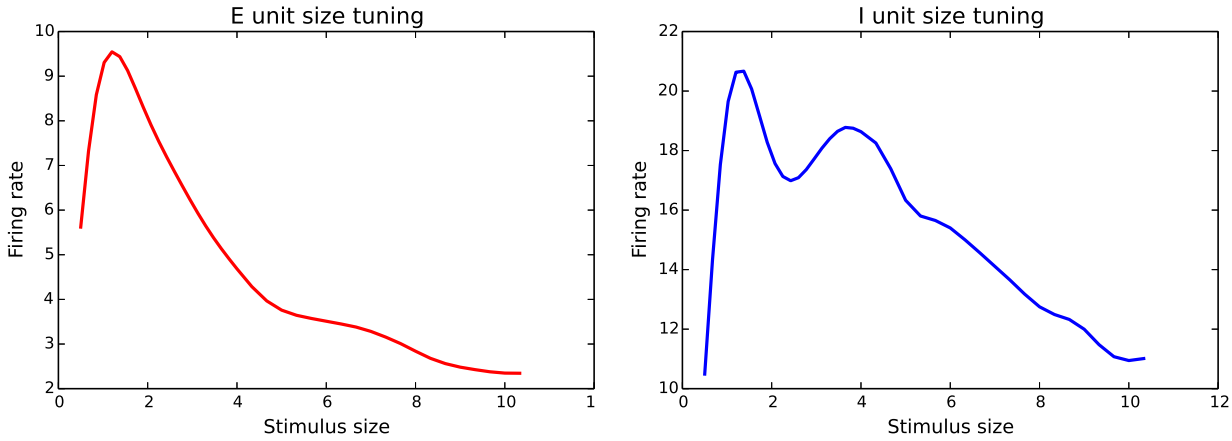

Figure 3.7: Mean size tuning for both E and I units, using the RF model as input. 40 pairs were probed, of which 40 E units and 11 I units were used for the plot (see text).

Figure 3.8: A sample stimulus for examining contrast sensitivity of surround suppression. The diameter of the inner stimulus was always the summation field size of the unit (the first peak in its size-tuning curve). 7 surround stimuli of increasing size were used (see figures 3.9 and 3.10). The surround stimulus contrast was held at 50, and the centre stimulus contrast was varied.

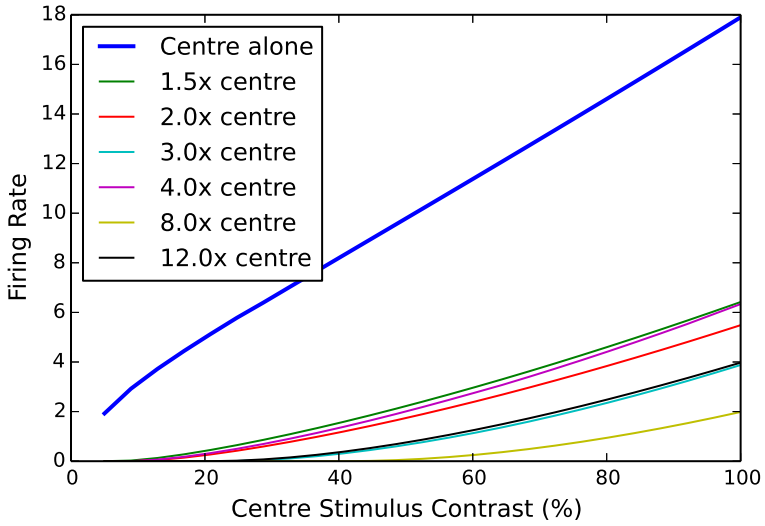

Figure 3.9: Contrast sensitivity of surround suppression for a single E unit. In this units case there is no facilitation for low contrast centres. Surround stimuli contrast was held at 50% while centre stimulus contrast was varied.

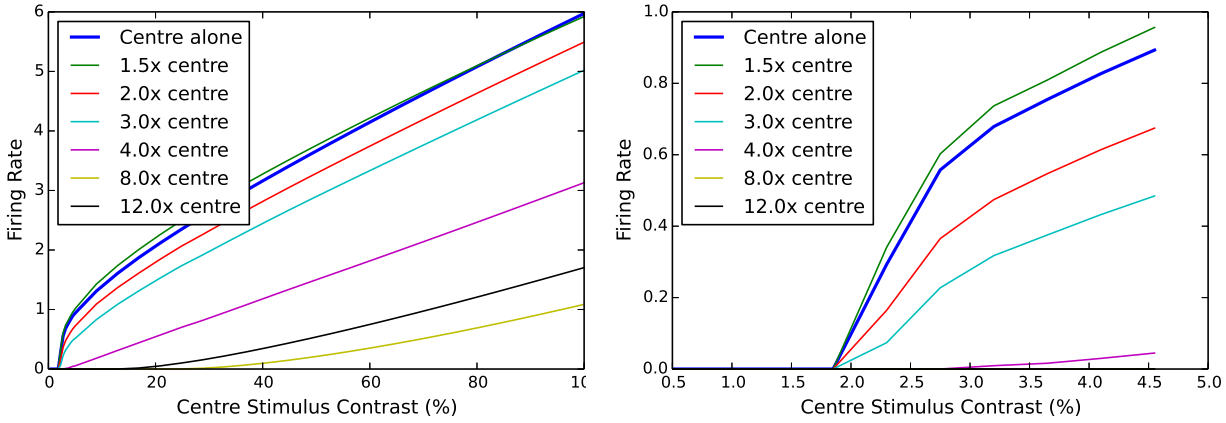

Figure 3.10: Contrast sensitivity for a neuron that shows slight facilitation for one surround stimuli (1.5x outer diameter). The left plot shows the contrast response for the full range of contrasts, while the left plot shows the same responses enlarged for low contrast. The facilitation is slight and does not transition into suppression at higher contrasts.

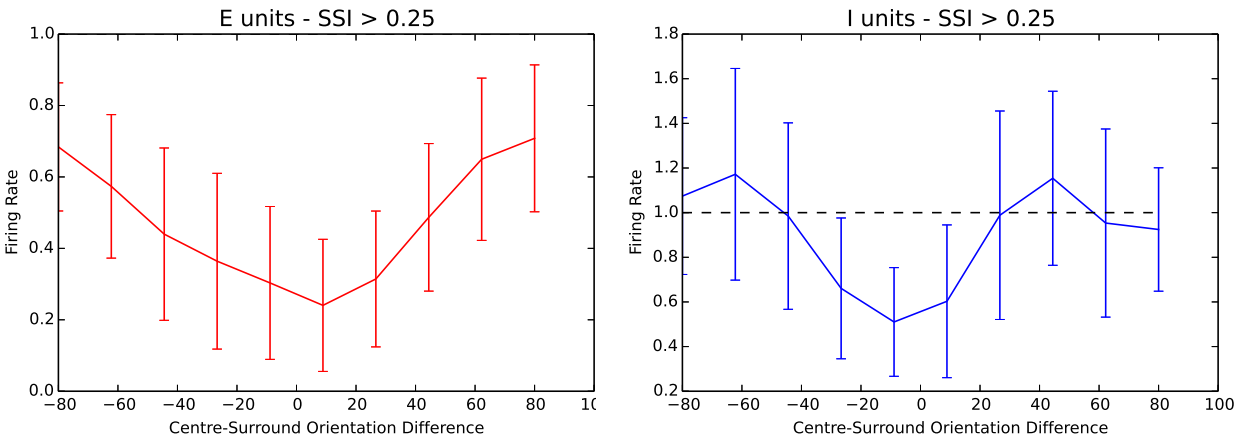

Figure 3.11: Orientation tuning of surround suppression averaged over 40 units. Firing rates are normalized to the response to preferred stimuli of the summation field size. Surround gratings had an outer diameter of 15.1°and inner diameter equal to the summation field size.

Orientation tuning of surround suppression

The integrated model shows orientation tuning of suppression similar to that of the original SSN. Figure 3.11 shows the results for 40 units. For each unit, the summation field size was found and then surrounds of various orientations were applied. Firing rates were normalized to the summation field size stimulus alone - values greater than 1 indicate facilitation while values less than 1 indicate suppression. As with size tuning, only units with SSI > 0.25 were plotted. I units showed suppression at surround orientations close to the preferred orientation, but were much less suppressed or even facilitated at orientations close to orthogonal. This is qualitatively similar to the results of the original SSN, as well as experimental results in the macaque (Cavanaugh, Bair, and Movshon, 2002).

3.2.4 Discussion

A pixel-computable model of the SSN was developed, and several major results of the SSN were reproduced using computer-generated images. These results include diverse surround suppression results: size tuning, summation field size dependence on contrast, and orientation-tuned surround suppression. These results are reproducible even with the saturating contrast response function of the RF model, which is not present in the SSN.

One shortcoming of the integrated model is the centre contrast sensitivity of surround suppression. E units probed did not show facilitation transitioning to suppression as observed in the original SSN and in the literature. This is likely a result of the sublinear contrast response function produced by the RF model. E unit facilitation can occur at low contrasts because the local inhibition is low, allowing for E→E connections from the surround to increase activity without being cancelled. Because the contrast response function of the integrated model is not supralinear, it is possible that the E component does not generally overcome the I component at low contrasts. Restoring a supralinear contrast function should rectify this.

The RF-integrated model used to make the SSN pixel-computable is not contrast invariant - at low contrasts the orientation bandwidth shrinks slightly. This is not observed in the physiology and can potentially be rectified by implementing a power-law in place of the LIF non-linearity (Priebe and Ferster, 2008). Preliminary experiments showed this to be effective, but the non-linearity resulted in unpredictable behaviour at non-preferred spatial frequencies.

The RF-unit layer used to provide input to the SSN does not currently contain preferences for other spatial frequencies. In the macaque there is a map of spatial frequency preference whose gradient tends to be orthogonal to orientation (Nauhaus et al., 2012). Because spatial frequency selectivity was not previously explored it was omitted here. Spatial frequency does affect surround suppression though, so developing a reasonable map of selectivity would allow for exploration of further tuning of suppression (Webb et al., 2005).

3.3 Extending the SSN

While the SSN has been shown to produce monocular surround suppression, it is interesting to examine whether it can account for other forms of suppression observed in V1.

3.3.1 Ocular dominance and interocular transfer of suppression

Surround suppression has been shown to have a binocular component. This "interocular suppression" was first identified in cat V1 by DeAngelis, Freeman, and Ohzawa (1994) (see figure 3.12), who were investigating the source of length tuning. Webb et al. (2005) showed in the macaque that for a preferred CRF stimulus presented to the dominant eye of a cell, a surround stimulus presented to the dominant eye is almost always more suppressive than a surround stimulus in the non-dominant eye. This binocularity implies that at least a component of suppression arises in cortex (earlier areas are monocularly driven). Since V1 is the first stage of visual processing that receives input from both eyes, and since surround suppression can be accounted for by the SSN, it is worth investigating if the SSN mechanisms can produce interocular transfer of suppression as observed in cats and monkeys.

There is no ocular dominance-related preference of horizontal connections in cats (Lowel and Singer, 1992) - that is, cortical neurons do not preferentially connect to one another based on eye preference. Thus the connections between SSN units do not need to be modified based on their eye preference. However, an ocular dominance map is required. Gradients of ocular dominance and orientation preference have been shown to be orthogonal (Obermayer and Blasdel, 1993). Therefore it is important that the spatial relationship between orientation preference and ocular dominance is maintained. Because the horizontal connections depend only on orientation, the observed interocular transfer behaviour may be a result of the structure of ocular dominance map and its relationship to orientation.

Obermayer and Blasdel (1993) presented a number of images of orientation and ocular dominance maps obtained through optical imaging in macaques. These images were registered about the vascular system and thus correspond to the same area of cortex. These images were extracted and then a histogram equalization was performed on each to produce reasonable distributions of preference values. Figure 3.13 shows the results. For the orien-

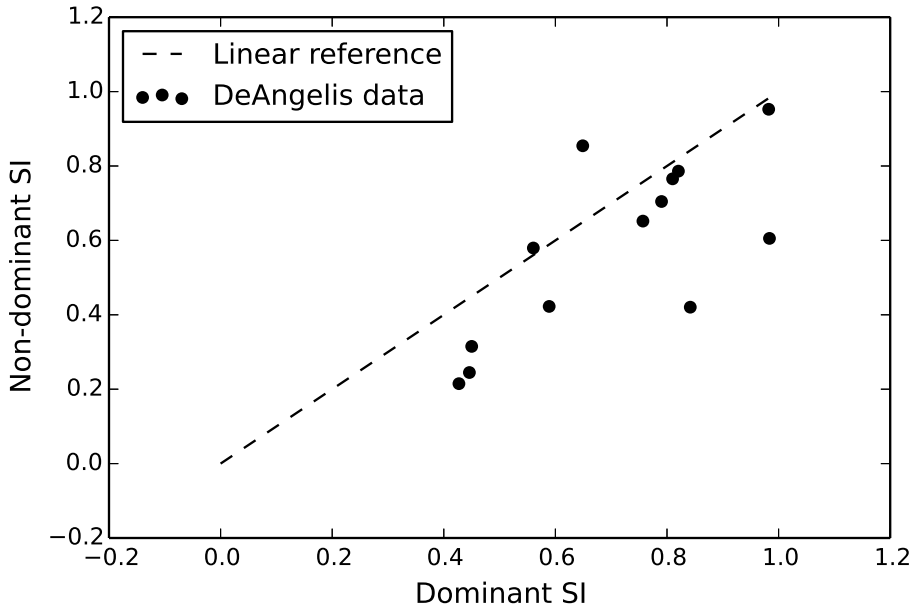

Figure 3.12: Interocular transfer of suppression in the cat. "SI" is suppression index, where 1 corresponds to complete suppression and 0 corresponds to no suppression. In 11 of 13 neurons studied, the surround stimuli in the dominant eye were at least as suppressive as those in the non-dominant eye, and were typically more suppressive. Data from DeAngelis, Freeman, and Ohzawa (1994). Similar results were found in the macaque (Webb et al., 2005), where 15 of 15 neurons studied showed stronger suppression in the dominant eye.

tation data, a uniform distribution of orientation preferences was used, matched to the hue channel in the image. The ocular dominance map was histogram-equalized to the distribution of ocular dominance preference in layer 2/3 found by Hubel and Wiesel (1968) (see figure 3.14).

Using this ocular dominance map, a new external stimulus $h(x)$ was developed for experiments with the SSN. It was assumed that binocular input was added linearly by the cell as has been shown in studies of ocular dominance (Hubel and Wiesel, 1968). The ocular dominance map was normalized to contain values from 0-1 with 0 corresponding to the contralateral eye and 1 corresponding to the ipsilateral eye. This was then used as a scale factor for the original $h(x)$ input specified in the SSN. For example, a unit in ocular dominance group 4 would have a normalized value of 0.5, so it would receive half its input from the contralateral eye, and half from the ipsilateral eye.

$$h_{total}(x) = h_{contra}(x) + h_{ipsi}(x) \tag{3.13}$$

$$h_{contra}(x) = h(x)O(x) \tag{3.14}$$

$$h_{ipsi}(x) = h(x)|1 - O(x)| \tag{3.15}$$

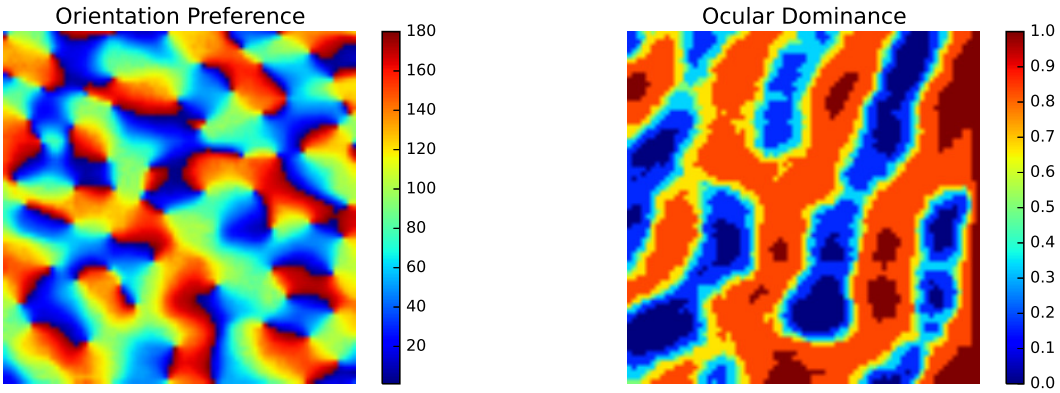

Figure 3.13: Extracted maps of orientation preference and ocular dominance from Obermayer and Blasdel (1993). Left: The raw image was histogram-matched to a uniform distribution from 0-180, to produce a map equally representing all orientations. Right: Extracted map of ocular dominance. The ocular dominance map was histogram-matched to a distribution (see figure 3.14), resulting in 7 discrete values which were normalized here.

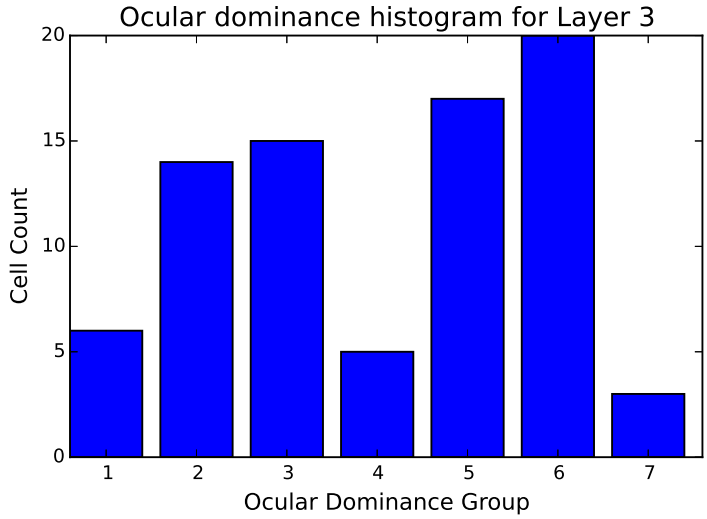

Figure 3.14: Distribution of ocular dominance preference in macaque monkeys. Adapted from Hubel and Wiesel (1968). Neurons were classified as belonging to one of 7 ocular dominance groups, with 1 corresponding to total preference for the contralateral eye, 4 to no apparent eye preference, and 7 to total preference for the ipsilateral eye.

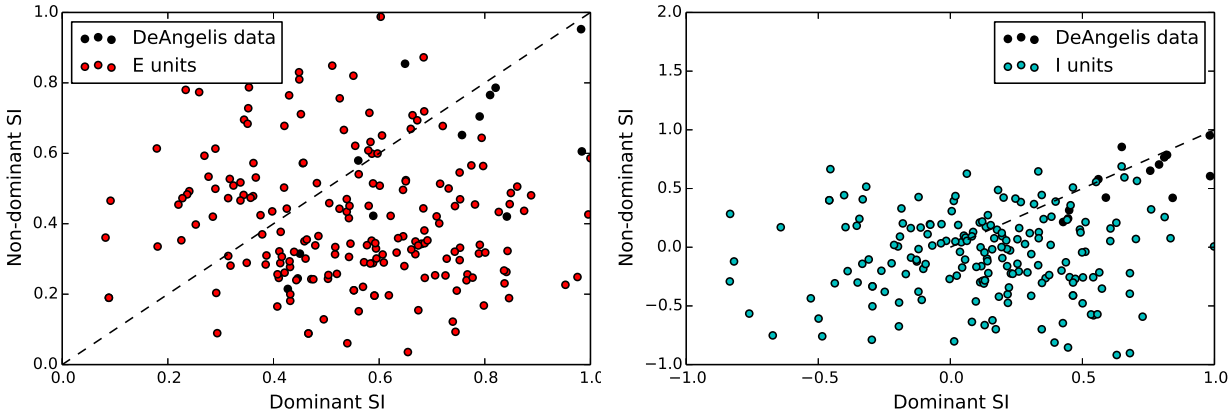

Figure 3.15: Interocular transfer as produced by the SSN with original parameters and added ocular dominance map. Suppression index calculated as in eq. 3.16. E unit p-value: 0.00518. I unit p-value: $< 1e-7$ (2D, 2-sample KS test comparing model results to observed data). Note that the dashed line in each figure is a linear reference, not a best-fit.

3.3.2 Initial evaluation

200 units (E and I pairs) were randomly selected and their summation field sizes were found. As described in DeAngelis, Freeman, and Ohzawa (1994). Each unit's dominance was determined by rounding its ocular dominance preference (units with of ocular dominance group 4 were not used). Surround stimuli were ring-shaped with inner diameter equal to that of the summation field size. The suppression index is defined as follows:

$$SI = \frac{R_{SFS} - R_{surround}}{R_{SFS}} \tag{3.16}$$

where R_{SFS} is the response rate to an optimally oriented stimulus of the summation field size and $R_{surround}$ is the surround-suppressed response rate. Surround stimuli were applied to the dominant and non-dominant eyes separately.

Figure 3.15 shows the results of this study. Neither E nor I units showed consistency with the observed data. Using only units which had SSI > 0.25 did not significantly improve the correlation. I units showed particularly poor consistency, with many units showing facilitation from either or both surround stimuli. This is understandable because I units show more periodicity in their size tuning curves, including some increases in firing rates above that of their summation field size. By adding an ocularity to the input stimuli, only half of the neurons are stimulated for a given monocular stimuli versus a binocular one. Because of this the total drive to the network is less than if the stimuli were monocular and therefore surround stimuli tend to be less suppressive.

3.3.3 Model optimization and evaluation

The SSN contains a number of parameters which may or may not affect the emergence of interocular transfer of suppression. Because a new orientation preference map is introduced,

Parameter	Min.	Max.
J_{EE}	0.09	0.2
J_{IE}	0.3	0.45
J_{EI}	0.089	0.105
J_{II}	0.08	0.105
σ_{EE}	$7\Delta x$	$9\Delta x$
σ_{IE}	$10\Delta x$	$16\Delta x$
σ_{EI}	$3\Delta x$	$5\Delta x$
σ_{II}	$3\Delta x$	$5\Delta x$

Table 3.2: SSN model parameters over which the hyperparamter optimization procedure was run. J_{ab} parameters affect the recurrent connection strengths in the network, and σ_{ab} affect the spatial span of these connections. These parameters all affect the emergence of surround suppression in the SSN.

whose scale and spatial structure differs from that of the original implementation, connection weights between units likely need to change to produce interocular transfer of suppression. Because it is difficult to predict which parameters will most affect the emergence of interocular suppression, an iterative optimization of model parameters was used. A procedure was developed using Hyperopt, a Python package for hyperparameter optimization (Bergstra, Yamins, and Cox, 2013). The parameters optimized over are summarized in table 3.2. Parameters were kept relatively close to those in the original model in order to avoid problems with network stability.

The Hyperopt procedure iteratively instantiated new SSN models with various combinations of the parameters, and ran the same interocular transfer experiment on each instance. In order to quantify the goodness of fit, a 2-dimensional, 2-sample Kolmogorov-Smirnov (KS) test (Press et al., 1996) was used. This is a statistical test which assumes no prior distribution of the data and can be used to compare whether two 2-D datasets are consistent with the same distribution. The KS test returns a p-value between 0 and 1, indicated the probability of the null hypothesis (that the two 2D datasets are consistent with the same distribution) is true. The goal is to find model parameters which produce interocular suppression similar to the observed data of DeAngelis, Freeman, and Ohzawa (1994). Thus the objective is to maximize the p-value produced by the 2D KS test.

The optimization procedure was run for 100 iterations, probing 50 E/I pairs at the same positions for each iteration. For both E and I units, no p-value larger than 0.001 was observed, indicating that for no combination of the selected parameters produced interocular transfer of suppression as observed in the cat or macaque. Figure 3.16 shows the best results (highest p-value) found during hyperparameter optimization. Again, many I units were facilitated by either or both the dominant and non-dominant surround stimuli. E units only showed facilitation for surrounds in the non-dominant eye. This may be because of insufficient drive from surrounding units which would typically cause suppression. Because these units do not show facilitation in the dominant eye, it is likely that the non-dominant stimulus was not very strong, thus adding local excitation without sufficient inhibition to the central units.

Another possibility is that while V1 neurons may not preferentially connect to neurons

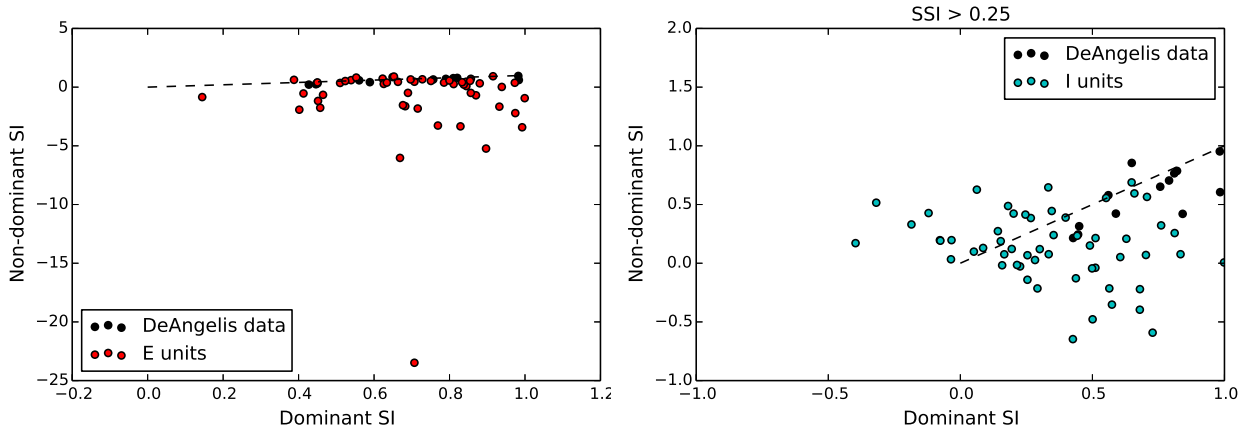

Figure 3.16: Results of hyperparameter optimization for both E and I units (p-value: 0.000809).

with similar eye preference, they may have stronger connection weights to those neurons than the connections to neurons preferring the opposite eye. Indeed, models of V1 based on Hebbian development have produced exactly this effect (Bednar and Miikkulainen, 2006). In order to test this, an additional parameter was introduced into the ocular dominance-enabled SSN which weakens the connection weights W_{ab} from neurons preferring the opposite eye by scaling them:

$$W_{ab} = Z \odot W_{ab} \tag{3.17}$$

where Z is some scaling matrix with values between 0 and 1, of the same dimensions as W_{ab}, is equal to 1 for neurons with the same eye preference (rounding the ocularity value to 0 or 1), and some constant value ζ between 0 and 1 elsewhere. The $\odot$ operator indicates element-wise multiplication.

Another hyperparameter optimization procedure was run for 50 iterations, yielding the results shown in figure 3.17. This time the cost function only calculated the fit of the E unit suppression. The I units previously showed little similarity to the data in all iterations, and so were possibly preventing the procedure from converging on an optimum for the E units. The procedure found a parameter combination yielding a p-value (for E units only) of 0.2418. This indicates that the E unit model data is not significantly different from the observed data from DeAngelis, Freeman, and Ohzawa (1994). The corresponding I units however yield a p-value of 8e-13, indicating no consistency. Table 3.3 shows the parameters found via optimization. By weakening connections from units which prefer the opposite eye by less than 5 ($\zeta = 0.959$) and slightly altering the connection weights, interocular transfer can be produced in the E units.

3.3.4 Discussion

The parameter search indicates that for parameters close to the standard SSN parameters presented in Rubin, Van Hooser, and Miller (2015), interocular transfer does not emerge as a

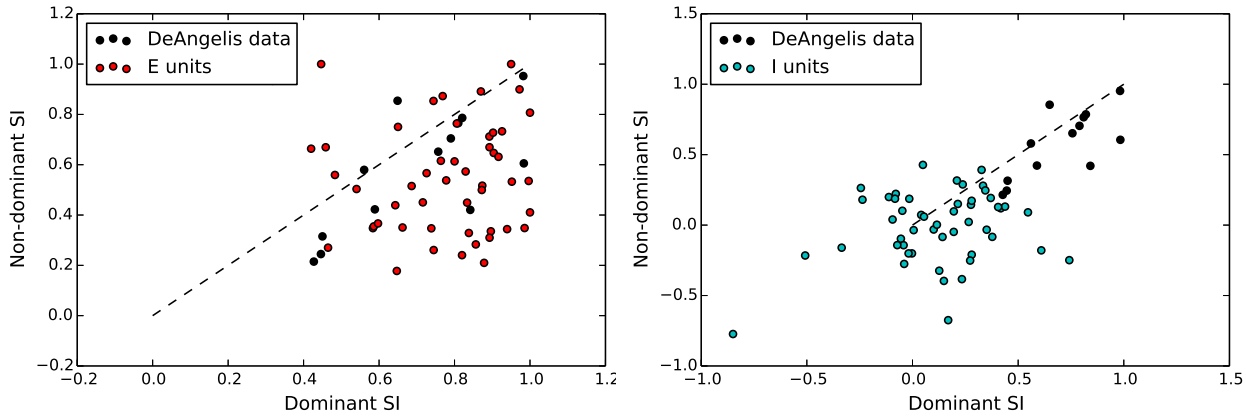

Figure 3.17: Results of hyperparameter optimization with connection weights modified by ocular dominance preference. E unit p-value: 0.2418, I unit p-value: 8e-13.

Parameter	**Optimal Value**	**Original Value**
J_{EE}	0.0931	0.1
J_{IE}	0.379	0.38
J_{EI}	0.0930	0.089
J_{II}	0.0841	0.096
σ_{EE}	$8.918\Delta x$	$8\Delta x$
σ_{IE}	$14.519\Delta x$	$12\Delta x$
σ_{EI}	$4.868\Delta x$	$4\Delta x$
σ_{II}	$3.982\Delta x$	$4\Delta x$
ζ	0.959	N/A

Table 3.3: SSN hyperparameters that yielded the interocular transfer results in figure 3.17.

common trait amongst SSN I units, but with minor weakening of connection weights based on eye preference, E units can show interocular transfer of suppression. I units tended towards lower suppression in general and were often facilitated by interocular surround stimuli, even with selective reweighting.

I units in the original SSN showed a broader range of surround suppression index, including several units showing facilitation (see figure 7B of Rubin, Van Hooser, and Miller (2015)). This is because the I units are more periodic in their size tuning due to E $\rightarrow$ I connections being the strongest with the longest range. Further examination of I units under interocular suppression is required to understand whether this inconsistency indicates a problem with the SSN.

DeAngelis, Freeman, and Ohzawa (1994) made no further distinction as to how recorded neurons were selected. Webb et al. (2005) introduced additional criteria for probing interocular transfer in the macaque, including ensuring that the non-dominant surround stimulus suppressed the response to $< 90\%$ of the CRF response. In preliminary trials, this only modestly improved results.

It is possible that interocular transfer of suppression does not arise from V1 mechanisms (Webb et al., 2005) and so the SSN does not necessarily need to account for this particular phenomenon. If this is the case, then it would be valuable to explore what feedback mechanisms cause interocular transfer to emerge in V1.

There may be other factors which were not sufficiently explored in the hyperparameter optimization. For example, there may be a more optimal range of connection lengths and weights for the extracted orientation and ocular dominance maps used. Because suppression in the SSN is emergent as a result of lateral interactions, it may be that the simplified connections of the SSN do not produce interocular transfer where a more realistic connection scheme may do so. More dense local connections with sparse long-range connections of higher weight for instance may produce more realistic suppression.

The SSN uses an even proportion of E and I units, but in reality only about 20% of cortical neurons are inhibitory. There are also several classes of inhibitory neurons found in cortex. In all trials of interocular transfer in the SSN, I units showed significantly more variation in behaviour and particularly showed facilitation for many non-dominant surrounds. Because of this, I units may not be adequately represented in the observed data. I units show non-dominant surround facilitation for many units which is somewhat expected because I units are more periodic in their size tuning. It is worthwhile to introduce a set of criteria similar to those of Webb et al. (2005) to see if there are constraints on the emergence of interocular transfer in the SSN.

Chapter 4

Design of a Macaque Vision Robot

In this section the requirements for a stereo camera orientation system are presented. The purpose of this system is to replicate the major components of the primate visual system including the high dynamic movements, visual acuity, and stereopsis. Note that only the requirements are presented here - the detailed design and implementation is being completed at the time of writing by a contractor at Axcendo Innovation Corporation. We explore the relevant physiology of the primate visual system and propose methods of emulating those properties. Because the camera system is to be used with models based on macaque monkey data, we try to restrict our data sources to those of macaques. Some data however is difficult to attain for macaques, such as monocular field of view. In these instances we refer to human data instead.

4.1 Motivation

One promising application of a visual cortex model is the area of machine vision. Tasks such as object identification, navigation, and visually-guided grasping are all areas in which robotics has yet to surpass human performance, and they are all made possible by the extensive network of the visual cortex. By studying the brain and its computations we may gain some insight in how to build new robotic systems with improved performance.

Biological vision is different from machine vision in many respects. Unlike cameras, which typically have sensors of uniform resolution across their fields of view, the primate retina is highly non-uniform, with high resolution in its centre (the *fovea*), and rapidly declining resolution in the peripheral regions. The image perceived by the primate viewer is constructed based on information primarily from this high resolution fovea, which is moved around the scene via rapid eye movements. By intelligently selecting where to focus next, primate eyes effectively fixate on potentially important parts of a visual scene, constructing and maintaining a representation of the surrounding world. This is computationally much more efficient than trying to process an entire scene at once at high resolution. Area V1 has be implicated in serving as a "saliency map" (Li, 2002), indicating to other brain areas where the eyes should move to next. This allows the fovea to image a region of interest with great detail, building understanding of the environment.

Because of this marked difference between primate eyes and traditional cameras used in

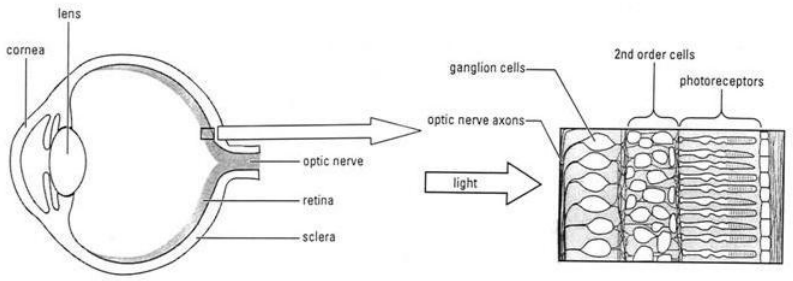

Figure 4.1: Primate eye and retina structure. Source: Dowling (2007)

machine vision, it is desirable to have a more biologically-grounded vision system to integrate with models of the visual cortex. Many features of the visual cortex are structured around the fast movement of the high-resolution fovea, such as cortical magnification (Wässle et al., 1990) and cortical areas controlling eye movements such as the superior colliculus.

Primate eyes move with great acceleration and velocity, and thus to emulate their function a custom camera orientation gimbal is desired. These eye movements and the brain areas that control them are well-studied. Therefore a platform which emulates eye moves provides an opportunity for the exploration of neural motor control and its integration with the visual system.

4.2 Background - Primate Eye Physiology

The primate eye is similar to a camera in its major components: it has a pupil for variable aperture and a lens for refracting and focusing light. In this sense the retina is analogous to a camera's sensor. The retina is the neural interface between the brain and the light of the outside world. It is made up of five types of neuron and two classes of photoreceptor cells, constituting three layers. In the outermost layer the photoreceptor cells absorb light and transduce it into a neural signal. This signal is then relayed through through "bipolar" neurons and then retinal ganglia cells whose axons form the optic nerve, carrying the signal to the brain (Meister and Tessier-Lavinge, 2013). Any light that is not transduced into neural signal is lost. Figure 4.1 illustrates some major components of the retina and its location in the eye.

4.2.1 Visual Acuity

There are many factors affecting the quality of the ocular image, including refractive error caused by lens abberation and the density of photoreceptors and neurons. Unlike a typical camera sensor, which has a uniform density of phototransducers across its surface, the retina has an area of highest photoreceptor and neuron density called the *fovea*. The fovea images the centre of gaze. When discussing the retina, the term *eccentricity* refers to the distance (typically in degrees, measured from the centre of the eye) from the fovea to some point on the retinal surface. Visual resolving power is often discussed in terms of *acuity*, which is

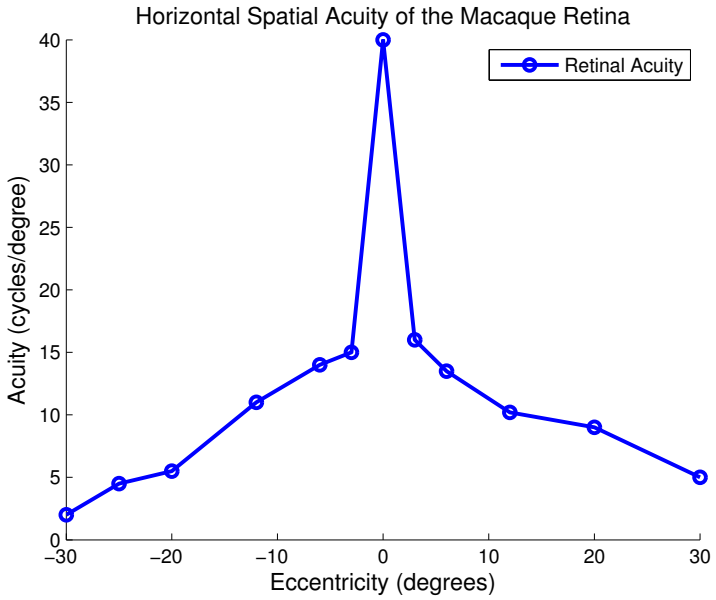

Figure 4.2: Macaque visual acuity versus horizontal eccentricity. Negative eccentricity corresponds to the nasal field, and positive to temporal field. Acuity here is measured in cycles of a sinusoidal grating per degree of visual field. Macaques show a greater asymmetry between the two fields than humans. Source: Merigan and Katz (1990)

sometimes measured in cycles of a grating/arc-minute. Merigan and Katz (1990) found that the visual acuity of the macaque monkey is highest in the fovea, and then falls off quickly with increasing eccentricity (see figure 4.2). They also showed that the acuity was dependent on photoreceptor (cone) density in the fovea, and on P-type ganglion cells at eccentricities greater than 10 degrees.

4.2.2 Motion Dynamics

In order to produce a detailed perception of the world with a narrow, high-resolution fovea, primate eyes move in rapid and well-characterised way. They also move in order to compensate for egocentric movements, particularly those of the neck. Eye movements can be broadly classified into six major categories (Leigh and Zee, 1999):

- Saccade: the sudden change of gaze direction to focus on another point. These are the fastest of all eye movements, and can be both voluntary and involuntary. In humans, saccades can reach speeds of 500 deg/s and accelerations of 25000°/s. In macaques, speeds of 600 deg/s and accelerations of 40000°/s have been observed (Leigh and Zee, 1999).

- Smooth Pursuit: the consistent tracking of moving objects. These movements can reach up to 100 deg/s.

- Fixation: small corrective movements are enacted when the eye is fixed on a single point.

- Vestibulo-ocular reflex (VOR): the equal and opposite rotation of the eye relative to the movement of the head, maintaining line of sight to a point of fixation. This can reach accelerations of 15000°/s and speeds of up to 100°/s (Huterer and Cullen, 2002).
- Vergence: the two eyes rotate inwards towards each other to focus on points at various depths.
- Optokinetic reflex: a general term for eye movements mediated by visual stimuli. In foveal eyes this is largely replaced by the smooth pursuit movement (Leigh and Zee, 1999).

Both the neck and eyes have 3 rotational degrees of freedom - pan, tilt, and roll. The range of motion for the eye is neurologically restricted to about ± 45°pan/tilt (Villgrattner, 2010). However, since there are redundant DOF in the eye and neck, the full range of eye motion is seldom used.

Humans generally gaze within 1-2 degrees of the target of fixtation, but it has been shown that they are capable of gaze precision up to 0.16 degrees.

4.2.3 Existing Solutions

While there are a multitude of camera orientation systems, only a small number of these have been designed to emulate the movement of primate eyes. Here we review the major past solutions for rapid, eye-like camera orientation.

Villgrattner (Villgrattner, 2010; Villgrattner and Ulbrich, 2011; Villgrattner and Ulbrich, 2010) designed both 2- and 3-DOF camera orientation systems which move with dynamics exceeding those of the human eye. The system was design for mounting on a human head for eye movement reproduction, and so had to be very lightweight and small. Because of this, human/monkey-like visual acuity was not reproduced.

The Agile Eye (Gosselin and Hamel, 1994) is another parallel camera orientation mechanism. The system meets the high dynamic requirements for camera orientation, however the geometry of the design prohibits stereo pairing with the desired baseline separation (see section "Stereopsis"). Also, the size and weight of the system would make mounting on a neck-gimbal difficult.

Bang et al. (2006) developed a modified version of the Agile Eye for mounting on a humanoid robot head. The design meets the dynamic requirements but has a restricted range of motion and does not attempt to reproduce the visual acuity of the eye.

Li et al. (2015) employ a stereo design based on the Agile Eye for VOR experiments on a mobile robot. The size of the two gimbals prohibits a realistic stereo baseline. The authors do not discuss the visual acuity of the system either, so it is assumed that this was not a factor in the design.

Limitations of Existing Solutions

While all of the above designs have met/exceeded the dynamic requirements of a human eye, neither has done so while reproducing the visual acuity of the eye. Both solutions use low-resolution, light weight cameras as their payload, and thus are limited in their ability

Requirement	Desired Value	Minimum Value
Eye Max. Velocity	600°/s	500°/s
Eye Max. Acceleration	40000°/s^2	25000°/s^2
Eye Angular Resolution	0.16°	0.2°
Eye Range of Motion (pan/tilt)	±45°	±28°
Neck Max. Velocity	100°/s	50°/s
Neck Max. Acceleration	15000°/s^2	10000°/s^2
Neck Angular Resolution	0.5°	1 °
Neck Range of Motion (tilt)	±90°	±45°
Neck Range of Motion (pan)	±90°	±45°
Neck Range of Motion (roll)	±45°	±30°

Table 4.1: Motion requirements for the camera orientation system.

to produce the desired level of acuity. Since the intended system application is for use with neurological visual system models, it is highly desirable to investigate what improvements can be made with regards to visual acuity.

4.3 Requirements

Outlined here are the requirements identified for producing a camera orientation system which can be used with software models of visual cortex.

4.3.1 Motion Dynamics

In order to reduce mechanism and control complexity, the 3 degrees of freedom of the macaque eye is reduced to two degrees of freedom - pan and tilt. The third, roll DOF can be compensated for in software.

The motion requirements of the neck and eye gimbals are summarized in table 4.1. Values are specified for all directions of movement unless otherwise specified.

The requirements for the neck gimbal are based around vestibulo-ocular reflex experiments that find the highest accelerations and velocities of the macaque neck while maintaining compensation by the eye (Huterer and Cullen, 2002).

The eye gimbal in particular must exhibit minimal backlash to ensure precise positioning.

4.3.2 Stereopsis

The range of disparity in a stereo vision system is determined by the separation of the eyes or cameras. Most adult humans have a interpupillary distance of 50-75mm, with a mean of 63mm (Dodgson, 2004). Macaque eye separation is approximately half of this (Harwerth, Smith, and Crawford, 1996), so a foveal camera separation range of 25-37mm should be appropriate for a robot.

4.4 Design

The high dynamic requirements of the camera orientation system will certainly demand a custom gimbal design. Because of this, it is reasonable to begin the design by selecting the appropriate camera arrangement as this will constrain the mechanical design.

4.4.1 Camera Selection

Frame rate and shutter type

The frame rate of the selected camera is constrained by the temporal frequency sensitivity of the primate eye.

The shutter type is also very important for high frame rate. A global shutter means that the entire sensor is read out at the same time, while a rolling shutter integrates each row of pixels with a slight time delay. Because of this delayed integration, a rolling shutter sensor can show motion blur in instances where the motion is relatively fast compared to the sensor integration time.

Camera resolution, field of view, and acuity

The stationary human eye has a horizontal monocular field of view of about 150 degrees (Howard and Rogers, 1995). The binocular region is about 114 degrees. The acuity data from Merigan and Katz (1990) is presented in cycles (of a grating) per arc-minute of horizontal visual field. It is reasonable to define a cycle as 2 pixels. From figure 4.2 it can be seen that the peak acuity is 40 cycles/arc-min., which is converted to 1.3 pixels/arc-min. A camera with a field of view of 150 degrees would require 11700 pixels in the horizontal direction to reproduce foveal acuity across the entire visual field. This is an extremely high pixel count which precludes cameras useful for motion imaging. Thus using a standard camera and lens will not meet the acuity requirements.

4.4.2 Camera arrangement

By trading off resolving power with field of view the retina can be emulated through arrangement of multiple cameras, custom optics, or non-traditional camera sensor readouts. However, while there are existing camera sensors that enable selective read-out such as the CMOSIS CHR70M, it is not possible to produce a fovea-like readout profile. Because of this, two design options were considered: a two- camera solution using standard lenses and a two camera solution using custom lenses.

Two-Camera Solution

In order to emulate the non-uniform acuity of the retina, two cameras could be used for each eye - one foveal and one peripheral camera per eye. This would allow one camera to have a smaller field of view with higher eccentricity, while the other has lower acuity and a larger field of view. To compare camera and lens arrangements, a simple cost function was developed. For regions of the visual space where either camera's acuity was less than desired,

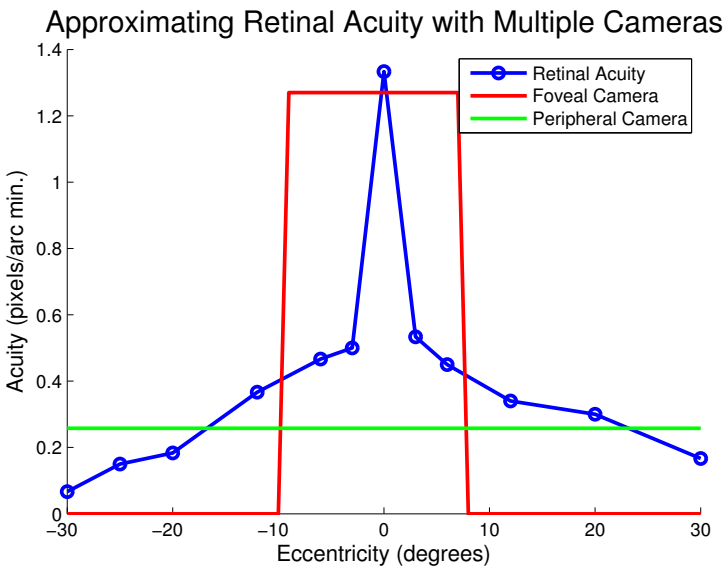

Figure 4.3: Acuity versus eccentricity curve, with overlayed camera resolution. Note that the total field of view of the peripheral camera (green line) is 108°(±54°). Retinal acuity data adapted from Merigan and Katz (1990).

the difference was integrated. Any regions with acuity higher than desired received a cost of zero. The arrangement which yields the lowest cost is used. After comparing several camera and lens combinations on the basis of field of view, acuity, frame rate and shutter type, the Point Grey FL3-U3-13S2M-CS was been selected for both peripheral and foveal cameras. The foveal lens is the Varioptic Caspian C-39N0-16, and the peripheral lens is the Fujinon YV2.8x2.8SA-2. These lenses were are compatible with the Point Grey camera in mounting (they are both CS-mount with acceptable back-flange distance), as well as their refractive power (both lenses support the camera's pixel size of $3.75 \times 3.75 \mu m$). This combination of camera and lenses produces acuity across the horizontal visual field shown in figure 4.3. The relative positioning of these two cameras and lenses is described in figures 4.4 and 4.5. The two cameras are fixed in relative position and rotate together about the specified point in both axes (pan and tilt). The centre of rotation was determined as half the diameter of an average macaque eye in front of the camera sensor, a distance of 6 mm from the front edge of the camera. This figure is used because the centre of rotation affects how images from the two cameras relate to each other at different fixations. Each camera requires a USB 3.0 cable attached to the back which may interfere with movement.

Custom Lens Solution

Another method of obtaining the non-uniform acuity and field of view is to use a custom lens which distorts the incoming light in a profile similar to that of the retina. A lens manufacturer (B-Con Engineering Inc.) is able to construct custom lenses with specified distortion profiles. In order to specify a distortion profile which provides the same magnification as the retina, the retinal ganglion cell density alone can be used. Wässle et al. (1990) showed that the cortical magnification factor, that is, the proportion of neurons representing a fraction of the

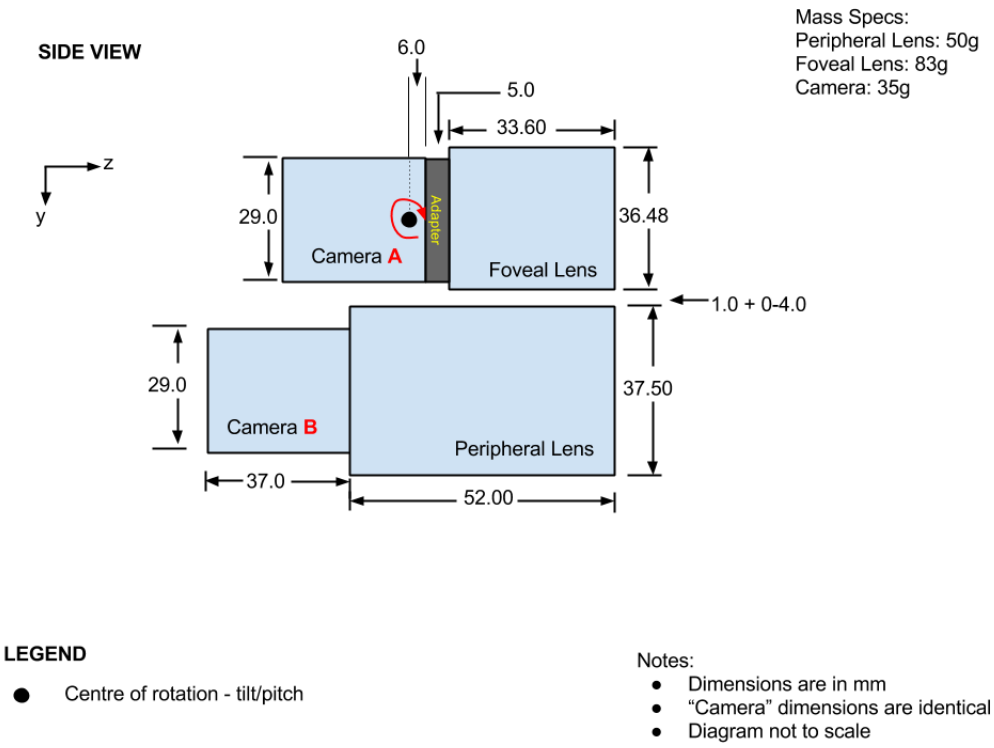

Figure 4.4: Two-camera solution: side view of relative camera positioning for one eye gimbal. The peripheral camera is positioned below the foveal one, and the outermost edge of each camera's lens are aligned. The two cameras would rotate together about the indicated centre of rotation, fixed relative to each other.

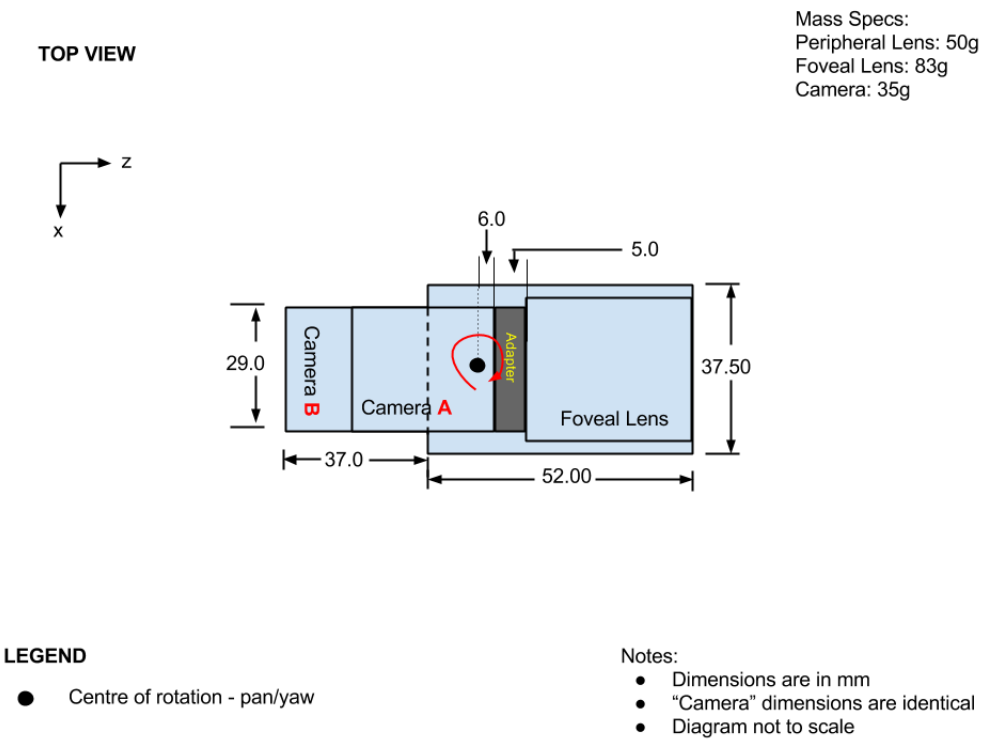

Figure 4.5: Two camera solution: Top view of relative camera positioning for one eye gimbal. The two cameras are vetically aligned, so that the wider field of view afforded by the peripheral camera extends equally on both sides.

visual field, is directly determined by the retinal ganglion cell density, without any further magnification. Therefore the proportion of cortical cells representing visual information from the fovea is directly proportional to the proportion of retinal ganglia in the fovea. It was decided to use a radially symmetric lens in order to avoid problems with installing the lens on the camera. Because the nasal field ganglion cell density declines less sharply than the temporal, we use the nasal retinal ganglia density (downsampling can be performed in software). Figure 4.6 illustrates the desired distortion profile. Figure 4.7 shows the expected image produced by a lens with the profile specified in figure 4.6.

The form-factor of the custom lens is also important because it constrains the relative positioning of the two eye gimbals (see "Eye Positioning" below). Because the Varioptic Caspian C-39N0-16 was considered an acceptable lens, we use it's dimensions to constrain the dimensions of the custom lens.

Final Selection

The two-camera solution requires the least data bandwidth as well as a simpler gimbal design and demands less powerful actuation to reach the dynamic requirements. The four-camera solution requires no custom optics and is thus less expensive to implement (and repair, should the need arise). The use of four cameras also significantly increases processing complexity - synchronization of four cameras is a computationally-intensive task. Preliminary simulations showed that the four-camera solution would require applied forces close to the limit of available actuators, so the total mass and arrangement of the system is of critical importance. After circulating lens requirements to various manufacturers, a quote for an appropriate custom lens was obtained that placed the two-camera option within budget. Because of the dynamic requirements and processing bandwidth, the design will incorporate two cameras using custom lenses of the profile described above.

4.4.3 Eye Gimbal Positioning

The eye gimbals will perform "vergence" movements of maximum 10°, meaning that they will rotate in opposite directions inwards to focus on an object to a maximum of 10°. This constrains the minimum separation of the cameras to avoid collision, including a safety buffer. See figure 4.8. Because of the length of the lens and required centre of rotation, the two gimbals holding each camera must be positioned at a separation of 58 mm to allow for a 5 mm safety buffer at 10° central vergence. This separation is approximately double that of the eye separation of the macaque, and much closer to that of humans. The separation will affect stereo vision, with a wider baseline and therefore greater discriminable depth than in macaques.

4.4.4 Actuator Selection

Villgrattner and Ulbrich (2011) made use of piezo actuators from *Physik Instrumente*. They were selected for their high force generation, light weight, and precise movement. Other designs have successfully used DC motors to produce the required velocities including Gosselin and Hamel (1994) and Dankers and Zelinsky (2003). As Villgrattner and Ulbrich (2011)

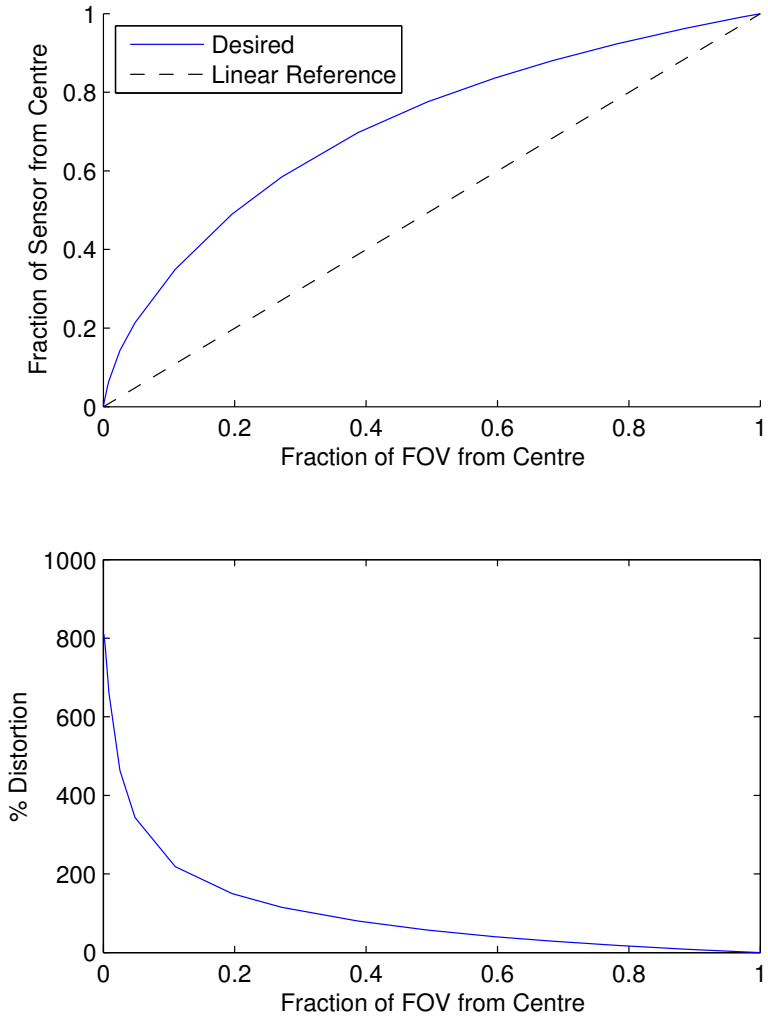

Figure 4.6: Top: Proportion of the camera sensor imaging the field of view. As distance from the centre of the image increases, the fraction of area on the sensor decreases. The linear reference is the profile for an ideal traditional lens. Bottom: Distortion profile for a lens replicating the retinal ganglion cell density of the retina. Distortion falls off exponentially at points increasingly far from the centre of the image. Figures produced by Bryan Tripp.

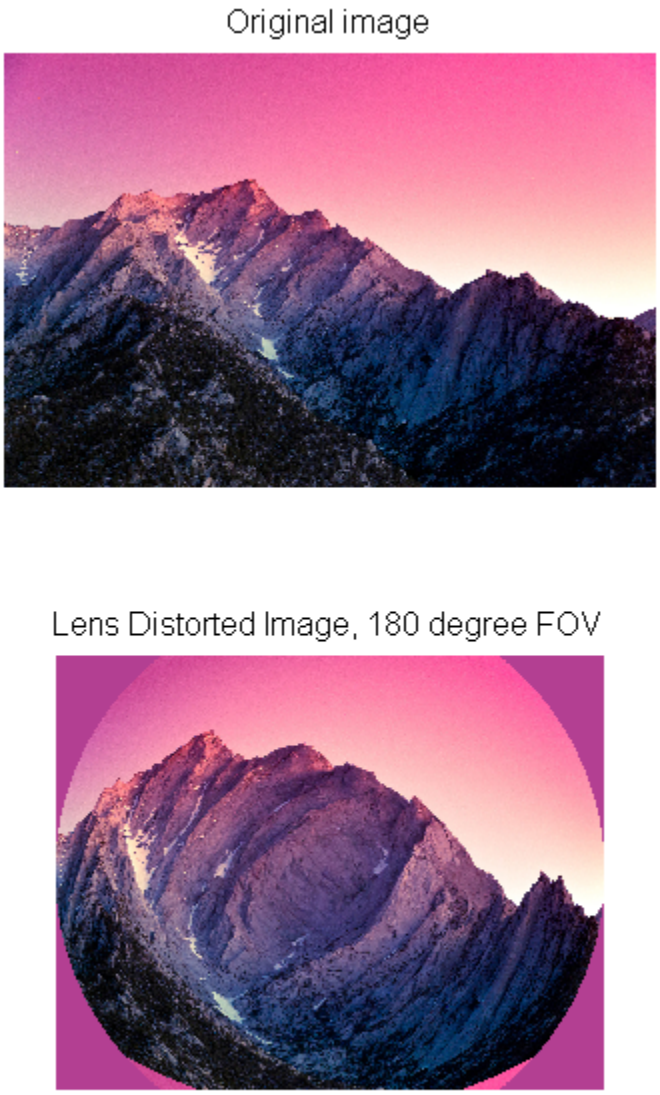

Figure 4.7: Expected image produced by a lens with the distortion profile specified in figure 4.6. Figure produced by Bryan Tripp.

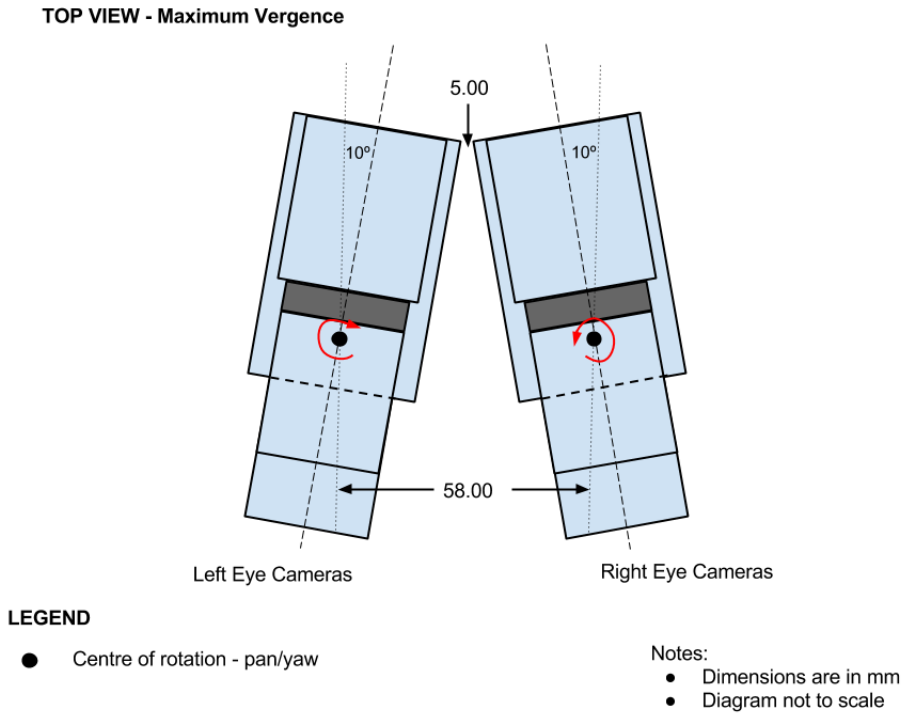

Figure 4.8: Top view of both eyes in central vergence. This position constrains the separation of the two gimbals as shown. Depending on the amount of backlash present in the two eye gimbals, the 5mm buffer may be adjusted. This buffer dimension is left to the designer.

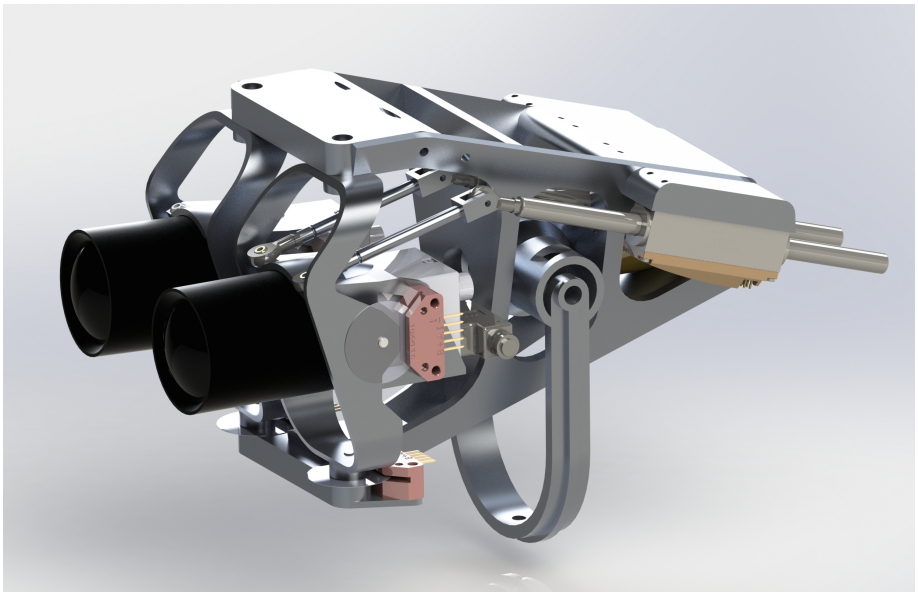

Figure 4.9: CAD rendering of a late revision of the robot, produced by Scott Huber of Axcendo Innovations.

notes, servo motors are likely inappropriate because the PWM control introduces additional dead-time and cannot be used with closed-loop control requiring actuator position.

4.4.5 Detailed Design and Validation

At the time of writing, neither the mechanical design or lens manufacturing is complete. A document summarizing these requirements was circulated amongst contractors, and Axcendo Innovations (`www.axcendo.com`) was selected based on their quote. Designers at Axcendo will complete the detailed design including component selection, and oversee fabrication of custom components and assembly. Figure 4.9 shows a CAD rendering of a late revision of the mechanical design.

Chapter 5

Conclusions

5.1 Contributions

This thesis lays a foundation for further modelling of visual neural networks. The head-robot specification presented in chapter 4 will provide a testbed for sensisorimotor modelling experiments and will be used to demonstrate the utility of such models for real-world applications. Specifically, this robot will enable embodied models of stereo vision, eye movements in response to visual stimuli, and motor control to name a few applications. The use of this robot demands that models perform in real time, which is a major obstacle for models of increasing biophysical realism.

The integrated model developed in chapters 2 and 3 provides a pixel-computable mechanistic model of V1 layer 2/3. This model produces a wide range of observed phenomena and does so in a biologically reasonable way. The model is based on the stabilized supralinear network and therefore a number of simplifications are made. Making the SSN pixel-computable allows for analysis of other regions of non-linear behaviour, for instance spatial frequency tuning of suppression.

The SSN was also modified to examine the emergence of interocular transfer of suppression. Excitatory units show interocular transfer of suppression, however the inhibitory units show behaviour not consistent with the observed behaviour. If interocular transfer is indeed a result of V1 mechanisms and not from feedback from higher areas, this inconsistent behaviour may implicate an alternate scheme of inhibition stabilization. The large variety of inhibitory neurons found in V1 may also play a role in this.

5.2 Future work

5.2.1 Supralinear neuron RF model

The SSN is predicated on a supralinear relationship between stimulus contrast and neuron firing rate in order to produce contrast-dependent nonlinearities. This is based on the work of Priebe and Ferster (Priebe et al., 2004; Priebe and Ferster, 2008) who found that a power law relationship between membrane potential and firing rate produces contrast-invariant response in a model of V1 layer 4 without any recurrence or feedback. This also requires

decreased variability of response with increasing contrast. Attempts to integrate this power-law relationship into the RF model produced superior fits to observed data when parameters were fit with an optimization procedure, but behaviour was unpredictable and unrealistic for non-preferred stimuli due to the non-linear nature of the power law. Further development of the RF model is required to produce contrast invariance as well as other phenomena described in the next section.

This power-law relationship begs the question: how does contrast saturation arise? It is possible that it is also a network effect, mediated by recurrent local inhibition. Katzner, Busse, and Carandini (2011) found that suppression of inhibitory neurons in cat V1 results in increase of responses at all contrasts (scaling the output), but no change in contrast sensitivity - neurons still saturate at the same contrast levels, but with higher response rates. This indicates that contrast saturation may in fact be an intrinsic property of neurons, or a result of earlier-stage normalization. In either case, saturation must be reconciled with the power-law relationship between contrast and spike rate that is required by the SSN. It is possible that the cortical network simply operates in the unsaturated region under normal conditions, and saturation is only observed in cases where small groups of neurons are stimulated.

5.2.2 Cross-orientation suppression

In addition to contrast invariance, Priebe and Ferster (2008) showed that cross-orientation suppression (sublinear response to superposed orthogonal gratings) can be accounted for with a purely feedforward model similar to the RF model developed here. This is due to the model's properties of spike threshold and contrast saturation, both of which are present in the RF model presented here. Katzner, Busse, and Carandini (2011) showed that cross-orientation inhibition was observed in V1 without cortical inhibition, and showed that the suppression was monocular, which is more characteristic of relay cells in the LGN. Because the RF model is effectively a model of processing up to V1 layer 4 (including LGN relay cells), with parameter tuning it should be possible to produce cross-orientation inhibition.

5.2.3 Top-down effects: attention and contour integration

There are a number of phenomena in V1 which are thought to be the result of feedback from higher cortical areas. One of the most important of these is attention, which has been shown to produce various types of response modulation to stimuli in neurons imaging the attended area. The integrated model currently has no provision for feedback influence. Modulation of normalization factors has been proposed as an explanation for some of these attentional modulations (Reynolds and Heeger, 2009; Carandini and Heeger, 2012). Rubin, Van Hooser, and Miller (2015) suggest that this could be achieved in the SSN via the "balanced amplification" of an inhibition-stabilized network - slight increases in network excitation results in largely increased E and I firing rates which may be similar to those observed in studies of attention. It is not entirely clear how attention influences the cortical circuit, so it is valuable to explore how the SSN may incorporate this.

While V1 has historically been suggested as the site of contour integration, Chen et al. (2014) showed that contour-related signals arise first in V4, up to 50 milliseconds before

they are detected in V1. This suggests that the contour integration demonstrated in V1 is at least partly the result of feedback connections from higher areas.

Bosking et al. (1997) showed that long-range excitatory connections in tree shrew V1 are preferential not only to neurons of similar orientation preference, but also those neurons whose receptive fields are co-axially aligned. This observation is resonant with the "association field" proposed by Field, Hayes, and Hess (1993) and the observed phenomenon of coaxial facilitation (Kapadia et al., 1995). Introducing this property to the SSN connection weights would certainly produce some degree of contour integration without feedback influence.

5.2.4 Large-scale network model

The integrated model presented here spans a relatively small scale. In order to work with real images of a useful size (such as those produced by the head robot) the model needs be scaled up considerably, eventually to a full-sized macaque V1. This large scale model must include realistic cortical magnification, retinotopy, and increasing receptive field size with eccentricity. The RF model also needs realistic variation in receptive field properties. In order to do this, simple sampling of distributions of selectivities is not enough because the maps of many tuning properties are spatially correlated. Some preference maps are well-studied, such as orientation and ocular dominance, and procedurally-generated versions of these maps may be sufficient for modelling. New methods of seeding selectivity maps with others will need to be developed in order to produce realistic topographic tuning properties.

Scaling up the integrated model is a major computational challenge which will be met through algorithm optimization and computational resources such as GPUs which allow for mass parallelization.

5.2.5 Integration with robot

The robot described in chapter 4 is designed for use with models of visual networks. Once the robot is fabricated and validated, systems-scale circuits can be built to drive it. These models will be fairly large, including early visual areas, sensory processing areas, and motor control areas. There have been several models replicating control of eye movements in response to visual stimuli including saccade (Girard and Berthoz, 2005) and smooth pursuit movements (Churchland and Lisberger, 2001). These models could be adapted to work with real images and used to control the robot in a closed-loop system. Receptive field analysis and development of each brain area in the circuit (as was done in this thesis for V1) in order to produce a visual network model. Some work to this end has already been performed for smooth pursuit eye movements (Sadat Rezai, 2014).

Bibliography

[1] Edward H Adelson and James R Bergen. "Spatiotemporal energy models for the perception of motion". In: *JOSA A* 2.2 (1985), pp. 284–299.

[2] Yashar Ahmadian, Daniel B Rubin, and Kenneth D Miller. "Analysis of the stabilized supralinear network". In: *Neural computation* 25.8 (2013), pp. 1994–2037.

[3] Duane G Albrecht and David B Hamilton. "Striate cortex of monkey and cat: contrast response function." In: *Journal of Neurophysiology* 48.1 (1982), pp. 217–237.

[4] John Allman, Francis Miezin, and EveLynn McGuinness. "Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons". In: *Annual review of neuroscience* 8.1 (1985), pp. 407–430.

[5] Jeffrey S Anderson et al. "Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex". In: *The Journal of Neuroscience* 21.6 (2001), pp. 2104–2112.

[6] Young-Bong Bang et al. "A three-degree-of-freedom anthropomorphic oculomotor simulator". In: *International Journal of Control Automation and Systems* 4.2 (2006), pp. 227–235.

[7] James A Bednar and Risto Miikkulainen. "Joint maps for orientation, eye, and direction preference in a self-organizing model of V1". In: *Neurocomputing* 69.10 (2006), pp. 1272–1276.

[8] James Bergstra, Daniel Yamins, and David Cox. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures". In: *Proceedings of The 30th International Conference on Machine Learning*. 2013, pp. 115–123.

[9] William H Bosking et al. "Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex". In: *The Journal of neuroscience* 17.6 (1997), pp. 2112–2127.

[10] Edward M Callaway. "Local circuits in primary visual cortex of the macaque monkey". In: *Annual review of neuroscience* 21.1 (1998), pp. 47–74.

[11] M. Carandini. "Area V1". In: *Scholarpedia* 7.7 (2012). revision 126411, p. 12105.

[12] Matteo Carandini and David J Heeger. "Normalization as a canonical neural computation". In: *Nature Reviews Neuroscience* 13.1 (2012), pp. 51–62.

[13] Matteo Carandini et al. “Do we know what the early visual system does?” In: *The Journal of Neuroscience* 25.46 (2005), pp. 10577–10597.

[14] Vivien Casagrande and Xiangmin Xu. “Parallel Visual Pathways: A comparative Perspective”. In: *The Visual Neurosciences*. Ed. by L.M. Chalupa and J.S. Werner. MIT Press, 2004, pp. 494–506.

[15] James R Cavanaugh, Wyeth Bair, and J Anthony Movshon. “Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons”. In: *Journal of Neurophysiology* 88.5 (2002), pp. 2547–2556.

[16] Minggui Chen et al. “Incremental integration of global contours through interplay between visual cortical areas”. In: *Neuron* 82.3 (2014), pp. 682–694.

[17] Mark M Churchland and Stephen G Lisberger. “Experimental and computational analysis of monkey smooth pursuit eye movements”. In: *Journal of neurophysiology* 86.2 (2001), pp. 741–759.

[18] Andrew Dankers and Alexander Zelinsky. “A Real-World Vision System: Mechanism, Control, and Vision Processing”. English. In: *Computer Vision Systems*. Ed. by JamesL. Crowley et al. Vol. 2626. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003, pp. 223–235. ISBN: 978-3-540-00921-4. DOI: 10.1007/3-540-36592-3_22. URL: http://dx.doi.org/10.1007/3-540-36592-3_22.

[19] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience*. Vol. 806. Cambridge, MA: MIT Press, 2001.

[20] Gregory C DeAngelis, Ralph D Freeman, and Izumi Ohzawa. “Length and width tuning of neurons in the cat’s primary visual cortex”. In: *Journal of Neurophysiology* 71.1 (1994), pp. 347–374.

[21] Russel DeValois, E. W. Yund, and N. Hepler. “The Orientation and Direction Selectivity of Cells in Macaque Visual Cortex”. In: *Vision Research* 22 (1982), pp. 531–544.

[22] Neil Dodgson. “Variation and extrema of human interpupillary distance”. In: *Stereoscopic Displays and Virtual Reality Systems*. Vol. 5291. 2004.

[23] J. Dowling. “Retina”. In: 2.12 (2007). revision 91715, p. 3487.

[24] Chris Eliasmith and Charles H Anderson. *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT press, 2004.

[25] Bard Ermentrout. “Neural networks as spatio-temporal pattern-forming systems”. In: *Reports on progress in physics* 61.4 (1998), p. 353.

[26] David J Field, Anthony Hayes, and Robert F Hess. “Contour integration by the human visual system: Evidence for a local “association field””. In: *Vision research* 33.2 (1993), pp. 173–193.

[27] Kent Foster et al. “Spatial and Temporal Frequency Selectivity of Neurones in Visual Cortical Areas V1 and V2 of the Macaque Monkey”. In: *Journal of Physiology* 15 (1985), pp. 331–363.

[28] Charles D. Gilbert. "Intermediat-level Visual Processing and Visual Primatives". In: *Principles of Neural Science*. Ed. by E. Kandel, Jessel, and Schwartz. McGraw Hill, 2013. Chap. 27, pp. 602–620.

[29] Charles D. Gilbert. "The Constructive Nature of Visual Processing". In: *Principles of Neural Science*. Ed. by E. Kandel, Jessel, and Schwartz. McGraw Hill, 2013. Chap. 25, pp. 556–576.

[30] BetABERTHOZ Girard and Alain Berthoz. "From brainstem to cortex: computational models of saccade generation circuitry". In: *Progress in neurobiology* 77.4 (2005), pp. 215–251.

[31] C.M. Gosselin and J.-F. Hamel. "The agile eye: a high-performance three-degree-of-freedom camera-orienting device". In: *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*. 1994, 781–786 vol.1. DOI: 10.1109/ROBOT.1994.351393.

[32] R. S. Harwerth, E. L. Smith, and M. L. J. Crawford. "Motor and sensory fusion in monkeys: Psychophysical measurements". In: *Eye* 10 (1996), pp. 209–216.

[33] David J Heeger. "Normalization of cell responses in cat striate cortex". In: *Visual neuroscience* 9.02 (1992), pp. 181–197.

[34] Horwitz and Hass. "Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing". In: *Nature Neuroscience* 15 (2012), pp. 913–919.

[35] Ian P Howard and Brian J Rogers. *Binocular vision and stereopsis*. Oxford University Press, 1995.

[36] D. Hubel and T. Wiesel. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of Physiology* 195 (1968), pp. 215–243.

[37] D.H. Hubel and T.N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex". In: *Journal of Physiology* 148 (1959), 574–591.

[38] Marko Huterer and Kathleen E Cullen. "Vestibuloocular reflex dynamics during high-frequency and high-acceleration rotations of the head on body in rhesus monkey". In: *Journal of neurophysiology* 88.1 (2002), pp. 13–28.

[39] J.P. Jones and L. Palmer. "The two-dimensional spatial structure of simple receptive fields in cat striate cortex". In: *Journal of Neurophysiology* 56 (1987), pp. 1187–211.

[40] Mitesh K Kapadia et al. "Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys". In: *Neuron* 15.4 (1995), pp. 843–856.

[41] Matthias Kaschube et al. "Universality in the evolution of orientation columns in the visual cortex". In: *science* 330.6007 (2010), pp. 1113–1116.

[42] Steffen Katzner, Laura Busse, and Matteo Carandini. "GABAA inhibition controls response gain in visual cortex". In: *The Journal of neuroscience* 31.16 (2011), pp. 5931–5941.

[43] Norbert Kruger et al. “Deep hierarchies in the primate visual cortex: What can we learn for computer vision?” In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.8 (2013), pp. 1847–1871.

[44] Stephen W Kuffler. “Discharge patterns and functional organization of mammalian retina”. In: *Journal of neurophysiology* 16.1 (1953), pp. 37–68.

[45] Victor Lamme. “Beyond the classical receptive field: Contextual modulation of V1 Responses”. In: *The Visual Neurosciences*. Ed. by L.M. Chalupa and J.S. Werner. MIT Press, 2004, pp. 720–732.

[46] R John Leigh and David S Zee. *The neurology of eye movements*. Oxford University Press, 1999.

[47] Hengyu Li et al. “Design and Control of 3-DoF Spherical Parallel Mechanism Robot Eyes Inspired by the Binocular Vestibule-ocular Reflex”. In: *Journal of Intelligent & Robotic Systems* 78.3-4 (2015), pp. 425–441.

[48] Zhaoping Li. “A neural model of contour integration in the primary visual cortex”. In: *Neural computation* 10.4 (1998), pp. 903–940.

[49] Zhaoping Li. “A saliency map in primary visual cortex”. In: *Trends in cognitive sciences* 6.1 (2002), pp. 9–16.

[50] Margaret S Livingstone and David H Hubel. “Anatomy and physiology of a color system in the primate visual cortex”. In: *J Neurosci* 4.1 (1984), pp. 309–356.

[51] Siegrid Lowel and Wolf Singer. “Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity”. In: *Science* 255.5041 (1992), pp. 209–212.

[52] Valerio Mante and Matteo Carandini. “Mapping of stimulus energy in primary visual cortex”. In: *Journal of neurophysiology* 94.1 (2005), pp. 788–798.

[53] Marjus Meister and Marc Tessier-Lavinge. “Low-Level Visual Processing: The Retina”. In: *Principles of Neural Science*. Ed. by E. Kandel, Jessel, and Schwartz. McGraw Hill, 2013, pp. 576–600.

[54] William H Merigan and Laurence M Katz. “Spatial resolution across the macaque retina”. In: *Vision research* 30.7 (1990), pp. 985–991.

[55] Vernon Mountcastle. “An organizing principle for cerebral function: The unit model and the distributed system”. In: (1978).

[56] Vernon B Mountcastle. “The columnar organization of the neocortex”. In: *Brain* 120.4 (1997), pp. 701–722.

[57] J Anthony Movshon, Ian D Thompson, and David J Tolhurst. “Spatial summation in the receptive fields of simple cells in the cat’s striate cortex.” In: *The Journal of physiology* 283.1 (1978), pp. 53–77.

[58] Ian Nauhaus and Kristina J Nielsen. “Building maps from maps in primary visual cortex”. In: *Current opinion in neurobiology* 24 (2014), pp. 1–6.

[59] Ian Nauhaus et al. “Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex”. In: *Nature neuroscience* 15.12 (2012), pp. 1683–1690.

[60] Klaus Obermayer and Gary G Blasdel. "Geometry of orientation and ocular dominance columns in monkey striate cortex". In: *The Journal of neuroscience* 13.10 (1993), pp. 4114–4129.

[61] Hirofumi Ozeki et al. "Inhibitory stabilization of the cortical network underlies visual surround suppression". In: *Neuron* 62.4 (2009), pp. 578–592.

[62] William H Press et al. *Numerical recipes in C*. Vol. 2. Cambridge university press Cambridge, 1996.

[63] Nicholas J Priebe and David Ferster. "Inhibition, spike threshold, and stimulus selectivity in primary visual cortex". In: *Neuron* 57.4 (2008), pp. 482–497.

[64] Nicholas J Priebe et al. "The contribution of spike threshold to the dichotomy of cortical simple and complex cells". In: *Nature neuroscience* 7.10 (2004), pp. 1113–1122.

[65] R Quian Quiroga et al. "Invariant visual representation by single neurons in the human brain". In: *Nature* 435.7045 (2005), pp. 1102–1107.

[66] John H Reynolds and David J Heeger. "The normalization model of attention". In: *Neuron* 61.2 (2009), pp. 168–185.

[67] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. "The Stabilized Supralinear Network: A Unifying Circuit Motif Underlying Multi-Input Integration in Sensory Cortex". In: *Neuron* 85.2 (2015), pp. 402–417.

[68] Seyed Omid Sadat Rezai. "A Neurocomputational Model of Smooth Pursuit Control to Interact with the Real World". MA thesis. University of Waterloo, 2014.

[69] James Schwartz, Ben Barres, and James Goldman. "The Cells of the Nervous System". In: *Principles of Neural Science*. Ed. by E. Kandel, Jessel, and Schwartz. McGraw Hill, 2013, pp. 71–97.

[70] F Sengpiel, Arjune Sen, and C Blakemore. "Characteristics of surround inhibition in cat area 17". In: *Experimental Brain Research* 116.2 (1997), pp. 216–228.

[71] Bernt C Skottun et al. "Classifying simple and complex cells on the basis of response modulation". In: *Vision research* 31.7 (1991), pp. 1078–1086.

[72] David C Somers et al. "A local circuit approach to understanding integration of long-range inputs in primary visual cortex." In: *Cerebral Cortex* 8.3 (1998), pp. 204–217.

[73] Michael W Spratling. "Predictive coding as a model of response properties in cortical area V1". In: *The Journal of Neuroscience* 30.9 (2010), pp. 3531–3543.

[74] N. Spruston. "Pyramidal neuron". In: 4.5 (2009). revision 130430, p. 6130.

[75] Misha V Tsodyks et al. "Paradoxical effects of external modulation of inhibitory interneurons". In: *The Journal of neuroscience* 17.11 (1997), pp. 4382–4388.

[76] D. Van Essen. "Organization of visual areas in macaque and human cerebral cortex". In: *The Visual Neurosciences*. Ed. by L.M. Chalupa and J.S. Werner. MIT Press, 2004, 507–521.

[77] T. Villgrattner and H. Ulbrich. "Design and Control of a Compact High-Dynamic Camera-Orientation System". In: *Mechatronics, IEEE/ASME Transactions on* 16.2 (2011), pp. 221–231. ISSN: 1083-4435. DOI: 10.1109/TMECH.2009.2039223.

[78] T. Villgrattner and H. Ulbrich. "Optimization and dynamic simulation of a parallel three degree-of-freedom camera orientation system". In: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. 2010, pp. 2829–2836. DOI: 10.1109/IROS.2010.5649720.

[79] Thomas Villgrattner. "Design and control of compact high dynamic camera orientation systems". PhD thesis. Technische Universität München, 2010.

[80] Heinz Wässle et al. "Retinal ganglion cell density and cortical magnification factor in the primate". In: *Vision research* 30.11 (1990), pp. 1897–1911.

[81] Ben S Webb et al. "Early and late mechanisms of surround suppression in striate cortex of macaque". In: *The Journal of Neuroscience* 25.50 (2005), pp. 11666–11675.

[82] Stuart P Wilson and James A Bednar. "What, if anything, are topological maps for?" In: *Developmental neurobiology* 75.6 (2015), pp. 667–681.