Models of Primate Supraretinal Visual Representations

Bedeho Mesghina Wolde Mender

Wadham College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary 2014

Abstract

The work in this thesis was an investigation into a set of non-classical visual receptive field properties of neurons observed in the primate brain. Two major phenomena were investigated. The first was neurons with head-centered visual receptive fields, which remain stationary in head-centered space across different eye positions. The second was a range of neuronal responses associated with the advent of a saccade that relocates a stimulus either into or out of the receptive field of an eye-centered visual neuron.

The two phenomena are independent, but both are considered to be involved in supporting perceptual stability and guiding spatially accurate eye movements. They have been proposed to achieve this in different ways, but both fundamentally rely on integrating visual and oculomotor information, and are therefore examples of what is called sensorimotor integration. The experimental tractability of the oculomotor system has led to a great deal of both empirical and theoretical work on such neural representations. This thesis attempts to extend such work by studying unsupervised neural network models of these representations, a topic which has received relatively little attention in the literature so far.

Chapters 1 to 5 cover the first of the two phenomena, the development of neurons with head-centered visual receptive fields.

Chapter 1 begins with a review of the conceptual, behavioural, physiological and theoretical foundations of head-centered visual representations.

Chapter 2 presents the hypothesis for the operation of a model of head-centered visual representations. The chapter then presents a neural network model based on this hypothesis, along with a methodological description of the network architecture, neuronal dynamics and input stimuli of the model.

Chapter 3 explores the behaviour of the model and the necessity of each component in the model. The model consisted of an input population of eye-centered visual neurons with eye position gain fields feeding into a competitive output population. In this chapter, input neurons had peaked eye position gain fields. Output neurons implemented a local associative learning rule which promoted binding of temporally proximal input patterns.

The model was required to make sequences of saccades while stimuli remained stationary in head-centered space. This, when combined with the learning rule, resulted in the development of output neurons which responded to the same location in head-centered space across different eye positions, i.e. neurons with head-centered receptive fields.

A set of follow up experiments were conducted to confirm the necessity of different model components. First, the visual and eye position components of the receptive fields of the input neurons were decoupled, leading to a collapse of the model performance and no head-centered neurons emerging. Second, the competition was eliminated in the output layer, leading to the development of output neurons with a very low selectivity in terms of receptive field size in head-centered space. Third, the synaptic learning rule was modified to decrease its efficacy in binding together temporally proximal input patterns, leading to a collapse of the model performance and no head-centered neurons emerging. Fourth, the relative frequency of eye movements compared to relocations of stimuli in head-centered space was reduced. This also lead to an initial decline, and eventual absence, of head-centered receptive fields in the output population. Model behaviour as a function of a handful of key parameters was also explored. Some of this work has been published in Mender & Stringer (2013).

Chapter 4 presented a set of experiments testing how well the model behaves when the training stimuli, and the model itself, better approximated natural training conditions.

The number of head-centered training locations within which the visual stimuli could be located during training was increased to approximate the continuous sampling of head-centered space during development, and this improved model performance.

The receptive field size of input neurons, in both the eye position and retinal space, was increased. This lead to a collapse in model performance due to interaction between the edges of the neural input representation and the output representation of head-centered spatial locations. The failure was ameliorated by altering the synaptic and neuronal dynamics.

Random dynamics in the training stimuli were increased to confirm the robustness of the learning process. The model performed well even for relatively large amounts of noise in the training dynamics. Long fixations were found to overwrite past learning, and therefore undermining model performance. This was ameliorated by explicitly bounding the maximum weight of individual synapses.

The eye position component or gain fields of input neurons were altered to have two peaks of varying random size in eye position space, and model performance remained qualitatively invariant.

Training was altered to involve multiple visual stimuli visible at all times, approximating scenes with multiple salient visual targets observed during development. The model still developed confined head-centered receptive fields to individual spatial locations due to the statistical decoupling between them. However, when tested on multiple visual targets simultaneously, the competitive interaction between the output representations of different stimuli led to the suppression of some otherwise well behaved output neurons.

In total, these findings showed that the model is robust to several important modifications approximating more ecologically realistic training conditions.

Some of this work has been published in Mender & Stringer (2014).

Chapter 5 presents a model with monotonic eye position gain fields among input neurons, and the consequences of this new form of input representation was explored. The gain modulation was in the form of a sigmoid curve in the eye position space, substituting the Gaussian curve used in previous experiments.

A manually hardwired model was first tested and shown to exhibit head-centered receptive fields, confirming the existence of a suitable synaptic weight structure for the given network architecture.

A model was then trained as in chapter 3, but with the visual input neurons modulated by monotonic eye position gain fields. This led to a dramatic failure of the learning process. Neurons were found to become more eye-centered, and less head-centered, with training. Weight vectors showed that with monotonic gain fields it was not possible to bind together input patterns corresponding to different fixation positions for the same head-centered stimulus location.

Critically, even the manually prewired network model collapsed in the same way when neural plasticity was introduced. This showed that, independently of whatever neurodevelopmental process established the working neural circuitry for the head-centered receptive fields, any remaining level of synaptic plasticity would undermine the functioning of the circuit.

It was, therefore, explored how well the model could perform when peaked and monotonic eye position gain fields were mixed at various rates in the input population. It was found that approximately 15% of the output neurons remained head-centered after training as long as less than one in five input neurons had monotonic gain fields.

Lastly, previously published models of coordinate transformation relying similar gain fields were replicated for comparison. These previous models relied on various forms of supervised learning, including biologically implausible error correction learning, with an external training signal to guide the activity of the output neurons during development. However, apart from the failure of these earlier modelling studies to adequately explain where such training signals may arise from in the brain, the performance of the trained models degrades quickly if the synapses are subsequently further modified by associative plasticity driven by feedforward visual signals in the absence of the supervisory training signals.

Chapters 6 to 8 cover the second of the above two phenomena, that is, the range of neuronal responses associated with the advent of a saccade that shifts a stimulus either into or out of the receptive field of an eye-centred visual neuron.

Chapter 6 presents a review of the neurophysiology and previous modelling studies of perisaccadic receptive field dynamics.

Chapter 7 presents a model of such perisaccadic receptive field dynamics, along with a description of the network architecture, neuronal dynamics, input stimuli, and testing procedures.

Chapter 8 explores the behaviour of the model. The main phenomenon the model was intended to explain was remapping, which refers to the response of a neuron around the time of a saccade due to the relocation of its receptive field to a head-centered location

where a visual stimulus was presented.

The model consisted of four neuronal populations. There were two input populations, one encoding the retinal location of visual stimuli and the other encoding the impending saccade. Both input populations had a given postsaccadic response latency before their activity was updated after a saccade to reflect the new retinal stimulus location. Both input populations projected, with diluted connectivity, to a competitive population, called the combination population. The combination population further projected onto a remapping population, which was where the remapping activity took place.

The learning in the model was hypothesized to work as follows. Before the saccade, the presaccadic retinal location of a stimulus would be encoded by the visual input population and the impending saccade would be encoded by the saccade input population. The combination of these inputs would cause neurons in the combination population to fire. The saccade would relocate the stimulus, assumed to be stationary in a head-centered reference frame, to a new retinal location. This location would be encoded by the remapping population after saccade completion. Due to the postsaccadic latency in both the visual and saccade input populations, the combination population would remain invariant for some period after saccade completion. This would cause the projections from the combination population to the remapping population to be modified by supervised pattern association learning. After this learning process, the combination population would always excite the neurons in the remapping population corresponding to the postsaccadic retinal location of the stimulus.

A model was first manually hardwired by assigning a unique combination of presaccadic stimulus location and saccade to each combination neuron, and setting up its afferent and efferent weights to reflect this. This led to correct remapping, and confirmed that a sufficient synaptic weight matrix existed.

A model with learning was able to do stimulus remapping by implementing the visually-guided learning mechanism described to update the synapses from the combination to the remapping population.

The impact of reducing the saccade aligned response latency in the visual and saccade input populations was explored. Reducing the period of latency in the input populations was found to reduce, and eventually eliminate, remapping. This was because it undermined the postsaccadic associative learning between the combination and remapping populations.

The model was then tested in trials where the stimulus was flashed at various times with respect to saccade onset, and also in either the presaccadic or postsaccadic retinal locations of the remapping neuron's receptive field. The response of the neuron was found to decrease as the stimulus onset time occurred later with respect to the saccade when the stimulus was flashed in the presaccadic location of the receptive field. Conversely, the response of the neuron was found to increase as the stimulus onset time occurred later with respect to the saccade when the stimulus was flashed in the postsaccadic location of the receptive field. These findings were in strong agreement with experimental observations.

The model was finally trained to remap multiple presynaptic stimulus locations into the receptive field of a given remapping neuron in order to investigate the spatial properties of remapping. Single neurons remapped from multiple presaccadic stimulus locations. There

was varying selectivity in the remapping strength of neurons across the multiple presaccadic locations, and neurons became more selective with decreasing connectivity between the combination population and remapping population. Activity could be remapped from the entire retina at the population level. Neurons remapping only a single location were not more responsive than those remapping multiple locations. Neurons with receptive fields across the visual field had comparable selectivity and remapping strength. These findings were also in strong agreement with experimental observations.

Some of this work has been published in Mender & Stringer (2015).

Chapter 9 presents a unified summary of the thesis and ends with some remarks on future work.

Models of Primate Supraretinal Visual Representations

Bedeho Mesghina Wolde Mender
Wadham College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2014

Acknowledgements

I would like to thank a number of people who have supported me during my work on this thesis.

First Id like to thank my supervisor Dr. Simon Stringer. You welcomed me from the moment we first got in touch, and along the way your unwavering optimism and encouragement always throttled me forward. More than any scientific finding, how wonderfully you treated all of us in the lab - more as peers and friends than students, is something I will never forget and will aspire to live up to myself. Also, thank you for arranging to have the Oxford Foundation for Theoretical Neuroscience and Artificial Intelligence support me during my research.

Likewise, I would like to thank the Norway Oxford Society for their very generous support through the Norway Scholarship.

I would also like to thank Dr. Mark Buckley for agreeing to be my joint supervisor.

A huge thanks goes to everyone in the lab: Ben, Dan, James, Juan, Irina, Hector, Akihiro, Erica, Harry, Julie and Mihaela. You all made it a pleasure to spend those countless hours in the lab. In particular, I want to thank you Ben for making our computer system run as well as it could, and for our many enjoyable late night contemplations about life and the future ahead. I'm still very optimistic on behalf of us both! Thank you to James for welcoming me to the lab when I first got here, and for our many OUC curry dinners, Giraffe breakfasts and Byron burgers. Your patience with my tardiness was on many occasions excessively gracious. Thank you Dan for always making your skills and support available to me and everyone else, your unparalleled generosity was always an inspiration. I could always count on you to make my weekends in the lab more enjoyable. Lastly, thank you Juan for putting up with my uncountably many foolish claims, and otherwise engaging in many of the most enjoyable conversational moments I can recall during my stay here in the city of dreaming spires.

A big thanks also goes to Richard Rormark for spending countless hours with me on Skype, talking about everything from the UFC to Feathercoin. I still maintain it is all about the community! Thank you to Sindre Aarsaether, speaking to you always made me optimistic about my post PhD future, and having you visit me here in Oxford was decidedly non-Lebanese!

However, no single person has been as instrumental in my life during this time, and to my ability to complete this work, as my dear Miss Camilla Rake, the love of my life. Your belief in me always outpaced my own, and for that I thank you. I so much look forward to coming home to you at this moment of writing.

Lastly, thank you to my family, you are the bedrock of everything I do and am, and I dedicate this work to all of you.

Abstract

This thesis investigates a set of non-classical visual receptive field properties observed in the primate brain. Two main phenomena were explored. The first phenomenon was neurons with head-centered visual receptive fields, in which a neuron responds maximally to a visual stimulus in the same head-centered location across all eye positions. The second phenomenon was perisaccadic receptive field dynamics, which involves a range of experimentally observed response behaviours of an eye-centered neuron associated with the advent of a saccade that relocates the neuron's receptive field. For each of these two phenomena, a hypothesis was proposed for how a neural circuit with a suitable initial architecture and synaptic learning rules could, when subjected to visually-guided training, develop the receptive field properties in question. Corresponding neural network models were first trained as hypothesized, and subsequently tested in conditions similar to experimental tasks used to interrogate the physiology of the relevant primate neural circuits. The behaviour of the models was compared to neurophysiological observations as a metric for their explanatory power. In both cases the neural network models were in broad agreement with experimental observations, and the operation of these models was studied to shed light on the neural processing behind these neural phenomena in the brain.

Contents

List of Figures

List of Tables

Chapter 1

Head-Centered Receptive Fields: Introduction

1.1 Introduction

The first part of this thesis is concerned with the development of neural network models of visual neurons in the primate parietal cortex that represent the locations of visual targets in a head-centered frame of reference. This chapter presents an introductory review of the conceptual, behavioural, physiological and theoretical work upon which the subsequent models are based.

The concept of a reference frame and coordinate system are discussed in the context of neural responses, and some well known examples of neural responses and their reference frames are presented. The potential behavioural significance of the reference frames of neural responses are discussed, and physiological and neuropsychological evidence supporting the relevance and potential role of such responses are presented. The main physiological, theoretical and modelling results relevant to later chapters on head-centered receptive fields

are reviewed.

1.2 Reference Frames

A *reference frame* is an object used to specify the relative linear and angular position of other bodies, and a position is given in terms of a particular set of basis vectors referred to as a *coordinate system*. In neural systems in the primate brain there exist a wide variety of reference frames, within which sensory and motor parameters are encoded (Andersen et al. 1997). The reference frames often arise as a result of the structure of the underlying receptor and effector driving the neural responses (Soechting & Flanders 1992).

A prevalent reference frame in primate visual neuroscience is the *head-centered* reference frame. The fixation position, i.e. position of the eyes in the head, is often given in this reference frame. This reference frame has a two dimensional spherical coordinate system that is centered on the head. It remains in fixed register with the orientation of the head. Positions in this coordinate system are specified by angles in the saggital and horizontal plane centered on the head, which are often referred to as the vertical and horizontal head-centered location respectively.

However, perhaps the most familiar reference frame in primate visual neuroscience is the *retinocentric* or *eye-centered reference frame*. The retinal locations of visual targets are often represented in this reference frame. It is similar to the head-centered reference frame, except that the coordinate system is always in fixed register with the retina and hence

fixation position as described above. This effectively means that eye-centered locations are given in terms of their offset from head-centered fixation position.

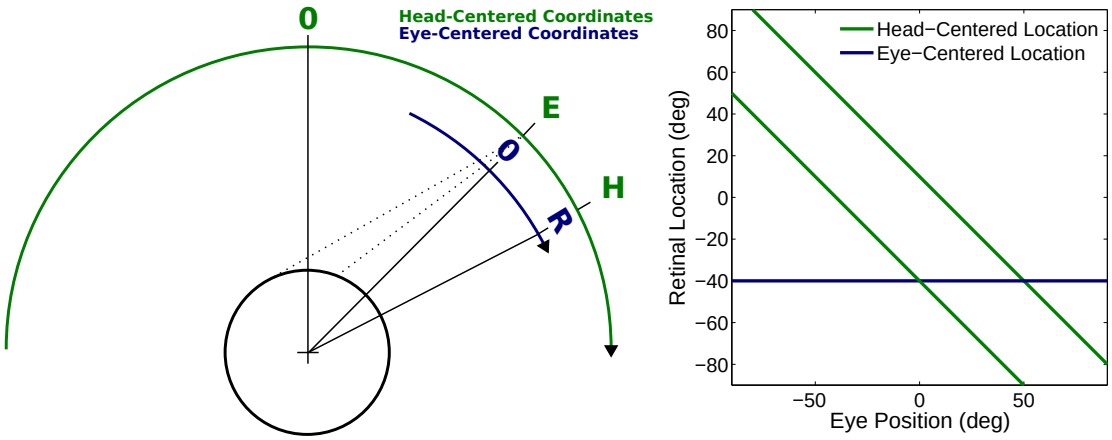
Figure 1.1a shows a one dimensional version of these reference frames. By definition, and as the figure shows, a given eye position E and target location in the eye-centered reference frame R may be related to a corresponding target location in the head-centered reference frame H by the relationship

$$H = E + R \tag{1.1}$$

It can be seen that a particular head-centered target location H thus corresponds to many possible pairs of eye position E and eye-centered target location R satisfying equation 1.1, which is also shown in figure 1.1b.

As an example, some eye-centered visual neurons have response functions that are gain modulated by eye position.

The response function of a sensory neuron in a given reference frame refers to the firing rate of the neuron as a function of stimulus location in the given reference frame. A sensory neuron is said to be compatible with a given reference frame if the response function within this reference frame remains invariant to changes in any secondary sensory or motor variable. Invariance typically refers to the stability of the locations of peaks and troughs in the response function, while absolute discharge magnitudes may or may not change. A population of neurons with responses in the same reference frame may be referred to as a



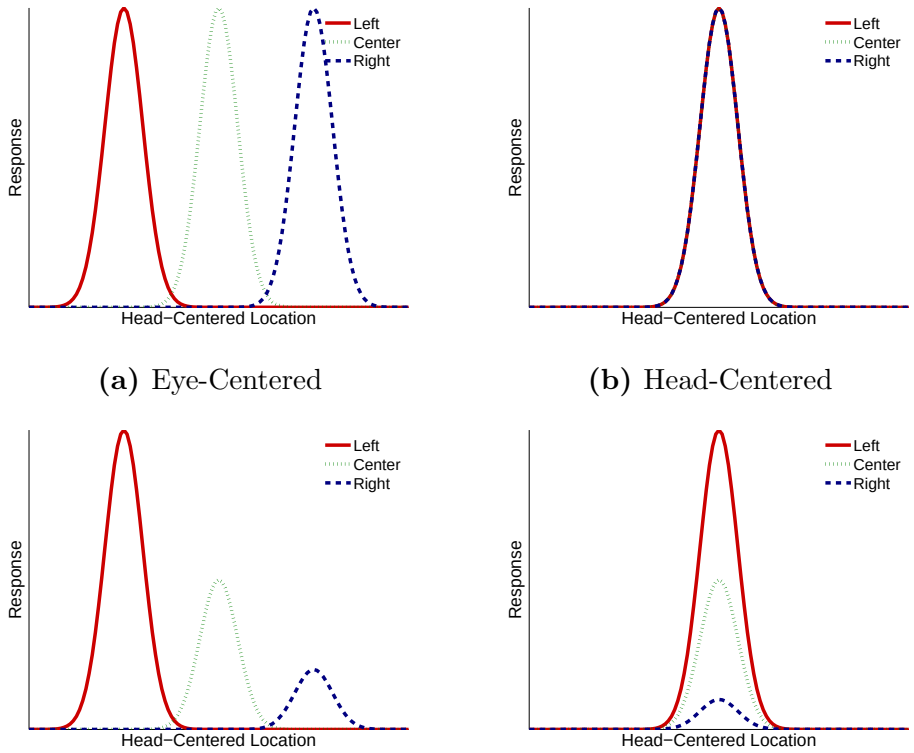
(a) Coordinate Systems **(b)** Reference Frame Locations

Figure 1.1: Plot (a) shows how the head- and eye-centered coordinate systems are related. The dark central circle is the head, and the two dashed lines show the line of sight for each eye, and they converge on location E in the head-centered coordinate system represented by the large green semicircle. The blue curved axis fixed to E is the eye-centered coordinate system, and location H in head-centered space is projected onto the eye-centered retinal location R. Plot (b) shows how the retinal location and eye position are related for different locations in the head-centered and eye-centered reference frames. The two green diagonals are for distinct head-centered locations, the upper one being $10°$ and the lower one being $-40°$. The horizontal blue line is the retinal location $-40°$.

representation in the given reference frame. In particular, a neuron is referred to as *gain modulated* in terms of a secondary variable when the amplitude of the response function depends on this variable (Salinas & Thier 2000). Often this modulation takes the form of some modulation function of the secondary variable interacting with a sensory response function in the given reference frame, in which case the the former function is referred to as a *gain field* (Salinas & Thier 2000).

Considering some example neuronal responses may help clarify the concept of the reference frames of a neuronal response. A visual neuron which has an eye-centered response function invariant across eye positions is referred to as *eye-centered* or *retinocentric*. This

is the natural reference frame for a visual response, given that retinal receptors are fixed to the eye, and so discharge to a given eye-centered location would not be expected to be influenced by changes in other variables. A visual neuron which has a head-centered response function invariant across eye positions is referred to as *head-centered* or *craneocentric*. This is not a natural reference frame for a visual response. Because the same head-centered location falls on different retinal locations for different eye positions, response invariance in head-centered space requires different retinal receptors to maximally drive neuronal discharge for different eye positions. An auditory neuron which has an eye-centered response function invariant across eye positions is referred to as eye-centered, and this is also not a natural reference frame for an auditory neuron. Because the same eye-centered location has different cochlear effects for different eye positions, response invariance requires taking this into account. Such neurons have indeed been identified in the macaque lateral intraparietal area (Cohen & Andersen 2000). In the visual domain, reference frames beyond the eye-centered reference frame, such as head-centered, arm-centered or body-centered reference frames, are referred to as *supraretinal* reference frames (Molotchnikoff & Ono 1993). Figure 1.2 gives examples of responses that are eye-centered, head-centered, eye-centered with eye position gain modulation, and head-centered with eye position gain modulation.



(a) Eye-Centered **(b)** Head-Centered

(c) Eye-Centered Gain Modulation **(d)** Head-Centered Gain Modulation

Figure 1.2: Typical response functions in eye-centered and head centered reference frames, with and without eye position gain modulation. All plots show the response of a neuron to a visual target located at 0° in head-centered space at three different eye positions, namely −30° (left),0° (center) and 30° (right). The following types of response function are shown: (a) eye-centered, (b) head-centered, (c) eye-centered with eye position gain modulation, and (d) head-centered with eye position gain modulation.

1.3 Behaviour

Since the location of a visual target is initially represented in a retinal reference frame, later stages of processing in the brain must carry out some form of coordinate transformation in order to recode this information into a head-centered reference frame. Logically, this transformation must involve the use of an eye position signal. This type of neural integration of sensory and motor signals is referred to as *sensorimotor integration*. The general

problem in sensorimotor integration is that sensory receptors, e.g in the retina, cochlea or soma, respond in different coordinate systems to the motor signal with which they are to be integrated and to the the motor effectors they may drive. For example, you may want to reach out in order to grasp a cup of coffee in the periphery of your visual field while reading a newspaper. This will require the transformation of an eye-centered visual signal into a representation suitable for guiding your limbs and posture (Andersen 1995). It has been suggested that such a transformation could happen in multiple stages, with each of which transforming the representation of the preceding stage into a new reference frame by integrating it with the appropriate extraretinal signal (Fig. 1.3). From this perspective there would be a range of distinct and complementary reference frames integrating sensory and motor variables. Sensory stimuli would be represented in a number of reference frames, each of which would be suitable for the particular subsequent processing or behaviour they support (Colby & Goldberg 1999).

The unity of visual perception and action in space suggests that there has to be a unified neural representation of space. This might seem incompatible with the existence of a multitude of different neural reference frames. However, the extraordinary physiological heterogeneity of sensorimotor areas like the posterior parietal cortex (PPC), as well as the specificity of neuropsychological impairments due to localised functional lesions, are consistent with the existence of a number of different reference frames situated in distinct brain areas (Colby & Goldberg 1999). A number of neuropsychological findings in human

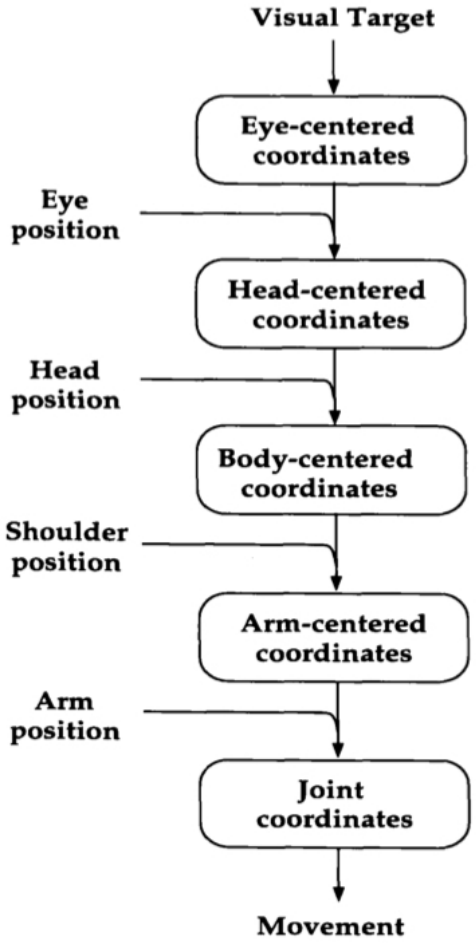


Figure 1.3: Hierarchy of neural representations integrating extraretinal motor variables, from Pouget & Sejnowski (1997).

inferior parietal lobule (IPL) support the hypothesis that space is represented by distinct neural representations with different action and sensory-oriented reference frames. Lesions in human IPL have been found to cause a variety of neglect behaviours where subjects, who are otherwise visually unimpaired, are unable to properly attend to the contralesional hemispace or lateral body half (Andersen 1987). Some subjects are unable to render the contralesional half of a visual scene that they are attempting to draw, which is called constructional apraxia. While others make mistakes when reaching or pointing to visual

targets. These very specific deficits, in the absence of general sensory or motor impairments, would be predicted as a result of lesioning specific neural representations with a narrow functional role in sensorimotor behaviour.

It has been suggested that the head-centered reference frame, in particular, may be involved in at least two other neural processes in the primate. First, a head-centered reference frame may contribute to the apparent stability of visual perception, in spite of frequent and rapid eye movements. Such perceptual stability may be mediated through visual responses in more stable supraretinal reference frames (Andersen 1989), a phenomenon that is also known as spatial constancy. Since responses in a head-centered reference frame are invariant to eye-position, they could provide a more stable visual experience. Secondly, a head-centered reference frame may support the execution of delayed double-step saccade tasks, where a subject is required to make an initial saccade to a visible target, and then execute a second saccade to another target that was only observed prior to the onset of the first saccade (Robinson 1975). The memory trace of the final saccade target will be incorrectly encoded after the first saccade if saccade targets are encoded in an eye-centered reference frame. However, the intervening saccade does not compromise the correctness of the representation if the targets are encoded in a head-centered reference frame that is independent of eye position. Such head-centered representations could, therefore, be used to drive the final saccade.

1.4 Physiology

Eye-Position Modulation

Andersen & Mountcastle (1983)

The earliest demonstration of the impact of gaze angle was by Andersen & Mountcastle (1983). They described how visual neurons, in a new area of the inferior parietal lobule called area 7a, were strongly influenced by the angle of gaze during attentive fixation. They also attempted to show how gaze angle had little influence during inattentive fixations. After a neuron had been isolated and confirmed to be light sensitive, they attempted to study it in both of the following tasks. The attentive task required the monkey to maintain fixation at some angle of gaze while a peripheral stimuli was flashed in the same eye-centered location. During this task the monkey was considered to be attentively fixating. The inattentive task, where the monkey freely oriented gaze while targets where flashed at random times and locations on the screen, was considered inattentive.

During the attentive task, 61% of examined neurons showed significant and frequently strong changes in their response to visual stimuli as the angle of gaze was varied. Among these neurons, responses could increase by more than 3 times when the gaze angle was shifted by 20° in the optimal direction. During the inattentive task, 10% of examined neurons showed significant change in their responses to visual stimuli as the angle of gaze was varied. Among neurons for which both tasks could be completed, none showed a gaze

effect in the inattentive task. There was no relationship between the hemisphere of the neuron and the direction of optimal gaze. The authors concluded that these modulation effects were not simply due to either the eye position or reduced strength of this effect in the inattentive task, but due to a convergence of an efference copy signal for eye movements and the pathways controlling excitability of parietal cortex.

Andersen et al. (1985)

The demonstration that light-sensitive 7a neurons where modulated by gaze led Andersen et al. (1985) to attempt to precisely characterize the spatial structure of this modulation. They hoped that the presence of gaze modulation meant that a head-centered encoding would exist in this area, and they hypothesized it would be useful for both visual stability across eye movements and motor planning and execution. The receptive fields of isolated neurons were initially mapped by requiring the head-restrained monkey to fixate straight ahead while a stimulus was presented systematically across a tangent screen. The responses allowed the localization of the most excitatory eye-centered location for the given neuron. In a subsequent task, the visual stimulus was maintained in this eye centered location while the animal had to fixate nine locations organized into a 3×3 grid. This allowed a systematic examination of the influence of eye position on the responses of the neuron.

Among neurons for which eye position effects were mapped, a regression analysis was conducted, and for 39% of these neurons the eye position effect was planar. For 38% there was a significant planar component, but a significant lack of fit which was mainly caused

by a single bump in the fit. The remaining 23% had no planar component, which was in general due to a peaked response in the central eye position. In order to assess how the modulation affected the entire receptive field, and not just the center of it, a subset of neurons had their receptive fields re-examined both at the most and least preferred eye position. This revealed that for any retinal location, the eye-position modulation from the least to the most preferred eye position was proportional to response at the least preferred eye position, in other words it was a multiplicative interaction. The response of these neurons could be described as the product between a two dimensional Gaussian visual receptive field and a two dimensional planar eye position modulation. Lastly the modulation was mapped for a subset of neurons in both the light and dark, and it remained unchanged across these two conditions.

The investigators expected to find an explicit head-centered representation at the single neuron level without an eye-position dependency. Therefore, they attempted to explain the absence of such head-centered neurons, and the potential function of the representation they did find, in terms of an implicit distributed encoding of head-centered target location. Specifically, they argued that an explicit representation of head-centered space may not be necessary since full information about the head-centered location of a visual target was already implicitly encoded by the neuronal responses actually observed in area 7a. However, this argument could also be applied to render the neuronal representations observed in area 7a redundant as well, as they also simply integrated existing information that was

distributed across preceding cortical areas. In fact, the argument could be applied to a variety of sensory processing pathways, since all relevant information must essentially exist in some form already at the receptor level. Indeed, later studies would find that explicit head-centered representations did indeed exist at the single neuron level in multiple brain areas.

Andersen, Bracewell, Barash, Gnadt & Fogassi (1990)

In a later study a more systematic analysis of gain modulation in both area 7a and the recently discovered area LIP (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990) was conducted separately for the different response periods at the single neuron level. This was the first work to suggest a possible role of LIP in sensorimotor integration. The optimal saccade direction for each neuron was initially mapped in a memory guided saccade task which consisted of four periods. The first period was the baseline period (B) where central fixation was maintained for either 800ms or 300ms. The second period was the light sensitive period (LS) where a peripheral saccade target was presented for 300ms. The third period was the memory period (M) were the saccade target was extinguished and fixation had to be kept for either 400ms or 700ms. The fourth saccade period (S) started when the fixation light was extinguished and a saccade to the remembered target was made. Once the optimal saccade direction was determined, and found to be the same across the different periods, the monkey was required to perform a memory guided saccade task. During the task, the monkey had to perform sacccades from different gaze angles

organized in a 3×3 grid, while the saccade target was always positioned in the optimal direction. This allowed the gain field to be mapped.

A majority of the cells showed significant effects of eye position for each period, and gain fields were usually planar or had a planar component for all periods in these cases. At the single neuron level, the gain fields were typically in agreement across all periods. Gain fields in LIP were found to have larger gradients and intercepts than in 7a. The authors concluded that the agreement across the different periods, and across two cortical areas, meant that such gain fields were an important phenomenon. They reiterated their contention that the lack of a single unit encoding of head-centered space meant that a distributed encoding was more likely.

Brotchie et al. (1995)

Prior to Brotchie et al. (1995), no physiological work on gain fields in the visual system had attempted to identify the presence of gain fields of a higher order than eye position modulation, for example, gain modulation by head position on the trunk. In this paper, the authors confirmed the existance of such higher order gain fields in the posterior parietal cortex, predominantly in LIP, and also showed that some eye position gain fields were actually gaze direction gain fields. Gaze direction refers to the direction with respect to the body resulting from combining the eye position with head position on the trunk.

A memory guided saccade task was used to reveal head position modulation. The monkey was required to fixate straight ahead at two different head positions. In each

case, peripheral stimuli appeared in the same set of eye-centered locations, and a saccade was made to the stimuli. This allowed the authors to investigate the effect of varying head position on the neuronal responses to visual stimuli in the same eye-centered locations with saccades made to those same locations. The visual response, as measured in a presaccadic time window, and the saccade response, measured in a postsaccadic time window, were analyzed separately.

They found that neurons had eye-centered responses that were modulated by head position, with amplification of the response in the contralateral direction. Modulation by head position of either visual or saccadic activity was found in 40% of neurons inspected. While 25% of neurons had only eye position modulation of either their visual or saccade activity.

Two further tasks were used to map the precise modulation of head and eye position independently for all neurons with significant head position modulation. In the first task, the head was oriented in one of five positions while fixating straight ahead in each one, and then a saccade to a visual target in the receptive field was performed. This revealed the head position gain on this one retinal location. In the second task, the head was oriented straight head while fixating each of the previous five head positions, and then a saccade to the same eye-centered visual target was performed. This revealed the eye position gain on the same retinal location.

The overwhelming majority of neurons had linear gain fields for both eyes and head,

as measured by both visual and saccade activity. A comparison of the slope of the two gain fields at the single neuron level found them to be extremely similar, for both visual and saccade activity, which showed that each gain field had the same effect while the other was held constant. This lead the authors to conclude that the gain fields could be best explained as being gaze direction gain fields, rather than purely eye- or head position gain fields. They speculated that the head position gain field may be driven by input signals from either proprioceptive neck muscles, integrated vestibular signals or integrated corollary discharge of head-movement commands. This finding lent further support to the hypothesis that multiple reference frames of visual space are encoded in parietal cortex. Moreover, the study provided the first evidence suggesting the existence of body-centered visual representations of space.

Snyder et al. (1998)

Snyder et al. (1998) studied whether the head position gain fields identified in Brotchie et al. (1995) were in a body-centered or a world-centered reference frame. That is, was the gain modulation of neuronal responses based on the position of the head with respect to the body or the world? The authors also wanted to determine what may be the driving signal of the gain field. They studied neurons from LIP and area 7a.

Two tasks were used to reveal the reference frame of the modulation. In a body-under-head rotation task, the head was world-fixed while the body was rotated passively to achieve a set of head-on-body configurations. For each configuration, the visual response

at a fixed eye-centered location was measured from fixation straight ahead. In a body-plus-head rotation task, the head and body were rotated together to achieve the same set of world-centered head orientations, and the same eye-centered responses were recorded as in the first task. It was found that more than 33% of inspected neurons had gain modulation in at least one reference frame, roughly at the same rate as both LIP and 7a. Almost without exception, LIP neurons had more body-centered gain fields, while 7a neurons had more world-centered gain fields.

To study the source of the gain field in LIP, they recorded neuronal responses for the same eye-centered stimulus location and set of positions of the head on the body, but where the head was actively oriented by the monkey. They found gain fields identical to the body-under-head task, ruling out vestibular signals as a source. Proprioception could obviously drive both gain modulations, and even efference copy signals due to active assistance by the monkey could not be ruled out. The 7a gain fields were clearly not dependant on either proprioceptive or efference copy signals. To confirm the vestibular origin of the signal, a few neurons were tested in a modified body-plus-head task, where rotations were just below and just above the vestibular threshold. In almost all cases, the gain fields only appeared above threshold, which provided strong evidence for the vestibular origin of gain fields in area 7a.

The authors finally speculated that the functional segregation across the two areas, LIP and area 7a, could be related to different computational constraints in the information

processing of the input signal.

Galletti et al. (1995)

The parietal occipital area (PO) was known to have both visual and non-visual neurons (Galletti et al. 1991), however the possible presence of eye position gain modulation effects had not been studied. Therefore, Galletti et al. (1995) attempted to identify the presence and characterize the nature of such an interaction in this area. The first task was performed in the dark, where the monkey was required to fixate nine screen locations organized in a 3×3 grid. The second task was performed for visual neurons, both in light and dark, where the visual receptive field was mapped at the same nine screen fixation locations.

About 48% of visual and 32% of non-visual neurons showed eye position-related activity in total darkness, while in 61% of visual neurons the visual response was modulated by eye position in the orbit. The fact that modulation was confirmed also in the dark was taken as evidence that it was an effect driven by eye position in orbit, not confounded with changes in background scenery due to changes in gaze. Some gain fields were planar, and therefore similar to those identified by Andersen, Bracewell, Barash, Gnadt & Fogassi (1990). However, other gain fields were peaked and confined to restricted regions of the field of view.

The authors speculated that the previously discovered head-centered visual neurons in PO (Galletti et al. 1993) could derive their responses from the identified eye-position sensitive neurons. They suggested that the lack of retinopic organization could be due

to the fact that local populations in PO in fact represented full retinotopic space. The eye-position and retinal preferences of all neurons within a pool could be coordinated so as to produce a downstream real-position neuron for a given head-centered location. Lesions in humans to the superior parietal lobule, where PO is located, produce visual localization and reaching impairments. This, and the PO's prospective role in coding extrapersonal space, led the authors to speculate that PO could be encoding visuo-spatial information for premotor cortex for reaching movements.

Breveglieri et al. (2009)

Breveglieri et al. (2009) conducted a follow up study to examine the prevalence of peaked versus planar eye position gain fields in area V6A, a subdivision of area PO, in which such modulation had been identified (Galletti et al. 1995). They also wished to compare the impact of each type of gain field on the performance of a well known neural network model of sensorimotor transformation developed by Salinas & Abbott (1995). As before, 9 locations organized in a 3×3 grid with 20° spacing had to be fixated, and the visual stimulus was presented only at the fixation point. This form of gain field mapping did not, therefore, take into account the receptive field structure of individual neurons by mapping the entire visual receptive field for each fixation point.

It was found that $\sim 56\%$ of neurons were modulated by eye position. Among neurons for which at least 5 fixation locations had been recorded, it was also found that only $\sim 27\%$ had planar gain fields. Moreover, many of these gain fields were of varying planar

fit. The majority of neurons ($\sim$ 73%) could not be classified as planar, but rather as peak-shaped. Among all neurons with any form of eye position modulation, $\sim$ 37% had a high selectivity for an individual fixation position. Lastly, all fixation positions were evenly preferred across the entire population of gain fields.

The investigators were also interested to know how the peak-shaped gain fields would affect the ability of the model of Salinas & Abbott (1995) to perform a transformation of reference frame from eye-centered to head-centered. Therefore, they trained two versions of the model, where one version used the clipped linear gain fields of area 7a and LIP and the other version used the peaked gain fields of area PO. Both versions of the model were trained over a range of different motor output population sizes. They found that as the population decreased, the models with peak-shaped gain fields gave a more accurate representation of the target in head-centered space compared to the clipped linear gain field models.

Head-Centered Neural Response

Galletti et al. (1993)

The earliest successful work investigating the presence of head-centered, or craneotopic, representations in macaque PO was in Galletti et al. (1993). It was thought that head-centered representations were useful for both perceptual stability and visually guided reaching, and the posterior parietal cortex was considered the best candidate for such neural

processing because of eye position gain modulation identified in 7a (Andersen et al. 1985), LIP (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990) and V3A (Galletti et al. 1991). Therefore, Galletti et al. (1991) hypothesized that single unit representations of head-centered space could possibly exist in area PO based on earlier work which showed both visual and eye-position information to be present in this area (Galletti et al. 1991).

The head-restrained monkey was required to fixate different screen locations. For each of these fixation positions, a visual target was then presented in a number of other screen locations. This permitted the receptive field to be mapped at different fixation positions. A neuron was classified as retinocentric if the screen location of maximal response within each fixation position depended on the fixation position, otherwise it was referred to as a 'real-position cell'.

They found that 11% of inspected neurons had receptive fields not systematically dependent on fixation position. Among these there were six neurons which had the canonical head-centered response that was completely independent of eye position, and they were called "real-position cells". Another four neurons had receptive fields which only depended on eye-position in one cardinal direction, but not the other, and were hence eye-centered and head-centered simultaneously, just in different receptive field dimensions.

The investigators pointed out that previous claims that there were no explicit head-centered representations were incorrect (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990), and that their study further emphasized the potential functional role of such rep-

resentations. They hypothesized that 'real-position cells' could derive their responses by pooling the output of gain modulated neurons, and that the mixed reference frame neurons may be a result of intermediate stages of integration.

Duhamel et al. (1997)

The second area in the PPC within which head-centered representations were identified was the ventral intraparietal area (VIP), an area located in the fundus of the intraparietal sulcus, which was known to have bimodal and congruent visual and tactile receptive fields (Colby & Goldberg 1999). Duhamel et al. (1997) were interested in examining the visual reference frames found in VIP, and in particular, to dissociate eye-centered and head-centered reference frames.

The visual receptive field of individual neurons was initially manually mapped, and the optimal stimulus location was determined. The receptive field was then mapped at 9 different fixation locations organized into a 3×3 grid. For each fixation location, a visual stimulus was flashed in 49 different screen locations organized into a 7×7 grid. Receptive field shifts due to different fixation positions were computed by cross-correlating the response matrices for different fixation positions, and comparing the largest correlation shift with the fixation shift between the two positions. A correlation shift much smaller than the fixation shift indicated a head-centered receptive field, and zero indicated perfect head-centeredness. Likewise, a shift close or identical to the fixation shift indicated eye-centeredness. This analysis was done independently for horizontal and vertical fixation

shifts.

It was found that among the VIP neurons investigated there was a continuum of reference frames, where perfect reference frames where identified at both ends of the spectrum. A clustering routine identified two distinct sub-populations of neurons based on their receptive field shifts in both directions, and the two populations could roughly be classified as being eye- and head-centered. An eye position effect, both on baseline and visual activity, was found in $\sim 53\%$ of neurons, and this effect was planar in $\sim 37\%$ of neurons.

This type of gaze encoding has been found in many other parietal areas, and has been hypothesized to constitute a distributed encoding of head-centered visual space. However, the presence of explicitly head-centered responses at the individual neuron level led investigators to speculate that both distributed and single cell encodings of head-centered space may have functional significance in the parietal cortex.

Mullette-Gillman et al. (2005)

It had been known that LIP and adjoining area 7a had eye-position gain fields for a long time (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990), but it had also been suggested that explicit head-centered visual representations did not exist there (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990). It was not until the work of Mullette-Gillman et al. (2005) that such reference frames were identified in area LIP. They investigated the reference frames of visual and auditory responses in areas LIP and MIP, and found among other things the first examples of purely visual head-centered responses in LIP. Their interest

in reference frames was motivated by the behavioural requirement for proper integration of visual and auditory signals, e.g. sound of speech and sight of moving lips. This would depend on some way of integrating information gathered by receptors in a head-centered and an eye-centered frame of reference, respectively.

A memory guided saccade task was used to physiologically distinguish LIP and MIP, where only the former has a memory response (Gnadt & Andersen 1988). An overlapping saccade task was used for assessing the reference frame of the response. In this task, one of three possible screen fixation locations had to be kept in memory, and after a random delay a secondary stimulus was presented in one of nine screen target locations. The target was either a visual or an auditory stimulus from trial to trial. After another random delay, a fixation light was extinguished and the monkey had to saccade and fixate the target location.

It was found that 72% and 51.4% of neurons were sensitive to visual and auditory target locations, respectively. The proportion of neurons with a target location sensitive response in both modalities ranged from 5% to 43% depending on the definition of responsiveness. An eye position effect was found for the visual and auditory response in 45.5% and 24.3% of neurons, respectively. Among target location modulated visual neurons, 33% were more consistent with an eye-centered reference frame, while 18% were more consistent with a head-centered reference frame. The remaining 49% of neurons could not be classified into either category. Likewise for auditory neurons, 10% had receptive fields more consistent

with an eye-centered reference frame, while 23% had receptive fields more consistent with a head-centered reference frame. The remaining 67% could not be classified into either category. Hence the general finding was that both modalities had both reference frames represented, while the most natural reference frame was the most predominant in each, and most neurons had complex reference frames not admissible to classification.

For bimodal neurons, the reference frame of the response was somewhat similar in the two modalities, although only a weak correlation was found. It was investigated if this weak correlation would still allow a downstream neuron to reliably read out the target location from the total neuron population found in the recording, irrespective of the sensory modality. This was done by training, using an error correction rule, a two layer neural network with a single output neuron, on the responses recorded from visual neurons. The network was tested on the training set and also on auditory responses from recorded neurons. The result was that the output neuron responded linearly in the eye-centered target location, irrespective of the modality of the input population. This was taken as evidence that modality-invariant downstream read out from the population was possible.

This work demonstrated visual head-centered reference frames in LIP for the first time, and also revealed the prevalence of complex reference frames, the functional purpose of which is not yet well understood.

1.5 Models

Zipser & Andersen (1988)

The classic paper of Zipser & Andersen (1988) demonstrated how a supervised backpropagation of error neural network model could be trained on independent eye-position and retinotopic visual inputs to produce head-centered output units. A key finding was that, during training, the hidden units within the intermediate layer developed retinocentric receptive fields with planar eye-position gain fields, similar to those found in area 7a (Andersen et al. 1985).

The input layer units of the network were partitioned into two separate populations, each corresponding to one of the different classes of neurons found in area 7a. The first input population consisted of purely visual neurons with retinocentric Gaussian response curves in two dimensions. The second input population consisted of purely eye-position encoding neurons with linear response curves of either horizontal or vertical eye position. The slopes and intercepts of these linear response curves varied among the units. Specifically, each unit preferred a particular eye position direction as determined by the slope sign, at a given rate as determined by slope magnitude, with given peak response as determined by slope intercept.

The hidden layer units received weighted synaptic connections from all input neurons, and the output layer received weighted synaptic connections from all hidden layer units. The output response of each hidden unit and output unit was computed by summing the

activities of the presynaptic neurons weighted by their corresponding afferent synaptic connection weights, and then passing this value through a sigmoid activation function.

All synaptic connections were initially set to have random weights. Then the network was trained using the back propagation of error algorithm for 1000 pairs of input and desired output patterns. The input was the 2D retinal location of a target $\mathbf{r}$ and the 2D eye position $\mathbf{e}$, while the output was the corresponding head-centered location of the target $\mathbf{h} = \mathbf{r} + \mathbf{e}$. These values were used to impose the correct activity in the input layer. Then the output responses were computed for the hidden and output layers as described. The computed activities of units in the output layer were then compared to the desired head-centered output representation. The desired response of each output neuron was set to a gaussian function of $\mathbf{h}$ centered on a head-centered location to which that individual neuron was assigned. The error between the computed and desired head-centered output patterns was used to adjust the synaptic weights within the network using the backpropagation of error algorithm. The effect of this was that, after training, the outputs successfully encoded the head-centered location of the target.

A key result of the training was the emergence of hidden units with responses similar to those found in area 7a. Hidden neurons displayed broad peaked retinotopic receptive fields, some single peaked and others multi-peaked or with more complex receptive fields. The peaks were distributed across a range of eccentricity on a population level. Perhaps most importantly, the eye-position gain modulation in these units was planar, similar to those

found in monkey 7a. This model constituted, not only the first model of head-centered responses in parietal areas, but also a significant early attempt to model the development of experimentally observed responses through the use of trained neural networks. However, the use of supervised error correction learning was not without some disadvantages, and the significance of this is considered in chapter 5 where the model is also reproduced.

Mazzoni et al. (1991)

In an effort to ameliorate concerns that the results of Zipser & Andersen (1988) were mere artefacts of the biologically implausible backpropagation learning paradigm, Mazzoni et al. (1991) went on to show that a model with a more plausible learning mechanism could replicate their original findings. They trained a stochastic version of the same network architecture described by Zipser & Andersen (1988) using a modified version of the associative reward-penalty rule from Barto (1985), rather than the backpropagation algorithm. This alternative learning procedure did not compute an error between the desired and actual output for each individual output neuron as in back-propagation, but instead computed a global error value, ϵ, for the entire output population. Specifically, for a desired output $d_1, \ldots, d_K$ and actual output $a_1, \ldots, a_K$, the following global error was computed

$$\epsilon = \left\{ \frac{1}{K} \sum_{k=1}^{K} |d_k - a_k| \right\}^{-n} \tag{1.2}$$

for some fixed n. Next, the term $r = 1 - \epsilon$ was computed, and this was used to update each synaptic weight w_{ij} from presynaptic unit j to postsynaptic unit i using the learning rule

$$\Delta w_{ij} = \rho r(x_i - p_i)x_j + \lambda\rho(1 - r)(1 - x_i - p_i)x_j \tag{1.3}$$

where ρ and λ were learning rates, and x_i and x_j were the activities of units i and j respectively. The term p_i was a logistic function of the sum of weighted inputs to neuron i. This was interpreted as the probability of firing of the given neuron, that is, the output of the neuron was 1 with probability p_i and 0 with probability $1 - p_i$.

During training, the network developed hidden units that displayed retinotopic receptive fields with eye-position gain fields comparable to those of Zipser & Andersen (1988). The network was also found to be able to generalize to new combinations of eye position and retinal target location corresponding to trained head-centered target locations, and also generalize to entirely new head-centered locations not experienced during training. The investigators therefore concluded that the network structure and function was invariant to learning mechanisms.

Mazzoni et al. (1991) rightly identified the backpropagation of error algorithm used by Zipser & Andersen (1988) as biologically implausible for a learning process in the cerebral cortex. In place of this, they employed a global error signal, which could conceivably be implemented by neuromodulator release. However, their approach requires that the cortex is able to compute the desired head-centered output $d_1, \ldots, d_K$ in the first instance, com-

pare the desired head-centered output with the actual output, compute a global error term ϵ at the population level, and then adjust the synaptic weights using a rather complicated function of this global error term. None of the neural machinery needed to support this has been identified in cortex. Moreover, the proposal would seem to require such a sophisticated neural framework that it might be more plausible for the synaptic weights to be coded genetically. Although, this is also most unlikely.

Salinas & Abbott (1995)

Salinas & Abbott (1995) pursued a qualitatively different approach to modelling reference frame transformation. The specific example in the paper was to transform a retinocentric visual representation into a head-centered motor output representation suitable for reaching tasks. Their approach was to assume that learning of the mapping between the visual representation and the motor representation was guided by the agent observing its own spontaneous motor movements, e.g. reaching, during some critical stage of development. The sight of these movements would induce a sensory representation in the input layer, and the execution of the command itself would induce the corresponding representation in the motor output layer. Associative Hebbian learning would then adjust the synaptic weights to encourage the network to produce the mapping between the two induced representations, so that a trained network would produce an accurate output motor representation corresponding to the given sensory input representation. They also derived a theoretically optimal synaptic weight matrix which allowed the network to perform the

correct transformation, which was in fact the limit of the learning mechanism described.

The authors first confirmed that the network could be trained to transform eye-centered stimuli to eye-centered motor output. Then eye position was introduced as an independent input variable, and the output motor representation was set to be head-centered. This also worked, and they found that performance improved as the size both layers was increased. Next, the number of dimensions of the sensory input was increased, and they tested how well the motor representation could be trained to extract a single dimension. It was found that the performance decreased as the number of dimensions increased. Lastly, they found that non-linear transformations of the sensory input could also be read out by the motor output.

Unlike previous work, this approach did not use error correction learning to develop the appropriate synaptic weight structure. This, in itself, was an improvement in terms of biological plausibility. However, with the associative learning mechanism of Salinas & Abbott (1995), the input layer did not drive the output layer in any way during training. Instead, activity was imposed by some external supervisory signal. The significance of this is considered in chapter 5 where the model is also reproduced.

Pouget & Sejnowski (1997)

Pouget & Sejnowski (1997) suggested an alternative perspective to reference frame transformation inspired by nonlinear function approximation theory. They contested the need for a hierarchy of egocentric reference frames of increasing complexity (eyes, head, body,

etc.), as much previous work had been based on. They suggested instead that a sufficiently large number of neurons behaving as basis functions in the relevant sensory modalities (vision, somatosensation, audition, proprioception etc.) would allow *any* desired nonlinear transformation of these variables to be achieved as a linear weighting of the appropriate basis function neurons. This would, for example, permit any particular egocentric reference frame to be read out in a single step given the correct weighting of basis function neurons.

As an application of their approach, they modelled a two layer feed forward neural network with a population of visual eye-centered input units with sigmoidal eye position gain modulation projecting to a two neuron output layer. The network was trained using an error correction rule to allow one of the output neurons to read out an eye-centered representation, and the other to read out a head-centered representation. This model is reproduced and discussed in further detail in chapter 5.

The authors conceded that most parietal neurons do not have this form of nonlinear response. However, they argued that parietal neurons do satisfy two weaker conditions that together enable a population of these neurons to operate as a set of basis functions. First, the eye position preference and retinal target location preference of individual neurons had to interact non-linearly. That is, linearly weighted sums of independent functions of each input variable were not admissible. Since the majority 43 % of neuronal responses in area 7a showed multiplicative interactions between eye position and retinal target location, and a further 28% were even more complex, this constraint was clearly satisfied in area 7a

(Zipser & Andersen 1988). Second, retinal receptive fields and gain fields had to be non-linear functions of each variable. This clearly excluded the traditional planar gain fields observed for many neurons in both 7a and LIP (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990). However, by reanalyzing the experimental data they found that the gain fields were indeed nonlinear within a normal range of eye positions, although short of conclusively demonstrating that they were sigmoidal. Based on this they concluded that parietal neurons do possibly constitute a basis set on which non-linear function approximations could be made downstream to effect coordinate transformations.

Xing & Andersen (2000*b*)

In a continuation of previous work, Xing & Andersen (2000*b*) explored how multiple distinct reference frames could be simultaneously produced from a set of input reference frames. In particular, they where interested in how the relationship between the input and output representations influenced the gain field structure of the hidden units of the network. They explored a range of models, with different combinations of input and output signals, but all the models shared major features of the work in Zipser & Andersen (1988). The models all had one or more input layers encoding an eye-centered visual signal, head-centered auditory and eye-position signal, and body-centered head-position signal. All models had an intermediate layer, which connected the input layers to the output layers in some way depending on the specific experiment. The models had one or more output layers encoding an eye-centered, head-centered or body-centered output. The network was trained using

the backpropagation of error algorithm while chosing random combinations of input signals until the performance converged. In the analysis, the authors mapped the receptive field and gain field of each intermediate layer neuron. The former was mapped by testing different retinal locations with stimuli, and the latter was mapped by stimulating the same retinal location across a range of eye positions. The relationship between the receptive field and gain field of individual intermediate neurons was assessed by inspecting the angle between the 2D vectors representing the center of mass in each field. The reference frame of intermediate neurons was also assessed by examining the change in receptive field centers with change in eye position. A variety of results were found for the different input-output representation combinations.

In a first set of experiments the network had eye position input, and either eye-centered visual input or head-centered auditory input. In the experiment with eye-centered visual input, the network was first trained to produce an eye-centered output. In this case, the network learned to produce eye-centered output as expected, but did not develop eye position gain fields in the intermediate layer. Then the model was trained to produce head-centered output, making it identical to Zipser & Andersen (1988). In this case, the output neurons developed responses in a reference frame between eye and head-centered, while eye position gain fields were found in the intermediate layer that were aligned with visual receptive fields. In the experiment with head-centered auditory input, the network was first trained to produce head-centered output. In this case, the network learned to

produce head-centered output as expected, but did not develop eye position gain fields in the intermediate layer. Then the model was trained to produce eye-centered output. In this case, the output neurons learned to respond in a mixed reference frame, while the intermediate layer neurons developed eye position gain fields in the opposite direction of the auditory receptive fields. This showed that the gain fields of intermediate layer neurons were either aligned or oppositely aligned with the retinotopic receptive fields of input neurons when the input signals had to be either added or subtracted to perform the desired transformation

In a second set of experiments the network had eye position input, eye-centered visual input and head-centered auditory input. The model was first successfully trained to transform all inputs to an eye-centered reference frame in the output layer. This produced oppositely aligned gain fields and receptive fields in the intermediate layer. Then the model was successfully trained to transform all inputs to a head-centered reference frame in the output layer. This produced similarly aligned gain fields and receptive fields in the intermediate layer. These results showed that the model was able to combine and map different sensory modalities onto the same output reference frame.

In a third experiment the network again had eye position input, eye-centered visual input and head-centered auditory input, but was now trained to produce both eye-centered and head-centered outputs. The model was able to learn the desired transformations. However, there was no clear structure in the relationship between the gain fields and

receptive fields in the intermediate layer. Nor did the intermediate layer neurons show responses clearly linked to a particular reference frame.

In a fourth and last experiment, the previous experiment was extended with head position as input, and body centered output. The results were similar to the previous model, except intermediate layer neurons now also had head position gain fields.

Spratling (2009)

The model architecture consisted of a hierarchy of stages, each made up of two populations of processing units, as seen in figure 1.4. The first stage received retinotopic visual input and an eye position signal, and therefore learned a head-centered visual representation. The second stage received this head-centered input in combination with a neck position signal, and therefore learned a body-centered visual representation.

Each of the two stages implemented the same form of processing and learning, however with different inputs. In each stage, the first population learned to represent conjunctions between its two inputs, and this happened through competitive learning with a Hebbian style synaptic learning rule. The second population learned to bind together these conjunctive representations across time through competitive learning with a temporally associative synaptic learning rule.

Due to the similarities between this model and work in subsequent chapters, a meaningful comparison in later discussion sections necessitates that it is presented here in some detail.

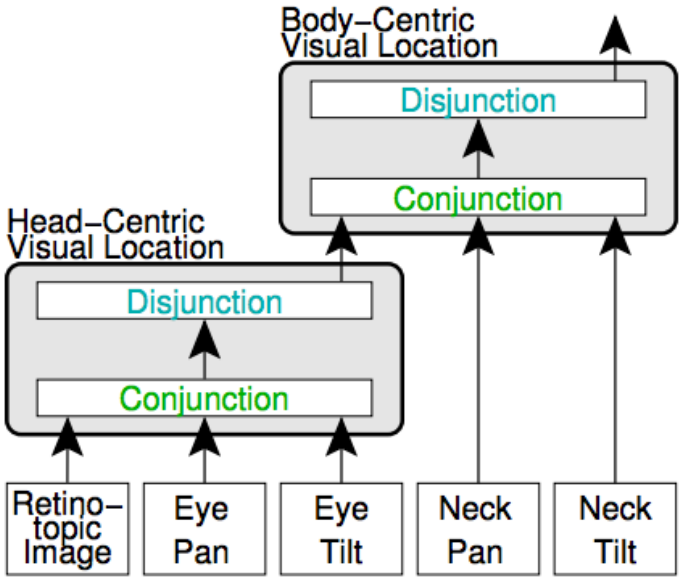


Figure 1.4: Model architecture in Spratling (2009)

The visual scene consisted of a two dimensional pixel array, where each pixel corresponded to an object location, and an object was considered present if the pixel had a positive intensity. The retina was therefore also a two dimensional pixel array of object locations. The eye and head position inputs were represented as additional one dimensional pixel arrays, where each pixel corresponded to a unique position.

For each of the two hierarchical processing stages shown in figure 1.4, learning was carried out for the conjunctive and disjunctive populations in the following manner.

First, consider a conjunctive population of n output units with activations $[y_1, \ldots, y_n]^T$. Each conjunctive unit receives an input pattern $[x_1, \ldots, x_m]^T$ from m input units. The output responses of the conjunctive population were computed by iteratively updating the

following coupled equations

$$e_j^{t+1} = \frac{x_j}{\sum_{i=1}^{n} \hat{w}_{ij} y_i^t} \tag{1.4}$$

$$y_i^{t+1} = y_i^t \sum_{j=1}^{m} w_{ij} e_j^t \tag{1.5}$$

where w_{ij} was the synaptic weight from input unit j to output unit i, and $\hat{w}_{ij}$ was the weight of the same synapse after the output weight vectors were normalized. After 100 updates of these equations, e_j^t and y_i^t settled to steady values e_j and y_i, respectively. The variable e_j was termed the reconstruction error. Then the following learning rule was applied

$$\delta w_{ij} = \beta y_i (e_j - 1) w_{ij} \tag{1.6}$$

where β was the learning rate for these connections. The reconstruction error e_j represented how well the output pattern corresponded to efferent synapses from input unit j. An error of $e_j = 1$ constituted a perfect correspondence, and a difference between e_j and 1 led to an error reducing change in efferent synapses from input unit j.

Second, consider a disjunctive population of n output units with activations $[y_1, \ldots, y_n]^T$. Each conjunctive unit receives an input pattern $[x_1, \ldots, x_m]^T$ from conjunctive population units at the same stage. For each disjunctive unit i, the most effective driving input unit

k_i was given by

$$k_i = \underset{j}{\operatorname{argmax}} \{\hat{w}_{ij}\check{w}_{ij}x_j\} \tag{1.7}$$

where $\check{w}_{ij}$ was the weight of the synapse from input unit j to output unit i after the efferent weight vector from input neuron j was normalized. The activation of output unit i was governed by

$$y_i = \hat{w}_{ik_i}\check{w}_{ik_i}x_{k_i} \tag{1.8}$$

All afferent synapses of output unit i, except the synapse originating from input unit k_i, were then updated according to the following learning rule

$$\delta w_{ij} = \gamma y_i(x_j^* - x_j)w_{ij} \tag{1.9}$$

where x_j^* was the activation of unit j for the previous input pattern, and γ was the learning rate for these connections. After each weight update, negative weights were clipped, and the efferent weight vector from each input unit was then normalized.

This learning rule encouraged a presently active output neuron to learn to respond to neurons that were active in the previous input pattern, thereby promoting temporal association. It also had the effect of making the same output neuron weaken its connections to presently active input neurons, thereby avoiding learning about multiple locations

in the visual scene. While this allowed the model to be trained on scenes with multiple simultaneously visible objects, it also required that the learning rule was not applied to the synapse from the most efficacious input unit because this would undermine the temporal binding. The normalization of the efferent synaptic weight vector from each input unit made synaptic strengthening rivalrous, and this effectively implemented a form of competition between the output units.

The model was trained on a sequence of visual scenes, where objects were gradually shifted around within the visual space. The retina of the network model shifted its fixation point within the scenes due to randomised movements of the eyes and head. The corresponding eye and head position signals were also supplied as inputs to the model as shown in figure 1.4. The relocation of objects within the visual space was sufficiently slow in order to allow retinal images of a relatively stable visual scene to be paired with a range of eye and head position signals, so that the reference frame transformation could be learned. The two stages of the network model were trained separately. The first head-centric stage was trained first on random eye position sequences, and then its weights were frozen while the second body-centric stage was trained on both random eye and head position sequences.

The model was tested on a single object in the visual scene for all combinations of eye and head position inputs, and an error metric for the output of the disjunctive populations was used as a measure of the quality of the learned visual representations.

The model successfully developed head-centered representations in the disjunctive pop-

ulation of the first stage, as well as body-centered representations in the disjunctive population of the second stage, through the hypothesised mechanism of unsupervised visually-guided learning. However, the following two observations were also made. First, increasing the number of simultaneously visible objects in the visual scene degraded both the head-centric and body-centric visual output representations. Second, increasing the instability of the visual scene, by effectively relocating visual objects more frequently during training, also led to a decreased quality in both visual output representations.

A limitation of the model of Spratling (2009) was that the encoding of the eye and head position signals was effectively peaked, and as a result so were the gain fields. This thesis investigates the effect of both peaked and monotonic gain fields on the head-centered representations.

Chapter 2

Head-Centered Receptive Fields: Methods

2.1 Introduction

This chapter presents a neural network model of head-centered visual representations using peak shaped eye position gain fields. The hypothesis on which the model is based is presented, and its basic premises are briefly explored. This chapter describes the network architecture, neuronal dynamics, training stimuli, testing stimuli and data analysis. These methods are used for the simulation studies presented in the following chapters 3, 4 and 5.

2.2 Hypothesis

It was hypothesised that the following four core model components would permit head centered visual representations to develop:

- There is a population of input neurons that encode both the retinotopic location of

visual targets and eye position at the single neuron level through coupled visual and eye position receptive fields.

- There is a population of output neurons that compete with each other through mutual inhibitory interactions mediated by inhibitory interneurons.
- The feedforward synaptic connections from the input cells to output cells are modified by a local synaptic trace learning rule that encourages individual output cells to learn to respond to subsets of input patterns that tend to occur close together in time.
- During natural self-motion, there are periods of time when the eyes are moving in the head while the head remains stationary with respect to the visual environment and visual objects also remain stationary within the environment.

Retinotopic visual neurons with eye position gain modulation, satisfying the first premise, have been identified in multiple primate cortical areas as reviewed in section 1, in particular in area PO. Premise two is a standard feature of cortical architecture, in which competitive interactions between excitatory neurons are mediated by inhibitory interneurons (Rolls & Treves 1998). Premises three and four are related, in that the latter is the ecological constraint providing the temporal structure which the former exploits.

Under the assumption that visual stimuli are relatively static in a world reference frame, a primate will more often adjust its gaze by moving its eyes rather than the head itself (Freedman & Sparks 1997). This behavioural strategy is preferable to making frequent

energetically costly and slow head movements to adjust gaze. Evidence for this has been found during exploration of natural environments with free eye, head and body movements (Einhäuser et al. 2007). It was found that when there was movement, isolated eye and isolated head movements occurred 33.1% and 13.3% of the time respectively, while the remaining time involved a mixture of movements. These experimental results confirm that the eyes move more frequently than the head. A consequence of this is that, during natural movement, there are periods when the eyes are moving while the head remains stationary with respect to the visual environment. This temporal structure can be exploited by a trace learning rule as follows.

A trace learning rule is a local associative learning rule that incorporates an exponentially decaying temporal trace of past neuronal activity. The effect of such a learning rule is to encourage individual postsynaptic neurons to learn to respond to subsets of input patterns that tend to occur close together in time (Földiák 1991, Wallis & Rolls 1997).

If the eyes are moving around a scene containing a visual target while the head remains stationary, then the visual system will receive a sequence of input patterns corresponding to the visual target in different retinal locations but the same relative position with respect to the head. That is, during this period, the visual target will change position in the eye-centered space but it will remain stationary in the head-centered space. The sequence of eye positions and resulting retinal locations of the visual target will be represented by the retinotopic eye-position gain modulated input neurons. The synaptic trace learning rule

will bind these input patterns onto the same output representation precisely because these input patterns frequently follow each other in time.

The visual system may be exposed on separate occasions to a number of such input pattern sequences of the visual target in the same head-centered location, but situated in many different retinal locations due to rapid movement of the eyes. Each of these sequences might represent the visual target in a different randomised subset of retinal locations. These randomised sequences would randomly overlap with each other, which would ensure that all possible patterns, corresponding to the same head-centered location but different retinal locations, are brought into temporal proximity with each other. In this way, all of the patterns corresponding to a single head-centered location but different retinal locations would tend to occur clustered together in time. This kind of randomised mixing of input patterns has previously been found to facilitate the temporal binding performed by trace learning (Wallis & Rolls 1997). Given this extensive training, a trace learning rule would eventually encourage a subset of postsynaptic output cells to learn to respond to this complete set of input patterns and thereby learn to respond to the visual target at a particular position in the head-centered frame of reference regardless of where the visual target occurs on the retina.

Occasionally the position of the head, itself, will be readjusted, whereupon this process continues with the visual target in a different position in head-centered space. That is, the location of the visual target would be shifted to new head-centered locations by the

natural head movements that occur between sequences of rapid eye movements. The learning process could thus be repeated with the visual target presented in many different head-centered locations. Due to the competitive interactions between the output cells, a new subset of postsynaptic output cells would learn to respond to the visual target in each different head-centered location. In this manner, the output layer would eventually develop neurons that cover the entire space of head-centered locations. The next three chapters will explore the feasibility of the above hypothesis.

2.3 Network Architecture

The architecture of the neural network model is shown in Fig. 2.1. The network consisted of two layers of neurons, one projecting to the other.

The first layer was a population of input neurons that simultaneously encoded the eye-position of the agent and the retinal location of the visual targets. These were modelled as retinotopic neurons with eye-position gain fields. The equations specifying the firing rate responses of these input cells are given below in section 2.6. The retinal and eye position spaces, representing the range of retinal locations and eye positions in orbit, covered $[-90°, 90°]$ and $[-30°, 30°]$ respectively. Neurons in the input layer sent feedforward synaptic connections to neurons in the second layer.

The second layer was a competitive population of N output neurons that competed to represent patterns in the input layer (Rolls & Treves 1998). Neurons in the output

layer all received the same number of afferent connections from the input layer, that is ϕ percent of the input population, but where each output neuron received connections from its own randomly assigned subset of the input neurons. The output neurons operated as a competitive layer and equations governing their firing rate responses are given below in section 2.6. The output cells developed their response tuning through visually-guided competitive learning.

Neither the input layer nor output layer were topographically organized.

At the start of each simulation, the strengths of the feedforward synaptic connections from the input layer to the output layer were initialised to random weights in the interval $[0, 1]$. Then the synaptic weight vector of each output neuron was renormalized as is typical in competitive networks (Rolls & Treves 1998).

During training of the network, the feedforward synaptic connection weights were modified by a trace learning rule as described below in section 2.6. After each training update with the trace learning rule, the dendritic weight vectors of the output neurons were again renormalized to ensure no single neuron learned to respond to all of the input patterns.

2.4 Training the Network

The network was trained on inputs representing a combination of visual signals and eye position signals. The visual signals represented scenes containing a number of static visual targets. The network was trained on these combined input signals while the eyes explored

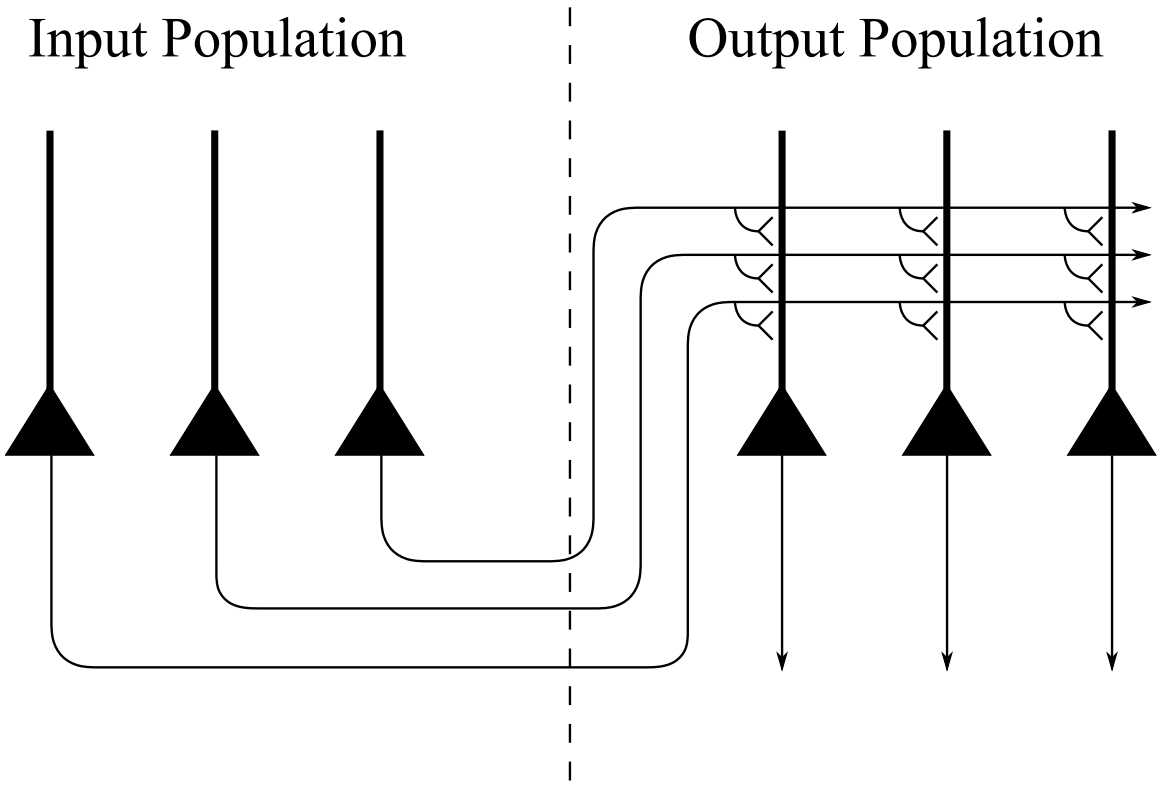


Figure 2.1: Architecture of 2-layer neural network model. The layer of input neurons on the left are projecting to the competitive output layer on the right. During learning, the strengths of the feedforward synaptic connections from the input layer to the output layer are modified by a trace learning rule.

a number of visual scenes by moving rapidly in their orbits with occasional movements of the head.

In order to reduce edge effects due to clipping of the input representations, the retinal locations of visual targets were kept within the interval $[-63°, 63°]$, while the position of the eyes was kept within $[24°, 24°]$.

In each experiment, M evenly spaced head-centered locations in $[-63°, 63°]$ were chosen, and all visual targets were located in one of these locations during training. This interval of head-centered locations ensured that the visual targets always remained in view as the

eyes moved.

Each training epoch was divided into a number of periods, where in each period there was a fixed unique subset of k head-centered locations occupied by visual stimuli. During such a period, the locations of the k visual targets with respect to the head remained fixed while the eyes saccaded through a randomised sequence of P different eye positions in the orbit. In most experiments the model was trained on only one visual target at a time, hence there were M periods within each training epoch, each corresponding to a visual target being located in one of the M head-centered locations. In other experiments the model was trained on two or more targets ($k > 1$) presented simultaneously within each period. In this case, since the model was trained on k visual objects presented in all possible subsets of the M head-centered locations, there were $\binom{M}{k}$ such periods within each training epoch. During each period in which k visual targets were presented in a fixed combination of head-centered locations, the model fixated P uniformly sampled eye positions in $[-24°, 24°]$. The duration of each fixation while the eyes remained stationary was usually set to 300ms, although in some experiments this was varied. The model saccaded between successive eye positions at a constant velocity of $400°/s$.

Thus, during training, the network was presented with sequences of combined visual and eye position input signals that represented the visual targets remaining in fixed head-centered locations while the eyes shifted through randomised positions in the orbit. This is the kind of temporal structure required by our hypothesis described above in section 2.2.

While the network is exposed to these training input sequences, the feedforward synaptic weights from the input layer to the output layer are adjusted by a trace learning rule described below in section 2.6. This learning rule binds together subsets of combined visual and eye position input representations that tend to occur close together in time. The effect of this, when combined with the temporal ordering of the input signals during training, is to encourage individual output neurons to learn to respond to subsets of input representations corresponding to the visual target situated in a particular head-centered location regardless of eye position. Furthermore, the competitive interaction between the output neurons ensures that different neurons learn to respond to different head-centered locations, with the entire set of M head-centered locations eventually represented by distinct subsets of neurons in the output layer.

2.5 Testing the Network

After training, the model was tested by recording the responses of the output neurons for all combinations of E different eye fixation positions and T head-centered visual target locations. The model was tested on a single visual target for the majority of experiments, and deviations from this are explained in detailed for the relevant experiments. The data from this testing was used to analyse the receptive field properties of the neurons, including the reference frame of response, receptive field size and receptive field location. For testing in these standard experiments, the model fixated a set of four eye positions $-18°, -6°, 6°$

and 18°. For each such fixation a single visual target was placed in each of eighty head-centered target locations within $[-79°, 79°]$ in increments of 2°. For each combination of eye position and head-centered visual target location the model fixated for $300ms$, and the firing rates of all neurons in the output layer were saved at the end of this period for analysis. The model was completely reset between different combinations of eye position and head-centered visual target location. For experiments departing from this protocol the results section will explicitly mention this.

2.6 Neuronal and Synaptic Dynamics

Input Layer

The neurons in the input layer were modelled by imposing a firing rate function that simulated the response properties of retinotopic neurons that are modulated by eye position gain fields. Such neurons have been found in a number of areas of the primate brain, including areas PO (Galletti et al. 1995) and the posterior parietal cortex (PPC) (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990).

We tested two alternative forms of neuronal response function for computing the firing rates of the input neurons. Both types of response function were used to map the eye position, denoted by e, and a set of retinal locations of targets, denoted by R, onto the instantaneous firing rate of each input neuron i, denoted by $v_i^I(t)$, within the range $[0, 1]$. Both of the response functions had the same form of dependency on the retinal location

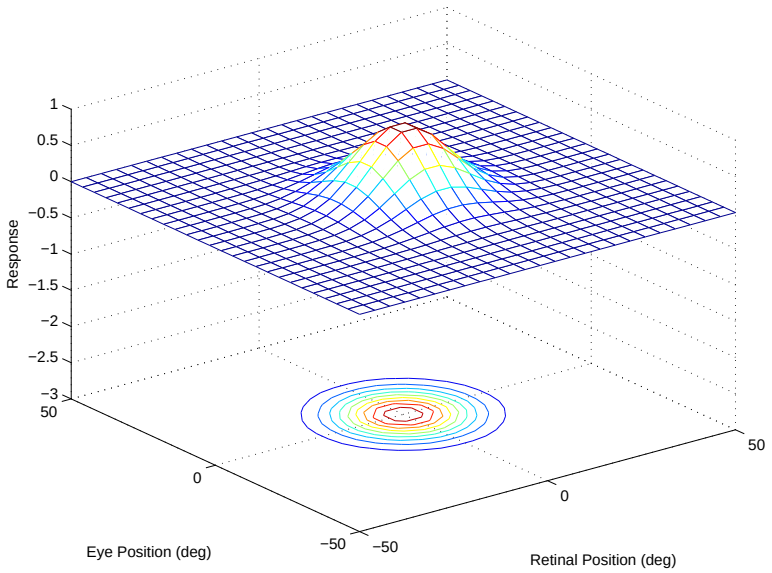
of the visual targets. However, the response functions differed in how the eye position modulated the responses. Examples of the two alternative response functions employing different forms of eye position gain modulation are shown in figure 2.2.

In some simulations multiple visual targets were presented to the network simultaneously. Therefore, the two alternative response functions needed to be defined when multiple visual targets were present within a scene. Therefore, the assumption that the response functions should respond to the presence of multiple visual targets by computing the sum of the responses to each visual target was made. This was a simple and plausible assumption facilitating the extension of the experimentally derived response functions (Andersen et al. 1985) to scenes with multiple simultaneously visible targets. Experiments have indeed shown that multiple visual targets separated on the retina do not disturb the receptive fields of LIP neurons to individual targets (Gottlieb et al. 1998), which is compatible with the superposition assumption stated above.

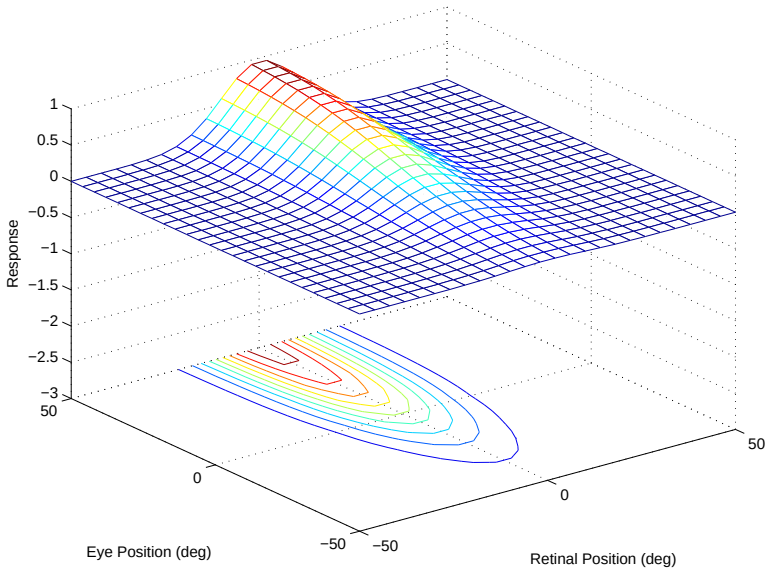
The first form of firing rate response function used to model the input neurons incorporated a peaked eye position gain field as shown in Fig. 2.2(a). This form of eye position modulation has been reported in cortical area PO by Galletti et al. (1995). The full response function is described by

$$v_i^I = \omega_i \exp\left(-\frac{\|e-\beta_i\|^2}{2\rho^2}\right) \times \sum_{r\in R} \exp\left(-\frac{\|r-\alpha_i\|^2}{2\sigma^2}\right) \tag{2.1}$$

This response function was composed of a product of two components: the first component



(a) Peaked Modulation



(b) Sigmoidal Modulation

Figure 2.2: Examples of the two alternative response functions used to compute the firing rates of input neurons. For each response function, the different responses of an individual input neuron as a function of eye position and retinal location of a single visual target is plotted. Subplot (a) shows an example of a peaked response function, in which the firing rate is modulated by a peaked eye position gain field as described by equation 2.1. Subplot (b) shows a response function that is modulated by a sigmoidal eye position gain field described by equation 2.2.

represented the eye position signal, while the second component encoded the retinotopic positions of the R visual targets in the scene. In equation 2.1, the neuronal response was modulated by a peaked Gaussian function of eye position. The parameter β_i represented the preferred eye position for each input neuron i, with the width of the corresponding Gaussian eye position tuning curve determined by the standard deviation ρ. The parameter α_i specified the preferred retinal location of a target stimulus for each input neuron i, and the standard deviation σ determined the width of the corresponding Gaussian retinal tuning curve. The parameter ω_i was an amplitude modulation coefficient determining the magnitude of maximal response of the neuron. In most simulations, ω_i was set to 1 for all input neurons. Each input neuron was set to respond maximally to a unique combination of retinal target location α_i and eye position β_i. The population of input neurons covered the entire two dimensional space resulting from combinations of eye position and retinal target location in integer steps of 1 degree in each dimension.

The second form of firing rate response function used to model the input neurons incorporated a sigmoid eye position gain field as shown in Fig. 2.2(b). This form of eye position modulation was monotonic in the eye position dimension, as has been identified in multiple areas with such modulation (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990, Andersen et al. 1985), and the form of the modulation was sigmoidal while most empirical work has described it as planar. However Pouget & Sejnowski (1997) showed the data was also compatible with a saturating sigmoidal gain formulation, hence the full

response function was given by

$$v_i^I = \frac{1}{1 + \exp\left(2\kappa_i(e - \beta_i)\right)} \times \sum_{r \in R} \exp\left(-\frac{\|r - \alpha_i\|^2}{2\sigma^2}\right) \tag{2.2}$$

In equation 2.2, the neuronal response is modulated by a sigmoidal function of eye position. For each input neuron i, the slope κ_i determined the gain direction and saturation rate of the modulation, and the inflection point β_i determined the eye position where a significant firing rate response began. The input neurons all had the same absolute saturation rate, that is $|\kappa_i| = |\kappa_j|$ for all i and j, but one half had a positive gain direction ($\kappa_i > 0$) while the other half had a negative gain direction ($\kappa_i < 0$). Each input neuron was set to respond maximally to a unique combination of retinal target location α_i, eye position β_i, and gain direction and saturation rate κ_i. The population of input neurons evenly covered the entire three dimensional space resulting from such combinations.

Output Layer

For each neuron i in the competitive output layer there were three dynamical quantities defined: a trace value $q_i(t)$ as discussed in section 2.2, an internal activation $h_i(t)$ and an instantaneous firing rate $v_i(t)$ (Dayan & Abbott 2001).

The activation $h_i(t)$ of each output neuron i was governed by

$$\tau_h \frac{dh_i}{dt} = -h_i + \sum_j w_{ij} v_j^I \tag{2.3}$$

where τ_h was a time constant common for all neurons in the output layer and w_{ij} was the synaptic weight of the synapse from input neuron j to output neuron i.

The firing rate of output neuron i was given by the sigmoid activation function

$$v_i = \frac{1}{1 + \exp\left(-2\varphi(h_i - p_\pi - \theta)\right)} \tag{2.4}$$

where φ was the sigmoid slope and θ was the threshold. The parameter p_π was used to regulate the level of competition between neurons in the output layer and so control the overall proportion of neurons that remained active. Specifically, p_π was set to the activation value at the π^{th} percentile point of the distribution of neuronal activations within the output layer. For example, if π was set to 90, then p_π was set to the top tenth percentile activation value. This was a practical means of implementing competition within the competitive output layer, which in cortex is likely to be implemented via inhibitory interneurons (Dayan & Abbott 2001). This way of implementing competition has been previously used in competitive neural network models of the primate visual system with trace learning (Rolls 2012). Although, in some experiments described later an alternative model of competition was also tested.

Trace Learning

Trace learning rules utilize a temporal trace of recent neuronal activity in order to encourage postsynaptic neurons to bind together subsets of input patterns that occur close together

in time. The trace value for neuron i in the output layer was denoted by $q_i(t)$ and was governed by the equation

$$\tau_q \frac{dq_i}{dt} = -q_i + v_i \tag{2.5}$$

where τ_q was a time constant common for all neurons in the output layer.

During training the strength of the synapse from input neuron j to output neuron i was governed by the trace learning rule

$$\frac{dw_{ij}}{dt} = \varrho q_i v_j^I \tag{2.6}$$

where ϱ was the learning rate, v_j^I was the firing rate of input neuron j, and ϱq_i was the trace value of output neuron i.

Finally, to prevent unbounded growth of the synaptic weights during training, the length of the weight vector for each output neuron i at time t, that is $\mathbf{w}_i = (w_{i1}, \ldots, w_{iN_I})$ where there are N_I input neurons, was renormalized by setting

$$\mathbf{w}_i := \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \tag{2.7}$$

after each weight update (Dayan & Abbott 2001). Experimental evidence for renormalisation of synaptic weights in the brain has been provided by Royer & Paré (2003).

In some experiments described below, alternative learning rule formulations were also

explored.

2.7 Numerical Simulation

The coupled differential equations 2.3, 2.5 and 2.6 were integrated numerically using the Forward-Euler scheme, where the numerical time step Δt was set to one tenth of the neuronal time constant τ_h. For all simulations the stability of the results was manually confirmed by checking that the qualitative nature of the results remained invariant at the single neuron level over reductions in time step and increases in the number of training epochs.

The combined visual and eye-position input signals described in the training section 2.4 and testing section 2.5 were simulated dynamically and sampled at 1kHz. Then, where necessary, linear interpolation was used to compute the numerical inputs to the discretized Forward Euler model equations, which required input values at every numerical time step $\Delta t = \tau_h/10$.

2.8 Analysis

Let $\mathbf{R}$ be a matrix containing the responses of a given neuron during testing, where $\mathbf{R}[i, j]$ denotes the firing rate when the model was fixating in the i^{th} eye position e_i and the visual target was in the j^{th} head-centered location t_j, as recorded during the testing protocol described above. The vector $(\mathbf{R}[i, 1], \ldots, \mathbf{R}[i, T])$ is referred to as the response vector at

the i^{th} eye position. The number of eye positions during testing is denoted by E, while the number of head-centered locations for visual targets during testing is denoted by T. The indexing of eye positions and head-centered target locations were ordered from left (negative) to right (positive), that is $e_1 \leq \ldots \leq e_E$ and $t_1 \leq \ldots \leq t_T$.

Reference Frames

To determine which reference frame an output neuron was responding in after training, two separate metrics were applied that reflected to what degree the neuronal response was compatible with either an eye-centered or head-centered reference frame, and then compared the values of these two metrics.

In section 1.2 a neuron was defined as responding to visual targets in either a head-centered or eye-centered reference frame when the neuron was visually responsive and the structure of this response was stable in either a head-centered or eye-centered space across different eye positions.

The head-centeredness metric measured the degree of such stability for a given output neuron by averaging the correlation between response vectors corresponding to all pairs of eye positions, that is

$$\Pi = \frac{1}{\binom{E}{2}} \sum_{1 \leq i_1 < i_2 \leq E} \frac{\sum_{j=1}^{T} \left(\mathbf{R}[i_1, j] - \overline{\mathbf{R}}[i_1]\right) \left(\mathbf{R}[i_2, j] - \overline{\mathbf{R}}[i_2]\right)}{\sqrt{\sum_{j=1}^{T} \left(\mathbf{R}[i_1, j] - \overline{\mathbf{R}}[i_1]\right)^2 \sum_{j=1}^{T} \left(\mathbf{R}[i_2, j] - \overline{\mathbf{R}}[i_2]\right)^2}} \tag{2.8}$$

where

$$\overline{\mathbf{R}}[i] = \frac{1}{T}\sum_{j=1}^{T}\mathbf{R}[i,j] \tag{2.9}$$

This yielded a metric which was referred to as the *head-centeredness* of the output neuron, and it was bounded between -1 and 1, where a perfect correlation of 1 indicated a perfectly head-centered response.

A very similar analysis was done to quantify the compatibility of the responses of the output neuron with an eye-centered frame of reference. That is, a visual neuron is judged to respond in an eye-centered frame of reference to the extent that its eye-centered response vectors remain stable across different eye positions. The eye-centered analysis proceeded as follows. To reiterate, each response vector $(\mathbf{R}[i,1],\ldots,\mathbf{R}[i,T])$ was the result of testing over the same set of head-centered locations, but with the model fixated in a distinct eye position. Therefore, each response vector also corresponded to a unique range of retinal locations. The *intersection* of these retinal ranges corresponded to different portions of each response vector, and it was these portions that were subject to correlation analysis. Specifically, f_i denotes the first vector position in the i^{th} response vector to be included, and the $V-1$ next positions are included as well such that the subvector $(\mathbf{R}[i,f_i],\ldots,\mathbf{R}[i,f_i+(V-1)])$ is the vector being used for the correlation analysis. The derivation of f_i and V are found in the supplementary information. This gave the metric

$$\Omega = \frac{1}{\binom{E}{2}} \sum_{1 \le i_1 < i_2 \le E} \frac{\sum_{j=0}^{V-1} \left(\mathbf{R}[i_1, f_{i_1} + j] - \overline{\mathbf{R}}[i_1]\right) \left(\mathbf{R}[i_2, f_{i_2} + j] - \overline{\mathbf{R}}[i_2]\right)}{\sqrt{\sum_{j=0}^{V-1} \left(\mathbf{R}[i_1, f_{i_1} + j] - \overline{\mathbf{R}}[i_1]\right)^2 \sum_{j=0}^{V-1} \left(\mathbf{R}[i_2, f_{i_2} + j] - \overline{\mathbf{R}}[i_2]\right)^2}} \tag{2.10}$$

where

$$\overline{\mathbf{R}}[i] = \frac{1}{V} \sum_{j=0}^{V-1} \mathbf{R}[i, f_i + j] \tag{2.11}$$

This was referred to as the *eye-centeredness* of the output neuron, and it was bounded between -1 and 1, where a perfect correlation of 1 indicated a perfectly eye-centered response. Response vectors which had no response for the extracted ranges were excluded from the correlation, and a neuron without a response within this range of retinal locations at any eye position was excluded from further analysis.

Both of the reference frame metrics were finally combined into a receptive field index (RFI) which classified each neuron along a spectrum from eye-centered to head-centered. The RFI was defined as

$$\text{RFI} = \begin{cases} \Pi - \Omega & \text{if } 0 \le \Pi \le 1 \text{ and } 0 \le \Omega \le 1 \\ \Pi & \text{if } 0 \le \Pi \le 1 \text{ and } -1 \le \Omega < 0 \\ -\Omega & \text{if } -1 \le \Pi < 0 \text{ and } 0 \le \Omega \le 1 \\ 0 & \text{if } -1 \le \Pi < 0 \text{ and } -1 \le \Omega < 0 \end{cases} \tag{2.12}$$

The index was a continuous valued function bounded between -1 and 1. An output neuron with either a positive, negative or nil RFI value was classified as head-centered, eye-centered or undetermined respectively. In general, a large positive value for the RFI indicated better compatibility with a head-centered reference frame than a smaller positive value. Similarly, a large negative value for the RFI indicated better compatibility with an eye-centered reference frame than a smaller negative value.

Receptive Field Location

The head-centered receptive field location of an output neuron was determined as follows. First, the head-centered receptive field location at each eye position e_i for $i = 1, \ldots, E$ during testing was computed. The head centered receptive field location for the i^{th} eye position was computed using the centre of mass of the head-centered response vector at this eye position. Next, the average of these head-centered locations over all eye positions was computed, and this averaging is shown to be optimal in Appendix B. This gave the final metric for each neuron

$$\frac{1}{E}\sum_{i=1}^{E}\frac{\sum_{j=1}^{T} t_j \mathbf{R}[i,j]}{\sum_{j=1}^{T}\mathbf{R}[i,j]} \tag{2.13}$$

Coverage

Among head-centered neurons, that is those with a positive RFI, it was determined how their receptive fields were distributed among the locations of the head-centered locations

were visual targets were located during training, denoted by $g_1, \ldots, g_M$. This was to establish whether head-centered space in general was represented evenly, and in particular whether each training location was preferred by atleast one head-centered output neuron.

This distribution was determined by first assigning each head-centered neuron to the closest training location. Let p_i denote the fraction of head-centered neurons assigned to training location g_i. To ensure that all locations had at least one neuron assigned to them, and also quantify the extent to which all locations were evenly represented, the normalized entropy of this distribution was computed by

$$-\frac{1}{\log_2 M}\sum_{i=1}^{M} p_i \log_2 p_i \tag{2.14}$$

and this was referred to as the *coverage* of the model. A perfectly uniform distribution would give a maximal value of 1, and if there was some $p_i = 0$ it was undefined and there was said to be no coverage.

Receptive Field Size

To quantify the receptive field size of a visual neuron, a firing rate threshold is often chosen to demarcate the responsive region or receptive field. In the following analysis, a neuron was considered responsive when it was firing above a threshold rate C no less than 50% of

its maximal rate across all eye positions and head-centered visual locations, that is

$$\frac{\max_{i,j} \mathbf{R}[i,j]}{2} \tag{2.15}$$

For each eye postion, this threshold was used to isolate the responsive head-centered regions of the given neuron. The total size of all such regions was added up, estimating the total receptive field size at the given eye position. The final receptive field size of the given neuron was the average of all these estimates from the different eye positions.

To isolate the responsive regions of a neuron at the i^{th} eye position, a head-centered piecewise linear response function $R_i : [-79°, 79°] \to [0, 1]$ was derived from the i^{th} response vector using linear interpolation. Responsive regions were isolated by first finding all solutions x^* to the equation $C = R_i(x^*)$, and then identifying the regions of head-centered space where the interpolated neuronal response was above threshold, denoted by $I_i^1, \ldots, I_i^{n_i}$. In practice there was hardly ever more than a single responsive region per eye position i, that is $n_i \leq 1$. Finally, the total size of all responsive regions $I_i^1, \ldots, I_i^{n_i}$ for each eye position was averaged across all eye positions, that is

$$\frac{1}{E} \sum_{i=1}^{E} \sum_{k=1}^{n_i} b_i^k - a_i^k \tag{2.16}$$

where $I_i^k = [a_i^k, b_i^k]$ was the k^{th} region at the i^{th} eye position, of which there were n_i in total. On a few rare occasions, a head centered neuron did not actually respond in one of

the eye positions. In this case, that eye position was excluded from the averaging procedure carried out in equation 2.16.

Chapter 3

Head-Centered Receptive Fields: Peaked Eye Position Gain Fields

3.1 Introduction

This chapter presents simulation results with a neural network model of head-centered visual representations using peak shaped eye position gain fields. Experimental results establishing the basic operation of the model are presented, and the necessity and sufficiency of the premises on which the learning hypothesis is based are explored. Additionally, the model behaviour as a function of a number of key parameters is explored.

3.2 Results

3.2.1 Hardwired Model

This experiment explored how a network with hardwired synaptic connections could compute head-centered visual responses based on peaked eye position gain modulation in the

input neuron population as described by equation 2.1, similar to representations found in area PO (Galletti et al. 1995). It served as a baseline experiment in two respects. Firstly, the hardwiring determined the neural responses at the single neuron level, and this allowed for validation of the values decoded by the metrics described in the analysis section 2.8. Secondly, the experiment served as baseline study establishing an ideal network which could be used for comparison with later networks that self-organised their synaptic connectivity. Beginning with a hardwired model also helped to emphasize the distinction between hardwiring the synaptic connectivity in order to achieve a desired output, and the network learning to compute the output representation, a distinction which was more deeply explored in chapter 5.

The parameters for the model can be found in table 3.1. There were 12261 neurons in the input population. Each input neuron was assigned to a unique combination of retinal location and eye-position, and the width σ of the retinal receptive field and the eye position tuning curve were both set to 3°. There were 900 neurons in the output population, each given a head-centered receptive field at one among eight head-centered locations, which were $-63°, -45°, -27°, -9°, 9°, 27°, 45°$ and $63°$. Each neuron in the output population received connections from a randomly assigned subpopulation of 1000 neurons in the input population. To hardwire the head-centered responses of each output neuron, the strength of each afferent synapse was set according to a Gaussian function in terms of the distance between the assigned head-centered location of the output neuron and the effective

preferred head-centered location of the input neuron. For each synaptic connection, this distance was given by

$$\frac{(\alpha + \beta) - h}{\sqrt{2}} \tag{3.1}$$

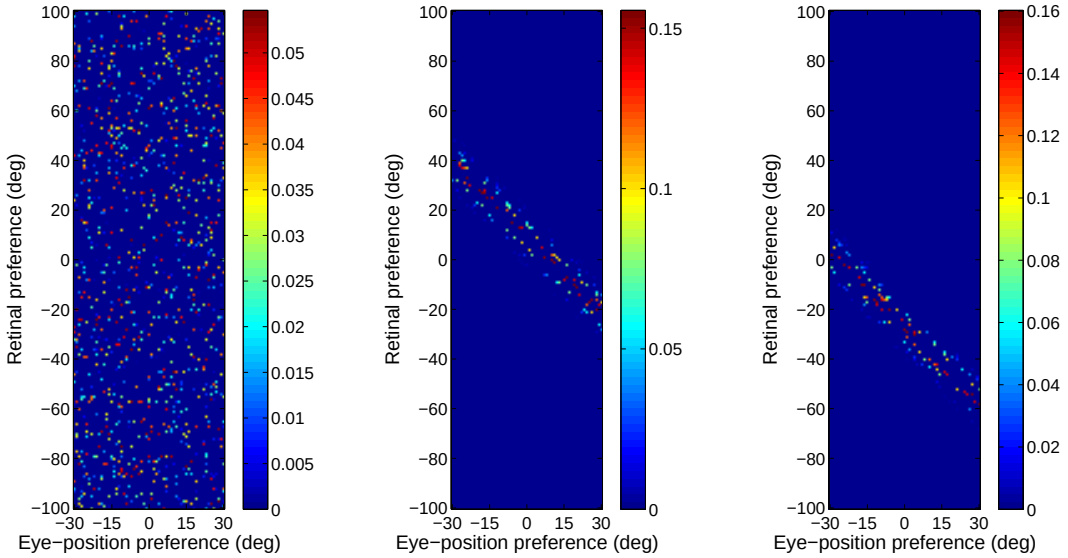
where the postsynaptic output neuron was assigned to head-centered location h and the presynaptic input neuron had preferred retinal location α and eye-position β. Specifically, the weight $w_{\alpha,\beta,h}$ assigned to such a synapse was set to

$$w_{\alpha,\beta,h} = \exp\left(-\frac{\|(\alpha + \beta) - h\|^2}{4\sigma^2}\right) \tag{3.2}$$

where σ was the width of the retinal receptive field. Then the weight vector for each output neuron was renormalized according to equation 2.7. Such a manually wired weight vector ensured that the degree of activation of the output neuron was related to the extent that a visual target was close to the preferred head-centered location of that neuron.

For the benefit of comparison, a network with random synaptic weights was tested the same way as a network with synaptic weights set according to equations 3.2 and 2.7, and these models are herein referred to as the randomly wired and manually wired models respectively.

Figure 3.1 shows the resulting weight vector of neuron #45 from the randomly wired model, as well as the weight vectors for the output neurons #62 and #530 from the manually wired model. The weight vector of the first neuron had no structure because



(a) Neuron #45 **(b)** Neuron #62 **(c)** Neuron #530

Figure 3.1: Synaptic weights for three different output neurons: (a) neuron #45 from the randomly wired model, (b) and (c) neurons #62 and #6 from the manually wired model that have been assigned to head-centered positions 9° and −27° respectively. Each plot show the afferent synaptic weights onto each output neuron from the population of input neurons. The synaptic have been arranged topographically by the effective preference of the input neuron for retinal location α_i and eye position β_i. The weight matrix of the first output neuron has no structure. However, the weight matrices of the latter two output neurons have large weights from input neurons with effective preferences for similar head-centered locations. This structure is apparent from the diagonal bands present in these plots.

the values of the weights were random. In contrast, the the weight vectors of the latter two neurons, each assigned to head-centered locations 9° and −27° respectively, had large weights from those input neurons that had preferences for similar head-centered locations as the output neurons.

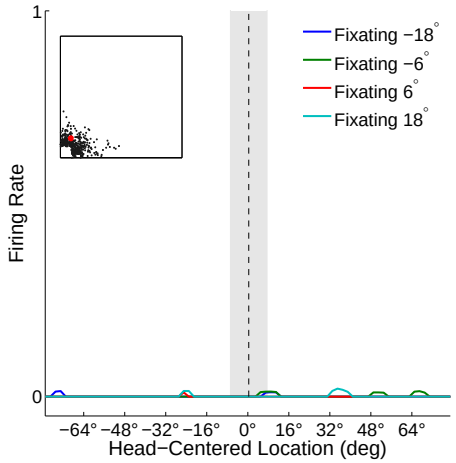
The response of these same three neurons during testing is seen in figure 3.2. While the response of neuron #45 from the randomly wired model exhibited no consistent structure in head-centered space across the different eye positions, neurons #62 and #6 from the

Parameter	**Symbol**	**Value**
Width of eye position tuning curve	ρ	3°
Width of retinal tuning curve	σ	3°
Output neuron population size	N	900
Input neuron population size		12261
Activation time constant	τ_h	100ms
Activation function slope	φ	6.5
Activation function threshold	θ	0.4
Sparseness percentile	π	70%
Connectivity rate	ϕ	8.16%

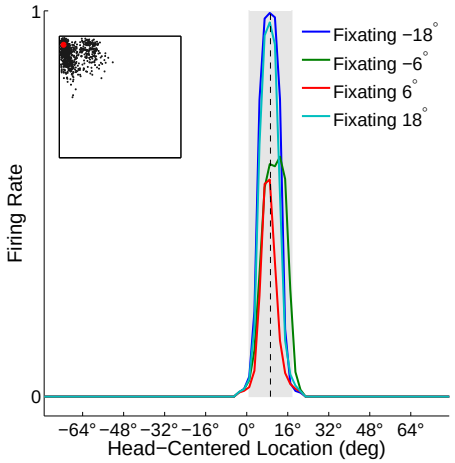
Table 3.1: Parameters of prewired model.

manually wired model responded maximally to the same head-centered location across all four eye positions. The performance metrics from section 2.8 were applied to these output neurons. Neuron #45 from the randomly wired model had head-centeredness ∼0.0757, eye-centeredness ∼0, head-centered receptive field location ∼0°, and size of head-centered receptive field ∼ 14°. While neurons #62 and #6 from the manually wired model had corresponding performance metrics ∼1, ∼0, ∼9°, ∼16° and ∼0.9, ∼0, ∼-25° and ∼19° respectively. This showed that the latter two neurons were responding in a head-centered frame of reference, while the former neuron did not. Moreover, there was good agreement about the receptive field location in head-centered space for these latter two neurons across their test responses, their weight vectors and how they were prewired. It can also be seen from figure 3.2 that there was also a degree of eye position modulation in the responses of these two head-centered neurons.

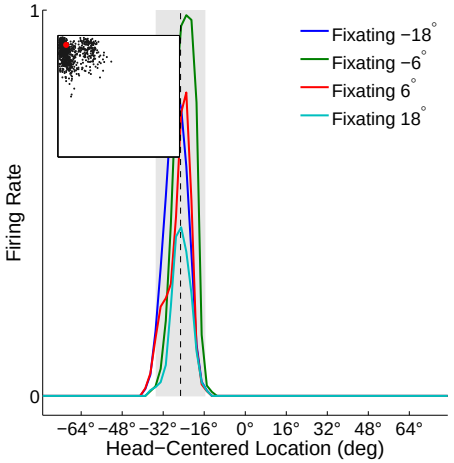
Figure 3.3 shows population analyses of response properties of output neurons in the



(a) Neuron #45



(b) Neuron #62



(c) Neuron #530

Figure 3.2: Firing rate responses during testing of three different output neurons: (a) neuron #45 from the randomly wired model, (b) and (c) neurons #62 and #6 from the manually wired model that have been assigned to head-centered positions 9° and −27° respectively. These correspond to the same neurons whose synaptic weights are shown in figure 3.1. Within each plot, each curve corresponds to a fixed eye position while a visual target is presented in different retinal locations. The vertical line shows the decoded head-centered receptive field location, and the grey bar shows the decoded receptive field size of the neuron. The miniature scatter plot shows the response characteristics of all neurons in the output layer, where each neuron is plotted as a point corresponding to that neuron's particular combination of head-centeredness (ordinate) and eye-centeredness (abscissa). The neuron whose firing rate responses have been plotted is shown in the scatter plot by a red mark.

randomly wired model and the manually wired model. Additional population statistics for these output cells are given in table 3.2.

Output neurons from the manually wired model had relatively large values of head-centeredness close to 1, while the values of eye-centeredness were near zero. In contrast, output neurons from the randomly wired model typically had both head-centeredness and eye-centeredness values close to zero (Fig 3.3a). The average receptive field index over all output neurons was larger in the manually wired model (0.92) compared to the randomly wired model (-0.01). The mode receptive field index in the randomly wired model was -1, reflecting an eye-centered response, which was reasonable since the output neurons were randomly wired to eye-centered input neurons. The prevalence of head-centered responses, as indicated by a positive index, was larger in the manually wired model ($\sim$99%) compared to the randomly wired model ($\sim$22%). Among these neurons, the average head-centeredness in the manually wired model (0.84) was also larger compared to the randomly wired model (0.06). In summary this showed that the manually wired model had more head-centered neurons, which were more compatible with a head-centered frame of reference.

During testing, it was found that the head-centered neurons in the manually wired model had decoded head-centered receptive fields that were clustered around one of the eight locations that the output neurons had been originally prewired to respond to, that is $-63°$,$-45°$,$-27°$,$-9°$,$9°$, $27°$,$45°$ and $63°$. There was also a similar distribution of head-

centered receptive field sizes for each of these eight head-centered locations. In contrast, the head-centered neurons in the randomly wired model had receptive fields covering head-centered space uniformly and a much wider range of receptive field sizes across head-centered space (Fig 3.3c,d). The average receptive field location was very close to zero in both models, both when considering the entire population of output neurons and when considering only head-centered neurons. This indicated that there was no lateralized bias in receptive field locations in either model. Among head-centered neurons in the manually wired model, there was a preference for every one of the eight prewired head-centered locations. Moreover, each head-centered location was represented so evenly in the manually wired model as to give near maximal coverage (0.998). While in the randomly wired model the two most eccentric locations ($-63°$ and $63°$) had no head-centered neurons preferring them. Lastly, among head-centered output neurons, the average receptive field size in the manually wired model was larger ($18.44°$) compared to the randomly wired model ($16.92°$). In summary, the manually wired model evenly represented the eight prewired head-centered locations with receptive field sizes clustered within the interval $[15°, 23°]$ (Fig 3.3c).

Lastly, figure 3.4 presents a further population analysis of the responses of the output neurons in the manually wired model. The plot shows the correlation between the head-centered receptive field locations that were assigned to the output neurons by hardwiring the afferent synaptic connection weights and the head-centered receptive field locations that were decoded during testing. These results showed an almost perfect correspondence

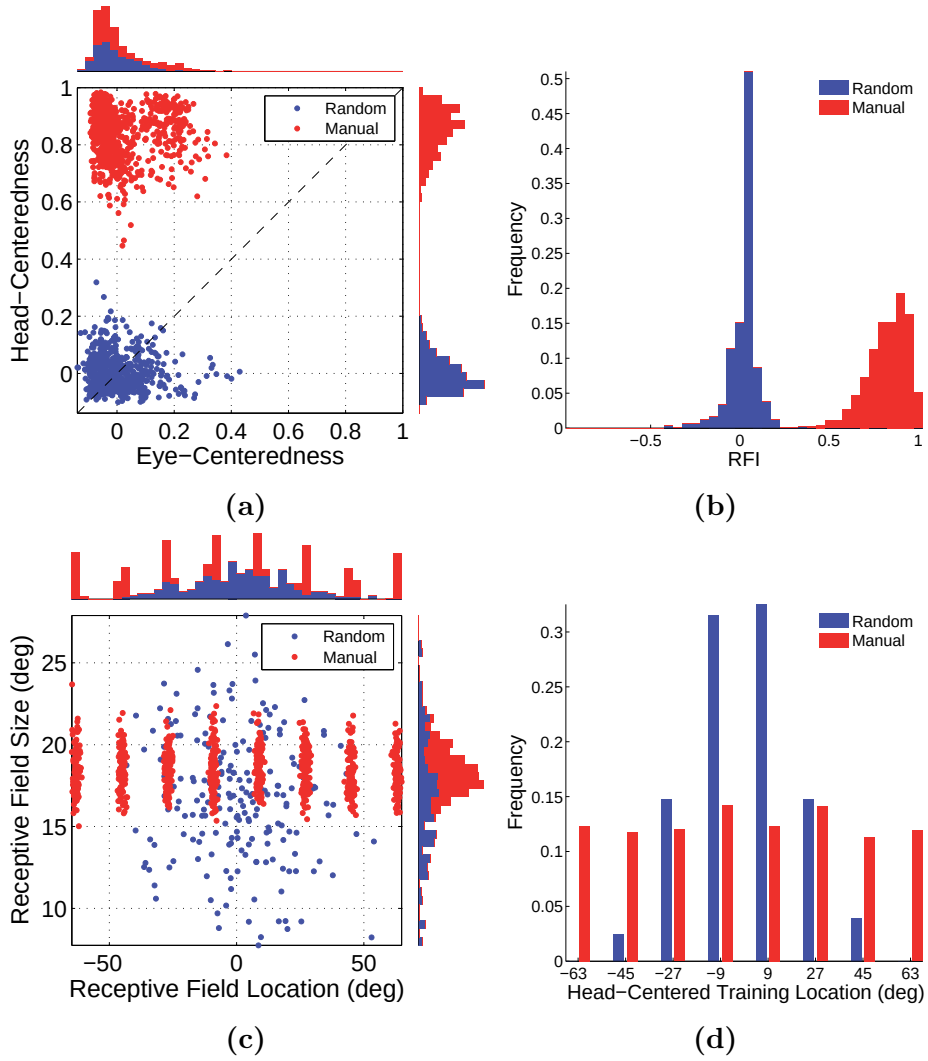


Figure 3.3: Population analyses of receptive field properties of output neurons in the randomly wired and manually wired models. (a) Scatter plot shows the reference frame response characteristics of all neurons in the output layer, where each neuron is plotted as a point corresponding to that neuron's particular combination of head-centeredness and eye-centeredness. Data points for the randomly wired model are plotted in blue, while results for the manually wired model are shown in red. (b) Distributions for receptive field index values in the two models. (c) Scatter plot showing the combination of head centered receptive field size and head-centered receptive field location of all head-centered output neurons for the two models. (d) Histograms showing the frequency distribution of the numbers of output neurons that responded preferentially to each of the head-centered locations which were used to hardwire the responses of the output neurons in the manually wired model.

Experiment 3.2.1				
	Random		**Manual**	
	All	RFI > 0 ($\sim$22%)	All	RFI > 0 ($\sim$99%)
Head-centeredness	0.02 (0.06)	0.06 (0.05)	0.84 (0.08)	0.84 (0.08)
Eye-centeredness	0.03 (0.09)	-0.04 (0.05)	0.01 (0.10)	0.01 (0.10)
RFI	-0.01 (0.08)	0.06 (0.05)	0.80 (0.11)	0.80 (0.11)
RF Location	0.04° (18.25°)	0.51° (19.50°)	-0.06° (40.55°)	-0.06° (40.55°)
RF Size	16.65° (3.78°)	16.92° (3.99°)	18.44° (1.28°)	18.44° (1.28°)

Table 3.2: Population summary statistics of response properties of output neurons in the randomly and manually wired models. Results for the randomly wired model are shown in the left two columns, while results for the manually wired model are shown in the right two columns. For each type of model, results are presented in two subcolumns: statistical measures computed over all output neurons are shown in the left subcolumn, while measures computed over neurons with a receptive field index greater than zero indicating head-centered responses are shown in the right subcolumn. Each row corrresponds to a different performance metric: head-centeredness, eye-centeredness, receptive field index, head-centered receptive field location, and head-centered receptive field size. Each cell of the table shows the mean and standard deviation (in parentheses) of the performance metric over the relevant population of output neurons.

of $r = 0.9998$, confirming that the receptive field location decoding worked well. This combined with the population analysis plots in figure 3.3, which produced outputs in perfect agreement with the deliberate hardwiring, showed that all metrics decoded the intended neuronal response property well.

3.2.2 Model with Learning

This experiment explored how neurons in the output layer may develop head-centered visual responses through updating of the synaptic weights by visually-guided learning with

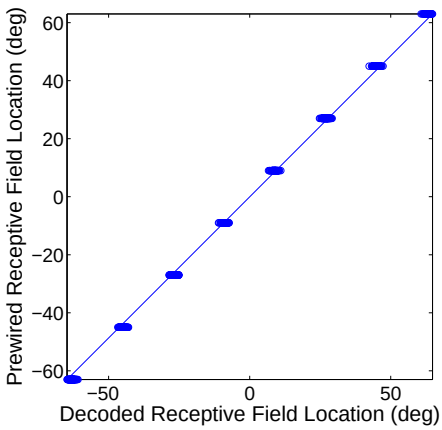


Figure 3.4: Scatter plot showing the correlation between the head-centered locations that output neurons in the manually wired model had been prewired to respond to, and the decoded head-centered locations of these neurons during testing. Each point in the scatter plot corresponds to an individual neuron in the output population, with the prewired head-centered location represented on the ordinate and the decoded head-centered location represented on the abscissa. It can be seen that the decoded head-centered receptive field locations of the output neurons were all very close to the head-centered locations that the neurons had been hardwired to respond to. The diagonal line has unit slope.

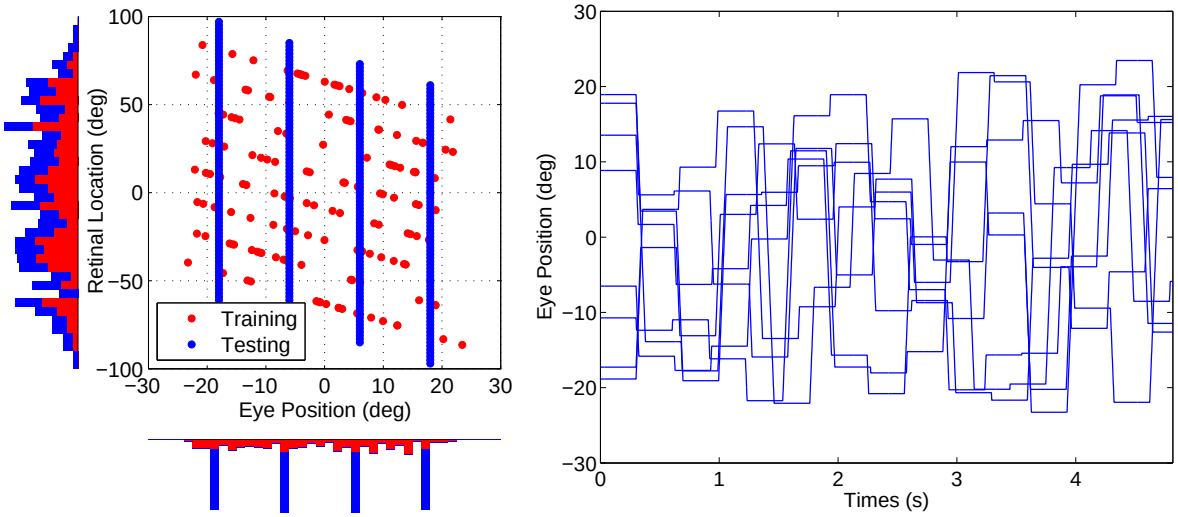
the trace learning rule 2.2 as hypothesised above in section 2.2. The firing rates of the input neurons were set with peaked eye position gain modulation as described by equation 2.1. The experiment provided a baseline investigation into the feasibility of the proposed hypothesis.

The model had 12261 neurons in the input population, and 900 neurons in the output layer. Each output neuron received 613 afferent synaptic connections from a randomly assigned subpopulation of the input population. At the beginning of training, the synaptic weights were set to random values. Then the synaptic weight vector of each of the output neurons was renormalized according to equation 2.7. The network was then trained for 20 epochs. During each training epoch, a visual target was presented for approximately

5s in each of the eight head centered training locations: $-63°, -45°, -27°, -9°, 9°, 27°, 45°$ and $63°$. For each period where the visual target was in a fixed head centered target location, the eye position was varied continuously through time as the model made a series of saccades and fixations. During each such period, the model performed 14 saccades interleaved with 15 fixations, where each fixation lasted 300ms. Each saccade was at a constant velocity of $400°/s$, and it was directed to a random eye position within the range $[-24°, 24°]$. Each training epoch thus lasted for approximately 40s, and the entire training of the network was completed after about 800s of simulated time. The model was tested as described above in section 3.2.1. Figure 3.5 shows the simulated movements of the eyes and head centered locations of visual targets during training and testing. The parameters for the model are given in table 3.3.

Figure 3.7 shows how the firing responses and synaptic weights of one the output neurons #79 develop during successive stages of training.

The responses of the output neuron prior to training exhibited no consistent structure in head-centered space across the different eye positions (Fig. 3.7a). However, at both 10 and 20 epochs there was a maximal response to the same head-centered location across all four eye positions (Fig. 3.7c,e), demonstrating more head-centered response characteristics. Before training, the neuron had head-centeredness $\sim$0.12, eye-centeredness $\sim$-0.2, receptive field location $\sim 10°$, and receptive field size $\sim 65°$. However, the corresponding metric values



(a) Fixations during training and testing **(b)** Eye movements during training

Figure 3.5: Simulated movements of the eyes and head-centered locations of visual targets during training and testing. (a) Scatter plot in which each point corresponds to a single fixation during either training (red) or testing (blue). The fixation points are plotted as a function of the eye position (abscissa) and the retinal location of the visual target (ordinate). Each of the diagonal lines of red points corresponds to a period during training when the visual target was fixed in one of the eight head-centered target locations while the eyes moved. The vertical lines of blue points correspond to the four eye positions in which the network was tested. (b) Multiple plots showing how the eye position is shifted through time in a randomised manner during training. Each plot corresponds to a different period during which the visual target is maintained in a fixed head centered location.

at 10 training epochs were ∼0.79, ∼0.13, ∼8° and ∼34°, indicating the development of head-centered responses. This became more pronounced at 20 training epochs, where the metric values were ∼0.78, ∼0.18, ∼11° and ∼25°, respectively.

There was also a correspondence between the weight vector of the output neuron and the response of the neuron during testing at each stage of training (Fig. 3.7b,d,f). Prior to training the afferent synapses had random values and no structure in terms of the relationship between the weight of a synapse and the characteristics of the presynaptic neuron. However, during training a clear diagonal structure developed in the synaptic

Parameter	Symbol	Value
Number of target locations	M	8
Fixation sequence length	P	15
Number of training epochs	-	20
Width of eye position tuning curve	ρ	6°
Width of retinal tuning curve	σ	6°
Output neuron population size	N	900
Input neuron population size		12261
Trace time constant	τ_q	400ms
Activation time constant	τ_h	100ms
Activation function slope	φ	4.5
Activation function threshold	θ	0.4
Sparseness percentile	π	80%
Learning rate	ρ	0.05
Synaptic connectivity	ϕ	5%

Table 3.3: Parameters of the model.

weights. The most potentiated synapses were those originating from input population neurons i which had a preference for a retinal location α_i and eye-position β_i corresponding to the head-centered location which the output neuron preferred. These particular input neurons lay on the diagonal line that is evident in the synaptic weights at 20 training epochs shown in Fig. 3.7f. This corresponded to the diagonal line in the with gradient -1 and retinal target location intercept r_0 at $e = 0$ equal to the head-centered location preferred by the neuron.

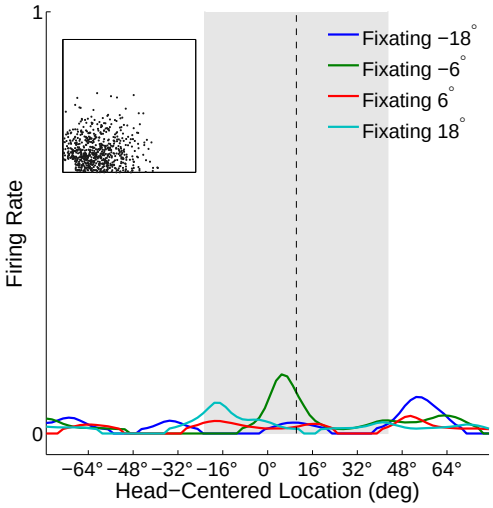
In summary these results showed that, after 10 and 20 epochs of training, the neuron responded in a head-centered frame of reference, while it did not prior to training. Moreover, at both 10 and 20 training epochs there was reasonably good agreement about the location of the receptive field in head-centered space according to the neuron's firing rate

responses and learned synaptic weights. There was also a degree of eye position modulation evident in the responses of this head-centered neuron.

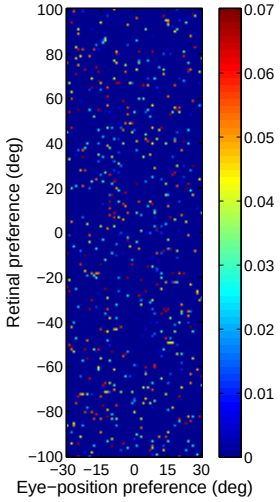
Figure 3.8 presents the population analyses of the receptive field properties of the output neurons after 0, 10 and 20 epochs of training, and population statistics are given in table 3.4. The model before training and after 20 epochs of training is herein referred to as the untrained and trained model respectively.

Most output neurons in the untrained model had head-centeredness values and eye-centeredness values clustered close to zero. However, many more neurons in the fully trained model had head-centeredness values clustered close to 1, with eye-centeredness values close to zero (Fig 3.8a). In particular, table 3.4 confirms that training the network led to an increase in the average head-centeredness over the population of output cells from 0.04 to 0.58. Also, the average RFI was increased from -0.38 to 0.27 with training. Before training the mode RFI was -1, or an eye-centered response. The proportion of neurons with a head-centered response, as indicated by a positive RFI, was increased from $\sim$26% to $\sim$69% during training. Among head-centered neurons with, the average head-centeredness was increased from 0.17 to 0.63 by training. In summary, these results showed that training the network had the effect of increasing the number of head-centered neurons, and also refined the response characteristics of individual neurons to be more compatible with a head-centered frame of reference.

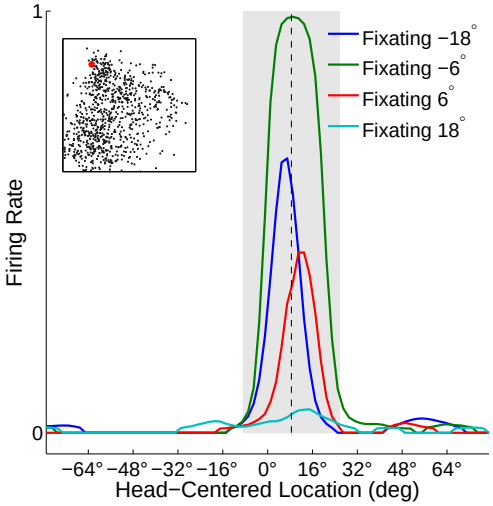
Head-centered neurons in the fully trained model had receptive fields clustered around



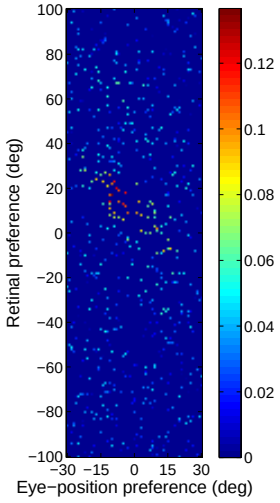
(a) Untrained



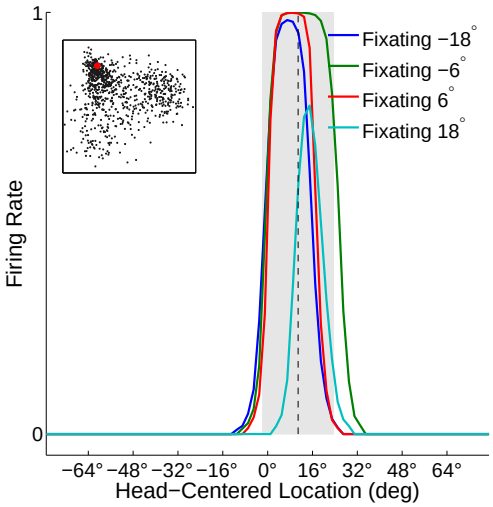
(b) Untrained



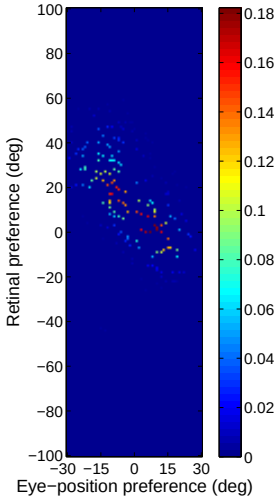
(c) Epoch 10



(d) Epoch 10



(e) Epoch 20



(f) Epoch 20

Figure 3.7: The development of the firing responses and synaptic weights of one the output neurons #79 during successive stages of training of the self-organizing model. Results are presented before training (top row), after ten training epochs (middle row), and after 20 training epochs (bottom row). Each plot in the left column shows the firing rate responses of the output neuron, where each curve corresponds to a fixed eye position while a visual target is presented in different retinal locations. Conventions as for figure 3.2. Each plot in the right column shows the afferent synaptic weights onto the output neuron from the population of input neurons. Conventions as in figure 3.1.

one of the eight head centered training locations, and there was a similar distribution of receptive field sizes for each of these head centered locations (Fig 3.8c,d). In contrast, the head-centered neurons in the untrained model had receptive fields covering head-centered space more uniformly and a much wider range of receptive field sizes across head-centered space. The average receptive field location was near zero for both the untrained and fully trained models, which indicated that there was no lateralized bias in receptive field locations in these models. However, while the fully trained model had head-centered neurons covering all eight head-centered training locations with a coverage of $\sim$0.96, the untrained model had no head-centered neurons for the two most eccentric head centered locations (-63° and 63°). Lastly, the average receptive field size decreased from 65.80° to 29.10° during training. Thus, in summary, training the network developed head-centered neurons covering head-centered space, and also increased neuronal selectivity by reducing the sizes of the head centered receptive fields.

Experiment 3.2.2						
	Untrained		**10 Epochs**		**20 Epochs**	
	All	RFI > 0 ($\sim$26%)	All	RFI > 0 ($\sim$59%)	All	RFI > 0 ($\sim$69%)
Head-centeredness	0.04 (0.14)	0.17 (0.12)	0.37 (0.25)	0.47 (0.22)	0.58 (0.19)	0.63 (0.16)
Eye-centeredness	0.18 (0.16)	0.04 (0.12)	0.29 (0.21)	0.21 (0.16)	0.36 (0.24)	0.25 (0.15)
RFI	-0.38 (0.71)	0.64 (0.37)	0.09 (0.53)	0.44 (0.29)	0.27 (0.37)	0.46 (0.23)
RF Location	0.09° (16.77°)	$-0.22°$ (18.46°)	$-0.59°$ (32.02°)	$-0.92°$ (30.59°)	$-1.76°$ (40.72°)	1.02° (34.45°)
RF Size	65.80° (7.15°)	66.43° (6.72°)	43.00° (14.19°)	40.69° (13.65°)	29.10° (7.78°)	28.61° (7.29°)

Table 3.4: Population summary statistics of response properties of output neurons in the model after successive stages of training. Results are presented before training (left two columns), after ten training epochs (middle two columns), and after 20 training epochs (right two columns).

The output neurons in the model update their afferent synaptic connections over the course of 20 training epochs by visually-guided competitive learning (Rolls & Treves 1998). In such a model, it is important to investigate whether the learning process converges and the response characteristics of the output neurons settle down asymptotically to some form of stable behaviour. The impact of each successive training epoch on the receptive field properties of output neurons was examined by plotting key summary statistics as a function of training epoch in figure 3.9. It could be seen that the fraction of head-centered output neurons, the average head-centeredness among head-centered neurons, and the coverage of the head centered training locations all increased close to monotonically during training. Also, the average head-centered receptive field size decreased monotonically during training. Most importantly, it was found that these summary statistics converged on steady values after further training. Thus, the key performance characteristics of the model developed close to monotonically in the desired way as the number of training epochs increased.

3.2.3 Control Experiments: Demonstrations of the Necessity of the Core Model Components

This set of experiments explored the contributions of the four core model components described in section 2.2 that were hypothesised to support the development of head-centered neurons.

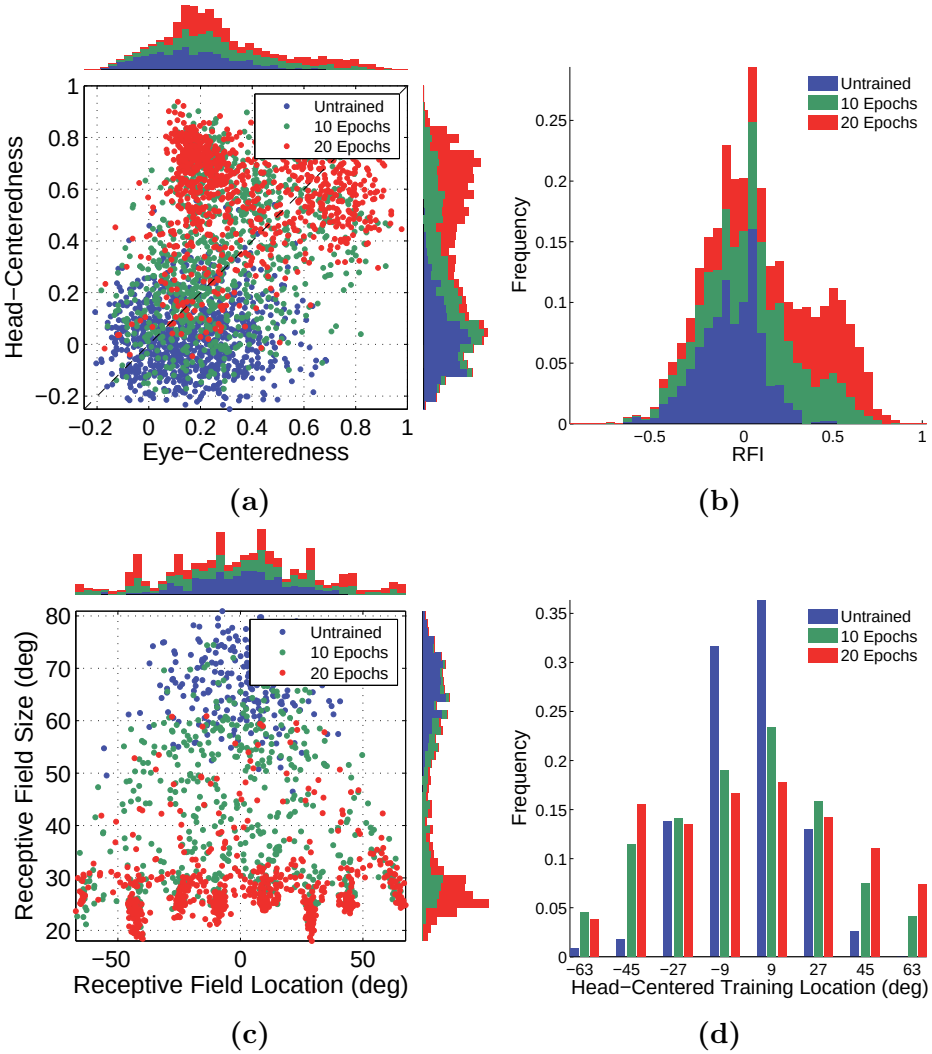


Figure 3.8: Population analyses of receptive field properties of output neurons during succcessive stages of training of the self-organizing model. Results are presented before training, after ten training epochs, and after 20 training epochs. Conventions as in figure 3.3.

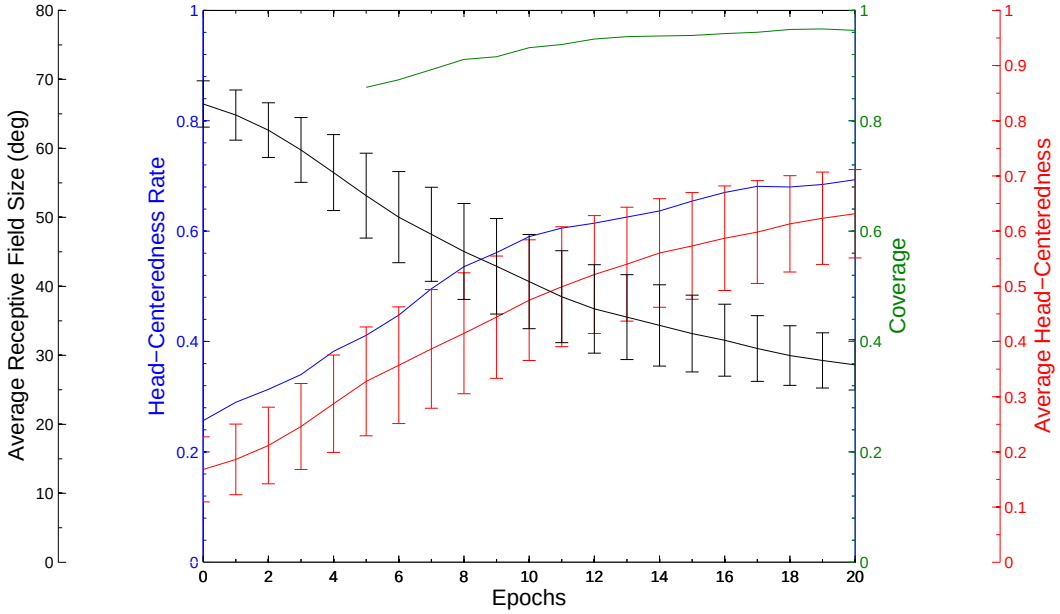


Figure 3.9: Population analyses of receptive field properties of output neurons in the model during successive training epochs. There are four plots as follows. The average receptive field size curve (black) shows the average size of the head centered receptive field among head-centered neurons, and the error bars represent the standard deviations. The head-centeredness rate (blue) was the fraction of output neurons that were head-centered. The coverage curve (green) was the coverage of the head-centered training locations by the output neuron population after the given number of epochs of training, where missing data points before epoch 5 were due to at least one of the eight head-centered training locations not being represented by the output cells. The average head-centeredness curve (red) was the average head-centeredness value among all head-centered neurons, and the error bars were the standard deviations.

Input Neurons with Coupled visual and Eye Position Receptive Fields

It was hypothesised that the model required input neurons with coupled visual and eye position receptive fields in order to be able to develop head-centered output neurons. Examples of such coupled receptive fields are given by equations 2.1 and 2.2. In this experiment the necessity of the this premise was investigated by exploring how decoupling

the visual and eye position components of the receptive fields of input neurons would affect the learning in the model. In order to decouple the visual and eye position components, half of the input neurons were set to respond purely to the retinotopic location of a visual target, while the other half of the input neurons were set to respond purely to the eye position. The experiment otherwise had the same parameters as used in section 3.2.2.

The expected result was that with decoupled receptive fields in the input population the output layer would be unable to form head-centered representations. The reason for this is that the single layer of synapses between the input layer and output layer would not be able to implement a suitable mapping because all of the input neurons would by definition participate in encoding any given location in head-centered space. This problem is solved in a model with input neurons with coupled visual and eye position receptive fields because the individual input neurons have more selective responses that correspond to particular locations in the head centered space. In this case, it is possible to find a set of synaptic weights that implement a mapping from such input neurons to head centered output neurons, which may be achieved manually as described in section 3.2.1 or set up by learning as demonstrated in section 3.2.2.

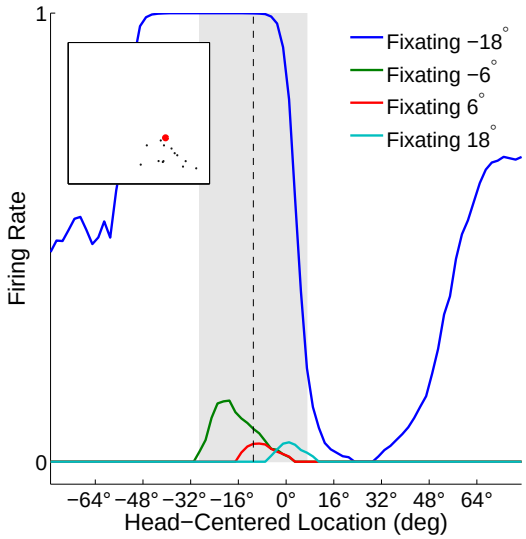
Table 3.5 presents typical population summary statistics of the response properties of output neurons from the model with input neurons with decoupled visual and eye position receptive fields. The statistics were computed over subsets of output neurons for which the head-centeredness and eye-centeredness metrics were mathematically defined. This meant

that 10% and ∼98% of neurons had to be discarded from further analysis in the untrained and trained model respectively. The table shows that the average head-centeredness did increase from 0.03 to 0.12 after training. However, the greatest head-centeredness value found among the trained output neurons was only ∼0.26. Moreover, all output neurons had a negative RFI, indicating an eye-centered response. Figure 3.11 presents the firing responses and synaptic weights of one of the output neurons #170 from the model. The results are shown after training. It can be seen that the neither the firing responses of the output neuron nor the afferent synaptic weights were consistent with a head centered receptive field.

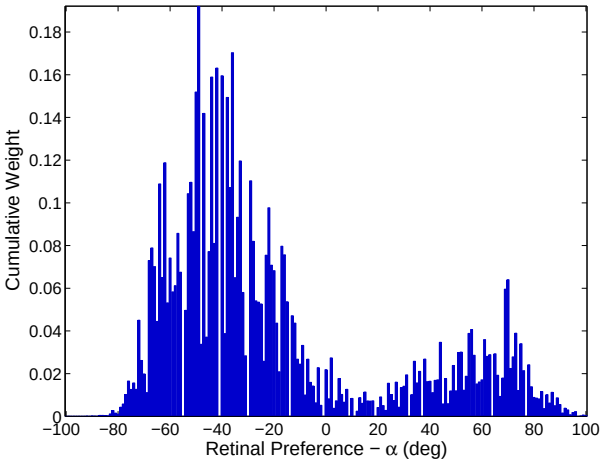
The results described here were typical over a broad range of model parameter values. Thus, these results demonstrated that decoupling the visual and eye position receptive fields of input neurons prevented the development of head-centered output neurons during training, and in fact had the opposite effect.

Competitive Interactions between Output Neurons

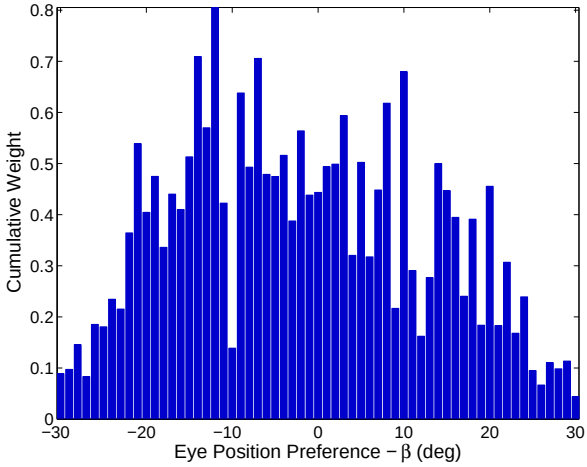
It was hypothesised that the model required competitive interactions between the output neurons in order to develop head-centered representations in the output population. The function of such competitive interactions is to ensure that only a small subset of output neurons remain active at any time. The effect of this, when combined with some form of associative synaptic learning, is to encourage individual output neurons to learn to respond highly selectively to distinct subsets of input patterns, with different output neurons learn-



(a) Response



(b) Synaptic weight distribution over visual input neurons



(c) Synaptic weight distribution over eye position input neurons

Experiment 3.2.3		
	Untrained	**Trained**
Head-centeredness	0.03 (0.14)	0.12 (0.08)
Eye-centeredness	0.92 (0.08)	0.68 (0.12)
RFI	−0.87 (0.18)	−0.69 (0.18)
RF Location	−0.66° (25.03°)	3.08° (34.26°)
RF Size	56.06° (11.23°)	28.79° (9.40°)

Table 3.5: Population summary statistics of response properties of output neurons in the model with decoupling of the visual and eye position components of the receptive fields of input neurons. Results are given before training (left column) and after training (right column). Each row corresponds to a different performance metric: head-centeredness, eye-centeredness, receptive field index, head-centered receptive field location, and head-centered receptive field size. Each cell of the table shows the mean and standard deviation (in parentheses) of the performance metric over a subset of output neurons described in the text.

Figure 3.11: Analysis of one of the output neurons #170 from the model with input neurons with decoupled visual and eye position receptive fields. Results are presented after training. (a) The firing rate responses of the output neuron, where each curve corresponds to a fixed eye position while a visual target is presented in different retinal locations. Conventions as for figure 3.2. (b) and (c) histograms of afferent synaptic weights onto the output neuron from input cells that represent the retinal target location and eye position, respectively. The histograms in (b) and (c) were produced by finding the sum of all synaptic weights from input neurons with the given retinal location (α) or eye position preference (β) respectively. Both histograms have bin sizes of 1°.

ing to respond to different subsets. The subsets of input patterns that the output neurons learn to represent reflect natural groupings within the space of input patterns, and may also depend on the form of associative learning rule used. This kind of learning process is known as *competitive learning* (Rolls & Treves 1998).

In this experiment, the necessity of competitive interactions between output neurons was investigated by exploring how turning off these competition interactions affected the

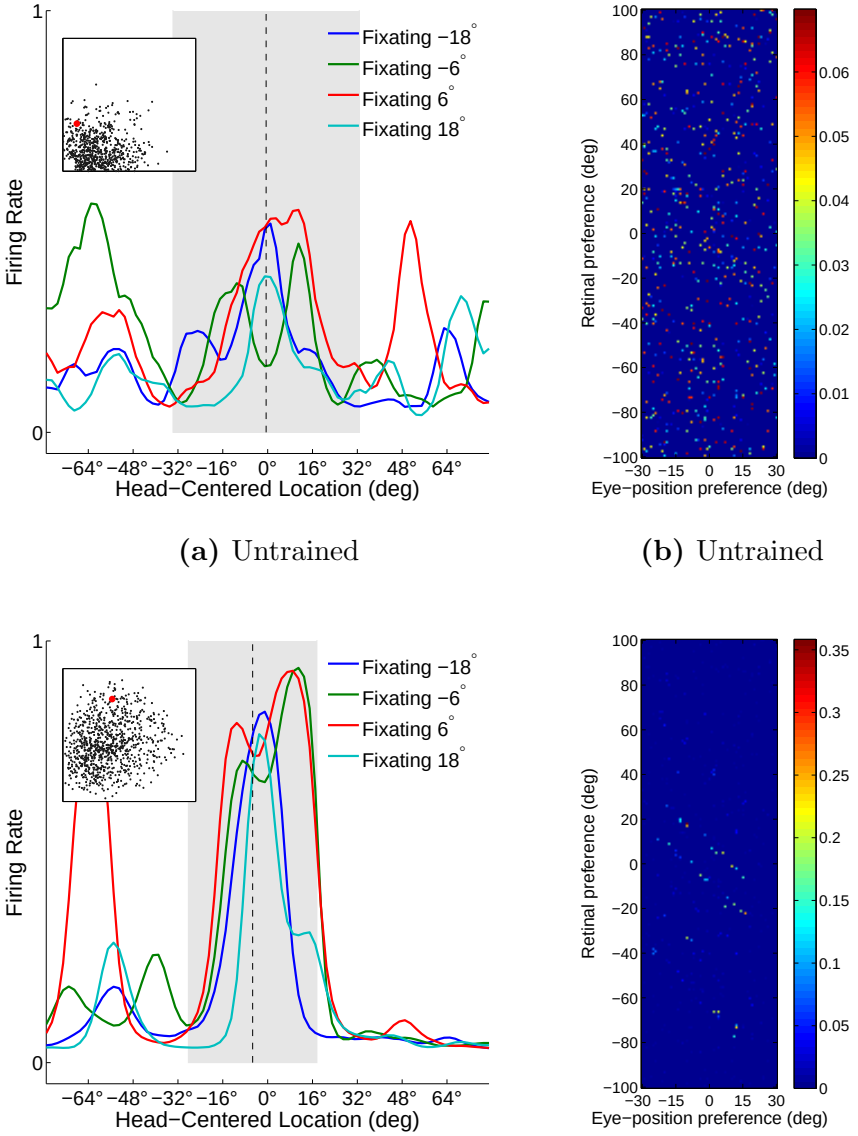
model. In previous experiments, the competitive interactions between output neurons were mediated by a dynamically adjusted response threshold p_π ensuring that all neurons with activation less than the π^{th} percentile of the activation distribution would fall below the sigmoidal response threshold, as specified in equation 2.4. In the simulations described next, turning off the competitive interactions between output neurons was achieved by setting the activation threshold to $\pi = 0$. This effectively permitted all of the output neurons to remain active. The experiment otherwise had the same parameters as used in section 3.2.2.

Results from a simulation without competitive interactions between output neurons are given in figure 3.12, which shows the firing responses and synaptic weights of output neuron #409 before and after training. Prior to training the neuron responded to a large portion of head-centered space at all eye positions (Fig. 3.12a). Although there was a weak response in the center of head space across all eye positions. This was reflected in head-centeredness and eye-centeredness values of ∼0.30 and ∼0.015 respectively, indicating a modest head-centered response. The size of the receptive field before training was ∼66°. After training the responses were more coordinated across all eye positions (Fig. 3.12c), with a much larger response localised in the middle of the head-centered space. The head-centeredness and eye-centredness values were ∼0.75 and ∼0.30 respectively, showing that training had significantly increased the head-centeredness of the neuron. The weight vector (Fig. 3.12d) reflected this as well. However, unlike previous experiments, the receptive field size of this

neuron after training remained very large, approximately 46°. Thus, although the neuron was head-centered, the head-centered receptive field was so large that the neuron would not convey much information about the location of a visual target in the head-centered space. This finding was also typical of other neurons in the output layer after training without competitive interactions between output neurons.

Further results from the same simulation are given in figure 3.13, which presents the population analyses of the receptive field properties of the output neurons before and after training. Population summary statistics for the simulation are given in table 3.5. The head-centeredness rate increased from ∼0.25 to ∼0.67 with training, and among head-centered neurons the average head-centeredness increased from 0.2 to 0.44. However, unlike previous experiments that developed highly selective head centered output neuron responses, the average receptive field size among head-centered neurons was 46.74°, which was ∼68% larger than the corresponding receptive field size 28° in experiment 3.2.2 which included competition between output neurons.

In summary, these results show that, while the model without competition between output neurons is capable of producing more head-centered output neurons after training, these output neurons actually develop much larger receptive fields in head centered space. This means that such output neurons would in fact convey much less information about the head-centered location of a visual target than output neurons from a model that incorporated competition within its output layer.



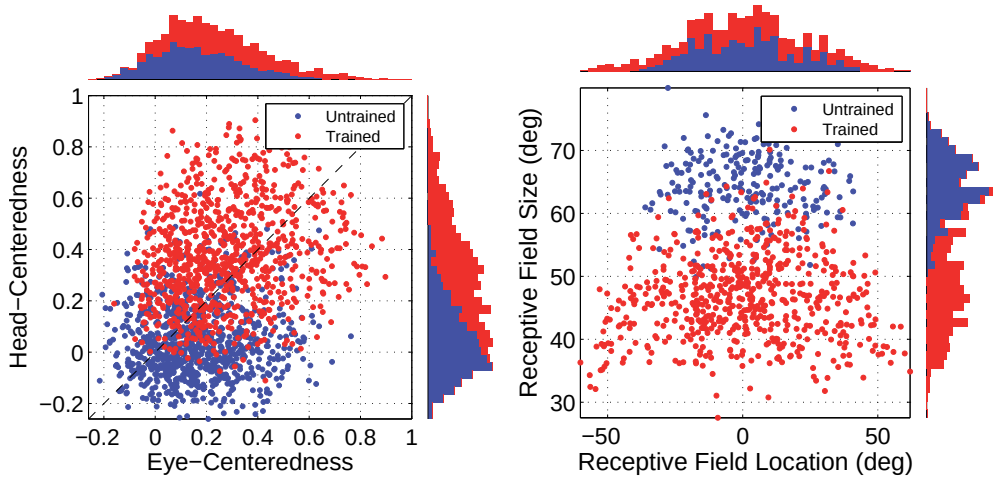
(a) Untrained **(b)** Untrained

(c) Trained **(d)** Trained

Figure 3.12: Results from a simulation without competitive interactions between output neurons. The figure shows the firing responses and synaptic weights of one the output neurons #409 before training (top row) and after 20 training epochs (bottom row). Each plot in the left column shows the firing rate responses of the output neuron, where each curve corresponds to a fixed eye position while a visual target is presented in different retinal locations. Conventions as for figure 3.2. Each plot in the right column shows the afferent synaptic weights onto the output neuron from the population of input neurons. Conventions as in figure 3.1.

Experiment 3.2.2				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$25%)	All	RFI > 0 ($\sim$67%)
Head-centeredness	0.06 (0.14)	0.20 (0.12)	0.38 (0.19)	0.44 (0.18)
Eye-centeredness	0.18 (0.16)	0.05 (0.12)	0.27 (0.20)	0.20 (0.15)
RFI	-0.38 (0.70)	0.61 (0.37)	0.19 (0.45)	0.43 (0.30)
RF Location	$-0.01°$ (16.06°)	1.23° (17.93°)	$-2.01°$ (24.18°)	0.35° (25.17°)
RF Size	64.47° (5.21°)	64.28° (5.06°)	46.96° (6.68°)	46.74° (6.77°)

Table 3.6: Simulation results without competitive interactions between neurons in the output layer. Population summary statistics of response properties of output neurons are shown before training (left two columns) and after training (right two columns). For each of these two stages of training, statistical measures computed over all output neurons are shown in the left subcolumn, while measures computed over neurons with a receptive field index greater than zero are shown in the right subcolumn. Conventions are similar to table 3.4.



(a) **(b)**

Figure 3.13: Simulation results without competitive interactions between output neurons. Population analyses of the receptive field properties of output neurons are presented before training (top) and after training (bottom). Conventions as in figure 3.3. (a) Scatter plot showing the reference frame response characteristics of all neurons in the output layer. (b) Scatter plot showing the combination of head centered receptive field size and head-centered receptive field location of all head-centered output neurons.

Temporal Binding

It was hypothesised that the model required a synaptic learning rule that incorporated a memory trace of recent neuronal activity in order to encourage output neurons to bind together input patterns that tended to occur close together in time. If, for much of the time, visual targets tend to remain fixed with respect to the head while the eyes move, then such a trace learning rule will encourage individual output neurons to learn to respond when the visual target is in a particular head-centered location regardless of the position of the eyes and hence the retinal location of the target. If the memory trace is removed from the learning rule, then this temporal binding cannot occur which should lead to a

failure of the output layer to develop head centered representations.

A memory trace can be incorporated into the synaptic learning rule in a number of alternative ways (Rolls & Milward 2000, Stringer & Rolls 2002). Equation 2.6 gives an example of a learning rule in which an explicit trace term $q_i(t)$ has been incorporated.

However, an alternative, and even simpler approach, is to use a standard Hebbian learning rule

$$\frac{dw_{ij}}{dt} = \varrho v_i v_j^I \tag{3.3}$$

where $w_{ij}(t)$ is the synaptic weight from presynaptic neuron j to postsynaptic neuron i, v_j and v_i^I are the firing rates of the pre- and postsynaptic neurons respectively, and ϱ is the learning rate. The hebbian learning rule 3.3 is combined with synaptic weight normalization 2.7 to prevent unbounded growth of the synaptic weights during training. If the time constant τ_h governing the activations of the pre- and postsynaptic neurons in equation 2.3 is increased, then this will lengthen the period of time taken for the activations and hence firing rates of these neurons to decay. In this case, the sustained neuronal activity effectively provides an implicit memory trace in the Hebbian learning rule 3.3 that can promote temporal binding of input patterns that occur close together in time.

The necessity of a memory trace in the synaptic learning rule was investigated in two sets of simulations. In the first set of simulations, the trace rule 2.6 was tested. Here the duration of the memory trace was varied by varying the time constant τ_q of the neuronal trace q_i in equation 2.5 over three orders of magnitude. In the second set of simulations,

the Hebbian learning rule 3.3 was tested. In this case, the duration of the effective memory trace was controlled by varying the neuronal activation time constant τ_h in equation 2.3 over the same range.

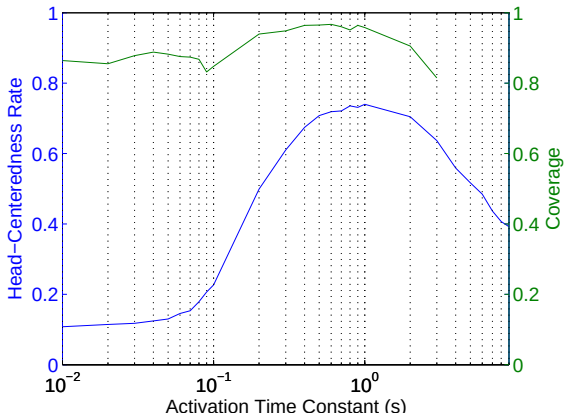
In both sets of simulations, it was expected that decreasing the relevant time constant, τ_q or τ_h, should reduce temporal binding by the output neurons. This is because the neuronal activity variables, that is the trace q_i and activation h_i, used in the two learning rules would reflect more recent neuronal activity and be less able to retain a memory of previous activity. This should lead to temporal binding over a shorter time window, and therefore retard the ability of output neurons to bind together and represent temporally proximal input patterns, which was hypothesized to be required for the development of head-centered output neurons. Thus, reducing the time constants τ_q or τ_h was expected to degrade the ability of the output layer to develop head-centered representations. The simulations otherwise had the same model parameters as in section 3.2.2.

The two sets of simulations also aimed to investigate the relative efficacies of the trace learning rule 2.6 and hebbian learning rule 3.3 as mechanisms for the temporal binding of subsets of input patterns that occur close together in time.

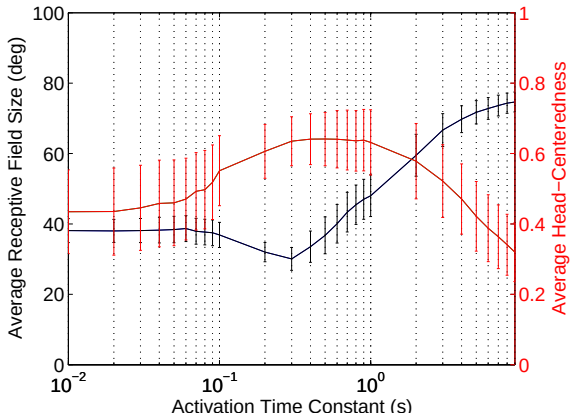
Figure 3.14 shows the effects of varying the length of the activation time constant τ_h and the trace time constant τ_q in the hebbian learning rule 3.3 and trace learning rule 2.6, respectively. The impact of varying the relevant time constant on the characteristics of the model was investigated by plotting key summary statistics as a function of the given time

constant.

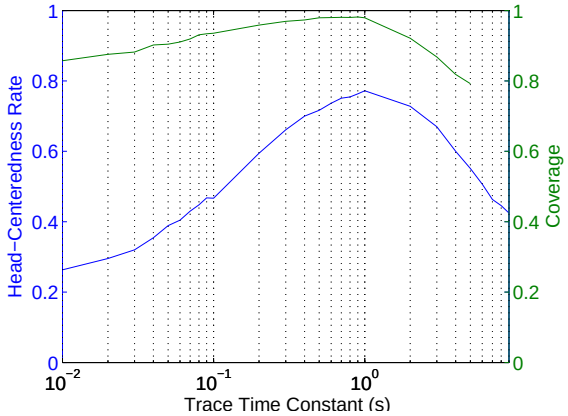
The observations from the simulations with the Hebbian learning rule were as follows. For time constant τ_h greater than 100ms, the head-centeredness rate was above the baseline rate in the untrained model, which was $\sim$25%. The untrained rate did in theory potentially depend on the time constant as well, but in practice it did not. The head-centeredness reached a maximum value of 74% at $\tau_h = 1$s. For τ_h less than 4s, the coverage did not drop below $\sim$0.81. There was no coverage for the untrained model at any time constant value. The average head-centeredness among head-centered neurons remained above the untrained average of $\sim$20% across the entire range of τ_h, peaking at $\sim$64% when $\tau_h = 500$ms. However, there was only a small difference among all after training values, irrispective of the time constant. In summary, this showed that across a wide range of values of the activation time constant τ_h the prevalence of head-centered neurons and the compatibility of their responses with a head-centered frame of reference were increased as a result of training. Although, for extreme values of τ_h in either direction the model performed worse. However, among the head-centered neurons that remained after training, shortening the time constant did not have a large negative impact on the average head-centeredness. The results of the simulations with a trace learning rule were qualitatively similar.



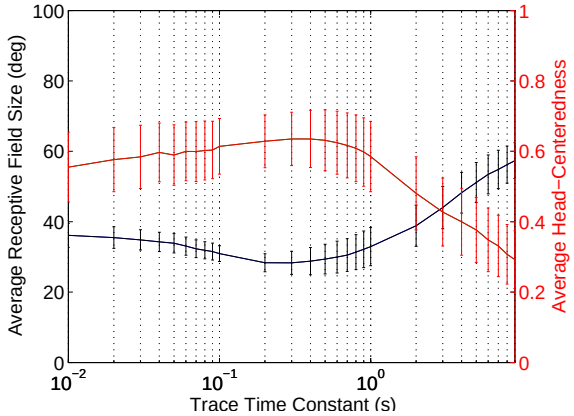
(a) Activation time constant



(b) Activation time constant



(c) Trace time constant



(d) Trace time constant

Figure 3.14: Simulations exploring the effects of varying the length of the activation time constants τ_h and the trace time constant τ_q in the Hebbian learning rule 3.3 and trace learning rule 2.6, respectively. The top row shows the series of simulations where a hebbian learning rule was used and the activation time constant was varied. The bottom row shows the series of simulations where a trace rule was used and the trace time constant was varied.

Movement Statistics of Eyes, Head and Visual Targets

It was hypothesised that the model required that for some periods of time the visual target remained stationary in head-centered space while the eyes moved in order for output neurons to develop head-centered responses using the trace learning rule. This would cause

different input patterns corresponding to a single head-centered target location to be clustered together in time. However, if the number of eye fixation positions P within each such period was reduced, then the trace learning rule would be prevented from binding together input patterns corresponding to the same head centered target locations. This would degrade the ability of the output layer of the model to form head-centered representations.

This experiment investigated the necessity of saccading between a sufficiently large number of successive eye position fixations for each fixed head-centered target location during training. It was expected that the output neurons would develop more strongly head-centered responses as the number of eye fixation positions P for each head centered target location was increased. More importantly, it was anticipated that the trained model would fail to produce output neurons with head-centered responses, in comparison to the untrained network, as P was reduced to 1. The sequence length was varied from $P = 1$ to $P = 11$, and the simulations otherwise had the same parameters as experiment 3.2.2.

The effects of varying the length of the fixation sequence for each fixed head centered target location on the performance of the trained model was investigated by plotting key summary statistics as a function of P in figure 3.15. Most importantly, the head-centeredness rate and the average head-centeredness increased almost monotonically with the length of the fixation sequence. Moreover, for all $P > 1$ the head-centeredness rate and average head-centeredness were greater than the corresponding values for the untrained model, which were $\sim$26% and $\sim$0.17, respectively. For all $P > 1$, there was always cover-

age, which was never less than ∼0.89, and which remained stable for all fixation sequence lengths. Like previous experiments, the average receptive field size among head-centered neurons decreased as head-centeredness rates increased, and remained stable for large values of P.

These simulations confirmed that the output neurons developed more head centered responses as the length, P, of the fixation sequence for each fixed head centered target location was increased. For the the shortest possible fixation sequence $P = 1$ the trained model had a lower head centeredness rate than the untrained model. These observations confirmed that for output neurons to develop head centered responses, the eye position must move through a sufficiently large number, $P > 1$, of successive fixations while the visual target remains in a fixed head-centered location.

3.2.4 Effects of Varying Network Parameters on Model Performance

The following experiments investigated how three key parameters of the model influenced the response characteristics of the output neurons after training. The simulations investigated the functional significance of these parameters and the robustness of the model's behaviour as the parameters were varied. The parameters that were explored were the learning rate ρ, sparseness percentile π which governed the level of activity in the output layer, and the number of neurons in the output layer N. The values of these three

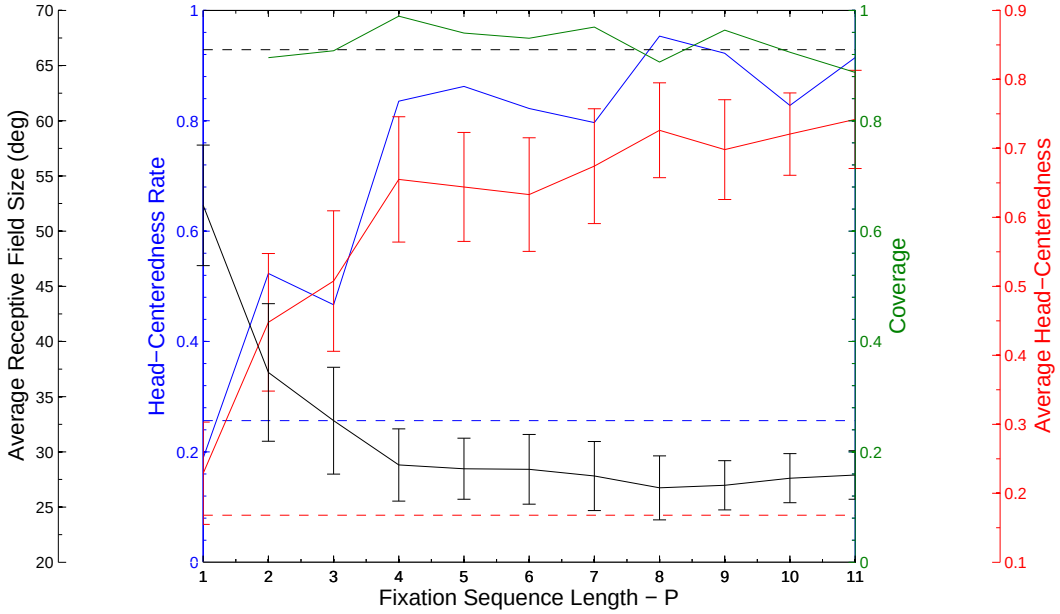


Figure 3.15: Simulations exploring the effects of varying the number, P, of eye fixation positions for each fixed head centered location of the visual target during training. Results are presented showing the response characteristics of the output neurons after 20 epochs of training. Conventions similar to figure 3.9. The dashed lines represent the corresponding values for the untrained network.

parameters were varied independently in turn while the remaining parameters remained fixed.

Learning Rate

This experiment investigated how varying the learning rate (ρ) influenced the learning in the model. The learning rate controlled the speed of synaptic plasticity, where a larger value promoted faster changes to the synaptic weights. It was expected that for sufficiently small learning rates, the network behaviour would remain fairly static and the output neurons would fail to develop head-centered responses during the finite amount of training time

available. On the other hand, with a sufficiently large learning rate, it was expected that the system would exhibit memoryless characteristics due to immediate overwriting of past learning. The process of forgetting would be enhanced by the long term depression (LTD) of synaptic weights due to the renormalization of synaptic weight vectors as described by equation 2.7. This, again, would prevent output neurons from developing head-centered responses. The model was thus expected to offer robust development of head-centered output neurons only when there was an appropriate intermediate balance between synaptic learning and memory stability. The learning rate was varied over three orders of magnitude. The simulations otherwise had the same parameters as experiment 3.2.2.

The effects of varying the learning rate ρ on the response characteristics of the output cells was investigated by plotting key summary statistics as a function of ρ in figure 3.16. The head-centeredness rate of the output neurons from the trained model was greater than the value of $\sim$26% for the untrained model over the entire range of learning rates. The head-centeredness rate initially increased monotonically to a peak value of $\sim$88% when $\rho = 1$, after which it collapsed. For the range of learning rates at which this collapse occurred there was no coverage. However, for all other values of $\rho > 0.01$, the coverage did not fall below $\sim$0.83. The average head-centeredness among head-centered neurons depended on the learning rate in a similar qualitative manner to the head-centeredness rate, peaking at $\sim$0.65 when $\rho = 0.2$. Like previous experiments, the average receptive field size among head-centered neurons was inversely related to their average head-centeredness.

In summary, the model was most effective at producing head centered output neurons for intermediate values of the learning rate. The performance of the model was fairly robust for learning rates spread over an order of magnitude from approximately 0.1 to 1. Below this intermediate range of learning rates, there was too little learning taking place to sufficiently alter the synaptic weight structure and thus modify the response properties of the output neurons. Alternatively, if the learning rate was too high, then the output neurons became forgetful of previously learned input patterns due to fast updating and overwriting of the synaptic weights with LTD due to weight renormalization.

Sparseness Percentile

This experiment investigated how varying the sparseness percentile, π influenced the learning in the model. The sparseness percentile governed the level of competition between output neurons. A dynamically adjusted response threshold p_π was used to control competition within the output layer. This ensured that all output neurons with an activation less than the π^{th} percentile of the activation distribution would fall below the response threshold of the sigmoid activation function 2.4. Thus, the sparseness percentile, π, controlled the degree of competition by silencing those output neurons with an activation u_i value below the given percentile value. Increasing the percentile π increased competition, while decreasing π decreased competition.

It was expected that the head-centerdness rate would decrease with increasing sparseness percentile because a large sparseness percentile imposes stricter competition, which

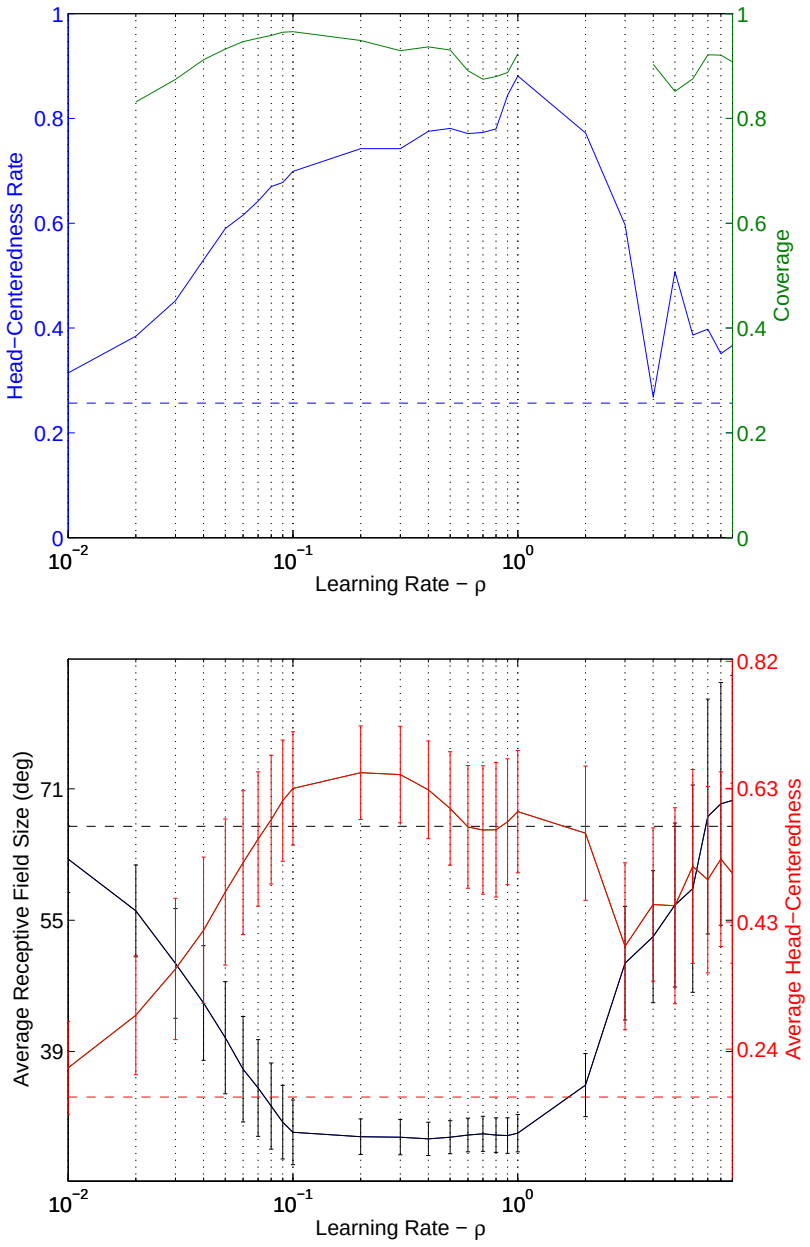


Figure 3.16: The effects of varying the learning rate, ρ, on the response characteristics of the output neurons. Results are presented after 20 epochs of training. Conventions similar to figure 3.9.

should reduce the number of neurons that are able to develop a head-centered response. For a similar reason, it was also anticipated that a larger sparseness percentile would lead to smaller head-centered receptive field sizes as individual neurons were inhibited from responding to such large regions of the head-centered space. The sparseness percentile was varied from 50% to 98% in increments of 2%. The simulations otherwise had the same parameters as experiment 3.2.2.

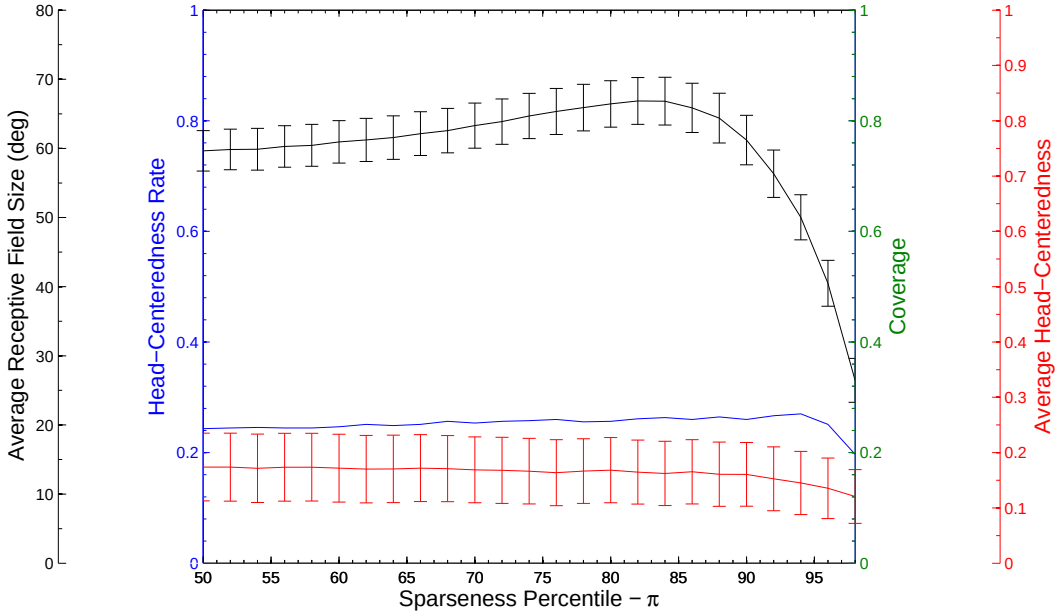
Figure 3.17 shows the effects of varying the sparseness percentile π on the firing properties of output neurons in both the untrained and trained models by plotting key summary statistics as a function of π. For the untrained model, the head-centeredness rate and average head-centeredness remained relatively low for all values of π, and there was no coverage at any value of the spareness percentile. The trained model performed better in terms of head-centered output representations than the untrained model across every performance metric for nearly all percentile values. For the trained model there was coverage at all percentiles, and the value of the coverage remained stable near the maximum value attained of $\sim$99%. The head-centeredness rate was larger in the trained model than the untrained model for all percentile values less than 95%. The average head-centeredness among head-centered neurons in the trained model was larger than in the untrained model for all percentile values. Lastly, the average head-centered receptive field size among head-centered neurons was smaller in the trained model compared to the untrained model at each percentile value.

The above results demonstrated that the trained model remained effective at producing head-centered output neurons over a broad range of sparseness percentile values, from 50% up to 95%. Indeed, the trained model performed better than the untrained model across every performance metric in this region. Thus, the model performance was robust with respect to the level of competition implemented between the output neurons. Although, as was shown in section 3.2.3, there does need to be some level of competition (i.e. $\pi > 0$) in order to support competitive learning. As the sparseness percentile is increased towards 95% and beyond, this will reduce the number of output neurons that can respond to each input pattern and also reduce the likelihood of a given neuron being able to respond to any particular input pattern. This was likely the reason why there were fewer head-centered neurons after training, and why these neurons had smaller head-centered receptive fields.

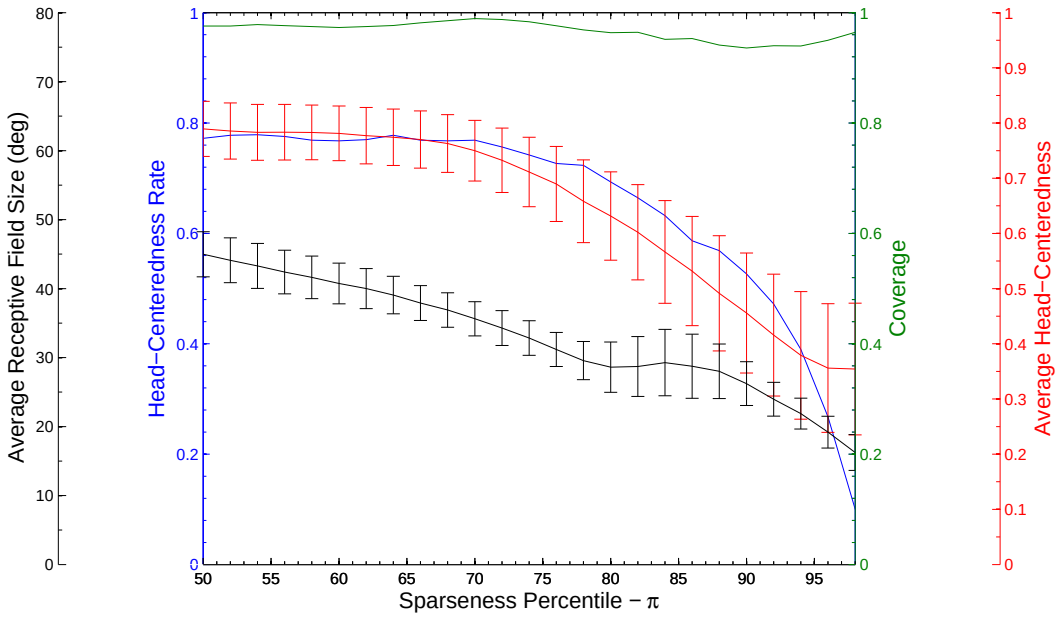
Size Of Output Population

This experiment investigated how varying the size of the output population, N, of the model influenced the response characteristics of the output neurons that developed during training. It was expected that the head-centeredness rate and average receptive field size of output neurons in the trained models would remain relatively stable across different population sizes, in so far as the model behaviour was robust, since these are both normalized metrics. The population size N was varied over $10^2, 20^2, \ldots, 70^2$.

The effect of varying the size of the competitive output population on the performance characteristics of the model was investigated by plotting key summary statistics as a func-



(a) Untrained



(b) Trained

Figure 3.17: The effects of varying the sparseness percentile, π, on the firing properties of output neurons in the untrained model (a) and after 20 epochs of training (b). Conventions are similar to figure 3.9.

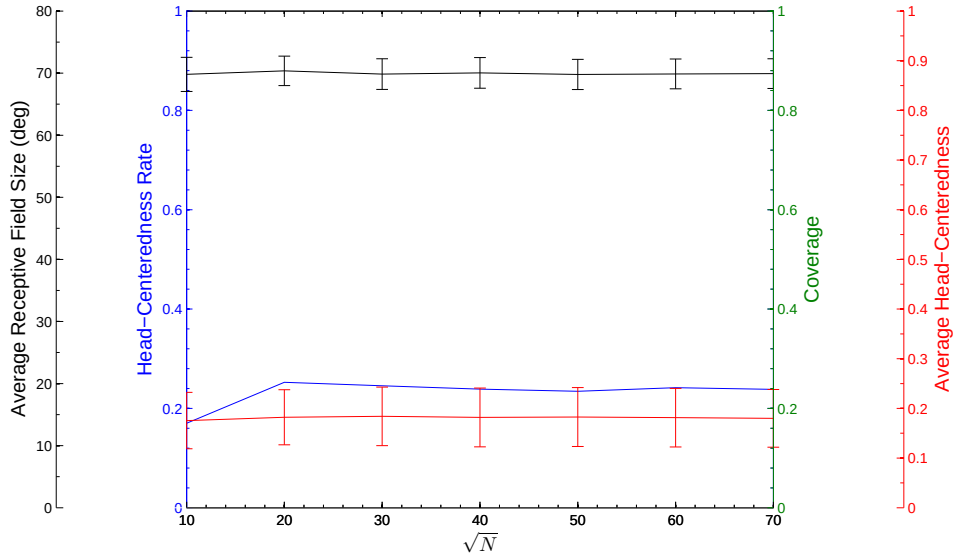
tion of $\sqrt{N}$ in figure 3.18. For the untrained model, the head-centeredness rate and average head-centeredness remained at relatively low values for all N, while there was no coverage at any output population size. For the trained model, the head-centeredness rate, average head-centeredness and average head-centered receptive field size all remained invariant to the size of the output population. The coverage was also invariant to output population size for $N \geq 20^2$. In particular, the trained model had a greater head-centeredness rate and average head-centeredness among head-centered neurons compared to the untrained model at any given population size. Also, the average receptive field size among head-centered neurons was smaller in the trained model compared to the untrained model.

In summary, the size of the output population had little impact on the performance of both the trained and untrained models. Moreover, the trained model was more effective at developing head-centered output representations than the untrained model over all simulated sizes of the output population.

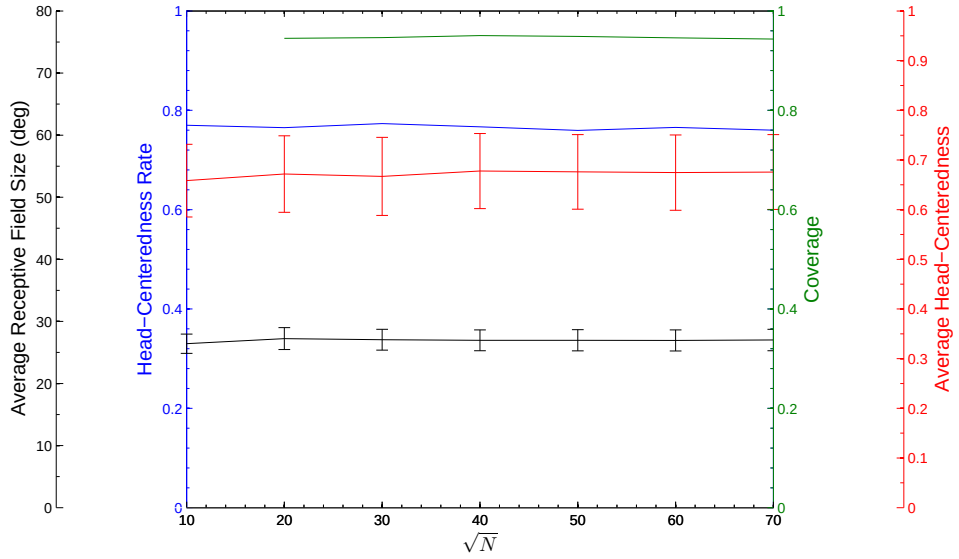
3.2.5 An Alternative Model of Competition within the Output Layer

This experiment investigated whether an alternative, more biologically accurate, model of competition within the output layer would still allow the model to develop head-centered output neurons.

In the previous simulations described above, competition between the output neurons



(a) Untrained



(b) Trained

Figure 3.18: The effects of varying the number of neurons in output population, N, on the firing properties of output neurons in the model before training (a) and after 20 epochs of training (b). Conventions are similar to figure 3.9.

was implemented by adjusting a dynamic activation threshold, p_π, which was continually reset to the π^{th} percentile of the current distribution of neuronal activations. All output neurons with an activation less than the π^{th} percentile would then fall below the response threshold of the sigmoidal activation function 2.4. Although this was an effective way of implementing competition between output neurons, it did not accurately mirror the inhibitory feedback mechanisms present in the brain.

In the following simulation, a more biologically accurate model of competition within the output layer was introduced. The competition was implemented as an additional inhibitory term inserted directly within the differential equations governing the activations of the output neurons. The magnitude of this inhibitory term was directly proportional to the total level of neuronal activity within the output layer. This new term thus represented inhibitory feedback from the activity of all of the output neurons, which in the brain would be mediated by inhibitory interneurons. This new way of implementing the competition was thought to be a more faithful representation of the biological feedback mechanisms implemented by inhibitory neurons in the brain. The activation and firing rate dynamics of the new competition model used in this experiment were as follows.

The activation $h_i(t)$ of each output neuron i was governed by

$$\tau_h \frac{dh_i}{dt} = -h_i + \sum_j w_{ij} v_j^I - \sum_j w^- v_j \tag{3.4}$$

where the first summation is over all presynaptic input neurons j, the second summation is

over all presynaptic output neurons j, and w^- is the inhibitory connection weight between all output neurons.

The firing rate of output neuron i was given by the sigmoid activation function

$$v_i = \frac{1}{1 + \exp\left(-2\varphi(h_i - \theta)\right)} \tag{3.5}$$

The more biologically accurate competition mechanism described by equations 3.4 and 3.5 does not so tightly constrain the number of neurons that remain active in the output layer. In the simulation described next, the new competition mechanism was implemented in a model with the same parameters as those used in experiment 3.2.2, except for $w^- = 0.0070$ and $\theta = 3.0$.

Figure 3.19 shows the receptive field properties of output neurons from the model before and after training, and population statistics are given in table 3.7.

The average RFI increased from -0.41 to 0.31 with training, and the prevalence of head-centered response, as indicated by a positive RFI, was increased from ∼24% to ∼77%. Among head-centered neurons, the average head-centeredness increased from 0.18 to 0.72. In summary this showed that training had the effect of increasing the number of head-centered neurons, and also increasing the compatibility of their responses with a head-centered reference frame of reference.

Head-centered neurons in the trained model had receptive fields clustered around one of the eight head-centered training locations, while receptive fields in the untrained model

where scattered in the center of head-centered space. Both trained and untrained models had an average head-centered receptive field location close to zero, indicating little lateralzed bias. However, the trained model had head-centered neurons corresponding to all of the eight training locations giving a coverage of $\sim$0.96, while the untrained model did not cover the two most eccentric head-centered locations.

The above results showed that the alternative, more biologically faithful competition mechanism described by equations 3.4 and 3.5 produced results very similar to the results found in experiment 3.2.2. This robust model performance suggests that the competitive interactions between output neurons, upon which the model depends, can be implemented in a variety of ways. Indeed, this important result provides further evidence for the robustness of the four core model components described in section 2.2 that are premised to underpin the development of head-centered visual representations in the brain.

3.3 Discussion

This chapter investigated the feasibility of the hypothesis, described in section 2.2, and it was successfully established, through computer simulation, that the combination of the described core model components did allow for the development of head-centered visual representations in the output layer of the network. After training, output neurons were found to have receptive fields well aligned with the head-centered target locations the model was exposed to during training, and their receptive fields were stable across different eye

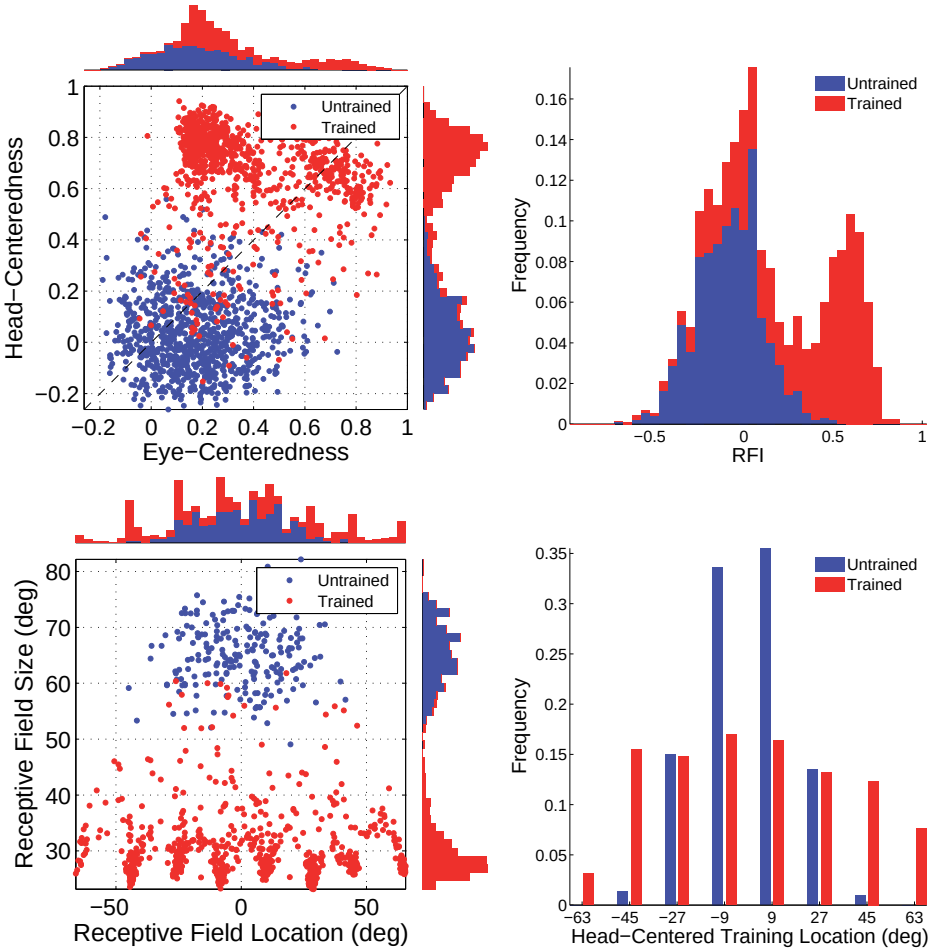


Figure 3.19: Population analyses of receptive field properties of output neurons from the model with a more biologically accurate competition model described by equations 3.4 and 3.5. Results are presented before training (blue) and after 20 training epochs (red). Conventions as in figure 3.3.

fixation positions as required. Control experiments were then conducted to investigate whether the four core model components described above, having been established as sufficient, were also individually necessary for the model.

It was shown that decoupling the visual and eye position dimensions of the receptive fields of input neurons prevented the model from developing head-centered output neurons

Experiment 3.2.5				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$24%)	All	RFI > 0 ($\sim$77%)
Head-centeredness	0.04 (0.14)	0.18 (0.12)	0.67 (0.18)	0.72 (0.13)
Eye-centeredness	0.19 (0.16)	0.03 (0.11)	0.37 (0.23)	0.28 (0.16)
RFI	-0.11 (0.18)	0.12 (0.10)	0.30 (0.32)	0.44 (0.20)
RF Location	$-0.57°$ (15.80°)	$-0.28°$ (16.63°)	$-1.63°$ (39.34°)	1.56° (34.25°)
RF Size	64.86° (5.51°)	65.26° (5.82°)	31.47° (7.60°)	30.62° (6.42°)

Table 3.7: Results of simulations with a more biologically accurate competition model described by equations 3.4 and 3.5. Population summary statistics of response properties of output neurons are shown before training (left two columns) and after training (right two columns). For each of these two stages of training, statistical measures computed over all output neurons are shown in the left subcolumn, while measures computed over neurons with a receptive field index greater than zero are shown in the right subcolumn. Conventions are similar to table 3.4.

during training. This result confirmed that input neurons with coupled visual and eye position receptive fields were indeed a necessary model component. Simply having the required input information, namely retinal target location and eye position, in a decoupled representation did not enable successful learning. This was because with input neurons that have decoupled receptive fields there does not exist any set of synaptic weights that can effect a mapping to head-centered output neurons. So no self-organisational synaptic learning process can solve this problem in a single layer of feed forward synapses, because there is in fact no solution. This can be understood by considering the following. For an output neuron to respond to a particular head centered location, it must respond to a set of many specific combinations of retinal target location and eye position that correspond

to that head-centered location. These combinations will together cover large portions of the retinal target location space and eye position space. However, the inputs from these two spaces are represented independently. This means that the output neuron must have strengthened synaptic connections from input neurons representing a broad region of the retinal target location space, as well as input neurons representing a broad region of the eye position space. Indeed, all retinal target locations will map onto most head-centered locations (depending on eye position), and all eye positions will map most head-centered locations (depending on retinal target location). In this case, the output neuron will receive equal stimulation from the many possible combinations that can be constructed from these large portions of the retinal target location and eye position spaces, not just combinations corresponding to one specific head-centered location. In this case, the output neuron cannot respond selectively to just one head-centered location. Thus, with input neurons with decoupled receptive fields, there is no set of synaptic weights that can effect a selective mapping to head-centered output neurons. These results may therefore help to explain the functional significance of neurons with coupled visual and eye position receptive fields that have been reported in cortical area PO by Galletti et al. (1995). This, in turn, would highlight the need for multiple stages of neural processing of the visual and eye position signals in the brain in order to develop neurons at an intermediate stage with coupled receptive fields, which then provide the required inputs for the model architectures presented in this chapter. These neurons with coupled visual and eye position receptive

fields may then facilitate subsequent stages of neural processing, such as competitive trace learning to produce head-centered output representations.

The general form of network architecture used in this chapter to develop head-centered output neurons is known as a *competitive neural network* (Hertz et al. 1991, Rolls & Treves 1998). Such a network implements competitive interactions between the output neurons during training and testing. The competition is needed to encourage individual output neurons to learn to respond selectively to particular subsets of input patterns, with different output neurons responding to different subsets of input patterns. The subsets (categories) of input patterns that the output neurons learn to represent depend on the structure of the space of input patterns, the temporal order in which the input patterns are presented during training, and the kind of learning rule used to modify the synaptic weights. The model simulations reported above confirmed that competitive interactions were needed within the output layer in order to force individual output neurons to learn to respond selectively to subsets of input patterns corresponding to particular head-centered locations. This result was consistent with standard theory of learning in competitive neural networks (Hertz et al. 1991, Rolls & Treves 1998).

It was shown that diminishing the efficacy of the synaptic learning rule to bind temporally proximal input patterns undermined the ability of the model to develop head-centered output neurons during training. Both the trace learning rule 2.6 and the hebbian learning rule 3.3 were sensitive to their relevant time constant, namely the trace time constant τ_q in

equation 2.5 and activation time constant τ_h in equation 2.3, respectively. **It was shown that the ability to increase the number of head-centered output neurons depended strongly on the time constant in both cases, where shorter values had a negative effect. However it was also shown that the head-centeredness among whatever head-centered neurons existed after training, was not sensitive to a shortened time constant.**

The fact that a hebbian learning rule could effect temporal binding of input patterns was an important result. Previous research had appeared to show that a hebbian learning rule could not effect such temporal binding (Rolls & Milward 2000, Wallis & Rolls 1997). However, this previous research had tested a more usual discrete time hebbian learning rule of the form $\Delta w_{ij} = k v_i v_j$. Such a discrete time version of a hebbian learning rule does not contain any memory trace of previous neural activity and so cannot perform temporal binding of input patterns. However, the hebbian learning rule 3.3 and activation equation 2.3 implemented in the simulations reported in this chapter were time-continuous differential formulations. The differential equation 2.3 simulated neuronal activations with an exponential decay governed by the time constant τ_h, which ensured the activation h_i effectively represented a memory trace of recent neural activity. This memory trace of activity was then incorporated into the Hebbian learning rule 3.3, which allowed the learning rule to perform temporal binding. An upshot of this result is that temporal binding in the brain may be performed by a simple Hebbian learning rule without the

need to invoke and explain additional mechanisms required for the explicit trace term q_i used in the trace learning rule 2.6. Indeed, temporal binding using a Hebbian learning rule has recently been demonstrated in a more biophysically detailed model with spiking neurons Evans & Stringer (2012). In these simulations, temporal binding was enhanced by increasing the time constant of the synaptic conductances, which controlled the flow of current into the postsynaptic neuron.

A core requirement for the model to produce head-centered output neurons through visually-guided learning is that there are periods of time during which the head and visual target remained fixed while the eyes move around a visual scene. This is a reasonable assumption because most visual stimuli remain static in the visual world for most of the time, and a primate will more frequently adjust its direction of gaze by moving its eyes rather than its head (Freedman & Sparks 1997). This will ensure that input patterns corresponding to a fixed head-centered location of the visual target will be clustered together in time. In this case, the trace learning rule is able to encourage individual output neurons to learn to respond selectively to subsets of input patterns corresponding to particular head-centered target locations. The simulations reported above confirmed the feasibility of this hypothesized mechanism for the development of head-centered visual representations in the primate dorsal visual pathway. Moreover, if these movement statistics were altered during training by reducing the number of eye fixation positions P for each fixed head-centered target location, then, consistent with the hypothesis, this prevented the model from form-

ing head-centered output representations. Of course, across different times, primates will experience a variety of different kinds of movement statistics of the eyes, head and visual targets. For example, sometimes a primate will move its head with respect to the static visual world, or a visual object will move while the eyes and head remain stationary. These are not the kind of movements required to build head-centered visual representations by binding of temporally proximal input patterns. However, this should not be a problem for the model, which should be capable of learning multiple different kinds of output representations. That is, when the eyes are moving while the head and visual target remain fixed, then some output neurons will learn to represent the head-centered location of the target. At other times, when the model is trained on different movement statistics, other output neurons may learn different kinds of representations such as eye-centered target locations. Indeed, evidence for this is provided by the fact that individual cortical regions such as LIP and PO do indeed contain a heterogeneous population of neurons with different response characteristics, including both eye-centered and head-centered responses. The computational feasibility this proposal is explored later in this thesis.

In most of the simulations, the competition within the output layer was implemented by continually resetting the activation threshold, p_π, to the π^{th} percentile of the current distribution of neuronal activations. All output neurons with an activation below p_π would then fall below the response threshold of the sigmoid activation function 2.4. In order to investigate the robustness of the model, we carried out further simulations with a more

biologically accurate form of competition mechanism described by equations 3.4 and 3.5. The new competition model had an additional term representing inhibitory feedback inserted into the equations governing the activations of the output neurons. In the brain, the inhibitory feedback between output neurons would be mediated by inhibitory interneurons, which are both driven by and inhibit the output neurons. However, as long as the dynamics of the inhibitory neurons are sufficiently fast, the model formulation given by equations 3.4 and 3.5 should remain reasonably accurate. Simulations with the more biologically accurate competition model produced qualitatively similar results, showing that the performance of the model was not sensitive to the precise mechanisms underpinning competition in the brain. This alternative form of competition is explored further in chapter 4.

The majority of past experimental and theoretical work in sensorimotor transformation has focused on parietal areas LIP and 7a. In these areas eye position gain modulated retinotopic neurons are easily isolated, and consequently the planar eye position gain fields in these areas were studied in detail (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990). Later work identified eye position gain fields in area PO that were peaked rather than planar. Subsequent theoretical work demonstrated that this type of input representation improves the efficiency of the coding scheme by reducing the number of neurons that are necessary to encode the visual target position (Breveglieri et al. 2009).

Some researchers speculated in early work that the difficulty in identifying head-centered neural representations at the single neuron level was evidence that eye position gain mod-

ulated neurons in parietal areas were the last stage of sensorimotor integration of signals representing the retinal location of visual targets and eye position, and that head-centered representations were only available at the population level of these eye position gain field neurons (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990). Subsequent experimental work did, however, reveal that multiple parietal areas, including PO, LIP and VIP, did have head-centered representations (Duhamel et al. 1997, Galletti et al. 1993, Mullette-Gillman et al. 2005).

So, in summary, neurophysiological studies have demonstrated the existence of the planar and peaked eye-position gain modulated retinotopic neurons and head-centered neurons that occur in the model explored in this and the following chapters.

However, it still remains to identify the flow of signals between the various relevant cortical areas, and it is possible that experimental studies will not show an obvious, simple progression from eye-position gain modulated retinotopic neurons to head-centered neurons between two successive visual areas in the brain. For example, neuroanatomical studies have shown that area PO and area LIP are reciprocally connected (Colby & Duhamel 1991), hence head-centered neurons in area LIP might develop using the trace learning principles described in this thesis applied to the afferent synaptic connections received from eye-position gain modulated retinotopic neurons present in area PO. However, these head-centered neurons in LIP might then project back to other neurons in PO, which would then inherit head-centered response characteristics from the LIP inputs. This would give

rise to a mixed population of both eye-position gain modulated retinotopic neurons and head-centered neurons in area PO, which is in fact what has been observed experimentally (Galletti et al. 1993). Nevertheless, the head-centered representations present in these areas would still initially develop by trace learning in the projections from eye-position gain modulated retinotopic neurons in area PO to area LIP. Another possible network architecture is that head-centered neurons in area PO develop by trace learning in the recurrent connections from eye-position gain modulated retinotopic neurons in area PO. Then, these head-centered neurons in area PO may project to neurons in area LIP that would then inherit these head-centered firing characteristics. There are many possible network architectures in which head-centered representations may develop without a readily apparent, simple progression from eye-position gain modulated retinotopic neurons in one visual area to head-centered neurons in a succeeding area. These more complex network architectures are characterised by the presence of either recurrent connections between neurons within an area, or the presence of both feedforward and feedback connections between different areas. Furthermore, these types of synaptic connectivity are indeed typical architectural features of the cortex (Rolls & Treves 1998). Nevertheless, the trace learning mechanisms described in this thesis may still operate in these more complex architectures in the manner described.

The model is distinguished from much previously published work by its emphasis on having synaptic connections modify through an unsupervised, visually-guided, competitive

learning process. This means that no artificial teaching signal is used to set the firing rates of the output neurons during training (Hertz et al. 1991, Rolls & Treves 1998). The model also utilises the natural movements of the eyes and head observed in primates (Freedman & Sparks 1997). The plausible way in which the required synaptic connections are set up in the model contrasts sharply with previously published work.

A highly influential early model of the self-organization of head-centered neural firing responses was developed by Zipser & Andersen (1988). Their model showed how a neural network trained on independent visual and eye position signals could develop head-centered output representations. However, the model relied on error-correction, backpropagation learning (Hertz et al. 1991, Rolls & Treves 1998). Such a learning algorithm is not biologically plausible because there is no way that the error signal used by the backpropagation algorithm to train the first layer of synapses could be present in the brain. A later model developed by Mazzoni et al. (1991) was able to develop head-centered output representations without the need for backpropagation learning. However, although the authors claimed their model was more biologically plausible, in fact the model still relied on a form of supervised global error correction learning. While this model constituted an improvement over the classical model of Zipser & Andersen (1988), synaptic learning was still dependent on an error term which is unlikely to be present in the cortex. Another more recent model of sensorimotor transformation by Pouget & Sejnowski (1997) also employed a supervised error-correction (delta) learning rule to modify the synaptic weights.

So error-correction learning has been for some time the dominant modelling approach in this field. However, an important effect of employing error-correction learning is that the same synaptic weights may become either positive or negative depending on training. This property, itself, is biologically implausible because it violates Dale's principle (Eccles et al. 1954), namely that a given neuron cannot simultaneously be both inhibitory and excitatory at different synapses. In section 1.5 an analysis was conducted on the models of both Zipser & Andersen (1988) and Pouget & Sejnowski (1997), which showed that both of them exhibit this problem for the vast majority of neurons. The implication of this is that the synaptic connectivity in these kinds of models cannot even qualitatively reflect of the patterns of synaptic connectivity present in the brain.

In contrast, while the standard hebbian learning rule 3.3 and trace learning rule 2.6 combined with synaptic weight normalization 2.7 permit both LTP and LTD, these learning rules still keep all of the feedforward synaptic weights positive. The inhibitory interactions that are needed to effect competition between output neurons are modelled separately, and in the brain would be mediated by a population of inhibitory interneurons. In this way the model simulations presented in this chapter do not violate Dale's principle, and consequently may actually embody the form of synaptic connectivity and neural processing present in the brain.

More recently Spratling (2009) explored the same basic temporal binding hypothesis for the development of head-centered representations as investigated in this chapter. Indeed,

Spratling (2009) extended the concept an extra step to produce body-centered reference frames. Moreover, this work showed that both the conjunctive gain field representations and then disjunctive reference frame representations could self-organize in an unsupervised manner within the same model. The work in this thesis complements and extends this earlier work in a number of important ways in this and subsequent chapters, especially by exploring how the learning process may operate with more biologically plausible input representations and ecological training conditions. In this chapter in particular, the use of a more realistic input encoding led to the development of proper visual receptive fields at the single neuron level in the output population, and these could be analyzed in terms of size, location and how well they were aligned with an eye-centered and head-centered reference frame. This reference frame analysis was also done in the exact same manner as in experimental studies (Mullette-Gillman et al. 2005).

3.4 Summary

The following is a summary of the main results and insights in this chapter.

- It was possible to manually hardwire the feedforward synaptic connections in a model with input neurons encoding retinal target location with peaked eye position modulation in order to endow the output neurons with head-centered firing response characteristics. This demonstrated the existence of a synaptic weight structure that implemented the desired mapping, and also permitted the head-centered performance

metrics to be tested and validated.

- The model can develop head-centered output representations by combining the following core model features: (i) the input neurons encode the retinal target location and eye position in a coupled manner, (ii) the output neurons compete with each other through inhibitory interactions, (iii) the synaptic weights are modified by a learning rule that incorporates a memory trace of recent neural activity in order to effect temporal binding of temporally proximal input patterns, and (iv) there are periods during training when the eyes are moving while the head and visual target remain stationary.

- If the input representations of retinal target location and eye position are decoupled, then there does not exist any set of synaptic weights that can effect a mapping to head-centered output representations. Consequently, the model is not able to develop head-centered output neurons during training.

- If competition is removed from the output layer, then the model will fail to develop output neurons that respond selectively with moderately sized head-centered receptive fields.

- **If the propensity of the synaptic learning rule to bind together temporally proximal input patterns is reduced, then this will degrade the increase the number of head-centered neurons after training, however average head-**

centeredness is not affected.

- A decrease in the frequency of eye movements with respect to head movements decreases, and eventually prevents, the development of head-centered output neurons.

- If the learning rate is too small then insufficient learning is able to take place in the available training time to produce head-centered output representations. Alternatively, if the learning rate is too high then the model suffers from continual memory loss due to overwriting of previous learning, and the model is again unable to develop head-centered output neurons.

- The model was effective at developing head-centered output representations over a wide range of sparseness percentile values, from 50% up to 95%. The performance of the model was thus robust with respect to the level of competition implemented within the output layer.

- The ability of the model to develop head-centered output representations is robust with respect to the size of the output layer. In particular, increasing the number of output neurons N beyond 400 had no impact on the head-centeredness rate or average head-centered receptive field size.

Chapter 4

Head-Centered Receptive Fields: Further Issues Regarding Biological and Ecological Plausibility

4.1 Introduction

This chapter presents experiments exploring further issues relating to the biological and ecological viability of the theory of the visually guided development of head-centered receptive fields as set out in the previous chapter. Simulations explored the performance of the underlying theoretical framework along a variety of stimulus and model dimensions as follows. The number of target locations during training was increased to approach natural conditions where a continuum of head-centered locations could be occupied. The visual receptive field size of the input population neurons was increased, whereupon the model failed, and solutions to this problem were proposed and explored. The stimulus dynamics were made more realistic during training by introducing additional eye and target movements that departed from the strict temporal statistics which the original theory had

assumed, and a range of such dynamics were explored. The duration of individual fixations was increased, which undermined learning in the model, and a solution is proposed and explored. Further simulations explored the effects of varying the length of fixation duration within individual training sessions to model more natural eye movements. The gain fields are altered to have both variable peak magnitudes, and also a mixture of unimodal and bimodal peaks with variable magnitude, to reflect the more varied mixture of gain fields actually found in the brain. Finally, the performance of the model is investigated when it is either trained or tested on multiple visual targets presented simultaneously, as would be the case in natural visual scenes.

4.2 Results

4.2.1 Number Of Target Locations

In natural visual scenes, visual targets can be seen in any location with respect to the head. Therefore, it is important to verify that the model can operate under these more ecological conditions. In this experiment it was investigated how varying the number of visual target locations in head centered space during training, M, would influence self-organization in the model. In particular, the asymptotic performance of the model was studied as the number of head centered target locations was increased towards an effective continuum, whereby visual targets may be seen anywhere with respect to the head.

Each simulation was performed with a fixed value of M. For each simulation, a stimulus

set was created as described in section 2.4. Each training epoch was divided into M periods, each of which corresponded to one of the head centered target locations. During each such training period, the location of the visual target remained fixed in the corresponding head centered location while the the model saccaded through a sequence of $P = 15$ eye positions. Three example stimulus sets are shown in figure 4.1 for values of M of 2, 4 and 12. The simulations otherwise had the same parameters as in section 3.2.2.

The impact of varying the number of head centered target locations during training on the performance of the model was inspected by plotting key summary statistics as a function of M in figure 4.2. The head-centeredness rate and the average head-centeredness increased approximately monotonically with increasing M, and were always above their corresponding untrained network values of $\sim$25% and $\sim$0.17 respectively. The coverage did not drop below $\sim$0.88 when $M \geq 7$ in the trained model, while there was no coverage for the untrained model. The average head centered receptive field size decreased steadily as M increased, and was always well below the untrained average of $\sim$69°. In summary, these results showed that model performance actually improved as the number of head centered training locations was increased during training.

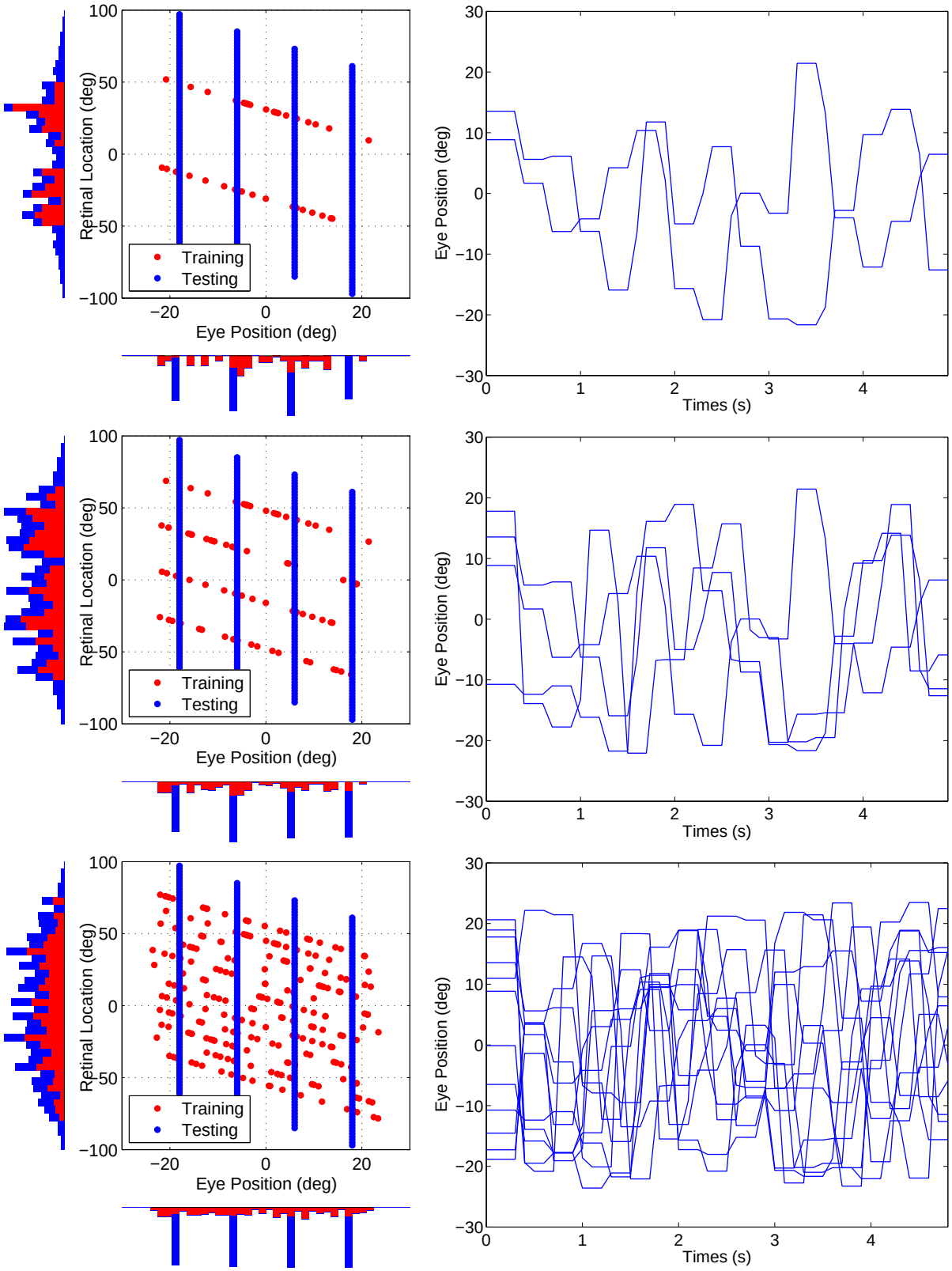


Figure 4.1: Each row shows the stimuli for a given value of M, that is 2, 4 and 12 from the top respectively. The left column shows the fixation positions in the testing and training sets, and the right column shows the eye movement dynamics during each period within a training epoch.

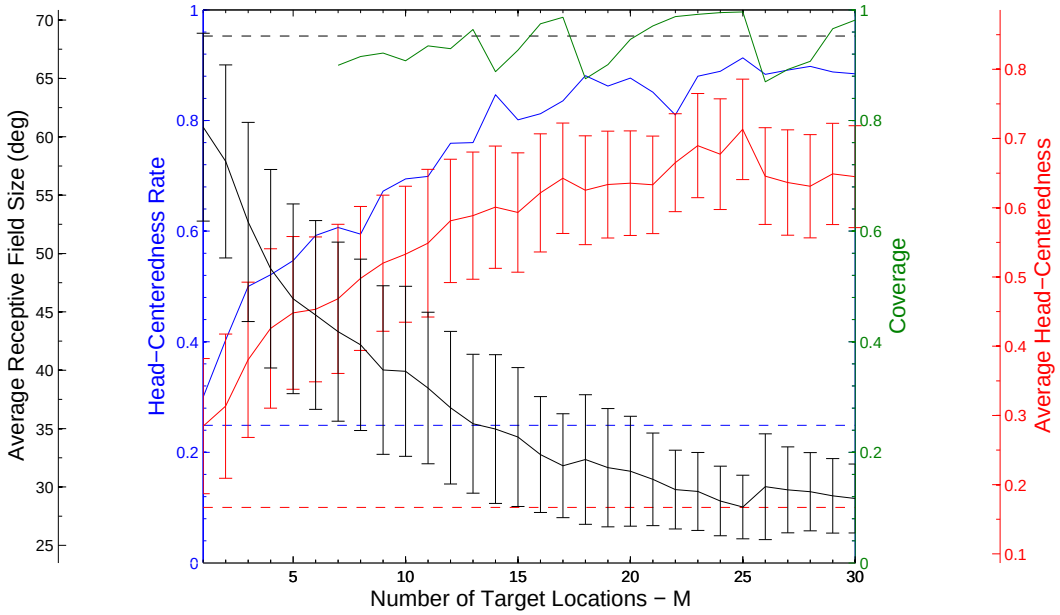


Figure 4.2: Varying number of target training locations - M

4.2.2 Varying Input Neuron Receptive Field Size

In chapter 3 the input neurons in the self-organizing model all had tuning curves in the retinal and eye position space of size $\sigma = \rho = 6°$. However, across the relevant brain areas neurons are found with a broad range of receptive field sizes. Therefore, it was explored what impact varying the receptive field size would have on the model. It was found that for larger receptive field sizes the model would increasingly produce eye-centered neurons, and a second simulation attempted, and succeeded, to remedy this by altering the activation and learning dynamics of the model.

Varying σ and ρ

In this experiment it was investigated how varying the receptive field size parameters σ and ρ would influence learning in the model. These parameters determined the size of the response function of the input neurons in the visual and eye position space respectively. They were both kept equal while they were varied from 1° to 30°, while the other model parameters were the same as in experiment 3.2.2.

The impact of varying the receptive field size in the input population on the characteristics of both the untrained and the trained model was inspected by plotting key summary statistics as a function of the receptive field size in each case (Fig.4.3).

For the untrained model there was no coverage for a receptive field size less than 9°. However, there was coverage at 9° and above. Furthermore, despite a low and declining head-centeredness rate, there was an increasing average head-centerdness among the head-centered neurons. This was a surprisingly good result for an untrained model. In contrast, for the trained model there was no coverage outside of the range $[5°, 9°]$, and beyond this the head-centeredness rate collapsed and remained close to 1%. Although, again, the average head-centeredness among head-centered neurons increased with receptive field size. A more detailed look at the reference frame properties at three receptive field sizes, namely 5°,10°,15°, are found in figure 4.4. What is clear in all three cases is that training has the effect of raising the head-centeredness value of the output neurons from the untrained

case. However with increasing receptive field size the increase in eye-centeredness becomes larger than the increase in head-centeredness, eventually resulting in only eye-centered output neurons.

Two clear candidate effects for why increasing receptive field size makes training increasingly promote eye-centeredness more than head-centeredness are edge effects and overlap effects. Edge effects refers to the influence of the pattern of neural activity being clipped by the finite size representation in the input population, and this clipping becomes more severe as the receptive field size increases. This clipping likely has independent effects during training and during testing. Overlap effects are those interactions between the learning dynamics at proximal head-centered training locations which undermine successful self-organization. These effects become more severe when the receptive field size increases because this causes greater overlap in the input patterns corresponding to proximal head-centered training locations.

In summary this experiment showed that the trained model collapsed outside of a narrow range of receptive field values in the input population, and it did worse with increases in this parameter while the untrained model did better. While head-centeredness did always increase with training for all receptive field sizes, increasing receptive field size caused larger increases in eye-centeredness, ultimately rendering the overwhelming majority of output neurons eye-centered for the largest receptive field sizes. Edge and overlap effects are hypothesized to explain the increasing distortion with increasing receptive field

size.

Nonlinear Synaptic Dynamics

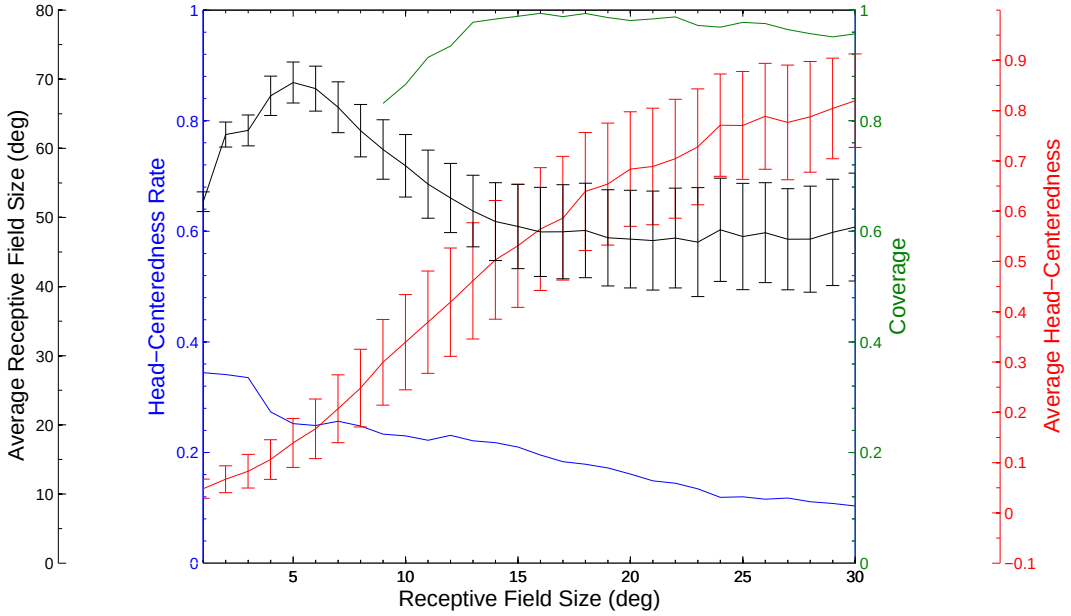
In this experiment it was investigated whether the behaviour of the model could be made robust with respect to changes in the receptive field size of the input population through altering the computation of the activation of the output neurons and the synaptic learning dynamics. The idea was to allow input neurons to have an arbitrarily large receptive field size, but then to mask the effect of a sufficiently large outer portion of the receptive field on the postsynaptic activation and synaptic modification. Firstly, the activation dynamics were altered so that only those synapses where presynaptic activity was above a given threshold would contribute to postsynaptic activation by introducing a nonlinearity in the weighting of the presynaptic firing rate as follows

$$\tau_h \frac{dh_i}{dt} = -h_i + \sum_j [v_j^I]_\gamma w_{ij} \tag{4.1}$$

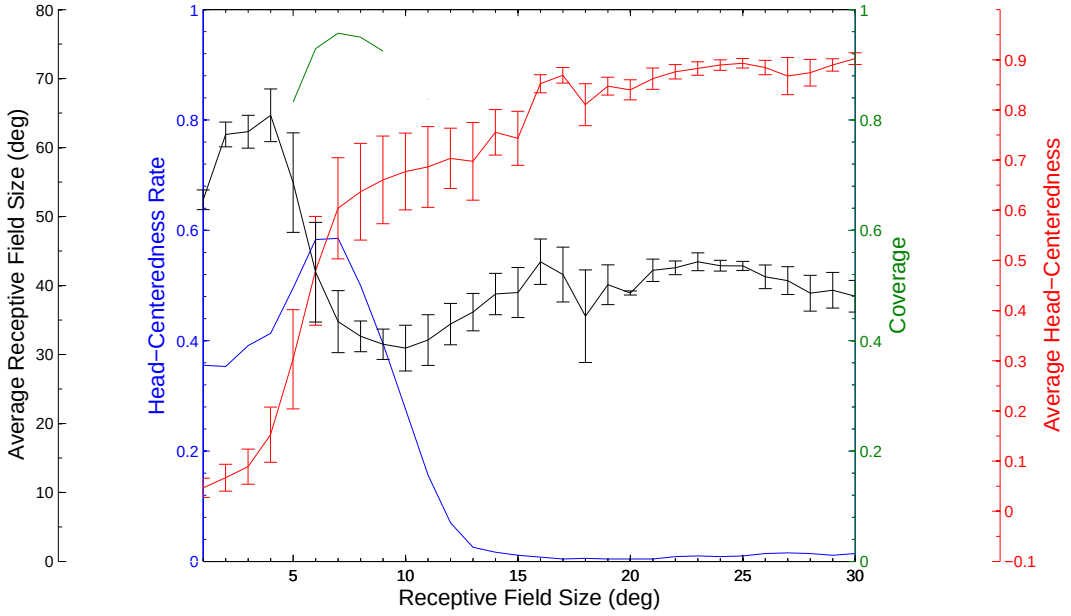
where

$$[v_j^I]_\gamma = \begin{cases} v_j^I - \gamma & \text{for } v_j^I > \gamma \\ 0 & \text{otherwise} \end{cases}$$

Other forms of threshold nonlinearity also sufficed. This change made neurons in the competitive output population excitable by a smaller subset of the input population. Sec-

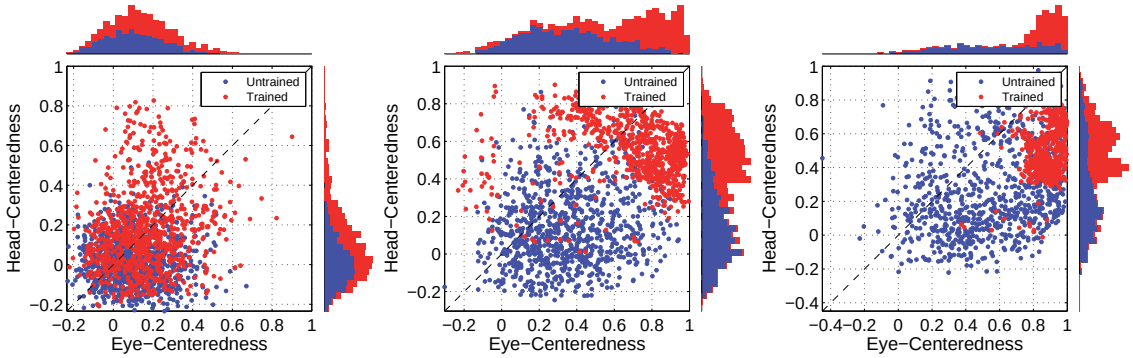


(a) Untrained



(b) Trained

Figure 4.3: Varying input receptive field size $\sigma = \rho$ in the untrained and trained model.



(a) $\sigma = \rho = 5°$ **(b)** $\sigma = \rho = 10°$ **(c)** $\sigma = \rho = 15°$

Figure 4.4: Population reference frames properties for three different receptive field sizes.

ondly, the learning rule was altered so as to require presynaptic activity to surpass the same threshold γ for potentiation to occur.

$$\frac{dw_{ij}}{dt} = \varrho q_i [v_j^I]_\gamma \tag{4.2}$$

A single experiment with a receptive field size of $\rho = \sigma = 19°$ was conducted, and the parameters for the experiment can be found in table 4.1.

The population statistics are given in table 4.2. The head-centeredness rate increased from ∼30% before training to ∼88% after training, and the average head-centeredness among head-centerd neurons increased from 0.1 to 0.63. The trained model also had a coverage of ∼0.99. This shows that introducing the threshold nonlinearity in the influence of presynaptic firing on the postsynaptic activation and synaptic learning dynamics allowed successful learning in the model when the receptive field size in the input population was a relatively large value of 19°.

Parameter	**Symbol**	**Value**
Number of target locations	M	8
Fixation sequence length	P	15
Number of training epochs	-	20
Width of eye position tuning curve	ρ	19°
Width of retinal tuning curve	σ	19°
Output neuron population size	N	900
Input neuron population size		12261
Trace time constant	τ_q	400ms
Activation time constant	τ_h	100ms
Activation function slope	φ	50
Activation function threshold	θ	0.02
Sparseness percentile	π	90%
Learning rate	ϱ	0.05
Synaptic connectivity	ϕ	5%
Excitability/Potentiation threshold	γ	0.9

Table 4.1: Parameters of model with a threshold γ introduced into the computation of the postsynaptic neuronal activation and synaptic learning dynamics.

4.2.3 Variation in Statistics of Stimulus Dynamics

The basic hypothesis for how head-centered output representations may develop assumes that during natural self-motion, there are periods of time when the eyes are moving in the head while the head remains stationary with respect to the visual environment and visual objects also remain stationary within the environment. In the simulations reported so far, the statistics of the stimulus dynamics have conformed to this assumption. However, in reality there will be times when the stimulus dynamics radically depart from this assumption. For example, sometimes the eyes might track a moving object while the head remains stationary, or the head may move while the eyes fixate a stationary object. Both of these alternative stimulus dynamics do not result in visual objects remaining in a fixed

Experiment 4.2.2				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$30%)	All	RFI > 0 ($\sim$88%)
Head-centeredness	0.00 (0.10)	0.10 (0.08)	0.58 (0.23)	0.63 (0.19)
Eye-centeredness	0.03 (0.12)	-0.02 (0.08)	0.14 (0.12)	0.12 (0.09)
RFI	-0.02 (0.11)	0.08 (0.07)	0.44 (0.30)	0.51 (0.22)
RF Location	$-0.00°$ (12.60°)	0.64° (13.91°)	$-0.92°$ (36.04°)	0.63° (37.47°)
RF Size	60.49° (4.60°)	60.27° (4.86°)	28.00° (9.05°)	26.64° (7.95°)

Table 4.2: Population summary statistics of the response properties of the output neurons in the model with a threshold γ introduced into the postsynaptic activation and synaptic learning dynamics, as described in section 4.2.2.

head-centered location for a continuous period while the object shifts on the retina, which is required for trace learning to form head-centered output representations. However, the hypothesis does not actually require the stimulus dynamics to conform to this assumption *all* of the time. Indeed, it was conjectured that as long as there were some periods of time when the visual objects remain in a fixed head-centered location while the eyes move, which may be interspersed with periods governed by alternative stimulus dynamics, then this would still permit head-centered output representations to develop. This key property is important to verify in order to support the biological and ecological validity of the model.

The following experiment investigated how altering the stimulus dynamics by including additional periods of randomised visual target and eye movements would influence the self-organization of the model. In previous experiments, the input stimulus was structured into regular training periods corresponding to sequences of eye movement while the stimulus

was kept stationary with respect to the head. In section 3.2.3 it was established that by reducing the length of these periods, effectively decreasing the relative frequency of eye versus head movements, the model performed worse in terms of the number of head-centered output neurons developed through learning. In the following experiment, the regular training periods during which the eyes move while the visual target remains fixed with respect to the head were interleaved with new periods consisting of completely randomised eye and stimulus movement sequences. It was investigated how learning in the model was influenced by varying the length of these new periods of randomised stimulus dynamics. The expectation was that increasing the length of these randomised stimulus sequences would eventually dominate the effects of the regular training periods promoting the development of head-centered neurons. How tolerant the model was to increasing the proportion of randomised stimulus dynamics during training would inform whether the model could self-organise head-centered output representations under more natural training conditions.

An epoch of training had $M = 8$ regular training periods where a single visual target was located in one of the eight head-centered locations $-56°$, $-40°$, $-24°$, $-8°$, $8°$, $24°$, $40°$ or $56°$ while a sequence of $P = 30$ fixations were performed. Additionally, the M regular training periods were interleaved by randomised training periods during which there was a fixed number, K, of random movements of both the eyes and the location of the stimuli in head-centered space. Separate simulations were performed for $K = 10, 15 \ldots, 75$ to explore the effects of gradually increasing the proportion of training that was governed by

Parameter	**Symbol**	**Value**
Number of target locations	M	8
Regular fixation sequence length	P	30
Randomised sequence length	K	$10, 15, \ldots, 75$
Number of training epochs	-	10
Width of eye position tuning curve	ρ	6°
Width of retinal tuning curve	σ	6°
Output neuron population size	N	900
Input neuron population size		12261
Trace time constant	τ_q	400ms
Activation time constant	τ_h	100ms
Activation function slope	φ	4.0
Activation function threshold	θ	-0.5
Sparseness percentile	π	80%
Learning rate	ϱ	0.05
Synaptic connectivity	ϕ	5%

Table 4.3: Parameters of experiment with additional training periods of randomised movements of the visual target and eyes as described in section 4.2.3.

randomised stimulus dynamics. Figure 4.6 shows the training and testing stimuli for three different values of K, namely 10, 20 and 40. The simulations parameters can be found in table 4.3.

The effects of varying the length of the randomised training periods on the characteristics of the model was inspected by plotting key summary statistics as a function of K in figure 4.7.

The head-centeredness rate remained above 67% across the explored range of K. While it did eventually decline with increasing K, it was still well above the head-centeredness rate of $\sim$22% in the untrained model. There was coverage across all explored values of K,

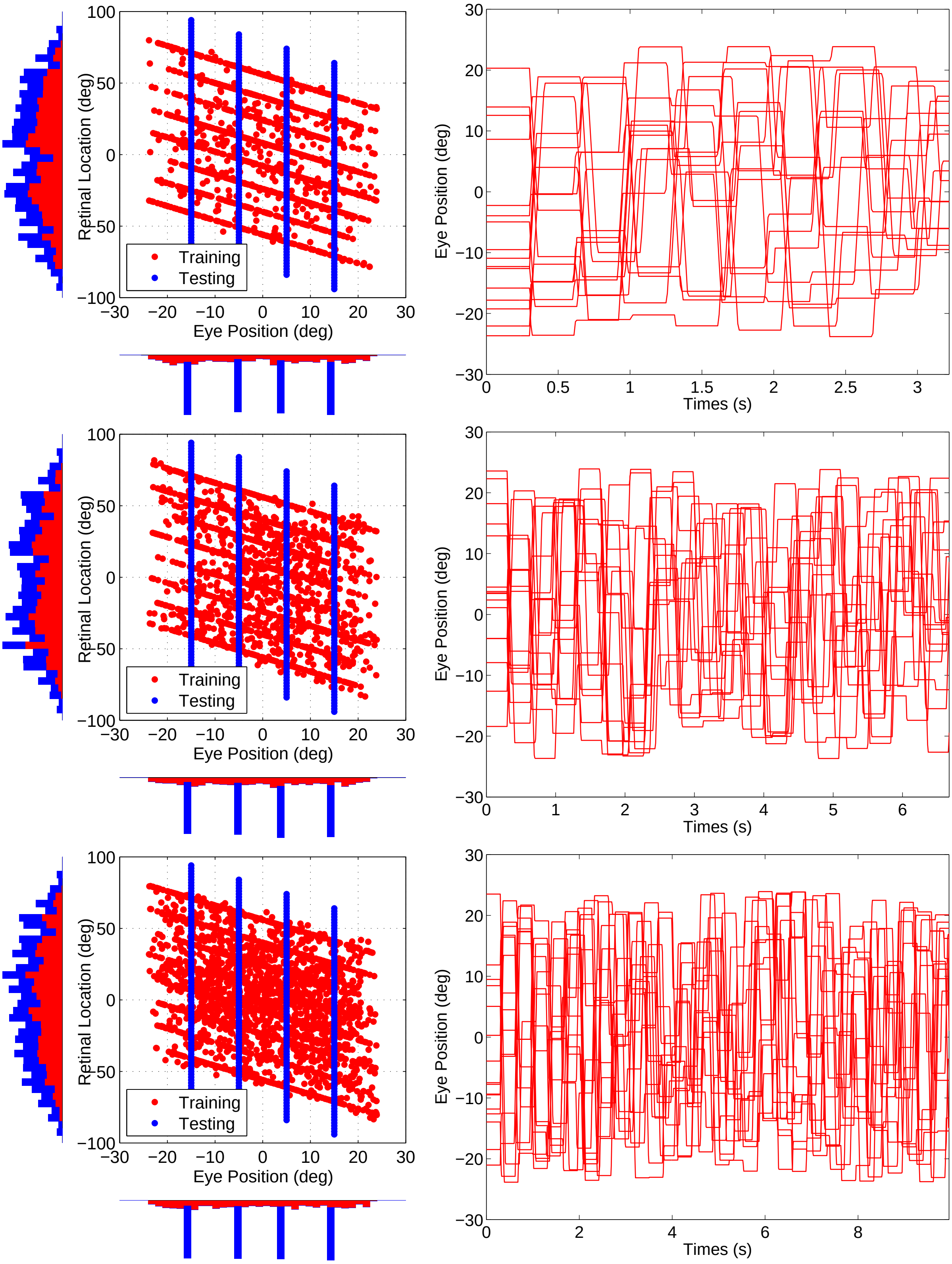

Retinal Location (deg)
Eye Position (deg)
Training
Testing
Eye Position (deg)
Times (s)
Retinal Location (deg)
Eye Position (deg)
Training
Testing
Eye Position (deg)
Times (s)
Retinal Location (deg)
Eye Position (deg)
Training
Testing
Eye Position (deg)
Times (s)

Figure 4.6: Movements of the eyes and locations of visual targets during simulations with additional training periods of randomised movements of the visual target and eyes. Each row shows the input stimuli for a given value of K, that is 10, 20 and 40 from the top respectively. The plots show the input stimuli during both periods of regular movements and periods of randomised movements. The left column shows the fixation positions in the testing and training set. The fixations off the main diagonals in the left column plots represent the fixations that occur during the training periods with randomised movements of the visual target and eyes. The right column shows the eye movement dynamics during each period within a single training epoch. The eye movement dynamics are shown only for the duration of the shortest period within each plot. That is, for $K = 10$ and $K = 20$ this corresponds to the periods of randomised movements, while for $K = 40$ this corresponds to the periods of regular movements for which $P = 30$.

and the coverage remained close to maximal, i.e. greater than $\sim$0.94, for all K. Among head-centered neurons the average head-centeredness remained very stable and never went below 0.67. Indeed, this measure showed minimal dependence on K. It was also well above the average head-centeredness among head-centered neurons in the untrained model which was $\sim$0.22. The average receptive field size among head-centered neurons also remained very stable and close to 30°, which was well below the untrained network value of $\sim$ 67°. In summary, this showed that model performance was robust to the introduction of a significant level of randomised eye-and head-movements during training.

4.2.4 Fixation Duration

In this section it was investigated how varying the duration of fixation would influence the learning in the model. In other experiments, all fixations during training were set to 300ms, largely in order to avoid simulations becoming too costly and for simplicity. However, during natural eye movement dynamics, fixation can potentially be much longer or shorter

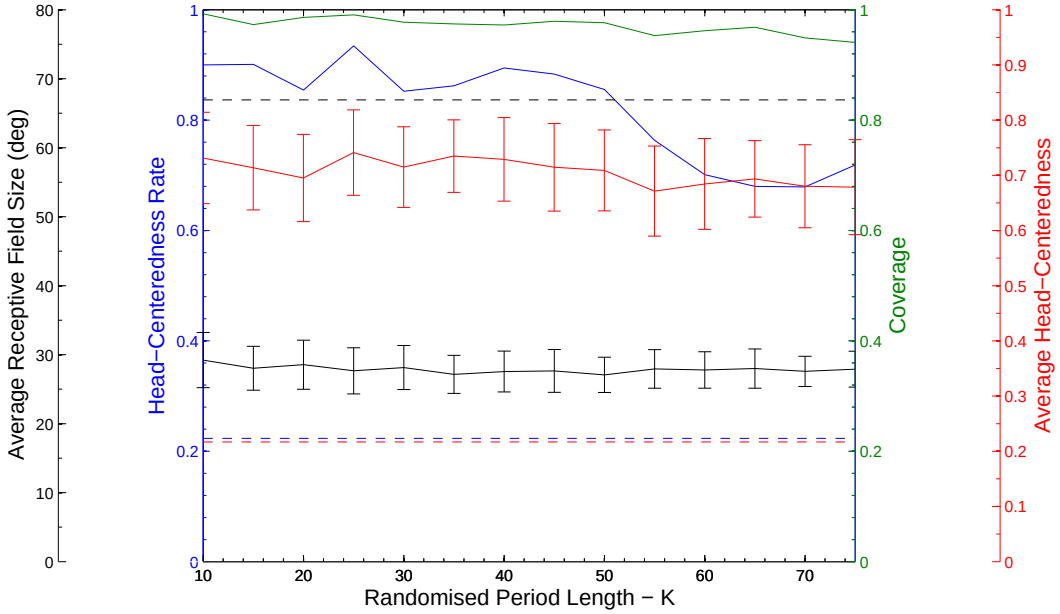


Figure 4.7: Population summary statistics of the response properties of the output neurons from experiments with additional training periods with randomised movements of the visual target and eyes. Results are plotted for simulations with different lengths of the randomised training periods K. The dashed lines show the corresponding quantity in the untrained model.

than this. The model included time constants and learning variables that potentially could be sensitive to fixation duration, and it was important to establish that the model was robust to such fixation duration variability that would be prevalent under natural conditions. To this end, the model was trained with a range of fixed fixation durations, going from 100ms to 1600ms. Here it was found that for larger durations the model failed to successfully develop head-centered output neurons. The trained weight vectors in these failed experiments showed that all potentiated synapses were focused on a specific position in the input space, reflecting a failure of trace learning to bind input patterns from a range of eye positions together. To remedy this, the synaptic learning rule was

modified to include an upper bound on the strength of each synapse, which reduced the rivalry between synapses due to the continuous renormalisation of each neuron's weight vector during training. This was found to restore the ability of the model to develop head-centered output representations when long fixations were implemented. Lastly, since natural conditions will include variability in fixation duration, a final set of experiments explored how the model would handle increasing levels of fixation duration variability within the same experiment. The new learning rule was found to be successful over the full range of variabilities.

Varying Fixation Duration

In this experiment it was investigated how varying the fixation duration from 100ms to 1600ms would influence the learning in the model.

Figure 4.8 shows key summary statistics as a function of fixation duration. The head-centeredness rate is maximal for shorter fixation periods, peaking at ~ 0.95 for 200ms, but then quickly decays and settles barely above the untrained rate for the longest fixation durations. Among the decreasing number of head-centered neurons the average head-centeredness decreased with increasing fixation duration. The coverage also decreases with increasing fixation duration, and beyond 1400ms there is no coverage.

An examination at the single neuron level revealed the nature of the failure. The weight vectors of a typical output neuron across three simulations with increasing fixation durations are seen in figure 4.9. For 100ms (Fig.4.9a), the weight vector exhibits the

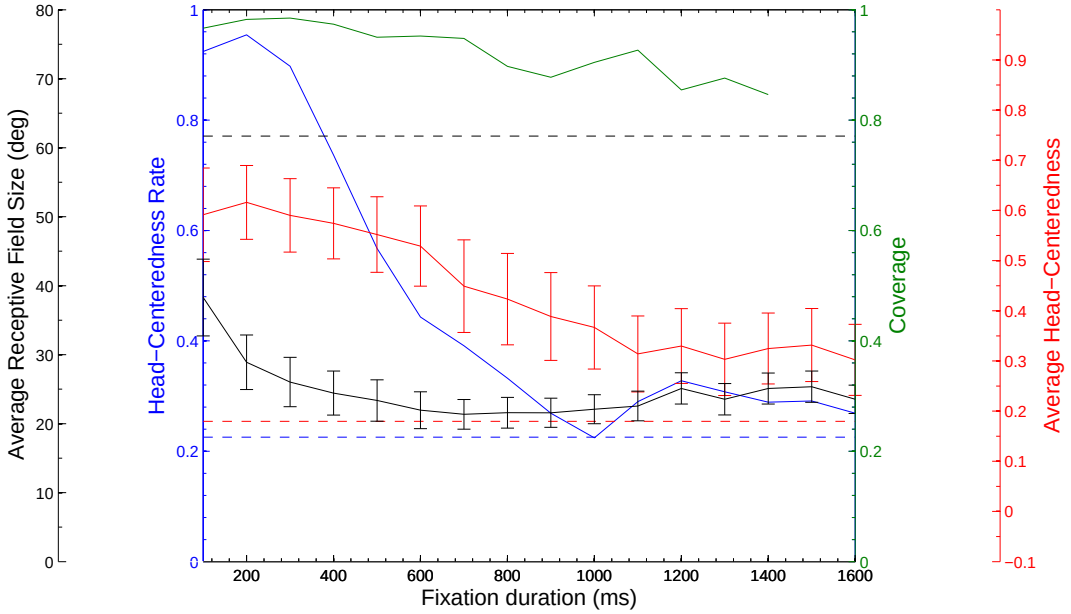
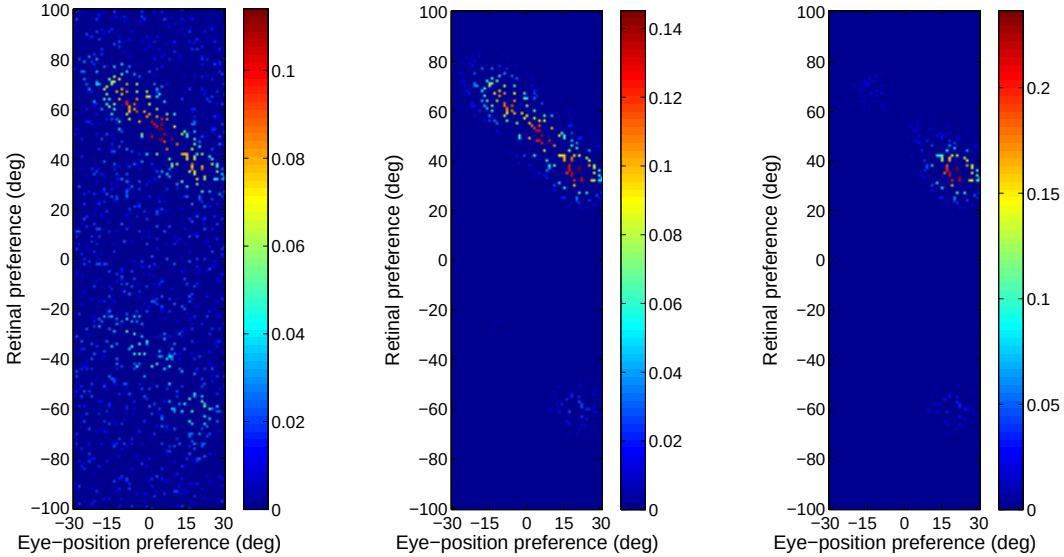


Figure 4.8: Population summary statistics of the response properties of output neurons from experiments in which the fixation duration during training is varied from 100ms to 1600ms. The dashed lines show the corresponding quanitity in the untrained model.

canonical diagonal structure in the input space similar to head-centered output neurons from previous successful experiments. For 600ms (Fig.4.9b) the weight vector was very similar, with even more synaptic weight on the diagonal feature. Both of these neurons were head-centered. However for 1200ms the weight vector was qualitatively altered. In this case only a small portion of the diagonal feature, corresponding to a small eye position range, remained. The reason for this change in the weight vector for a long fixation period was because the prolonged fixation caused the trace learning to focus all the synaptic weight on a single fixation. Comparing the weight of the strongest synapses across the three cases clearly confirms this. In summary, increasing the fixation duration during training severely undermined successful self-organization of head-centered output representations because



(a) 100ms **(b)** 600ms **(c)** 1200ms

Figure 4.9: Weight vector for output neuron #583 across three simulations with different fixation durations during training.

the longer fixation caused trace learning to focus the synaptic weights of an individual output neuron onto a single fixation point.

Bounding the Strength of Individual Synapses

In this experiment it was explored whether altering the learning dynamics could remedy the above failure of the model with longer fixations. The problem in simulations with longer fixations is that the synapses from the input neurons that are active for the current fixation continue to be strengthened for a longer period of time at the expense of synapses that were potentiated during previous fixations, which are depressed by the continuous global renormalisation of each output neuron's synaptic weight vector. This process depresses the previously potentiated synapses in favour of the present input to such a degree that output

neurons are unable to develop and sustain strong synapses to multiple input patterns with different combinations of eye position and retinal target location corresponding to a particular head-centered target location. The remedy investigated in this experiment was to introduce an upper bound on individual synapses which prevented long periods of fixation from causing indefinitely long potentiations at the single synapse level at the expense of the rest of the weight vector. Specifically, the learning rule used was

$$\frac{dw_{ij}}{dt} = \varrho(w^* - w_{ij})q_i v_j^I \tag{4.3}$$

where w^* is an effective upper bound on each synapse. That is, as w_{ij} approaches w^* from below, the rate of change of w_{ij} tends to zero due to the factor $(w^* - w_{ij})$ and hence further potentiation ceases.

The weight vector normalization performed at each time step in previous experiments was also included. The parameters for this simulation were identical to those used above, except $w^* = 0.15$ and $\varrho = 2$. The fixations during training lasted for the relatively long period of 1200ms, which had caused the previous model simulations to fail.

Figure 4.10 shows the population results for the model before and after 20 epochs of training, and population statistics are given in table 4.4.

The head-centeredness rate increased from ∼23% to ∼63% after training (Fig. 4.10a). In particular, this was much greater than the head-centeredness rate of ∼0.8% in the trained model from the previous experiment without an upper bound on each synapse.

Experiment 4.2.4				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$23%)	All	RFI > 0 ($\sim$63%)
Head-centeredness	0.08 (0.15)	0.24 (0.13)	0.56 (0.22)	0.64 (0.18)
Eye-centeredness	0.23 (0.18)	0.11 (0.13)	0.45 (0.26)	0.33 (0.18)
RFI	-0.13 (0.20)	0.12 (0.10)	0.10 (0.33)	0.31 (0.19)
RF Location	$0.04°$ ($17.63°$)	$-0.13°$ ($21.57°$)	$-1.15°$ ($39.46°$)	$4.95°$ ($34.81°$)
RF Size	$65.65°$ ($7.48°$)	$66.55°$ ($6.83°$)	$29.15°$ ($4.77°$)	$28.87°$ ($4.64°$)

Table 4.4: Population summary statistics of the response properties of output neurons when an upper bound w^* is implemented on the strength of each synapse during training, as described in section 4.2.4.

Among head-centered neurons the average head-centeredness increased from 0.24 to 0.64 with training. This was reflected in the average RFI increasing from -0.42 to 0.12 (Fig. 4.10b). Head-centered neurons in the trained model had receptive fields loosely clustered around five head-centered locations (Fig. 4.10c), and while the this clustering was not as even as in previous experiments, coverage was at $\sim$0.91 (Fig. 4.10d).

In summary, the above results confirmed that training with the modified learning rule 4.3, in which an upper bound was implemented on the strength of each synapse to reduce rivalry between synapses during weight vector renormalisation, dramatically increased the number of head-centered output neurons after training with long fixations.

Variability in Fixation Durations

In this experiment it was investigated how variability of fixation duration within the same model simulation would influence learning in the model. Each previous model simulation

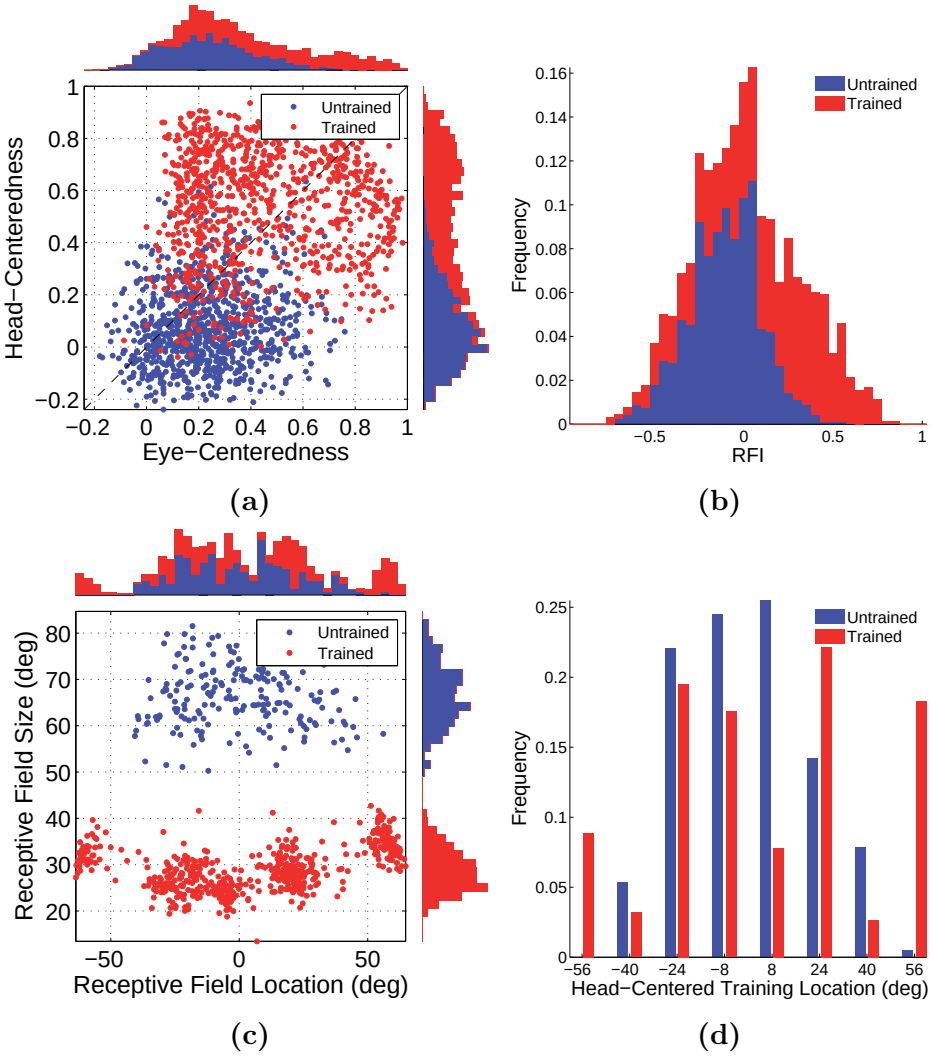


Figure 4.10: Population analyses of the receptive field properties of output neurons when an upper bound w^* is implemented on the strength of each synapse during training. Results are presented before training (blue) and after 20 epochs of training (red). Conventions as in figure 3.3.

had a fixed fixation duration during training. However, under natural conditions there would be substantial variability in the duration of individual fixations. Therefore, it was important to establish whether the model could still function successfully under these conditions. The training dynamics were exactly as above, except that all fixation durations were independently sampled from $\mathcal{N}(300, \sigma)$ were σ was varied and negative samples were discarded and resampled. A set of experiments were conducted where σ was varied from 0ms to 2000ms in increments of 100ms, resulting in a total of 21 experiments.

The impact of the variability of fixation duration on the characteristics of the model was inspected by plotting key summary statistics as a function of fixation duration standard deviation σ in figure 4.11. The head-centeredness rate and the average head-centeredness initially decreased with increasing σ. However, the head-centeredness rate and average head-centeredness eventually stabilised just below 60% and above 0.6, respectively. In particular, the head-centeredness rate of the trained model was well above the untrained rate of $\sim$22%. Among head-centered neurons the average head-centeredness and receptive field size both remained very stable across the full range of standard deviations σ, and also well above and below corresponding values in the untrained condition respectively. Finally there was coverage, at no less than $\sim$0.95, across the full range of explored standard deviations. In summary, these results showed that model performance was robust to variability in fixation duration across a wide range of σ.

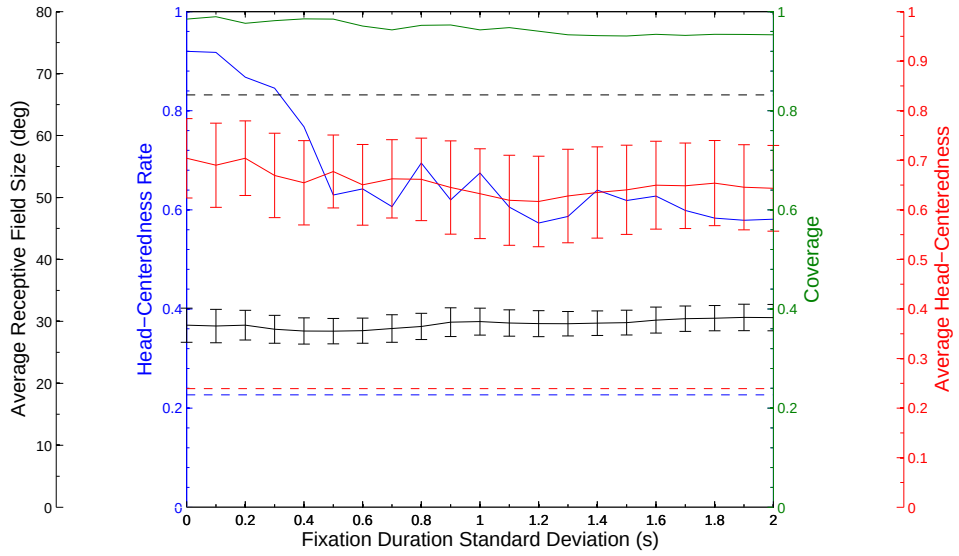


Figure 4.11: Population summary statistics of the response properties of output neurons from experiments in which the fixation durations during training are continually sampled from a normal distribution $\mathcal{N}(300, \sigma)$. The dashed lines show the corresponding quanitity in the untrained model.

4.2.5 Eye Position Gain Fields with more Varied Functional Forms

Previous experiments employed a form of eye position gain modulation which only had a single local maxima along the eye position dimension of the response function. However experiments have shown that the peaked gain encoding found in area PO is not always best described as a single peaked preference (Breveglieri et al. 2009). In the following experiments it was investigated what impact two kinds of variability in the input population responses would had on the learning in the model. Firstly, it was investigated whether the model could successfully develop head-centered output neurons if the input neurons had maximal firing rate responses that varied across the population. Secondly, the impact of introducing multiple gain modulation peaks in some fraction of the input population was

investigated.

Variability in Peak Response of Input Neurons

In this experiment it was investigated how introducing variability in the response magnitudes of neurons in the input population would influence the learning in the model. This was done by stochastically drawing the coefficient ω_i in the response curve 2.1 from the uniform distribution $U[0, 1]$. The parameters used were the same as in experiment 3.2.2, except $\theta = 0.05$ and $\psi = 7.0$.

Figure 4.12 shows the population results from testing the model before and after training, and population statistics are given in table 4.6.

The head-centeredness rate increased from ∼19% to ∼60% after training (Fig. 4.12a), and among these neurons the average head-centeredness increased from 0.20 to 0.61 with training. Head-centered neurons in the trained model had receptive fields clustered around the eight training locations in head-centered space (Fig. 4.10c), and the coverage attained a relatively high value of ∼0.94 (Fig. 4.10d).

In summary, these results showed that the model could successfully develop many more head-centered neurons after training, and that these neurons had a larger degree of head-centeredness, even when there was stochastic variability in the response magnitudes of the input neurons.

Experiment 4.2.5				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$19%)	All	RFI > 0 ($\sim$60%)
Head-centeredness	0.05 (0.14)	0.20 (0.12)	0.51 (0.22)	0.61 (0.17)
Eye-centeredness	0.21 (0.18)	0.06 (0.11)	0.43 (0.26)	0.28 (0.15)
RFI	-0.14 (0.19)	0.12 (0.11)	0.09 (0.36)	0.33 (0.20)
RF Location	0.72° (16.00°)	1.61° (19.56°)	0.19° (37.63°)	6.93° (32.73°)
RF Size	65.14° (4.86°)	64.84° (4.79°)	29.60° (4.50°)	28.73° (4.37°)

Table 4.5: Population summary statistics of the response properties of output neurons for simulations with stochastic variability in the peak responses of input neurons.

Gain Fields with Multiple Peaks

In this experiment it was investigated how increasing the number of peaks in the gain fields of input neurons would influence the learning in the model. For simplicity, only single and double peaked gain fields were explored. In the latter case both peaks were located along the same retinal location, while on different eye position locations. For each input neuron, the number of peaks was stochastically drawn from a Bernoulli distribution, with a $p = 20\%$ propensity in favour of double peaks. The magnitude of the second peak was set by stochastically drawing the coefficient ω_i from $U[0, 1]$ in the double peak response curve described by

$$v_i^I = \left(\exp\left(-\frac{\|e - \beta_i^1\|^2}{2\rho_i^2} \right) + \omega_i \exp\left(-\frac{\|e - \beta_i^2\|^2}{2\rho_i^2} \right) \right) \times \exp\left(-\frac{\|r - \alpha_i\|^2}{2\sigma_i^2} \right) \tag{4.4}$$

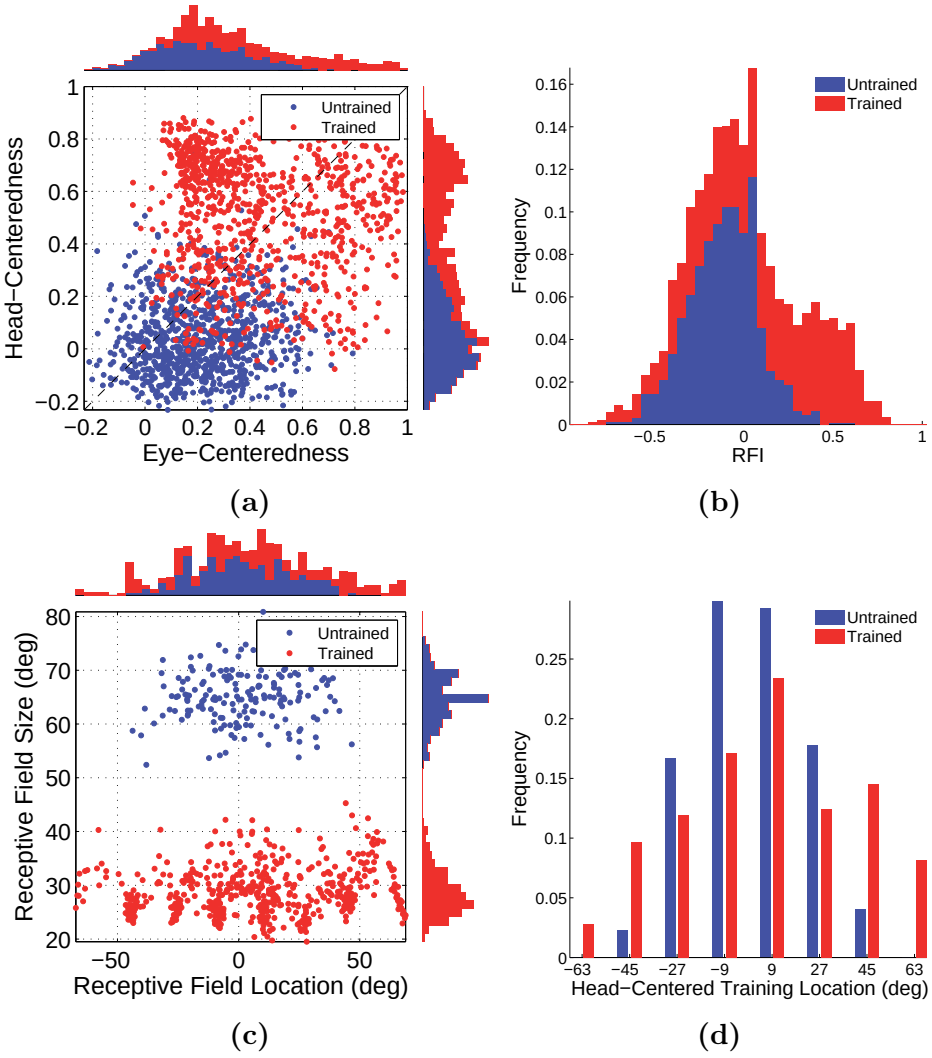


Figure 4.12: Population analyses of the receptive field properties of output neurons for simulations with stochastic variability in the peak responses of input neurons. Results are presented before training (blue) and after training (red). Conventions as in figure 3.3.

where β_i^1, β_i^2 were independently drawn from $U[-30, 30]$. Observe that the same retinal preference is associated with both eye position preferences. An example ensemble of input population response curves can be observed in figure 4.13. The model parameters were otherwise as above.

Figure 4.14 shows the population results from testing the model before and after train-

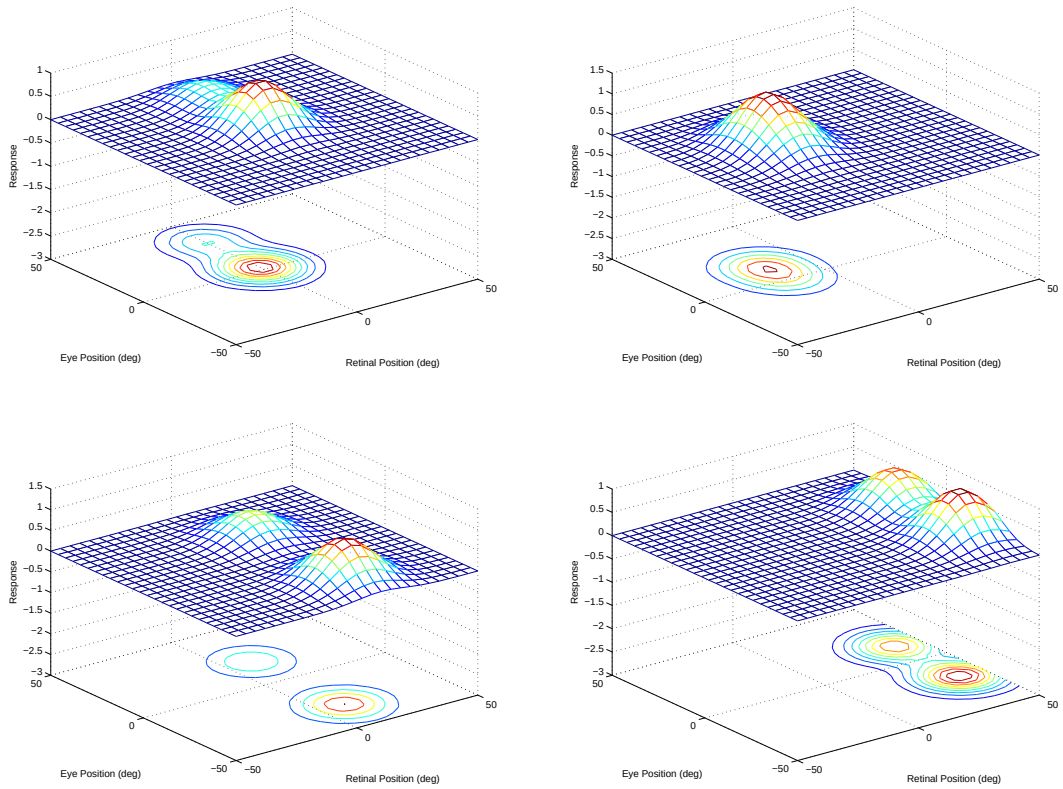


Figure 4.13: Examples of single and muli-peaked response functions for four different input neurons, where the responses are a function of eye position and retinal target location.

ing, and population statistics are given in table 4.6.

The head-centeredness rate increased from ∼17% to ∼56% after training (Fig. 4.14a), and among these neurons the average head-centeredness increased from 0.18 to 0.63 with training. Receptive fields clustered around the training locations (Fig. 4.14c), though training location −63° was under-represented (Fig.4.14d). These results showed that the model could successfully develop more head-centered neurons after training, and these neurons had a larger degree of head-centeredness, even when the gain fields of the input neurons had stochastically assigned multiple peaks with stochastic magnitudes. The model

Experiment 4.2.5				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$17%)	All	RFI > 0 ($\sim$56%)
Head-centeredness	0.04 (0.14)	0.18 (0.12)	0.50 (0.23)	0.63 (0.14)
Eye-centeredness	0.24 (0.17)	0.07 (0.11)	0.41 (0.27)	0.24 (0.14)
RFI	-0.53 (0.59)	0.52 (0.37)	0.11 (0.49)	0.47 (0.24)
RF Location	$-0.47°$ (16.71°)	0.15° (19.34°)	$-1.76°$ (40.19°)	7.57° (33.12°)
RF Size	62.32° (5.13°)	62.67° (5.46°)	25.48° (3.63°)	24.41° (3.14°)

Table 4.6: Population summary statistics of the response properties of output neurons for simulations with multiple peaks in the eye position gain fields of input neurons.

was thus found to offer robust performance when eye position gain fields with a diverse range of multi-peaked functional forms were implemented.

4.2.6 Multiple Simultanously Visible Targets During Training

In this experiment it was investigated whether the model could develop neurons with head-centered receptive fields despite always being exposed to multiple simultanously appearing visual targets during training. This was an important issue to explore because the primate visual system will typically be exposed to multiple objects within a visual scene at any one time.

Stimuli were presented during training in *two* target locations at a time while the eyes moved through a sequence of saccades. Showing stimuli in all possible pairs of target locations was intended to break the statistical coupling between different target locations. This effective statistical decoupling between different target locations can be exploited by

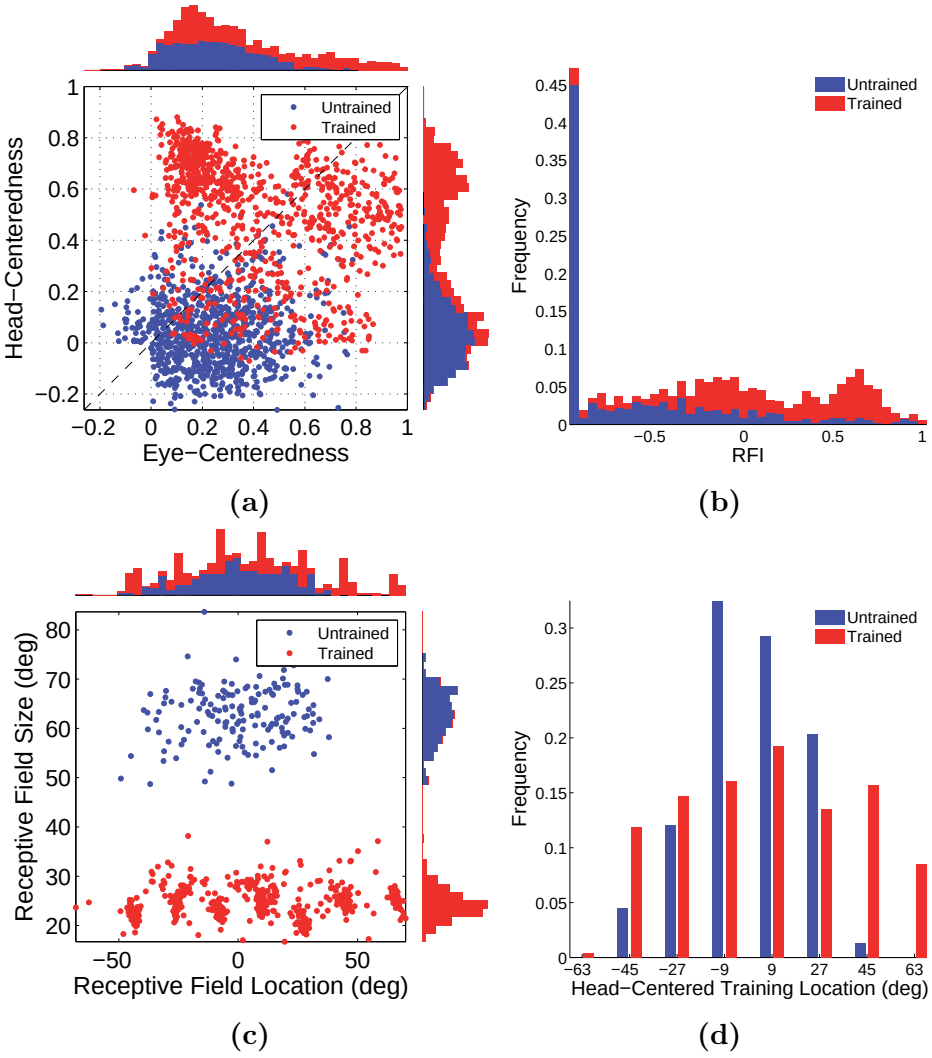


Figure 4.14: Population analyses of the receptive field properties of output neurons for simulations with multiple peaks in the eye position gain fields of input neurons. Results are presented before training (blue) and after training (red).

the architecture and operation of a competitive neural network, which will be forced to learn to represent the individual target locations rather than the pairs of target locations that are actually presented during training. That is, the competitive output layer is forced to develop output neurons that only respond to one target location. A number of papers have previously explored how this general phenomenon of statistical decoupling between

multiple visual objects can allow a competitive neural network model of the primate visual system to develop separate representations of the individual objects that have been shown together in different combinations (e.g. pairs or triples) during training (Stringer & Rolls 2008, Stringer et al. 2007). In particular, Stringer et al. (2007) showed that this mechanism of statistical decoupling in a competitive neural network architecture can be successfully combined with trace learning to simultaneously perform temporal binding of input patterns that occur in temporal proximity.

In the following simulation there were eight head-centered target locations that the visual stimuli could occupy during training. During training, stimuli were presented in two of these locations at a time while the eyes shifted through a sequence of saccades. Therefore, each epoch of training consisted of $\binom{8}{2} = 28$ periods, where the network was exposed to a unique pair of target locations for each such period. In particular, none of the target locations were ever presented singularly to the network during training. Unlike previous experiments, the sequence of eye movements was identical across periods, which was found to permit statistical decoupling with fewer fixations.

Figure 4.15 shows the population results from testing the model before and after training, and population statistics are given in table 4.7.

The head-centeredness rate increased from ∼35% to ∼97% after training (Fig. 4.15a), and among these neurons the average head-centeredness increased from 0.08 to 0.59 with training. The receptive fields were clustered around the eight training locations (Fig.

Experiment 4.2.6				
	Untrained		**Trained**	
	All	RFI > 0 ($\sim$35%)	All	RFI > 0 ($\sim$97%)
Head-centeredness	0.00 (0.08)	0.08 (0.07)	0.58 (0.16)	0.59 (0.15)
Eye-centeredness	0.01 (0.09)	-0.03 (0.06)	0.13 (0.12)	0.12 (0.11)
RFI	-0.01 (0.09)	0.07 (0.07)	0.45 (0.21)	0.47 (0.19)
RF Location	$-0.26°$ (14.64°)	$-0.06°$ (16.31°)	$-0.22°$ (35.71°)	0.10° (35.24°)
RF Size	36.06° (5.05°)	35.99° (5.10°)	24.24° (4.07°)	24.19° (4.07°)

Table 4.7: Population summary statistics of the response properties of output neurons for simulations where multiple visual targets are simultaneously visible during training, as described in section 4.2.6.

4.15c). Neurons did not have multimodal receptive field peaks corresponding to multiple target locations, but rather neurons were tuned to single training locations as desired. This was also reflected in a very even distribution of head-centered receptive fields among the eight locations, which gave a coverage of $\sim$0.99. In summary this showed that the model could successfully develop head-centered output neurons with single peaked receptive fields in head-centered space after being exposed only to *multiple* (i.e. pairs of) visual targets appearing simultaneously during training.

4.2.7 Multiple Simultanous Targets in Testing

In this experiment the influence of multiple simultaneously presented visual targets on the structure of the receptive fields of head-centered neurons was investigated.

For the sake of visualization of the neural responses, only two targets were presented

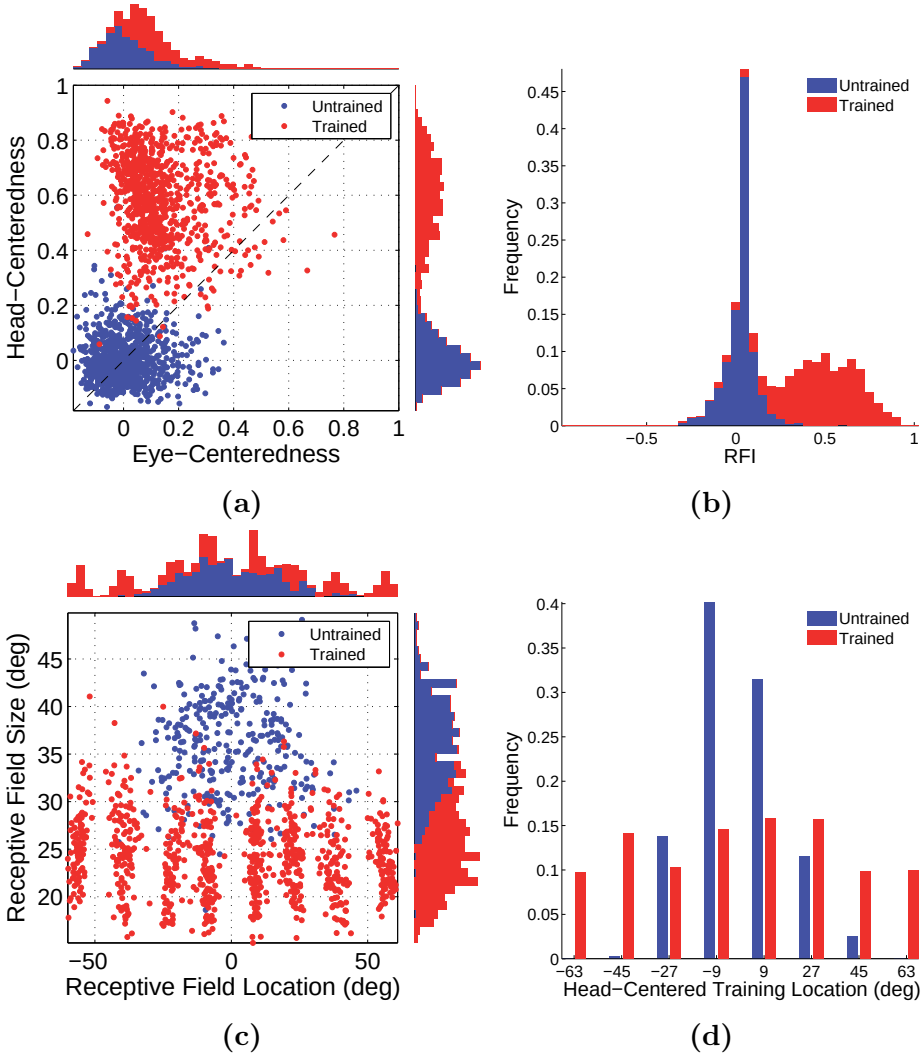


Figure 4.15: Population analyses of the receptive field properties of output neurons for simulations where multiple visual targets are simultaneously visible during training, as described in section 4.2.6. Results are presented before training (blue) and after training (red).

simultaneously during testing. The testing procedure constituted a series of eye position fixations, where for each fixation all possible pairs of head-centered locations, chosen from a fixed set of locations, were occupied by visual targets. More specifically, the model was tested by having it fixate the $E = 4$ eye positions $\{-12°, -4°, 4°, 12°\}$, and for each fixation visual stimuli were presented in all possible pairs of the $T = 40$ head-centered target locations $\{-77°, -73°, \ldots, 79°\}$. This constituted $E \times \binom{T}{2} = 3120$ stimulus presentations during testing, with each presentation corresponding to a unique combination of eye position and pair of target positions in visual field. The firing rates of all output neurons were recorded for each such stimulus presentation. The response of an output neuron to fixation in the i^{th} eye position when the first and second targets were located in the j^{th} and k^{th} target locations was denoted by r_{ijk}.

Standard Model

The responses of head-centered output neurons to the presence of two visual targets was investigated by examining how the neuronal responses varied as a function of the difference between the centre of the receptive field and the nearest target. A scatter plot showing such data points for all combinations of head-centered output neuron, fixation position and pair of target locations is shown in figure 4.16. Large firing rates occurred exclusively in response to targets being located close to the receptive field centre. This confirmed that when the firing rate of a head-centred neuron in the competitive population was large, it was a reliable indication of where a target was located in head-centered space, namely

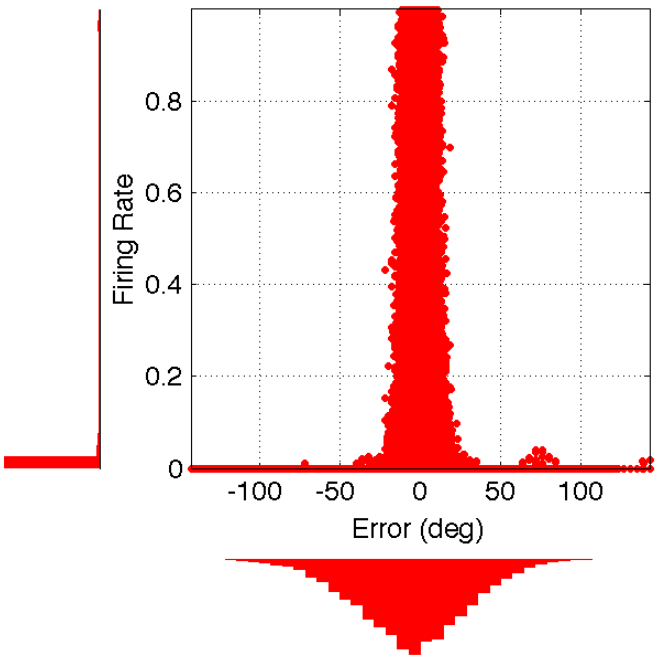


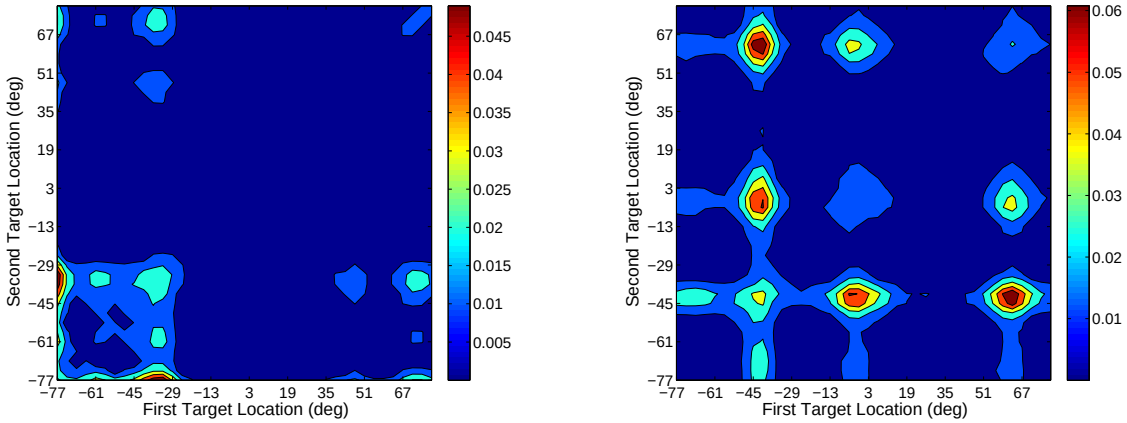
Figure 4.16: Population results for the model tested on two simultaneous targets. Each point in the scatterplot corresponds to a particular combination of output neuron, fixation point and visual target positions. The ordinate is the firing rate of the output neuron, and the abcissa is the difference between the centre of the receptive field of the output neuron and the closest visual target.

near its receptive field centre. However, the results also indicated that sometimes a head-centered neuron did not respond well to a visual target close to its receptive field. In other words, the firing rate of an individual neuron could be, in the face of multiple stimuli, a false negative but not a false positive signal of target presence in the receptive field.

The response of neuron #846 during testing can be seen in the in figures 4.17 and 4.18, corresponding to the untrained and trained condition respectively. In the untrained condition the response of the neuron lacked any apparent regular structure, however in the trained condition the response exhibited several properties. Most clearly, there was

substantially increasing response suppression as the fixation position shifted from right (12°) to left (−12°). At fixation position 12° the response appeared stable to the second target. That is, the location and size of the region of head-centered space within which the first target would elicit a response was unaffected by the location of the second target. However, as the fixation position shifted to the left, the neuronal response was modulated by the fixation position.

At a fixation position of −12° the neuron only responded when both targets were within the neuron's head-centered receptive field. That is, the neuron was unable to respond to a target presented within the neuron's head-centered receptive field if the other target was presented elsewhere. This meant that the presence of multiple targets within the visual field could render the neuron susceptible to false negative responses for some fixation positions. To better understand this effect, figure 4.19 shows the responses of neuron #846 to a single target in different head-centered locations as well as the neuron's synaptic weight vector. These results reavealed that the neuron had a variable response at different fixation positions, and also had a variable amount of synaptic weight devoted across the region of the input population corresponding to it's head-centered receptive field. Moreover, the weaker responses were at fixation positions −18° and −6° (a), which were the fixations closest to the fixations for which there was suppression in the multiple target test, that is −12° and−4°, see Fig. 4.18a,b. This was also in agreement with the weight vector which was severely biased away from these fixation positions (Fig. 4.19b). The source



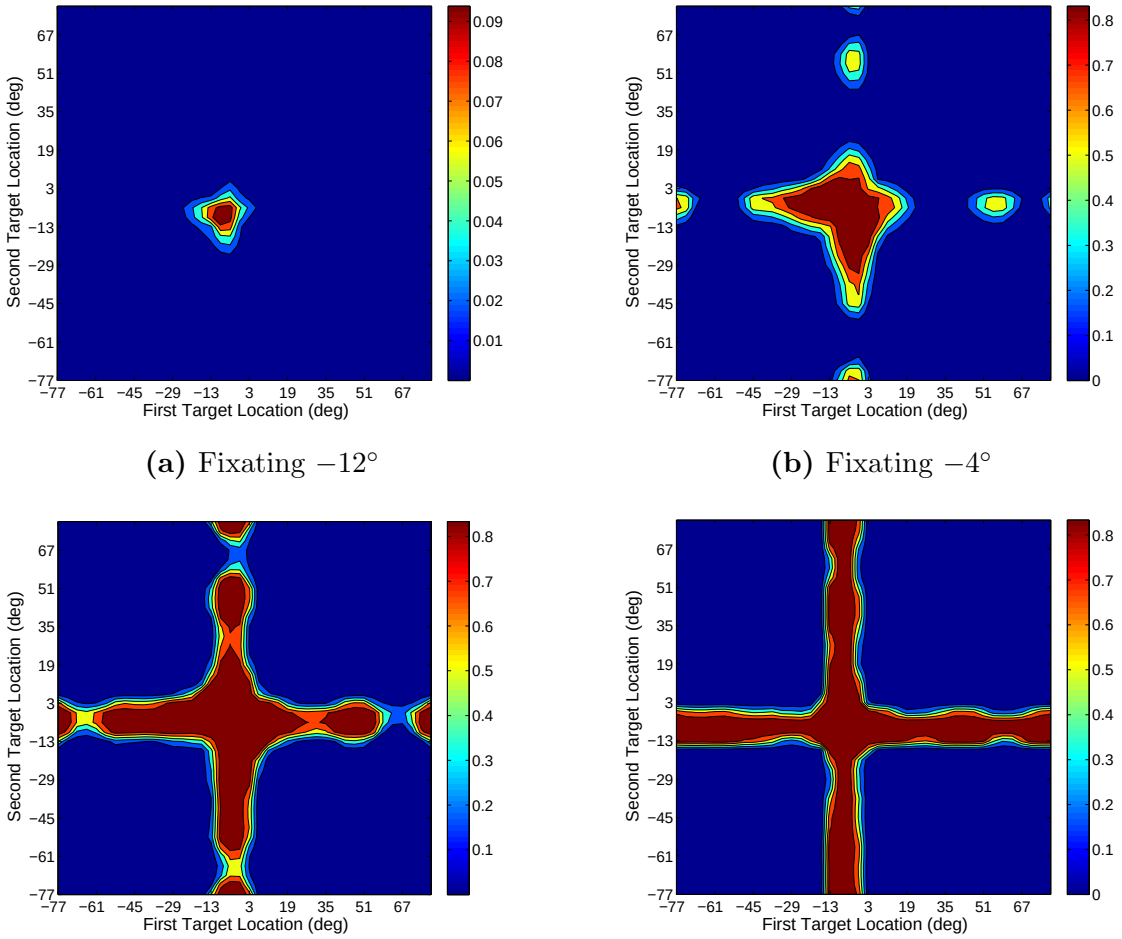
(a) Fixating $-12°$

(b) Fixating $-4°$

(c) Fixating $4°$

(d) Fixating $12°$

Figure 4.17: Each contour plot shows the response of neuron #846 during testing prior to training, at fixation positions $-12°$ **(a)**, $-4°$ **(b)**, $4°$ **(c)** and $12°$ **(d)** respectively. Each plot shows the firing rate of the neuron in response to the simultaneous presence of the first and second targets, whose locations are represented on the horizontal and vertical axes respectively. All plots are symmetric about the diagonal from the origin due to the neuronal response being identical when the two targets switch locations.



(a) Fixating −12°

(b) Fixating −4°

(c) Fixating 4°

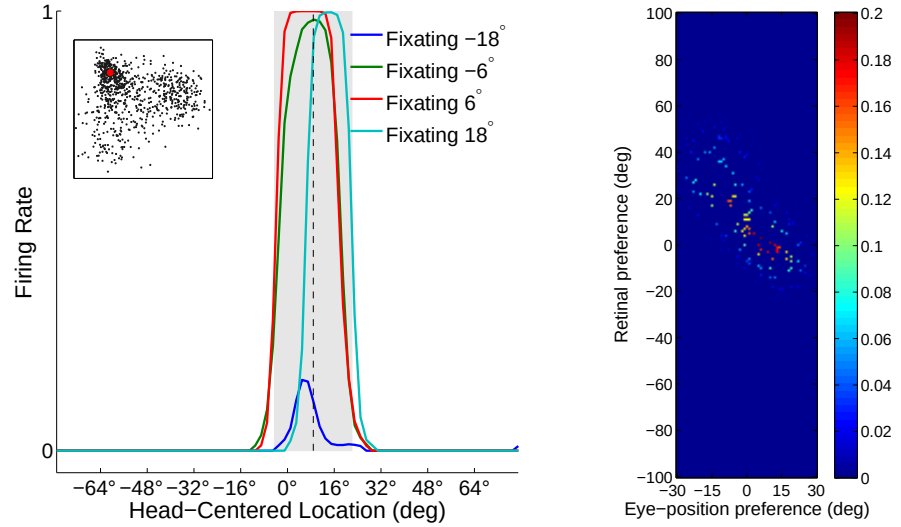
(d) Fixating 12°

Figure 4.18: Each contour plot shows the response of neuron #846 during testing after training, at fixation positions −12° **(a)**, −4° **(b)**, 4° **(c)** and 12° **(d)** respectively.

this synaptic weight variability, which gave rise to the response variability, was firstly the stochastic initial connectivity which limited the postsynaptic neuron's connectivity rate to the input population to 5%. Moreover, there was variability in the degree of synaptic potentiation, the source of which was the complex interaction between the network and stimuli structure. This means that not only was there variability within the same weight vector, but if the weight vectors of two distinct neurons were considered, then there would

be variability between the two as well. This would apply both for neurons with the same head-centered preference, and for neurons with different preferences, the latter of which being relevant to problem of receptive field suppression. This suppression was caused by the combination of across neuron weight vector variability at a given fixation position and the competition model used in the output layer of the network. When a given output neuron was presented with targets in both the neuron's preferred head-centered location and a second different location, and this was at a fixation position which was weakly potentiated for the given output neuron, then the presence of the second target would stimulate other output neurons that competed with and suppressed the output neuron tuned to the first target location. In more detail, the activity of the output neurons stimulated by the second target would render the level of presynaptic stimulation of the given output neuron tuned to the first target location insufficient to be included in the required percentile (π) of the activation distribution. This displacement would be caused by the population of neurons responding to the second target, and this displacement would locate the given neuron below the percentile threshold, and in some cases silence the neuron entirely as seen in figure 4.18.

A control experiment was conducted to confirm the hypothesis that the inflexible competition was causing the suppression by retesting the same model at a lower sparseness percentile of $\pi = 60\%$ rather than $\pi = 90\%$. The results for the same neuron in this new test are shown in figure 4.20. With the reduced sparseness there was an increased robust-

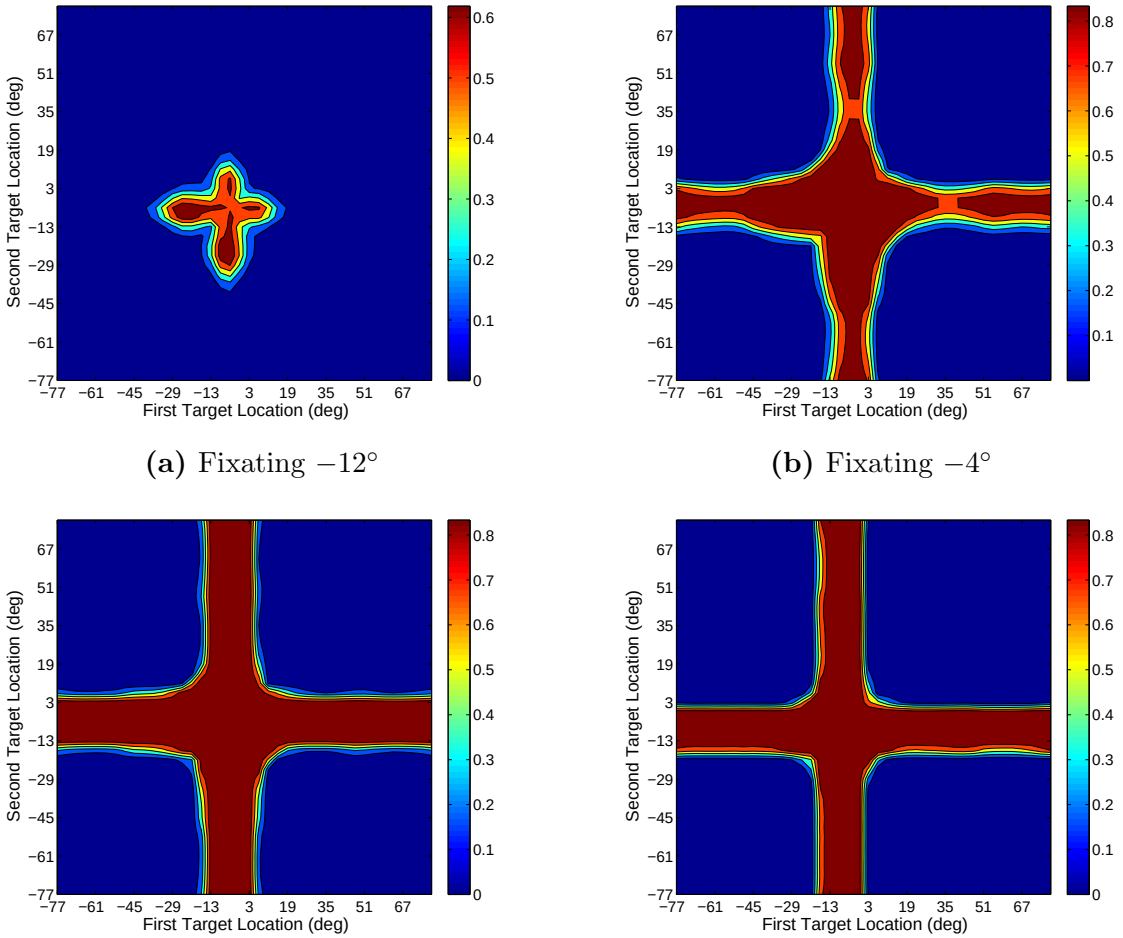


(a) Response **(b)** Weight vector

Figure 4.19: Plot (a) shows the response of neuron #846 to a single target at different head-centered locations during testing, and the vertical dashed line marks the decoded receptive field of the neuron. Plot (b) shows the weight vector of the same neuron. Observe that there is a strong potentiation bias towards positive eye positions.

ness in the response, in particular at fixation position $-4°$ **(b)**, confirming that relaxing the competition allowed the neuron to remain above the firing rate threshold more frequently.

Thus, when the competition model in the output layer was set to a relatively high sparseness such as $\pi = 90\%$, it was observed that output neurons frequently failed to respond when a visual target was presented within their head-centered receptive field. The number of head-centered output neurons that were susceptible to false negative responses when multiple (i.e. two) targets were presented during testing was calculated as follows. The analysis was based on the eight original head-centered locations in which visual targets were presented during training. For each of these training locations, all output neurons with a head-centeredness no less than 0.7 and with a receptive field no more than $5°$ from



(a) Fixating −12°

(b) Fixating −4°

(c) Fixating 4°

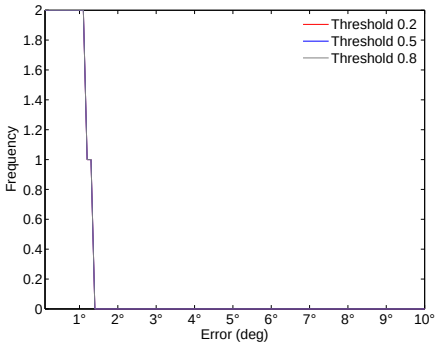
(d) Fixating 12°

Figure 4.20: Response of neuron #846 to two targets presented simultaneously during testing at reduced sparseness $\pi = 60\%$. Results are shown for fixation positions −12° **(a)**, −4° **(b)**, 4° **(c)** and 12° **(d)**.

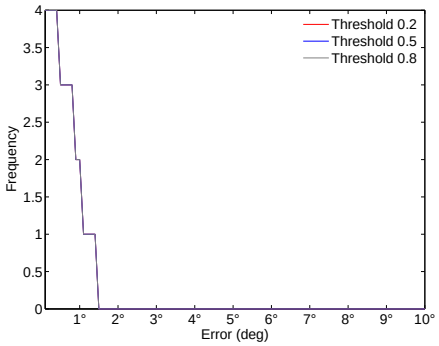
the given training location were grouped for analysis. For each head-centered training location, the number of neurons in the corresponding group which always responded above a given threshold (e.g. 0.2, 0.5 or 0.8) was computed as a function of the distance from the visual target to the receptive field centre. Figure 4.22 shows these functions for the different head-centered training locations and thresholds. The analysis shows that, for

each head-centered training location, only a few (i.e. 2 to 14) output neurons were able to respond reliably to the presence of a visual target within $1-2^{\circ}$ of the receptive field centre of the neuron. Furthermore, no neurons responded reliably to a visual target which was further away than 2° from the neuron's receptive field centre. The results were identical for all three thresholds, 0.2, 0.5 and 0.8. These findings confirmed that the problem of false negative responses due to strong inter-neuron competition within the output layer afflicted all eight head-centered training locations. In particular, the presence of multiple visual targets rendered all head-centered output neurons susceptible to false negative responses when the visual targets were any more than $1-2^{\circ}$ from the receptive field centre of the neuron. Thus, the level of competition in the output layer has a significant effect on the robust performance of the model.

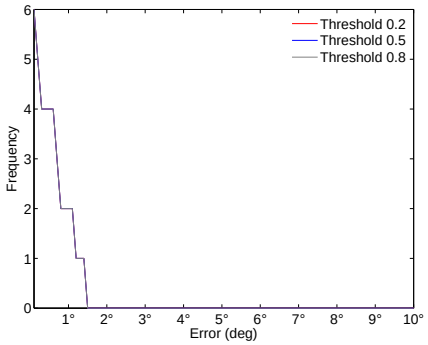
Above it was shown that, with high levels of competition between neurons in the output layer, the responses of individual neurons may not provide reliable indicators of the presence of a visual target within their respective receptive fields across all possible eye fixation positions when other targets are also present within a visual scene. However, one possible solution might be to utilise a more distributed encoding of target location across the population of head-centered output neurons. Consider, for example, neuron #846. For a relatively high level of competition, the suppression of the responses of this neuron was most prevalent at fixation positions -12° and -4°, as seen in figure 4.18. Consequently, neuron #846, by itself, cannot give a reliable indication of target location across all fixation



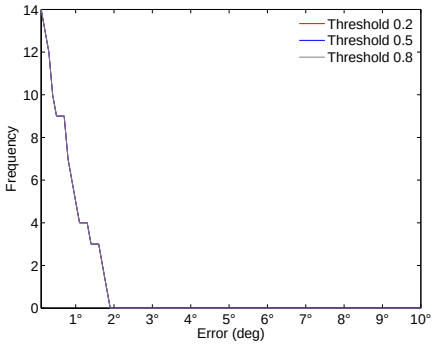
(a) $-63°$



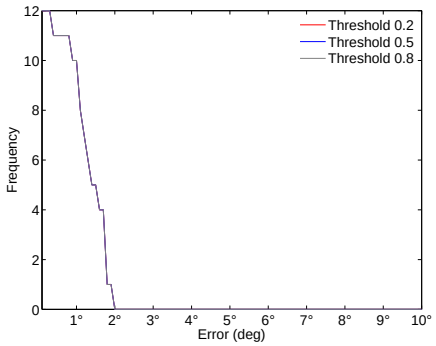
(b) $-45°$



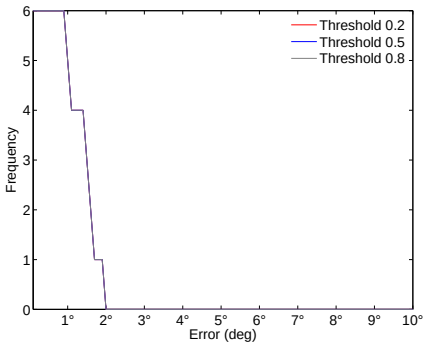
(c) $-27°$



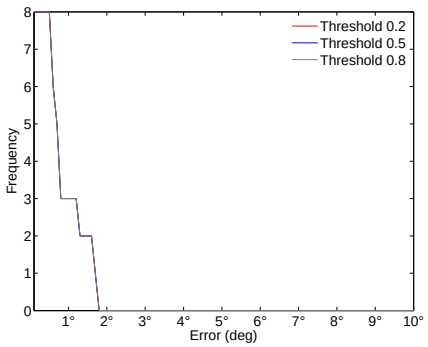
(d) $-9°$



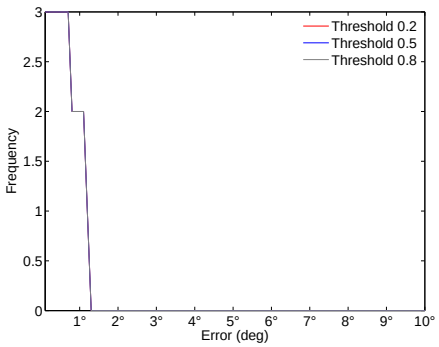
(e) $9°$



(f) $27°$



(g) $45°$



(h) $63°$

Figure 4.22: The number of output neurons that respond reliably to a visual target within their receptive field despite the presence of a second visual target at other head-centered locations. Results are given for the case where there is a relatively high level of competition implemented within the output layer by setting the sparseness $\pi = 90\%$. Each of the eight plots corresponds to a particular group of head-centered output neurons with receptive fields that are no more than 5° from one of the eight original head-centered training locations: (a) −63°, (b) −45°, (c) −27°, (d) −9°, (e) 9°, (f) 27°, (g) 45°, (h) 63°. Each plot shows the number of output neurons which always responded above a given threshold (0.2, 0.5 or 0.8) plotted as a function of the distance from the visual target to the neuron's receptive field centre, despite the presence of a second target presented at various other locations. Specifically, for each threshold, the y-axis represents the number of head-centered neurons that always respond above the given threshold for all head-centered target locations that are equal or closer to the neuron's receptive field centre than the distance represented on the x-axis.

positions. However, if we consider a population of head-centered output neurons, then such false negative responses may not occur consistently on the same subset of fixation positions for all neurons. In particular, which fixation positions suffer suppression may be evenly distributed across the population of head-centered output neurons, with no particular subset of suppressed fixation positions being significantly more prevalent across the output population. In this case, a population of neurons could potentially still provide a robust signal, free of false negatives about the target location when multiple targets are present. Therefore, it was investigated how the suppression was distributed across different fixation positions over the entire population of head-centered output neurons. Figure 4.23 shows the results of this analysis. It was clear that the suppression of neuronal responses, that is low response despite close proximity of a visual target to the neuron's receptive field centre, was approximately equally prevalent across all four fixation positions. This meant that that the suppression of responses was not consistently focussed on a particular subset

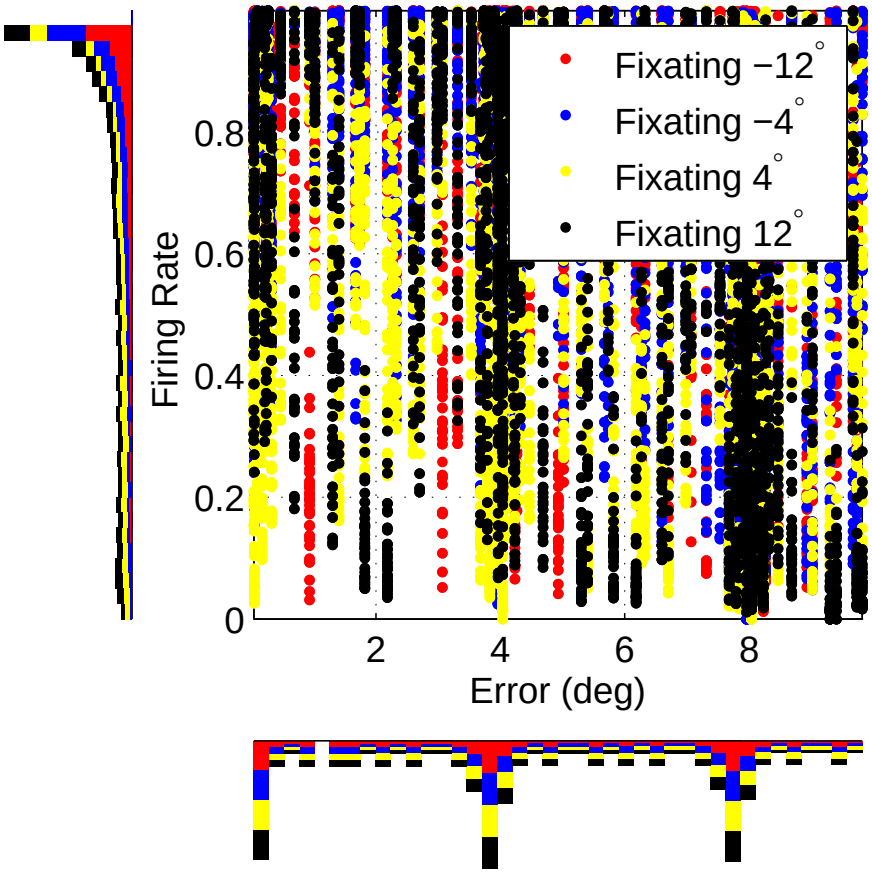


Figure 4.23: Analysis of how eye fixation position affects the responses of head-centered output neurons when the model is tested on two simultaneous targets. Each point in the scatterplot corresponds to a particular combination of output neuron, fixation point and visual target position. The ordinate is the firing rate response of the output neuron, and the abscissa is the distance between the centre of the receptive field of the output neuron and the closest visual target. Results are given for the following four eye fixation positions: $-12°$, $-4°$, $4°$ and $12°$.

of fixation positions. In this case, it is possible that target position might be more reliably represented by a distributed encoding across the entire population of head-centered output neurons.

An Alternative Model of Competition within the Output Layer

Given the sensitivity of the network to the level of competition implemented within the output layer, it was decided to test the performance of the network on the more biologically accurate competition model previously introduced in section 3.2.5. The standard competition model tested above imposes a fixed percentile threshold on the population of output neurons, which strongly prescribes the number of neurons that are activated regardless of the number of visual targets that are present. In contrast, the biologically accurate model implements inhibitory feedback, which is directly proportional to the current total level of activity within the output population. Such a competition model does not so strongly prescribe the number of neurons that are activated in the output layer, and hence may permit the level of activity in the output layer to reflect the number of targets within the visual scene. Therefore, it was hypothesised that this alternative, more biologically accurate, form of competition could potentially ameliorate the suppression of target representations when multiple targets are presented during testing.

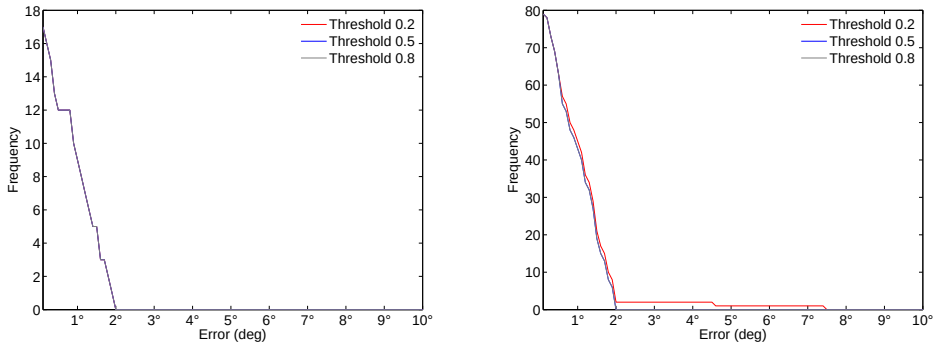
We examined the performance of the network with the more biologically accurate competition model over a range of values of the inhibitory feedback scaling parameter w^-, as described in equation 3.4. Typical network performance is shown in figure 4.24. First, it can be seen that resetting the response threshold now affects the network performance because the number of active output neurons is no longer tightly prescribed with the biologically accurate competition model. Consequently, for head-centered training locations

$-27°$ (c), $-9°$ (d), $9°$ (e) and $45°$ (g), there was an improved result for the relatively low threshold values of 0.2 and 0.5 when compared with results from the standard competition model shown in figure 4.22. Specifically, there were more neurons that were free of false negative responses over a wider range of distances of the target from the neuron's receptive field center. However, for training locations $-63°$ (a), $-45°$ (b), $27°$ (f) and $63°$ (h) there was little or no improvement compared to the previous results, and there was little or no difference between the different response thresholds.

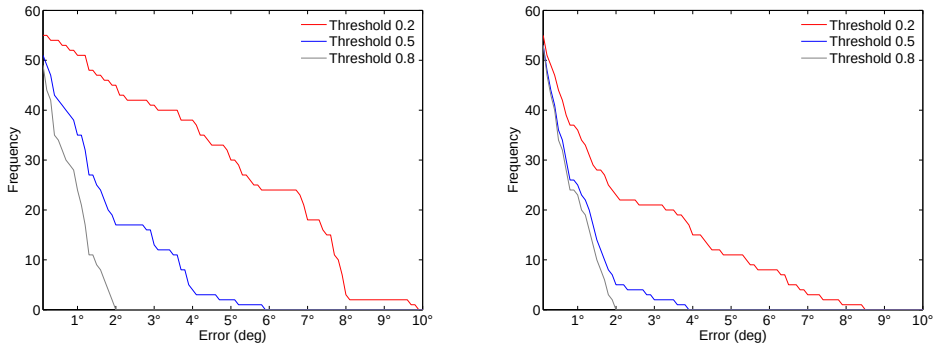
These results showed that while there was a mild improvement from the standard competition model, the more biologically accurate competition did not solve the problem of representing scenes with multiple targets perfectly at the single neuron level.

4.3 Discussion

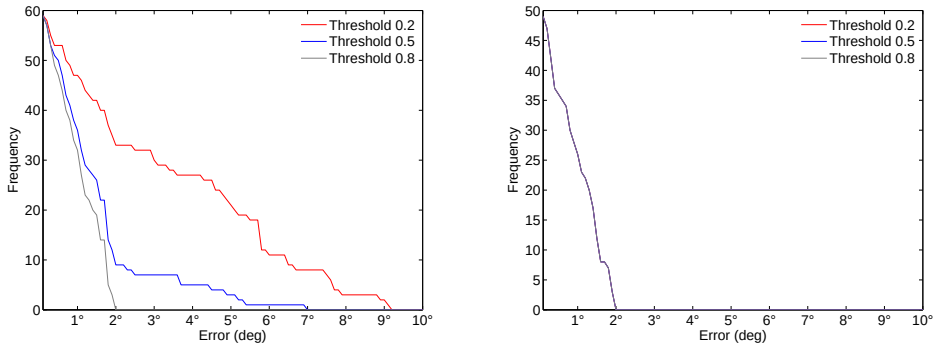
This chapter investigated further issues related to the plausibility of the visually guided development of head-centered neurons by trace learning combined with the natural statistics of eye and head movement. The aim was to extend the more idealized experiments of chapter 3 with more biologically and ecologically constrained model features, thus lending further support to the viability of the model. A wide range of more realistic model features were successfully accommodated without further difficulty. These model extensions included increasing the number of head-centered target training locations, implementing more realistic stimulus dynamics during training, incorporating input neurons with more



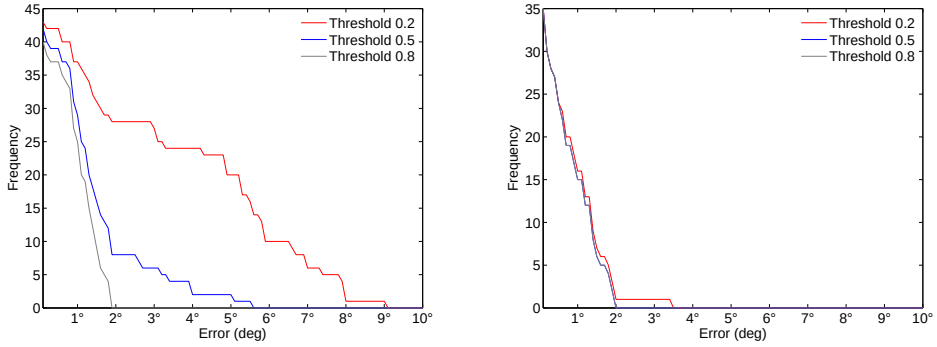
(a) $-63°$ **(b)** $-45°$



(c) $-27°$ **(d)** $-9°$



(e) $9°$ **(f)** $27°$



(g) $45°$ **(h)** $63°$

Figure 4.24: Performance of the network with the more biologically accurate model of competition within the output layer. The plots show the number of output neurons that respond reliably to a visual target within their receptive field despite the presence of a second visual target at other head-centered locations. Note that subfigures **(a)**and **(f)** both only show one curve do to overlap. Conventions as in Fig. 4.22.

realistic forms of gain field, and training on multiple simultaneously visible targets. This work is the first time a model of the development of head-centered neurons has been examined along these dimensions. **However, the model also had some important limitations with respect to biological plausibility. First, representation of multiple targets during testing was imperfect, as a given target would for some eye positions not be well represented because of competition with the other target. Second, the transition between visual scenes involved the simultaneous appearance and disappearance of one or both visual targets, rather than the more natural case where targets appear and disappear at independent times. Third, the competition model used less biologically plausible compared to models where inhibition is mediated through synaptic interaction.**

Increasing the receptive field size in the standard model undermined successful development of head-centered representations because it made input stimulus patterns from distant head-centered locations so overlapping that the output neurons began to generalize in their weight space in an incoherent way across different fixation positions. The proposed solution, which also worked, introduced a nonlinearity in the influence of presynaptic firing on both the postynaptic activation and synapse potentiation. The former is qualitatively similar to short term facilitation, in which large presynaptic spike rates increase underlying vesicle release probabilities at the synapse. This, in turn, increases postynaptic EPSPs and hence postsynaptic spike rates (Zucker & Regehr 2002). The modified learning rule was

coordinated to agree with the nonlinear activation dynamics, making it a homosynaptic LTP rule (Dayan & Abbott 2001), somewhat similar to covariance learning. It should be noted that a nominal comparison of receptive field sizes in this model, as given by σ and ρ for the input neurons and the decoded output receptive field size, with those measured in the relevant brain areas is not appropriate. This is because these values in the abstracted model have no direct relation to quantitative methods for defining receptive field sizes for real neural data. All one can aim to do is assess the qualitative impact of varying σ and ρ on model behaviour, or in the case of output neuron receptive field sizes, to assess the impact of various model changes on the output receptive field sizes.

Increased fixation duration in the standard model undermined successful self-organization of head-centered output neurons when associative synaptic potentiation was combined with global normalization of synaptic weight vectors. This was because neurons then forgot past input patterns in favour of the most recent one presented for a relatively long period of time. To remedy this, an upper bound was implemented on individual synapses, which prevented long periods of fixation from causing indefinitely long potentiations at the single synapse level at the expense of the rest of the weight vector. This allowed the model to successfully self-organise with relatively long fixation durations.

There is a large space of potentially biophysical solutions to the problem of normalizing the synaptic weight vector, and the solution explored is only one successful example among them. For example, an unexplored alternative might be to associate a blocking variable,

b_ij with each synapse, which integrates past potentiation and determines the degree to which subsequent potentiation could take place based on this. Specifically

$$\frac{dw_{ij}}{dt} = \varrho q_i v_j b_{ij} \tag{4.5}$$

where the blocking variable is governed by

$$\frac{db_{ij}}{dt} = \tau_+ (1 - b_{ij}) - \tau_- \frac{dw_{ij}}{dt} \tag{4.6}$$

$$= \tau_+ (1 - b_{ij}) - \tau_- \varrho q_i v_j b_{ij} \tag{4.7}$$

with the initial condition $b_{ij}(0) = 1$. For $b_{ij} = 1$ the learning rule 4.5 would operate as a standard trace rule. However, as the synapse is potentiated with a positive $\frac{dw_{ij}}{dt}$ this would cause 'leak' from b_{ij} at a time scale governed by τ_-. As b_{ij} reduces to 0, this would in turn diminish or even zero the effective learning rate of the synapse. The blocking variable b_{ij} could be simultaneously driven by a 'charging' term that seeks to restore b_{ij} to a resting value of 1 at a time scale driven by τ_+. Future modelling studies could usefully explore this and other potential solutions to the problem of preventing long fixation durations with synaptic normalisation from overwriting learning of previously encountered stimulus patterns.

Training the standard model on multiple simultaneous visual targets produced head-centered receptive fields analogous to those found when training with a single target.

However, testing the model on multiple simultaneous targets led to head-centered output neurons responding unreliably to targets within their receptive fields across different eye fixation positions. In particular, head-centered output neurons gave false negative responses for some fixation positions. This failure was attributed to variable levels of synaptic potentiation to input neurons across different eye positions, in combination with the standard competition model in the output layer that only permitted a fixed percentile of neurons to reach firing threshold. The presence of multiple visual targets caused more severe competition between the output neurons. In this case, a head-centered neuron with a target within its receptive field would receive strong competition from other neurons that are responding to other targets. This was found to suppress the responses of individual head-centered output neurons for a subset of eye fixation positions. This suppression happened to all of the head-centered output neurons. In particular, all neurons gave false negative responses for some eye positions when the target was more than 2° from the neuron's receptive field centre. However, the suppression of neuronal responses in the output population did not consistently take place for the same subset of eye fixation positions. In particular, no particular subset of fixation positions was more frequently suppressed across the output population. Hence it was proposed that the entire population of head-centered output neurons could collectively provide a robust signal, free of false negatives, across all eye fixation positions.

Spratling (2009) was also able to learn head-centered representations using scenes with

multiple simultaneously visible objects. This was achieved by selectively blocking the anti-Hebbian learning update on the single synapse which gave the largest normalized drive to the postsynaptic neuron. Since the corresponding input neuron would encode the retinal location of a single object, this would cause the postsynaptic neuron to only learn about a single object location. However, it is not clear how such selective treatment of synaptic plasticity in the synapse from the maximal input unit could be implemented biophysically. If more than a single synapse were to be excluded from the learning update, then it could not be guaranteed that both corresponded to inputs encoding the same object location.

Alternative stimulus dynamics were explored by including additional periods of randomised visual target and eye movements, as well as varying the length of eye position fixations. However, the stimulus dynamics in these experiments were still substantially less rich than natural dynamics. Both the retinal and eye position spaces in these simulations were one dimensional, and including a second dimension in both is likely to introduce a new range of dynamics. Perhaps more importantly, there was no relationship between the movement dynamics of the head, eyes and the visual stimulus that the model was presented with. However, this is certainly not the case under natural conditions considering that animals frequently reorient gaze to acquire visual targets.

4.4 Summary

The following is a summary of the main results and insights in this chapter.

- An increase in the number of target locations the model is exposed to during training improves the head-centeredness rate and decreases the average receptive field size.

- An increase in the receptive field size of the input neurons prevents the development of head-centered output neurons for all target locations. However, this can be ameliorated by introducing a non-linearity in the contribution individual synapses make to the activation of the postsynaptic neurons, similar to facilitation in short term plasticity.

- An increase in the degree of randomised movement dynamics during training results in a decreased head-centeredness rate after training, but the rate did remain above 20% for values explored.

- An increase in fixation duration induced forgetting of past input patterns at the single neuron level, and consequently individual neurons ended up responding to only a single combination of eye fixation position and retinal target location. However, both prolonged fixation and fixation variability within the same experiment can be accommodated by introducing an upper bound on the magnitude of each synaptic weight.

- Gain fields with variable peak size across the input population, or even multiple peaks for some fraction of the input population, still allows the model to properly develop head-centered output neurons.

- Training on multiple visual targets simultaneously still permits successful development of head-centered output neurons as long as the targets are presented in different combinations, such as pairs in the above simulations, on different occasions.

- Testing on multiple visual targets produces typical head-centered receptive fields, however suppression of neuronal responses takes place at some eye fixation positions for all neurons.

Chapter 5

Head-Centered Receptive Fields: Monotonic Gain Fields

5.1 Introduction

This chapter presents neural network simulations investigating the ability of output neurons to develop head-centered visual receptive fields when the network is driven during learning by input neurons with sigmoidal eye position gain fields as described by equation 2.2. That is, the same methodology and hypothesis from chapters 3 and 4 is investigated, with only the gain modulation in the input population altered from peaked to monotonic. Experimental results establishing model failure are presented, and the cause of the failure is analyzed both by investigating the learning process in detail and conducting a theoretical analysis of the nature of the different types of gain fields which have been used. These theoretical results explain precisely the nature of the failure and how it is attributable to the presence of monotonic gain fields, and it is shown that even a prewired model which does have head-centered responses is subject to the same failure when subjected to synaptic

plasticity. The implications of this failure on past theoretical work and contemporary understanding of the role of gain fields is discussed.

5.2 Gain Modulation

In chapters 3 and 4, the basic hypothesis set out in Section 2.2, was tested and validated in computer simulations using peaked shaped gain modulated input neurons. However, the most frequently identified form of eye position gain modulation in the macaque cerebral cortex and consequently most frequently studied form of modulation in theoretical sensorimotor integration models, has been monotonic gain modulation. The location of most of these monotonic gain fields suggests that the head-centered neurons found in area LIP are likely derived from them (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990).

Monotonic eye position gain modulation refers to the trend of increasing discharge in an eye-centered visual neuron as the eye fixation position shifts along one of the two cardinal axes. This is often modelled as a multiplicative interaction between a Gaussian retinotopic receptive field with a sigmoidal or linear function of eye position along a cardinal axis. Most theoretical work has used linear gain modulation, with the exception of Pouget & Sejnowski (1997) which used a sigmoidal modulation interaction based on empirical evidence suggesting this as more appropriate. This work also uses sigmoidal modulation, primarily due to the convenience of the saturation behaviour, but the qualitative nature of the results are not contingent on this. Figure 2.2(b) shows a response function that is

modulated by a sigmoidal eye position gain field described by equation 2.2.

5.3 Results

5.3.1 Prewired Model

In this experiment we explored the performance of a manually hardwired network model, where the weight the vectors were set to have a structure corresponding to the outcome of an ideal learning process. The manually wired model was tested on the same stimuli that were to be used later in the simulations described in Section 5.3.2 with learning. The purpose of this experiment was to establish a baseline successful network peformance, which would inform subsequent experiments.

There were 24522 neurons in the input population, each corresponding to a unique combination of retinal-position preference (α_i), eye-position preference (β_i) and slope (κ_i). There were 900 neurons in the output population, each given a head-centered receptive field at one among nine head-centered locations, which were $-68°$, $-51°$, $-34°$, $-17°$, $0°$, $17°$, $34°$, $51°$ and $68°$. Each neuron in the output population was postsynaptically connected to a randomly assigned subpopulation of the input population. There were only two synaptic weight values across all synapses, simply referred to as elevated and depressed. The strength of a synapse was elevated if the presynaptic input neuron responded maximally to a combination of eye-position and retinal location corresponding to a head-centered location that was closest to the head-centered location assigned to the output neuron.

Otherwise the synapse was depressed. Therefore, the output neuron received strong driving input to the extent that a visual target was near its assigned head-centered location. Specifically, the weight assigned to a synapse with a postsynaptic neuron assigned to head-centered receptive field location h and a presynaptic neuron having retinal-preference α, eye-position preference β and $\kappa > 0$ was given by

$$w_{h,\alpha,\beta} = \begin{cases} 10 & h - \beta \leq \alpha \leq h + E/2 \quad \text{(Elevated)} \\ 1 & \text{else} \qquad\qquad\qquad\quad \text{(Depressed)} \end{cases}$$

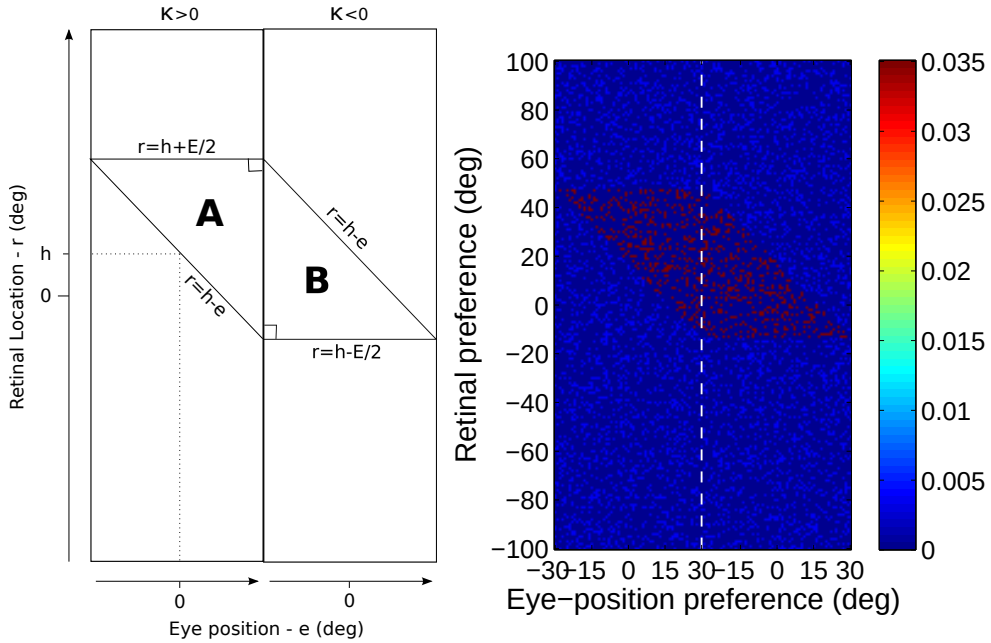
where $E = 60^\circ$ was the size of the eye-position dimension. Likewise when $\kappa < 0$ the weight was given by

$$w_{h,\alpha,\beta} = \begin{cases} 10 & h - E/2 \leq \alpha \leq h - \beta \quad \text{(Elevated)} \\ 1 & \text{else} \qquad\qquad\qquad\quad \text{(Depressed)} \end{cases}$$

Figure 5.1 shows the structure of the canonical weight vector and the actual weight vector (with diluted connectivity) of output neuron #139 from the manually wired model. Before testing the network, the synaptic weight vectors of all output neurons underwent the normalization step described by equation 2.7.

To provide a baseline for comparison, a network with randomly wired synaptic connections was also tested in the same way. The parameters for the randomly wired model can be found in table 5.1.

Figure 5.2 shows population analyses of the response properties of output neurons in



(a) **(b)**

Figure 5.1: Synaptic connectivity in the manually wired network model. The schematic (a) shows the regions of the retinal position $\times$ eye position space represented by the two subpopulations input neurons that are connected with elevated synaptic weights to an output neuron that has been assigned head-centered location h. Each of the two rectangles in (a) represent the topographic organization of one half of the input population in terms of retinal-preference (α_i) and eye position preference (β_i), with the input neurons in the left rectangle having $\kappa > 0$ (positive gain) and the right rectangle having $\kappa < 0$ (negative gain). A neuron in the competitive output population which has been assigned a head-centered receptive field at location h will have elevated connections from input neurons with preferences located in the right-angle triangles of the input space, labelled **A** and **B**. The scatterplot (b) shows the synaptic weight vector of output neuron #139 from the manually wired model, with the input neurons arranged in the same topographic manner.

the randomly wired model and the manually wired model. Additional population statistics for these output cells are given in table 3.2.

There were no head-centered neurons in the randomly wired model. However, in the manually wired model the head-centeredness rate was ∼77%. This is evident in the scatterplot shown in Fig. 5.2b. The average RFI also increased from −0.77 in the randomly

Parameter	**Symbol**	**Value**
Width of retinal tuning curve	σ	6°
Output neuron population size	N	900
Input neuron population size		24522
Activation time constant	τ_h	100ms
Activation function slope	φ	4
Activation function threshold	θ	0
Sparseness percentile	π	90%
Connectivity rate	ϕ	8.16%

Table 5.1: Parameters of the manually hardwired model.

wired model, which indicates a response more compatible with an eye-centered reference frame, to 0.14 in the manually wired model, which indicates a response more compatible with a head-centered reference frame (table 3.2). The manually wired model also had an average head-centerness among head-centered neurons of 0.70. These results showed that the manually wired model, with sigmoidal modulation in the input population, had neurons responding in a head-centered frame of reference, in contrast to the randomly wired model which did not.

Head centered receptive field properties were only computed for head-centered neurons in the manually wired model since the randomly wired model had no head-centered neurons. The receptive fields were distributed quite uniformly across the eight head-centered locations to which the output neurons had been prewired (Fig.5.2c,d). Furthermore, the manually wired model had a coverage near the maximum possible value of 1.

In summary, these results demonstrated that the underlying network architecture with sigmoidal eye position gain modulation of input neurons could effect head centered output

Experiment 5.3.1				
	Random		**Manual**	
	All	RFI > 0 (0%)	All	RFI > 0 ($\sim$77%)
Head-centeredness	0.07 (0.17)	-	0.66 (0.14)	0.70 (0.12)
Eye-centeredness	0.88 (0.08)	-	0.52 (0.19)	0.47 (0.17)
RFI	-0.77 (0.15)	-	0.14 (0.22)	0.23 (0.14)
RF Location	$-0.53°$ (20.87°)	-	$-1.20°$ (44.07°)	0.25° (42.80°)
RF Size	68.97° (7.16°)	-	28.13° (2.80°)	28.31° (2.58°)

Table 5.2: Population summary statistics of response properties of output neurons in the randomly and manually wired models described in Section 5.3.1.

neurons when given an appropriate synaptic weight structure. In the simulations reported in this section, the weights were manually hardwired. Next we investigate the performance of the network with synaptic weights set up through learning.

5.3.2 Model with Learning

Learning

In this experiment it was explored how well the model described in chapter 3 could develop head-centered neurons when the neurons in the input population had sigmoidal eye position gain modulation as described by equation 2.2 and shown in Fig. 2.2(b). The model parameters are shown in table 5.3.

Figure 5.3 presents the population analyses of the receptive field properties of the output neurons before and after training, and population statistics are given in table 5.2.

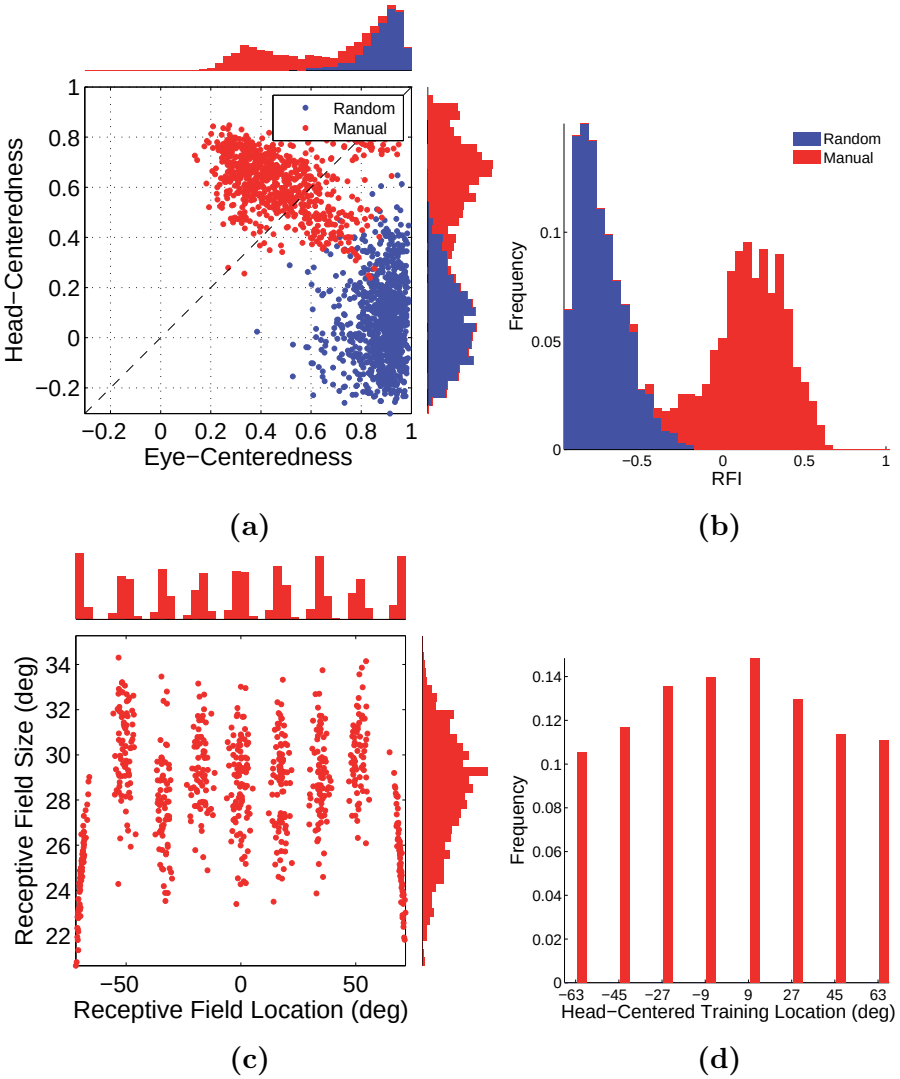


Figure 5.2: Population analyses of receptive field properties of output neurons in the randomly wired and manually wired models described in Section 5.3.1.

The untrained model had only eye-centered neurons and an average eye-centeredness of 0.88, while the trained model had a single head-centered neuron but an average eye-centeredness of 0.96. This meant, as is confirmed in figure 5.3, that training had the effect of promoting more eye-centered rather than head-centered responses.

Parameter	**Symbol**	**Value**
Number of target locations	M	8
Fixation sequence length	P	15
Number of training epochs	-	10
Width of retinal tuning curve	σ	6°
Output neuron population size	N	900
Input neuron population size		24522
Trace time constant	τ_q	400ms
Activation time constant	τ_h	100ms
Activation function slope	φ	4.5
Activation function threshold	θ	0
Sparseness percentile	π	90%
Learning rate	ϱ	0.05
Synaptic connectivity	ϕ	5%

Table 5.3: Parameters of self-organizing model with sigmoidal eye position gain modulation of input neurons.

Input Population Covariance Patterns

The previous experiment showed that the model failed when the input population had sigmoidal eye position modulation, despite succeeding with peaked eye position modulation in previous experiments. This raised the question of what the difference was between the two different forms of input encoding from the perspective of competitive learning. It is well known that standard competitive networks develop weight vectors that reflect the covariance between the activities of input neurons. In particular, there is a tendency for output neurons to learn to respond to subsets of input neurons whose activities are highly correlated. Hence inspecting the covariance between input neurons across all input patterns would reveal what structure the weight vectors should converge towards under standard competitive learning conditions with a Hebbian learning rule.

Experiment 5.3.2				
	Untrained		**Trained**	
	All	RFI > 0 (0%)	All	RFI > 0 ($\sim$0.24%)
Head-centeredness	0.07 (0.17)	-	0.22 (0.15)	0.14 (0.00)
Eye-centeredness	0.88 (0.09)	-	0.96 (0.08)	-0.04 (0.00)
RFI	-0.77 (0.15)	-	-0.75 (0.14)	0.14 (0.00)
RF Location	0.11° (21.16°)	-	0.58° (40.95°)	0.39° (0.00°)
RF Size	66.09° (7.51°)	-	24.93° (7.46°)	9.53° (0.00°)

Table 5.4: Population summary statistics of response properties of output neurons in the model with sigmoidal eye position gain modulation of input neurons as described in Section 5.3.2. Results are presented before training (left two columns) after training (right two columns).

The covariance between input neurons may be computed in the same way for both input neurons with peaked gain fields described by equation 2.1 and input neurons with sigmoidal gain fields described by equation 2.2. In the case of an input neuron with retinal preference α and eye position preference β, the covariance between it and a second input neuron with corresponding preferences α^*, β^* is given by

$$\mathrm{cov}_{\alpha,\beta}(\alpha^*, \beta^*) = \iint_{R \times E} \left(R_{\alpha,\beta}(r,e) - \overline{R}_{\alpha,\beta}\right)\left(R_{\alpha^*,\beta^*}(r,e) - \overline{R}_{\alpha^*,\beta^*}\right) dr de \tag{5.1}$$

$R_{\alpha,\beta}(r,e)$ is the response of a neuron with preferences α, β to a visual target at retinal location r and eye position e, as given by one of the equations 2.1 or 2.2. The term $\overline{R}_{\alpha,\beta}$ is the average response of the same neuron across all possible inputs in the $E \times R$ space,

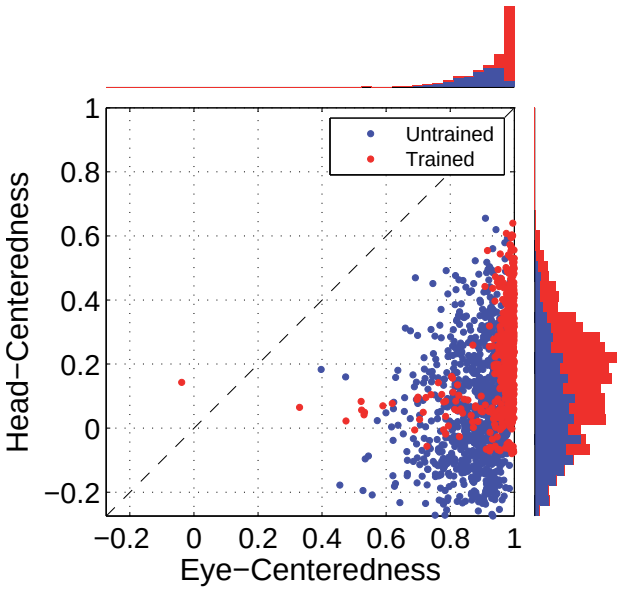


Figure 5.3: Population analyses of receptive field properties of output neurons in the model with sigmoidal eye position gain modulation of input neurons as described in Section 5.3.2. Results are presented before training (blue) after training (red).

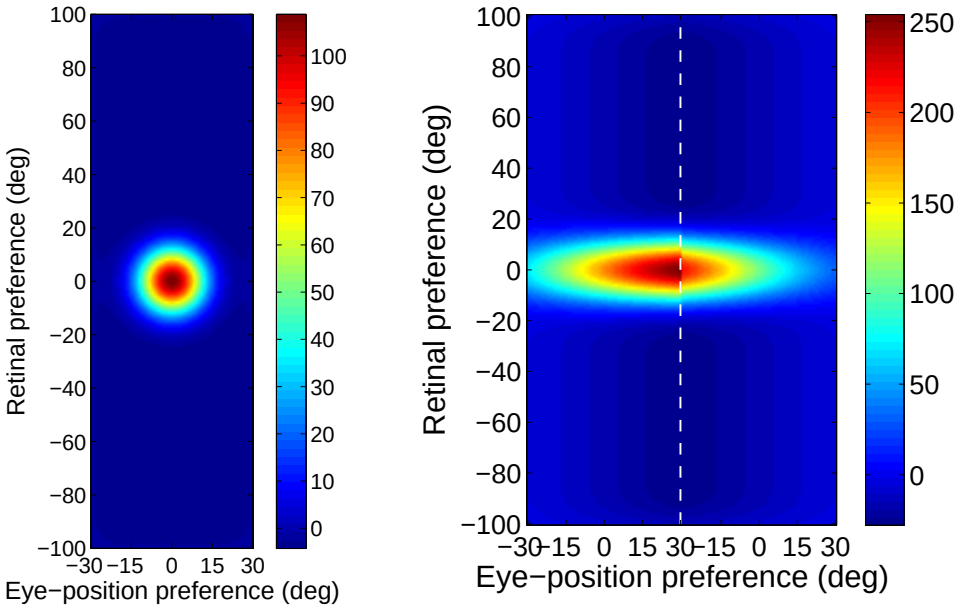
that is

$$\overline{R}_{\alpha,\beta} = \frac{1}{|R \times E|} \iint_{R \times E} R_{\alpha,\beta}(r,e) dr de \tag{5.2}$$

Figure 5.4a shows this covariance map for an input neuron with a peaked gain field and preferences $\alpha = \beta = 0°$, and Figure 5.4b shows the covariance map for an input neuron with sigmoidal gain field and preferences $\alpha = \beta = 0°$ and $\kappa > 0$.

The structure of covariance in the peaked modulation case is functionally identical to the response function of the input neuron, namely a two dimensional Gaussian tuning curve. The form of this covariance function is obvious by considering the correlations between the activities of input neurons with peaked gain fields.

In the sigmoidal modulation case, the situation is more complicated. The strong co-



(a) Peaked modulation **(b)** Sigmoidal modulation

Figure 5.4: Each plot shows the covariance between a given input neuron and the rest of the input neuron population in the form of a topographic map analogous to the weight vector maps shown above. Parts (a) and (b) show results for input neurons with peaked gain fields and sigmoidal gain fields, respectively. In both cases the input neuron has preferences $\alpha = \beta = 0°$, and in the sigmoidal case the input has positive gain ($\kappa > 0$).

variance is localised within the retinal preference dimension, but elongated within the eye-position dimension. This can again be understood by considering the response functions of the input neurons. Firstly, because all input neurons have a sharp, peaked tuning profile in the retinal preference dimension, any two input neurons need to have similar retinal preferences in order to have the possibility of being coactive. This explains the localisation of strong covariance in the retinal preference dimension. Secondly, the elongated form of the covariance function in the eye-position dimension results directly from the sigmoidal gain fields as follows. In the subpopulation of input neurons with a positive

gain direction, similar to the reference input neuron $(0°, 0°)$ itself, it is clear that other neurons with a similar retinal preference and with an eye position preference to the right (i.e. larger than $0°$) cofire more frequently with the reference neuron. This is because a positive gain implies that an input neuron responds to all eye positions to the left of (i.e. smaller than) the eye position preference of the neuron. Conversely, a negative gain implies that an input neuron responds to eye positions to the right of (i.e. greater than) the eye position preference of the neuron. Hence, in the subpopulation of input neurons with a negative gain direction, it can be seen that neurons with a similar retinal preference and with an eye position preference to the left of (i.e. smaller than) $0°$ cofire more frequently with the reference neuron.

The covariance maps shown in Figure 5.4a and 5.4b predict the structure of the weight vectors that we would expect to see develop in a competitive network with a standard Hebbian learning rule trained over all input patterns with either peaked gain fields or sigmoidal gain fields, respectively. These predictions were tested by running simulations with a Hebbian learning rule with both peaked and sigmoidal gain modulated input neurons. For each simulation, there were 200 training patterns corresponding to random locations in the $E \times R$ space. The activation time constant was reduced to $\tau_h = 30ms$ to avoid any trace effect during learning. Figures 5.5a and 5.5c show the synaptic weight vectors of two typical output neurons that developed after training with the Hebbian learning rule when the input neurons were modulated by either peaked gain fields or sigmoidal gain fields,

respectively. It is clear that these synaptic weight vectors have a very similar structure to the corresponding covariance maps shown in Figure 5.4. Thus, with the Hebbian learning rule, the underlying correlations between the activities of the input neurons with either peaked gain fields or sigmoidal gain fields shape the synaptic weight structure that develops during training. Most importantly, with the sigmoidal gain fields, the synaptic weights are localised within the retinal preference dimension, but elongated within the eye-position preference dimension. This kind of synaptic weight structure leads to eye-centered output responses.

Next, comparison simulations were run with the trace learning rule. Figures 5.5b and 5.5d show the synaptic weight vectors of two typical output neurons that developed after training with the trace learning rule when the input neurons were modulated by either peaked gain fields or sigmoidal gain fields, respectively. Figure 5.5b shows a diagonal band of potentiated synaptic weights, which correspond to input neurons representing the same head-centered location but with different combinations of retinal location and eye-position. Thus, with peaked gain fields, the trace learning rule is able to simply bind together clusters of input neurons along a diagonal line in the (retinotopic preference $\times$ eye-position preference) input space corresponding to a particular head-centered location. Output neurons will then respond to particular head-centered locations regardless of eye-position or the retinal location of a visual target. However, the situation is quite different with sigmoidal gain modulated input neurons. Figure 5.5d shows a very similar weight structure to that

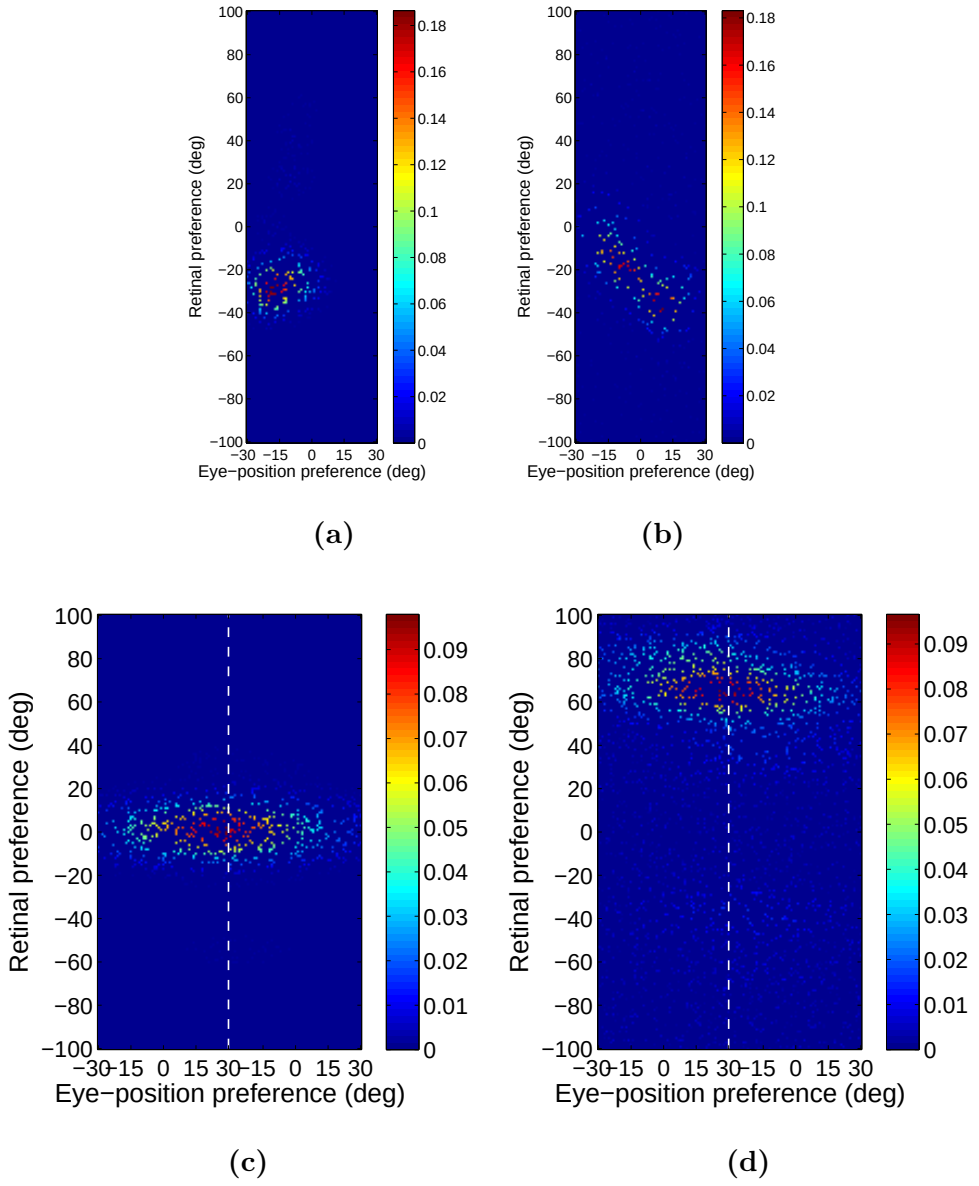


Figure 5.5: The top row shows the weight vectors of two typical output neurons that develop when the input neurons have peaked eye position gain modulation and the network is trained with either the Hebbian learning rule **(a)** or the trace learning rule **(b)**. The bottom row shows the weight vectors of two typical output neurons when the input neurons have sigmoidal eye-position gain fields and the network is trained with either the Hebbian learning rule **(c)** or the trace learning rule **(d)**.

obtained with the Hebbian learning rule. In particular, in both of these last two cases, the weight vector was very similar to the covariance structure found among the input neurons with sigmoidal eye position modulation (Fig.5.4b). Thus, with sigmoidal gain fields, even if a trace learning rule is implemented, the output neurons still learn to represent eye-centered rather than head-centered locations. This is because developing head-centered output responses would require the trace learning rule to do more than simply bind input patterns together. With sigmoidal gain fields, trace learning must also disrupt and break apart output representations corresponding to clusters of highly correlated input neurons, which are localised in the retinotopic preference dimension but elongated in the eye-position preference dimension. However, in practice the trace learning effect is not strong enough to achieve this. Consequently, even with trace learning, these elongated clusters of input neurons with correlated activities continue to drive the development of eye-centered output neurons, as was observed in the simulations reported in Section 5.3.2.

5.3.3 Introducing Learning into the Prewired Model

In this experiment it was investigated what impact introducing synaptic plasticity would have on the manually prewired model in section 5.3.1. This would inform to what extent the problem outlined in section 5.3.2 was solely an obstacle in developing the circuit from scratch, or whether it would undermine a circuit which was already functioning well, for example due to genetic prewiring. The prewired model was trained for 10 epochs with the

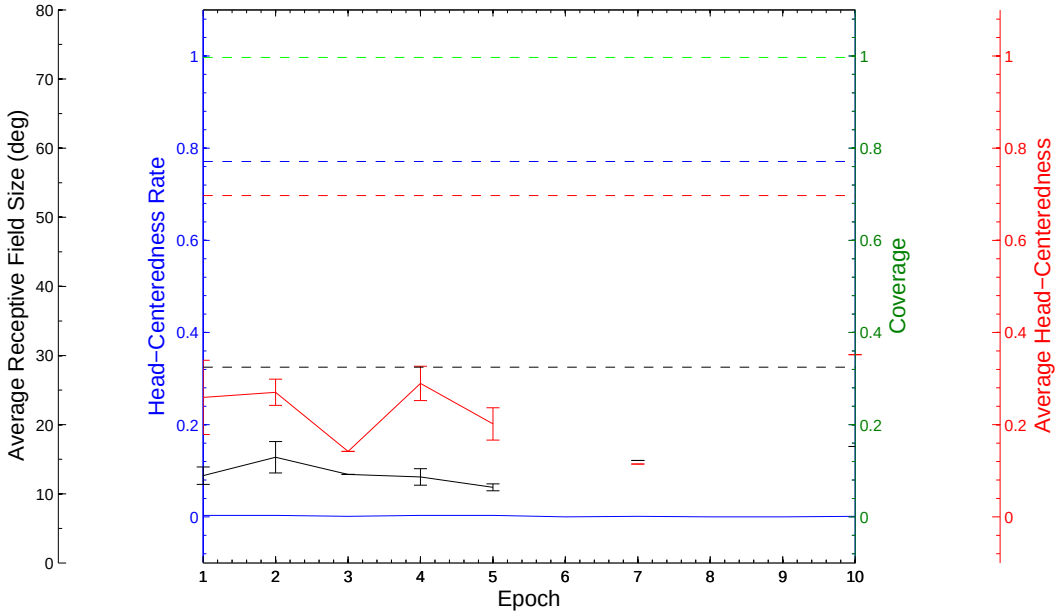


Figure 5.6: Population analyses of receptive field properties of output neurons when learning was introduced. Results are shown through successive training epochs. The dashed lines correspond to the prewired model before training. Observe that after 7 epochs there were no head-centered neurons, and consequently the average head-centeredness and receptive field size among head centered neurons were undefined.

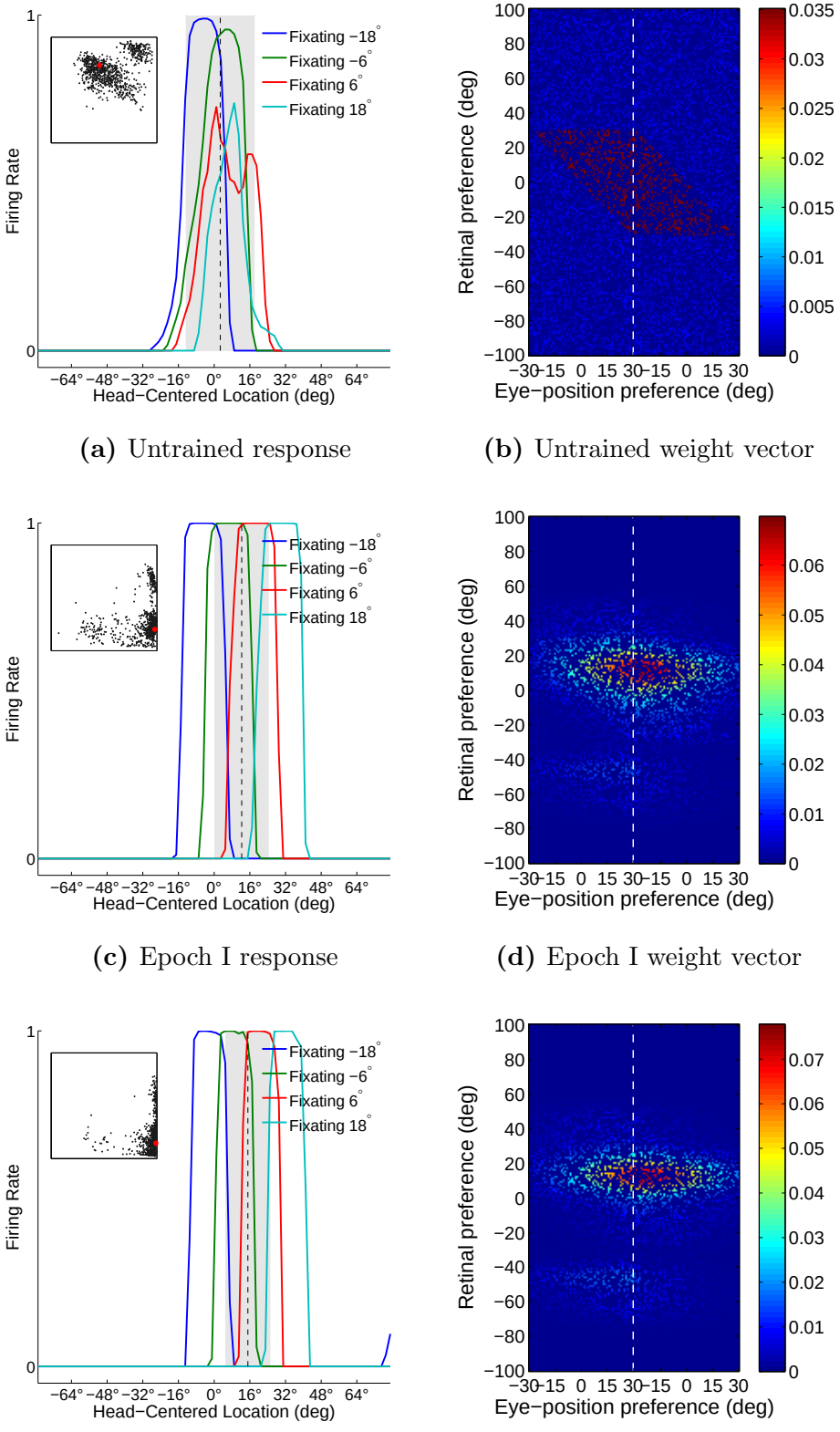
same parameters as in section 5.3.1.

The impact of the training on the characteristics of the prewired model was inspected by plotting key summary statistics as a function of the number of epochs in Figure 5.6. There was a catastrophic drop in model performance after only the first epoch of training, where the head-centeredness rate decreased from ∼77% to ∼0.3%, average head-centeredness among head-centered neurons decreased from ∼0.7 to ∼0.26, and there was no longer coverage. Subsequent training epochs remained around these levels, although after epoch 5 there was not a single head-centered neuron.

The firing responses and weight vectors of the output neurons were inspected through

successive training epochs. Figure 5.7 shows results for a typical output neuron #492 after 0, 1 and 10 training epochs. The response of neuron #492 before training (Fig. 5.7a) was very compatible with a response in a head-centered frame of reference, since there was a clear trend across all four fixation positions to respond maximally to locations close to the decoded head-centered location $\sim 2.7°$, and the weight vector was in its prewired condition (Fig. 5.7b). After a single epoch of training the response had developed a clear eye position dependency in the peak response location in head-centered space across fixation positions, and as such had become more compatible with a response in an eye-centered frame of reference (Fig.5.7c). Indeed the head-centeredness decreased from $\sim$0.7 to $\sim$0.12, and eye-centerdness increased from $\sim$0.40 to $\sim$0.98. The weight vector looked very similar to the weight vectors observed in section 5.3.2, with potentiated synapses focused along a narrow range of retinal preferences (Fig.5.7d). After an additional nine epochs of training the firing response and weight vector were qualitatively similar to what was observed after one epoch, just as in section 5.3.2.

In summary, it was found that just a single training epoch switched most of the output neurons from being head-centered to eye-centered. We hypothesised that this was due to the same visually-guided learning dynamics described above in section 5.3.2, which come into operation when the retinotopic input neurons are modulated by sigmoidal gain fields. Thus, even if the synaptic weights were initially prewired to effect head-centered output responses, which might be suggested to happen in the brain through genetic specification,



(a) Untrained response **(b)** Untrained weight vector

(c) Epoch I response **(d)** Epoch I weight vector

(e) Final response **(f)** Final weight vector

Figure 5.7: Firing rate responses and afferent synaptic weights of output neuron #492 in when learning was introduced. Results are shown before training (top row), after 1 training epoch (middle row), and after 10 training epochs (bottom row).

the introduction of just a limited amount of synaptic plasticity and visually-guided learning led to the output neurons rapidly switching to eye-centered responses. The presence of even modest levels of synaptic plasticity, which is ubiquitous in primate cortex, will quickly overwrite head-centered representations that have been set up through structured (e.g. genetic) prewiring. Thus, any explanation for the development of head-centered responses in primate cortex must not only take into account, but in fact be based on, some form of visually-guided learning.

In order to investigate this hypothesis further, the weight vector (Fig. 5.8) and reference frame metrics (Fig. 5.9) of output neuron #492 were tracked through the eight training periods of the first epoch, where for each training period the visual target was situated in different fixed head-centered location while the eyes moved through a series of saccades and fixations.

After the first two training periods, during which the visual target was located at 63° and -27° in head-centered space respectively, the head-centeredness of neuron #492 decreased to 0.58 despite the weight vector being unaffected (Fig. 5.8a,b). This change in response, despite a static weight vector, was due to the change in the responses of other output neurons that did have their weight vectors altered through training, and which consequently altered the responses of neuron #492 through competition within the output layer.

The third training period, when the visual target was located at -9°, was the first time

where there was a clear impact on the weight vector of neuron #492 due to learning (Fig. 5.8c). This learning took place because the neuron was prewired to respond to 0°, which was close to the visual target location, and it was in this period when the neuron became eye-centered for the first time as its eye-centeredness surpassed its head-centeredness.

The fourth period saw no salient change in the weight vector (Fig. 5.8d). However, there was a small movement towards a head-centered response, which again was driven by the change in the response of other neurons with which the given neuron competed.

The fifth period, when the visual target was located at 9°, saw the largest change in RFI towards an eye-centered response and simultanously the biggest change in the qualitative structure of the weight vector so far (Fig. 5.8e). Again this large change occurred when the visual target was located in a head-centered location close to the initial preference of the neuron.

The sixth period also caused a large change in the synaptic weight vector, with the potentiated weights now becoming more tightly clustered within a localised region of the retinal preference dimension indicating an eye-centered response (Fig. 5.8f).

For the last two training periods, there was no material change in the weight vector or responses of the neuron (Fig. 5.8g,h).

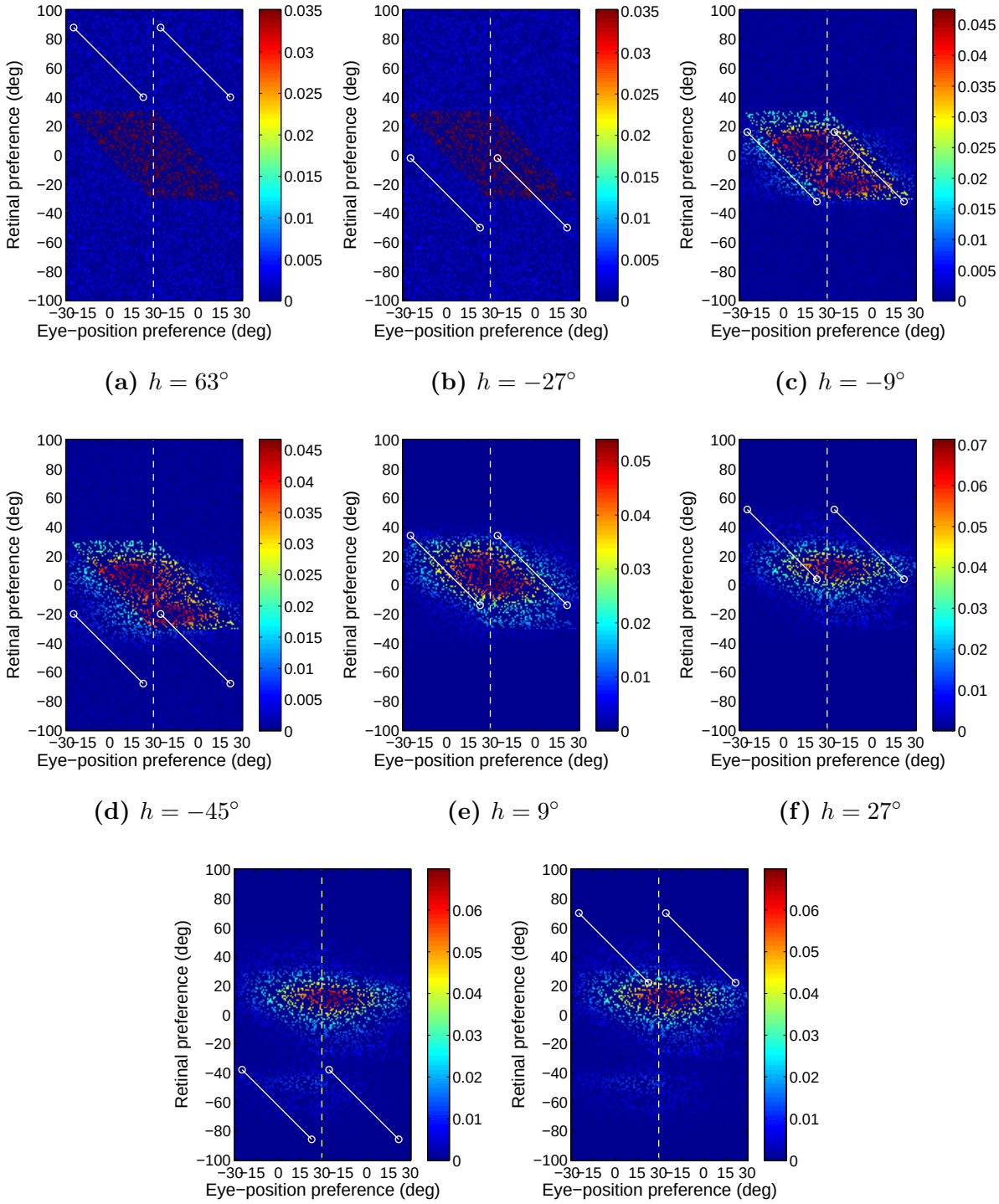
In summary, the evolution history of the synaptic weight vector of neuron #492 during the first training epoch, after which there was little qualitative change to the weight vector, showed that even weight vectors which initially produced head-centered responses could

not remain head-centered if visually-guided learning was subsequently introduced. This restructuring of the synaptic weights was due to the overlap in the input patterns that represented different head-centered target locations, as observed before.

5.3.4 Incorporating a Mixed Population of Input Neurons with both Peaked and Sigmoidal Eye Position Gain Fields

In this experiment it was investigated how mixing peaked and sigmoidal gain fields in the input population in varying proportions would influence the learning in the model. This was an important issue since all cortical areas with eye-position gain modulation exhibit a mixture of different gain field structures (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990, Andersen et al. 1985, Galletti et al. 1995). A series of simulations were conducted where each neuron in the input population was independently and randomly set to have either a peaked or sigmoidal eye position gain field. Specifically, each input neuron was changed from having a peaked to a sigmoidal gain field with a probability p, called the sigmoid modulation rate, and values of $p = 0, 0.1, \ldots, 1.0$ were explored.

The impact of varying the sigmoid modulation rate on the characteristics of the model was inspected by plotting key summary statistics as a function of p in Figure 5.10. As expected, the head-centeredness rate decreased as the sigmoid modulation rate increased, both in the trained and untrained models. Although, the trained model had a higher head-centeredness rate than the untrained model for all values of the sigmoid modulation



(a) $h = 63°$ **(b)** $h = -27°$ **(c)** $h = -9°$

(d) $h = -45°$ **(e)** $h = 9°$ **(f)** $h = 27°$

(g) $h = -63°$ **(h)** $h = 45°$

Figure 5.8: Evolution of the weight vector for output neuron #492 during one epoch of training of the prewired model. Results are shown for each of the eight training periods, where each period corresponds to the visual target kept in a fixed head centered location while the eye position is shifted. Each of the plots (a) to (h) shows the weight vector after the corresponding training period, 1 to 8, respectively. The white line in each plot shows a set of points corresponding to the head-centered location where the visual target was located in the given training period.

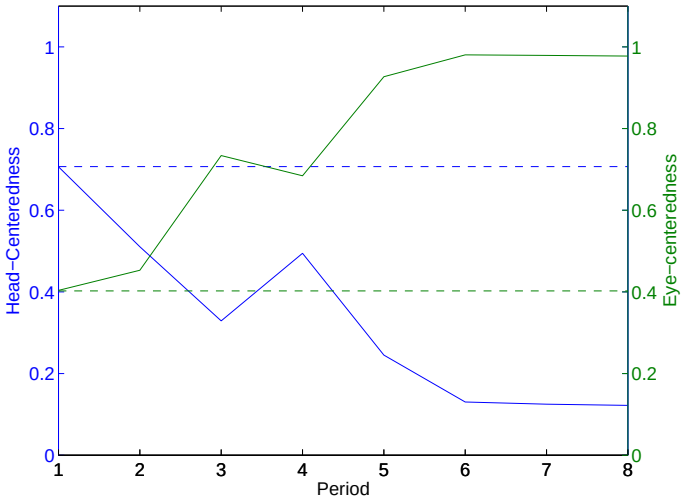


Figure 5.9: The evolution of the head-centeredness and eye-centeredness metrics for output neuron #492 during one epoch of training of the prewired model. Results are shown for each of the eight training periods, where for each period the visual target is kept fixed in a different head centered location while the eye position is shifted.

rate. The untrained model never had coverage, while the trained model had coverage for sigmoid modulation rates no greater than 20%. While both the head-centeredness rate and coverage decreased over this range of p, they both remained no less than $\sim$15% and $\sim$0.85 respectively. The average head-centeredness among head-centered neurons was also greater in the trained than the untrained model over this range of p, and it remained no less than $\sim$0.58.

In summary, these results showed that when there was a large proportion of input neurons with peaked eye position gain fields, say with $0 \leq p \leq 0.2$, then the model was still capable of developing a significant proportion, i.e. no less than $\sim$15%, of head-centered ouput neurons during training. However, as the sigmoid modulation rate increased, the performance of the model deteriorated with far fewer head-centered output neurons present

in the trained model.

5.3.5 Replicating Zipser & Andersen (1988)

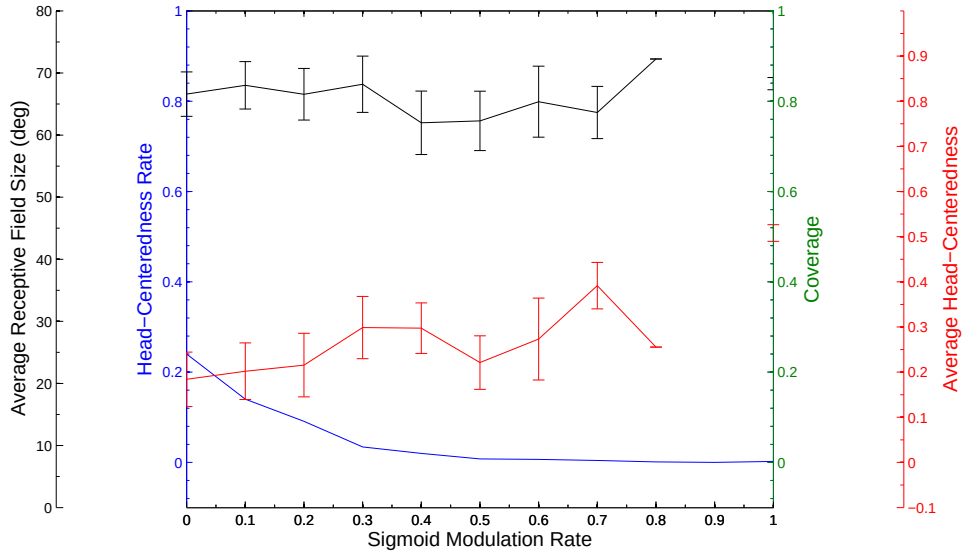
In this experiment the the classic model of Zipser & Andersen (1988) was reproduced and studied. The input layer consisted of two populations of units. The first was 8×8 grid of retinocentric units spaced $10°$ apart and centered on $(0°, 0°)$ in terms of their retinal preference. Specifically, retinocentric input unit i had horizontal and vertical preference r_i^h and r_i^v respectively, and had response r_i to visual target at retinal location r^h and r^v given by

$$r_i = \exp\left(-\frac{\left\|(r_i^h, r_i^v) - (r^h, r^v)\right\|^2}{2\sigma^2}\right) \tag{5.3}$$

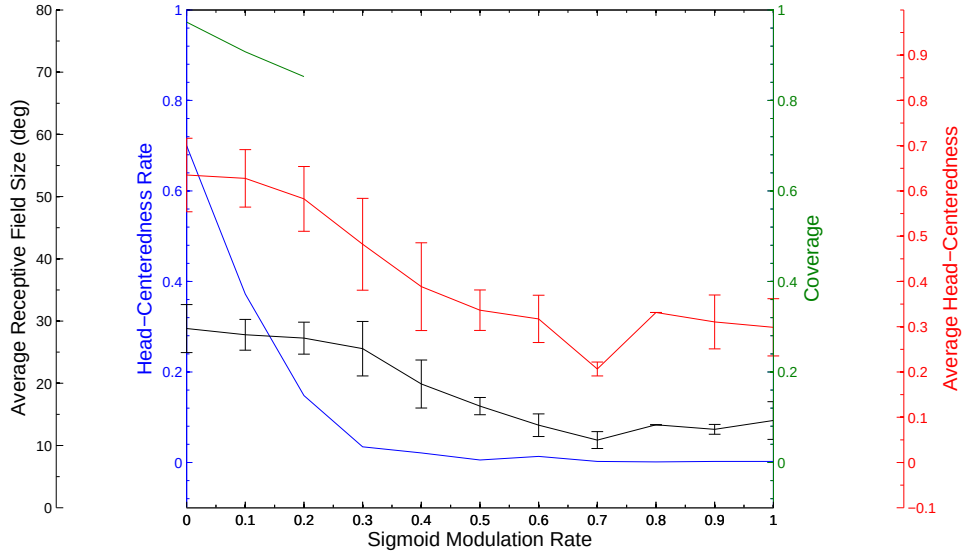
where $\sigma = 15°$. The second population consisted of four subgroups of planar eye-position units, each group corresponding to a unique combination of slope sign (positive or negative) and cardinal axis (horizontal and vertical). Each unit i had a randomly assigned slope in $|e_i^*| \in [0, 1]$ with appropriate sign and an intercept $b_i \in [-0.5, 0.5]$, and had response e_i to eye-position e^* in corresponding cardinal axis was given by

$$e_i = e_i^* e + b_i \tag{5.4}$$

This gave a total input layer of size $8 \times 8 + 4 \times 8 = 96$. The hidden layer consisted of 9 units, and the response h_i of unit i was given by



(a) Untrained



(b) Trained

Figure 5.10: Simulations exploring the effects of incorporating a mixed population of input neurons with both peaked and sigmoidal eye position gain fields. The plots show how the performance metrics vary with the sigmoid modulation rate, p, which is the probability of each input neuron having a sigmoidal eye position gain field. Results are presented showing the response characteristics of the output neurons before training (a) and after training (b). Conventions are similar to figure 5.6.

$$h_i = \sum_{j=1}^{64} w_{ij}^r r_j + \sum_{j=1}^{32} w_{ij}^e e_j \tag{5.5}$$

where w_{ij}^r and w_{ij}^e was the synaptic weight of the synapse from j^{th} retinocentric input unit and the j^{th} eye-position input unit to the i^{th} hidden unit respectively. The output layer was 8×8 grid of craneocentric units spaced $10°$ apart and centered on $(0°, 0°)$ in terms of their craneocentric preference. Specifically, output unit i had horizontal and vertical preference a_i^h and a_i^v respectively, and had desired output a_i^d to visual target at craneocentric location $a^h = e^h + r^h$ and $a^v = e^v + r^v$ given by

$$a_i^d = \exp\left(-\frac{\left\|(a_i^h, a_i^v) - (a^h, a^v)\right\|^2}{2\sigma^2}\right) \tag{5.6}$$

where $\sigma = 18°$. During training and testing, the feed forward output - or actual output, of unit i was given by

$$a_i = \sum_{j=1}^{9} w_{ij} h_j \tag{5.7}$$

where w_{ij} was the synaptic weight of the synapse from j^{th} hidden to the i^{th} output unit. The network was trained on all combinations of $r^h = [-30, -10, 10, 30]$, $r^v = [-30, -10, 10, 30]$, $e^h = [-13, 0, 13]$ and $e^v = [-13, 0, 13]$ for 18 epochs using the MATLAB (The MathWorks, Inc.) neural networks toolkit employing the Levenberg-Marquardt version of the backpropagation algorithm . The training converged to the desired perfor-

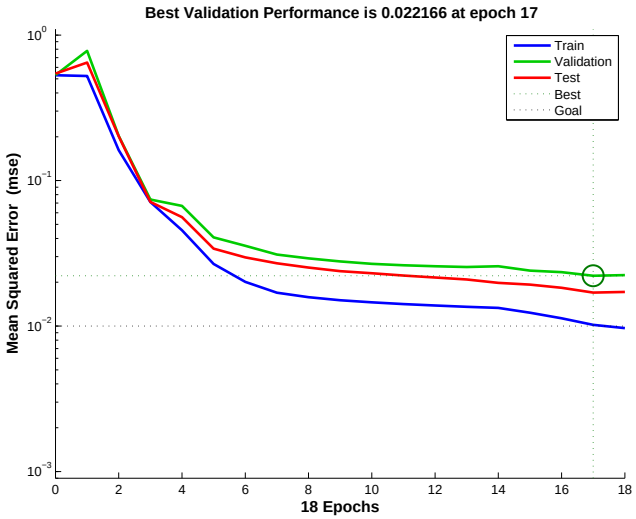
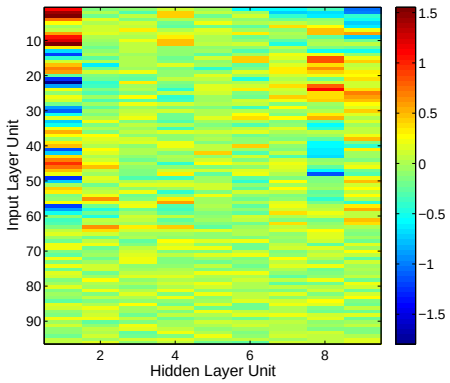


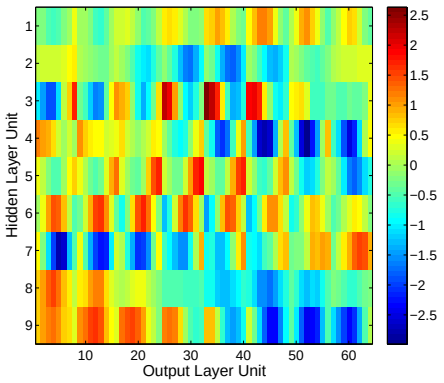
Figure 5.11: Performance of neural network as a function of time is shown. Performance is measured by mean squared error between desired (a_i^d) and actual output (a_i). The training performance is measured across all input-output pairs, while testing and validation are done across disjoint subsets of input-output pairs.

mance within the 17$^{\text{th}}$ epoch 5.11.

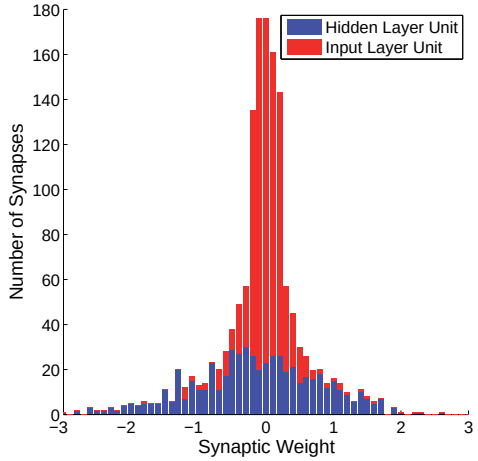
Figure 5.12 shows the weight matrices for the synapses from the input layer to the hidden layer 5.12a and the hidden layer to the output layer, as well as an analysis of both. Two salient features stand out. Firstly, there is a near equal total number of excitatory and inhibitory synapses in both layers of synapses 5.12c. Secondly, given a specific input layer unit projecting to the hidden layer or a given hidden layer unit projecting to the output layer, it is is clear that most neurons simultaneously excite and inhibit postsynaptic neurons 5.12d. Both of these findings contradict widely accepted facts about cerebral cortical anatomy, hence in summary the Zipser & Andersen (1988) model is confirmed to develop craneotopic output neurons using planar eye-position encoding, however feature of the final neural circuit is not compatible with known cortical anatomy.



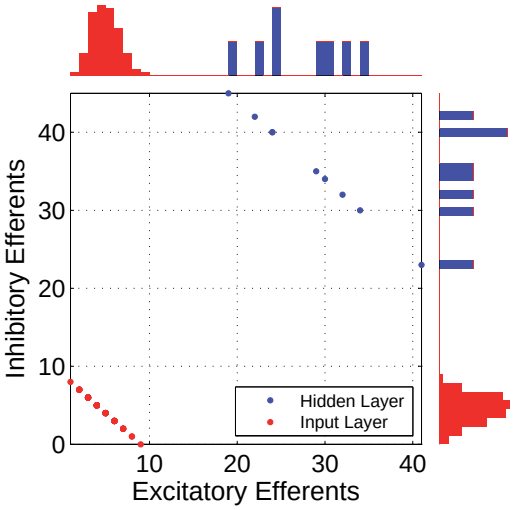
(a) Input-to-hidden layer weight matrix



(b) Hidden-to-output layer weight matrix



(c) Prevalence of excitation and inhibition



(d) Dale's Law violation

Figure 5.12: Plots (a) and (b) show the weight matrixes between for the synapses from the input layer to the hidden layer (a) and the hidden layer to the output layer (b). Plot (c) shows a distribution of the synaptic weights in both weight matrices. Each point in plot (d) corresponds to an input unit and hidden unit, and for each point shows the number of excitatory and inhibitory efferents from the given unit. Observe that units from the same layer are on the same unit slope diagonal because all neurons in the same layer have the same number of efferent synapses.

5.3.6 Replicating Pouget & Sejnowski (1997)

The second model which was replicated was Pouget & Sejnowski (1997). This model proposed that sensorimotor transformation could be understood as non-linear function approximation based on a population of parietal neurons behaving as a basis set, each neuron response denoted by $B_i(\mathbf{s}, \mathbf{m})$, in terms of a sets of sensory and motor input variables, denoted by $\mathbf{s}$ and $\mathbf{m}$ respectively. Each transformation M, for example from a head-centered auditory signal into an eye-centered reference frame, was a result of taking and appropriate weighting of across the set of parietal neurons responses, that is

$$M = \sum_i c_i B_i(\mathbf{s}, \mathbf{m}) \tag{5.8}$$

The key idea was that this population of parietal neurons allowed for simulatanous one step read-out of multiple distinct sensory transformations, in so far as the input population constituted a basis set in term of the appropriate sensory and motor variables. In the paper they show an example of how both an eye-centered and a head-centered visual response can be simultaneously derived from the same input population.

The experiment they demonstrated was of sensorimotor transformation from a retinocentric gain modulated input layer with 121 units to two output units, one being cranetopic and one being purely retinocentric. The retinal location of the visual target r, the eye position e and head position h were all one dimensional. The i^{th} input neuron had a unique combination of retinal preference $x_i \in [-60, 60]$, spaced 12°, and eye-position preference

$e_i \in [-40, 40]$, spaced 8°, and its response b_i was given by

$$b_i = \exp\left(-\frac{(r - x_i)^2}{2\sigma^2}\right) \times \frac{1}{1 + \exp - \left(\frac{e - e_{x_i}}{T}\right)} \tag{5.9}$$

The output of the k^{th} output neuron was given by

$$a_k = \sum_{i=1} w_{ik} b_i \tag{5.10}$$

The network had two output units, each fully connected with all input units. It was trained on 441 distinct retinal location and eye position combinations within $[-40, 40] \times [-20, 20]$, confined to 21 points in each space, using a supervised error correction scheme (Widrow & Hoff 1960) which for the two output units minimized the squared error between their respective desired and actual, specifically the rule attempts to minimize

$$E = \sum_{k=1}^{441} (d_k - a_k)^2 \tag{5.11}$$

by updating the synaptic weight from input unit j to output unit i by

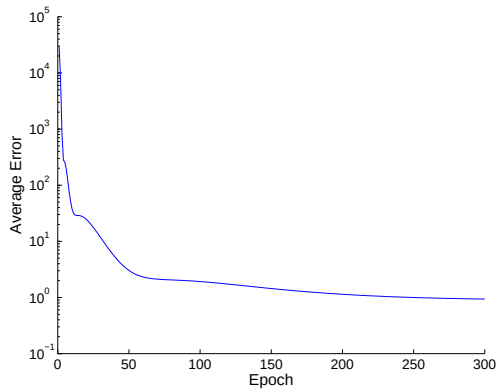
$$\Delta w_{ij} = \alpha(d_i - a_i) b_j \tag{5.12}$$

The result was two output units responding simultaneously in two distinct egocentric reference frames based on the same input population of basis units.

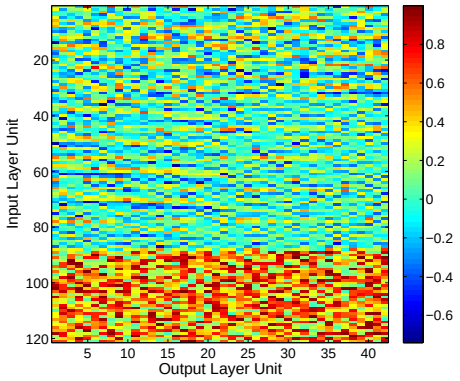
In this replication experiment, the output layer had 40 units, half being eye-centered and the other half being head-centered. Both groups had neurons with preferences $-20°, -18°, \ldots, 20°$ in the corresponding space. The network was trained for 500 epochs.

Figure 5.13a shows the decline of the net error (Eq.5.11) averaged across all output units, it was 0.93721 after epoch 300. Unlike the replication of the Zipser & Andersen (1988) model, a clear majority of synapses where excitatory (3328) and not inhibitory (1754), where a synapse was defined as excitatory if it was non-negative, and inhibitory otherwise. However, just like the previous replication, all input neurons violated Dales Law (Fig.5.13d). Hence, reading out multiple simultanous reference frame from the same population of units can impose contradictory constraints on the postsynaptic effect a presynaptic neuron should have. From the perspective of nonlinear function approximation via basis sets this simply means that two sets of coefficients $c_1^1, \ldots c_n^1$ and $c_1^2, \ldots c_n^2$, from equation , may not leave all coefficients sign invariant, that is $\text{sign}(c_j^1) \neq \text{sign}(c_j^2)$ for some j. In such a case, as the experiment is an example of, the equivalent circuit is neuroanatomically impossible.

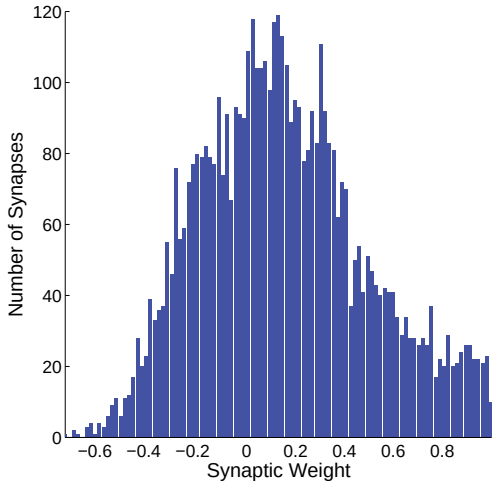
In summary the Pouget & Sejnowski (1997) model is confirmed to develop craneotopic output neurons using planar eye-position encoding, however feature of the final neural circuit is not compatible with known cortical anatomy.



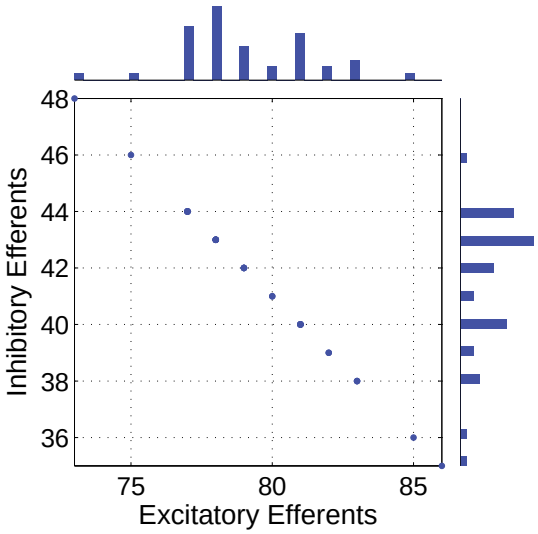
(a) Average error during learning



(b) Input-to-output layer weight matrix



(c) Synaptic weight distribution



(d) Dale's Law violation

Figure 5.13: Plot (a) shows the average error across all output neurons after each epoch of training, the y-axis is logarithmic. Plot (b) shows the weight matrixes between for the synapses from the input layer to output layer. Plot (c) shows a distribution of the synaptic weight. Each point in plot (d) corresponds to an input unit and hidden unit, and for each point shows the number of excitatory and inhibitory efferents from the given unit.

5.3.7 Replicating Salinas & Abbott (1995)

The final model to be replicated was Salinas & Abbott (1995). This was a particularly interesting model because, unlike the previous two models, it was not a result of a supervised error correction learning algorithm, but rather it proposed a novel hypothesis for how the model may self-organize through a biologically plausible learning mechanism. It became particularly critical to understand the success of this model given the fundamental problems revealed in section 5.3.2, since it had been shown that it could successfully self-organize when the input sensory layer used a linear, and therefore monotonic, eye position encoding. The main computational result of relevance was successful development of an output motor layer which could respond in a head-centered frame of reference based on the activity in an input sensory layer which encoded retinal location of the visual target x and the eye position y through sensory input neurons with retinotopic visual receptive fields with linear eye position gain modulation. The learning hypothesis was that during development, the agent would make spontaneous motor movements, for example reaching, and that these spontaneous movements commands would be represented by the output layer in a head-centered frame of reference. Simultaneously, these movements would be observed visually, and therefore be represented by the sensory layer in an eye-centered frame of reference. The addition of a simple Hebbian learning rule modifying the synapses between the sensory input layer and the motor output layer so as to eventually produce a weight structure allowing the sensory layer to drive the motor layer to encode a visual

target in a head-centered frame of reference.

The sensory representation was an input layer of retinocentric sensory units, where the response R_j^s of the j^{th} unit to a visual stimuli in retinal location x and eye position y was given by

$$R_j^s = \frac{R_{\max}}{M} \exp\left(-\frac{(a_j - x)^2}{2\sigma^2}\right)[\pm(b_j - y) + M]_+^M + \eta_j^s \tag{5.13}$$

where a_j and b_j was a unit-specific retinal preference and eye-position preference respectively, M was a normalization term setting the maximum response of the unit and η_j^s was an unit-specific noise term. An input neuron with a positive sign to the linear gain term $(b_j - y)$ was said to have positive gain, and one with a negative sign had negative gain. This was essentially input neurons with retinocentric receptive fields that had planar eye-position gain fields. The motor representation was an output layer of head-centered motor units, where the response R_i^m of the i^{th} to a motor command in head-centered location $z = x + y$ was given by

$$R_i^m = R_{\max} \exp\left(-\frac{(c_i - z)^2}{2\sigma^2}\right) \tag{5.14}$$

where c_i was a unit-specific head-centered preference. The model started with a fully connected but randomly weighted weight matrix. It was trained by randomly varying the eye position y and the head-centered motor target z as if the agent was performing a series of spontaneous motor actions. This resulted in a corresponding pattern of activity in the input and output layer. Observe that the input layer did *not* drive the output layer during learning. After one iteration of a set of observed motor commands, the weight matrix we

adjusted using a covariance learning rule of the form

$$w_{ij} = \langle R_i^m R_j^s \rangle - k \tag{5.15}$$

where w_{ij} was the weight of the j^{th} sensory neuron to the i^{th} motor neuron and k was a scalar. The averaging was done across all observed motor commands. During testing the sensory layer did drive the motor layer as described by

$$R_i^m = \left[\sum_j w_{ij} R_j^s \right]_+ + \eta_i^m \tag{5.16}$$

and they found that output motor neurons responded correctly in a head-centered frame of reference.

In this replication experiment, the randomeness in equation 5.13 and equation were taken out for the sake of simplicity, but clearly without loss of generality. The retinal preference a_j and eye position prefernce b_j of units in the input layer where $-10, -9.5, -9, \ldots, 10$, and for each unique combination of preferences there were to input neurons, one with a positive gain and one with a negative gain. Consequently, there were a total of $41 \times 41 \times 2 = 3362$ units in the input layer. The preferred head-centered locations c_i of units in the output layer where $-4, -3, \ldots, 4$, consequently there were a total of 9 units in the output layer. The model was exposed to all combinations of eye positions x, varied over $-6.5, -6, \ldots, 6.5$, and retinal location y, was varied over the same range. The other parameter settings where

$M = 3$, $R_{\max} = 1$, $\sigma = 3$ and $k = 0$.

In figure 5.14a the resulting distribution of synaptic weights is shown, and it is quite distinct from the weight distributions found in the previous two replications. There are no inhibitory synaptic weights, as seen by the fact that there are no negative synapses. This implies that this circuit does not violate Dale's Law, unlike the two previous models. A look at synaptic weight vector of a specific output unit, namely the unit with head-centered preference $c_i = -4$ as seen in figure 5.14b,c , shows that units have developed a weight vector very similar to the prewired circuit from experiment 5.3.1, and not subject to the failure in experiments 5.3.2 and 5.3.4, despite being derived from input units with monotonic gain fields.

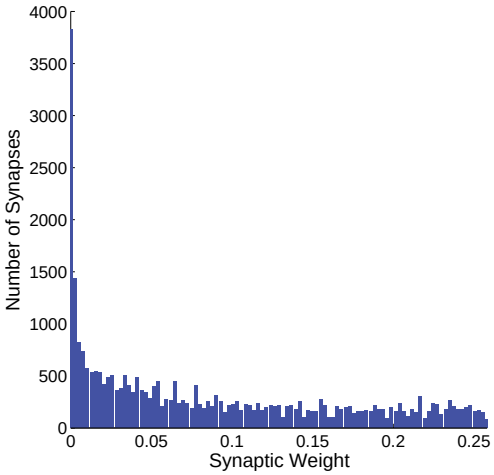
The reason that the self-organization process of Salinas & Abbott (1995) is immune to the previous failures of the trace learning mechanism described in 5.3.2 is because the output neurons in the former model have their activities imposed in a supervised manner during training. Effectively, the output layer implicitly uses the imposed motor tuning curves as a supervisory signal to force output neurons to only strengthen synaptic connections with presynaptic input neurons that respond for the appropriate region of head-centered space. This makes the performance of the model more robust. On the other hand, the competitive trace learning process has no such supervisory constraint on its output response, and its synaptic connections develop freely on the basis of the spatiotemporal dynamics in the input layer as described. However, although the model

of Salinas & Abbott (1995) offers more robust performance in terms of developing head-centered output representations, it has not yet been explained where such a supervisory motor signal may come from in the brain. In contrast, the trace learning mechanism does not require the existence of such a supervisory motor signal.

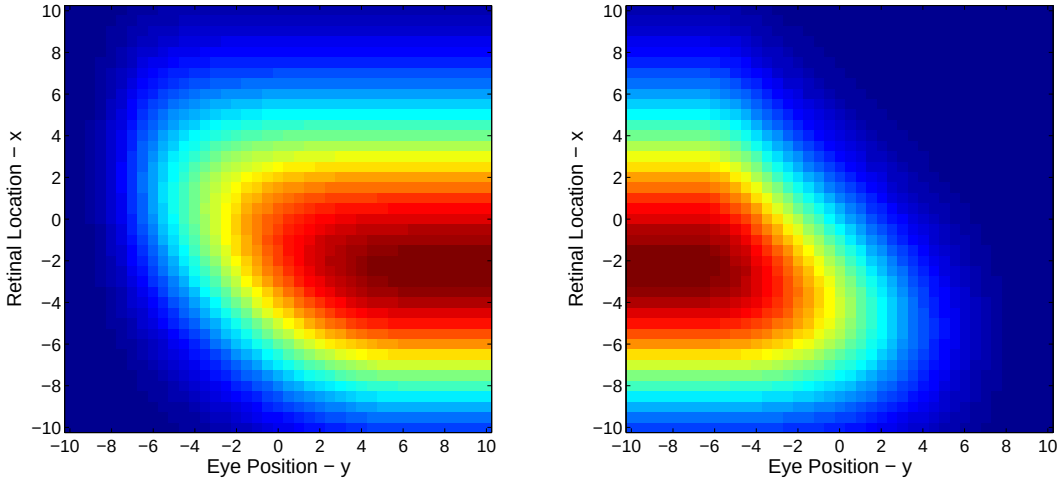
5.4 Discussion

It was demonstrated that a prewired neural network model with sigmoidal eye-position gain fields in the input population could compute head-centered representations in the output population. This was achieved by elevating the weight of synaptic connections to all input neurons which responded strongly, for some eye position, to a visual target in the head-centered location to which the postsynaptic output neuron was assigned. This was a valuable baseline to contrast with the severe failure which was observed when a randomly wired network was trained with the trace learning rule, in the same way as previous successful experiments in chapters 3 and 4, yet completely failed to develop a head-centered output representation. In fact, the trained model had a substantial increase in the eye-centeredness as compared to the untrained model, hence the training had the opposite effect to prior experiments and to the desired outcome.

In chapter 3, it was shown that the model with peaked eye-position gain modulated input neurons was able to develop head-centered output representations during training with the trace learning rule. However, the model with purely sigmoidal eye-position gain



(a) Synaptic weight distribution



(b) Neuron $c_i = -4$ synaptic weight vector from negative gain input neurons

(c) Neuron $c_i = -4$ synaptic weight vector from positive gain input neurons

Figure 5.14: (a) shows the synaptic weight distribution across all synapses from the sensory to the motor layer. (b) and (c) show the synaptic weight vector of motor output neuron with $c_i = -4$ with presynaptic sensory neurons separated in terms of their preferred gain direction, where (b) shows weights of synapse from input neurons with a negative gain direction preference and (c) shows weights of synapse from input neurons with a positive gain direction preference. Each of the two plots show the synaptic weights topographically organized in terms of the retinal preference a_i and eye position preference b_i of the presynaptic sensory neuron of the synapse.

modulated input neurons develops eye-centered rather than head-centered output neurons. The explanation for this behaviour is due to the correlations between the activities of the input neurons across all of the input patterns during training. It is well understood that in a standard competitive neural network, individual output neurons learn to respond to subsets of input neurons that tend to be most frequently co-active (Rolls & Treves 1998). In the model with peaked gain fields, the subsets of input neurons that are frequently co-active correspond to circular clusters that are highly localised in both the retinotopic preference and eye-position preference dimensions. With a standard Hebbian learning rule with no significant memory trace of recent neuronal activity, individual output neurons will learn to respond to these localised circular clusters of input neurons. If a trace learning rule is implemented, it is a relatively easy task for individual output neurons to simply bind together these clusters of input neurons along a diagonal line in the (retinotopic preference $\times$ eye-position preference) input space corresponding to a particular head-centered location. Output neurons will then respond to particular head-centered locations regardless of eye-position or the retinal location of a visual target. However, the situation is quite different with sigmoidal gain modulated input neurons. Due to the way in which the input patterns representing different head-centered locations overlap with each other, the subsets of input neurons that are most frequently co-active are localised in the retinotopic preference dimension but elongated in the eye-position preference dimension. With a standard Hebbian learning rule, individual output neurons will learn to respond to these elongated

clusters of input neurons, which will give rise to eye-centered output responses. However, if a trace learning rule is implemented, the output neurons still learn to represent eye-centered rather than head-centered locations. This is because developing head-centered output responses would require the trace learning rule to disrupt and break apart output representations corresponding to the elongated clusters of input neurons with correlated activities. However, in practice the trace learning effect is not strong enough to achieve this. Consequently, even with trace learning, the elongated clusters of input neurons with correlated activities continue to drive the development of eye-centered output neurons.

The severity of the identified problem was further investigated by examining the impact of introducing synaptic plasticity into the prewired model, with a head-centered output representation, from the first experiment in the chapter. The result was that just a single epoch of synaptic plasticity was sufficient to dramatically reduced the prevalence of head-centered neurons in the output population, and by the 7^{th} epoch all neurons were eye-centered for all subsequent epochs of training. This showed that monotonic gain fields would, in so far as the synapses were plastic, even undermine the functioning of a cortical circuit which was somehow genetically prewired to produce the head-centered representations. This was, therefore, a substantial challenge to the feasibility of a circuit based purely on input neurons with monotonic eye position gain fields.

However, it has been shown that most relevant brain areas have a mixture of gain field structure, not all of which are planar. In fact 45% had no planar component in the classic

paper of Zipser & Andersen (1988), and 41% and 56% are not planar in areas LIP and 7a, respectively (Andersen, Bracewell, Barash, Gnadt & Fogassi 1990). Consequently, it was explored how various degrees of prevalence of monotonic gain modulation would influence the learning model. It was found that so long as there was no more than approximately 30% monotonic gain modulation in the input population then the trained model still had more head-centered output neurons than the untrained model. However, the decrease in performance, and eventual collapse of the model, was very severe when the presence of monotonic gain modulation in the input population was increased. These results showed that monotonic gain modulation could not constitute the majority of input neurons to the output population. Consistent with this, 73% of neurons have peak shaped gain modulation in area PO (Breveglieri et al. 2009). In this paper they also found that replacing the planar gain fields in the model of Salinas & Abbott (1995) with peak shaped gain fields produced a superior head-centered representation of target location. This suggests that head-centered representations in LIP and 7a are perhaps derived partially from peak-shaped gain modulated input neurons in other areas such as PO. This is possible since, for example, PO is known to project to LIP. Alternatively, it may be that the neurons in LIP and 7a which have a significant planar component, but which are not perfectly planar, together with other peaked responses, constitute a sufficiently large proportion of neurons that are not strictly monotonic. Such a mixed population of neurons with non-monotonic gain modulation may be sufficient to allow the development of head-centered

responses using the trace learning mechanism demonstrated. An alternative ameliorating effect could be the combined effect of diluted synaptic connectivity and a large output population. A sufficiently large output population would statistically include neurons which by chance received afferent synaptic connections predominantly from neurons with peaked modulation. This would allow these individual neurons to avoid the problem of monotonic gain modulation.

The failure of the learning in the model to develop head-centered output representations when the majority of the input neurons were modulated by sigmoidal eye position gain fields prompted an investigation of how other important coordinate transformation models dealt with the problem.

The first model investigated was the classic model of Zipser & Andersen (1988). This model used a backpropagation of error algorithm (Rolls & Treves 1998) to train the output units to respond in a craneocentric (i.e. head-centered) reference frame. The model was simulated and the synaptic weight structure of the two layers of synapses was inspected after training. The structure of the synaptic weights was found to be inconsistent with two broad features of cerebral cortical anatomy that are widely accepted. Firstly, it was found that the ratio of excitatory to inhibitory synapses was roughly 1:1, despite the fact that it is known to be closer to 5:1 in the cortex (Kandel et al. 2000). Secondly, and perhaps more importantly, almost all presynaptic neurons operated both as excitatory and inhibitory neurons simultaneously, which violates Dales principle stating that a single

neuron must have the same postsynaptic effect (e.g. excitatory or inhibitory) in all of its synaptic terminals (Kandel et al. 2000). Moreover, inhibition is generally thought of as a postsynaptically non-specific counterbalance to local excitatory activity (Kullmann et al. 2012), while this is not the case in the model of Zipser & Andersen (1988). These substantive violations of widely accepted features of cortical architecture undermine the possibility of this model being implemented in the cortex. The authors do concede the importance of learning. However, their model relies on a backpropagation of error training algorithm, which is not biologically plausible due to the non-local terms used to adjust the synaptic weights in the hidden layer (Rolls & Treves 1998). It is this biologically implausible learning rule that gives rise to a synaptic connectivity that is inconsistent with known cortical architecture.

The second model investigated was the function approximation model of Pouget & Sejnowski (1997). This model proposed that sensorimotor transformations could be understood as a non-linear function approximation based on a population of parietal neurons behaving as a basis set. In their model, different subsets of output neurons simultaneously learned to represent either an eye-centered or head-centered response derived from the same population of input neurons using a supervised error correction scheme (Widrow & Hoff 1960). While they made a plausible case for why parietal neurons potentially provide a suitable basis set, their trained model also violated Dale's Law in that individual input neurons sent both excitatory and inhibitory connections to different output neurons. Thus,

the model of Pouget & Sejnowski (1997) is also inconsistent with known cortical anatomy.

The last model investigated was that proposed by Salinas & Abbott (1995), which was also able to develop head-centered output representations through learning. This was the only model among the three that was making a claim for a biologically plausible mechanism of self-organization, which did not rely on error correction learning. Instead, the output neurons learned to respond in a head-centered reference frame by an associative Hebbian learning rule. The trained model did not contain inhibitory synapses, and so did not violate Dale's Law. The lack of inhibitory synapses is not necessarily a fundamental problem for the model since the model could in principle be extended to incorporate inhibitory interneurons underpinning lateral inhibition among the output neurons. Furthermore, the synaptic weight vectors of output neurons looked very similar to the manually hardwired weight vectors from experiment 5.3.1, and not to the weight vectors from the failed trace learning experiments of section 5.3.2. However, the success of the model could be attributed to supervised learning. That is, it was assumed that during training the agent would make spontaneous motor movements, which would be represented by the output layer in a head-centered frame of reference. This meant that the output neurons effectively had their activities imposed in a supervised manner during training. This makes the performance of the model robust even with planar eye-position gain modulated input neurons. However, although the model of Salinas & Abbott (1995) successfully develops head-centered output representations, it has not yet been explained where such a supervisory motor signal may

come from in the brain. In contrast, the unsupervised trace learning mechanism does not require the existence of such a supervisory motor signal.

A possibly deeper problem with the associative learning model of Salinas & Abbott (1995) is as follows. Although it is proposed that the output neurons will be driven in a supervised manner by motor-related signals during early learning, subsequently activity in the output layer is intended to be driven by the visual input signals. However, if the synapses remain plastic while the output neurons are driven by visual inputs, then the synaptic weight structure will deteriorate as demonstrated in section 5.3.3 where synaptic plasticity was introduced into the prewired model. Hence, it seems that the only viable way for this model to operate in cortex would be for there to be an initial learning phase during which the output neurons are driven in a supervised manner by motor-related signals, which is then followed by an abrupt change in which the synaptic plasticity disappears and the motor-related signals simultaneously cease to dominate the activity of the output neurons. This proposal seems unlikely.

Spratling (2009) also studied the development of head-centered and body-centered representations using a biologically plausible learning mechanism. However, this study used peaked encoding of both eye position and head position. The findings in this chapter are particularly relevant to this work because, while there exists peaked encoding of eye position in the brain, encoding of head position on the trunk has only been found to be monotonic (Brotchie et al. 1995). This means that the second stage of processing in the

model of Spratling (2009) is likely afflicted with the same problem identified in this chapter.

In conclusion, the lack of focus in the scientific literature on biologically plausible hypotheses for how head-centered representations may develop in the primate dorsal visual pathway has not only left this central question unaccounted for, but has also resulted in a family of models being accepted despite suffering from some substantial incompatibilities with widely accepted neuroanatomical principles. Furthermore, the work presented here shows that the mere *existence* of a synaptic architecture that produces head-centered output representations, as shown in the experiment with a manually hardwired network, does not in general guarantee that it can *emerge* through a process of self-organization such as trace learning. Indeed, the lack of emphasis on biologically plausible self-organising models, employing unsupervised competitive learning, has left hidden the substantial problem of how input neurons with monotonic eye-position gain fields may drive the development of eye-centered rather than head-centered output representations when learning takes place in the absence of a supervisory training signal. In the simulations reported in this chapter, the ability of the trace learning model to produce head-centered output representations deteriorated markedly when the proportion of sigmoidal eye-position gain modulated input neurons rose above 30%. This implies that either there are sufficient numbers of parietal neurons in the primate brain with either peaked or other forms of non-monotonic eye-position gain fields, or substantive changes to the current form of the trace learning model will need to be made.

5.5 Summary

The following is a summary of the main results and insights in this chapter.

- It was possible to manually hardwire the feedforward synaptic connections in a model with sigmoidal eye-position gain modulated input neurons in order to endow the output neurons with head-centered firing response characteristics.

- The model with purely sigmoidal eye-position gain modulated input neurons developed eye-centered rather than head-centered output neurons.

- The reason that the model with purely sigmoidal gain modulated input neurons developed eye-centered output neurons was due to the correlations in the activities of the input neurons during training. The output neurons learned to respond to subsets of input neurons that were most frequently co-active. With sigmoidal gain fields, these co-active subsets of input neurons are clustered around particular locations in the retinal preference dimension. Hence, the output neurons developed eye-centered responses that were localised in retinotopic space.

- If synaptic plasticity is introduced into the prewired model, then the synaptic architecture swiftly deteriorates and the head-centered output representations disappear.

- It was shown that the model was able to develop head-centered output representations when there was a mixed population of input neurons with both sigmoidal and peaked eye-position gain fields. In particular, when there was a large proportion of input

neurons with peaked eye position gain fields, say with $0 \leq p \leq 0.2$, then the model was still capable of developing a significant proportion, i.e. no less than $\sim$15%, of head-centered ouput neurons during training.

- Previously published models of how head-centered visual representations may develop in the primate dorsal visual pathway are not biologically plausible; they typically rely on biologically implausible learning mechanisms such as backpropagation of error or employ unexplained supervisory training signals.

- Future research must focus on biologically plausible models that rely on unsupervised competitive learning mechanisms. An important challenge for these models will be to show how the learning may operate with sigmoidal eye-position gain modulated input neurons.

Chapter 6

Perisaccadic Receptive Field Dynamics: Introduction

6.1 Introduction

The remaining results chapters of this thesis now switch to modelling the visually-guided development of perisaccadic receptive field dynamics among eye-centered neurons in areas such as the lateral intraparietal sulcus (LIP), frontal eye fields (FEF), superior colliculus (SC), extrastriate and striate cortex.

This chapter presents a review of the behavioural, physiological and theoretical work upon which chapters 7 and 8 are based. The concepts of predictive response, presaccadic response, trace response truncation and remapping response are introduced, along with the potential behavioural significance of such phenomena. The main physiological, neuropsychological and theoretical results relevant to later chapters on perisaccadic neuronal dynamics are reviewed.

6.2 Terminology

6.2.1 Perisaccadic locations

Consider a visual scene with a single visual target fixed in a head-centered location h while fixation is kept at head-centered location e_{PRE}. Prior to saccade onset, the visual target is projected onto retinal location $r_{\mathrm{PRE}} = h - e_{\mathrm{PRE}}$. At some time, the eye-centered saccade s is initiated, relocating fixation to head-centered location $e_{\mathrm{POST}} = e_{\mathrm{PRE}} + s$, after which the visual target is projected onto retinal location $r_{\mathrm{POST}} = h - e_{\mathrm{POST}} = h - (e_{\mathrm{PRE}} + s)$. These quantities are shown in their respective reference frames in figure 6.1a.

6.2.2 Remapping

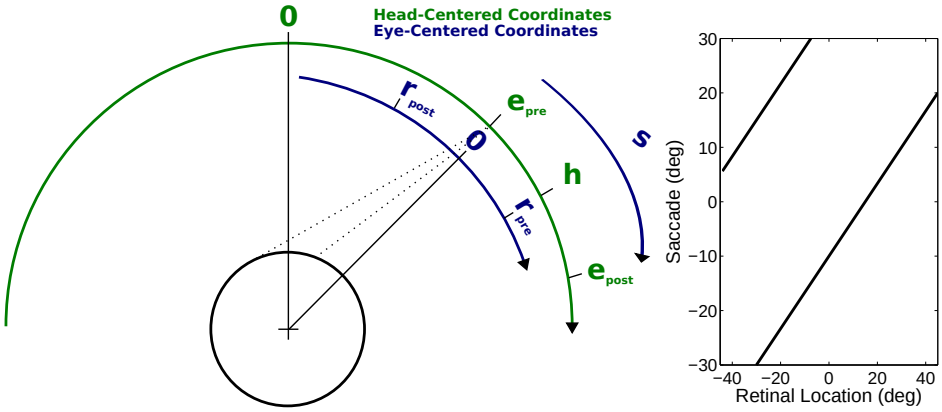
Consider a neuron with a retinal receptive field at eye-centered location r. Let the agent initiate the saccade which brings this receptive field in alignment with h, that is the saccade $s = r_{\mathrm{PRE}} - r$. If the visual target is still present in h, one would assume that the neuron would begin to respond some time after the receptive field had settled. If the target was extinguished in advance of saccade initiation, one would assume that the neuron would not respond at any time. If a neuron responds in the latter scenario, this is known as *memory trace remapping*, or just remapping. For each retinal location $r^{'}$ there is a corresponding saccade $s = r^{'} - r$ that may reveal a remapping response if initiated while a stimulus is flashed in $r^{'}$ presaccadically (Fig.6.1b).

6.2.3 Latency

When a visual stimulus is flashed in the receptive field of the neuron, the time between stimulus onset and the beginning of the neuronal discharge is known as the *visual onset latency*. Likewise, during a saccade as in the previous section, the time between the onset of the saccade and the beginning of discharge is known as *remapping latency*. Remapping latency can be found both for situations where the stimulus is still present in h at the onset of the saccade, and situations where it has been extinguished prior to saccade onset. In either case, if the remapping latency found is less than the visual onset latency, then the response is referred to as a *predictive remapping*. A predictive response where the neuronal discharge occurs before the saccade onset, i.e. the remapping latency is negative, is known as a *presaccadic response*.

6.2.4 Truncation

When a visual stimulus is flashed in the receptive field of the neuron, there may a long period of sustained neuronal discharge after stimulus offset. This is known as a *trace response*. If a saccade moves the receptive field of the neuron away from a stimulus, one would expect a trace response similar to one observed when a stimulus is simply extinguished. If the saccade has the effect of promptly cutting the duration of the trace response short, then this is known as *trace response truncation*.



(a) Coordinate Systems **(b)** Remapping space

Figure 6.1: Plot (a) shows the spatial relationships between various locations which are relevant for remapping, in their respective spaces. The dark central circle is the head, and the two dashed lines show the line of sight for each eye, and they converge on location e_{pre} in the head-centered coordinate system represented by the large green semicircle. The blue curved axis fixed to 0 is the eye-centered coordinate system. Plot (b) shows for two neurons with different retinal receptive field locations, 35° and −10°, the corresponding combinations of presaccadic retinal stimulus location r_{PRE} and saccade s that could give rise to remapping responses, in which the neurons respond even though the stimulus is extinguished before entering the neurons' given receptive fields.

6.3 Behaviour

It has been proposed that the various perisaccadic neuronal phenomena introduced in the previous section are useful in at least two different ways. First, these mechanisms may provide a neural substrate for perceptual stability across eye gaze shifts. Secondly, these mechanisms may help to guide spatially accurate shifts in gaze as the animal performs a sequence of saccades between visual stimuli.

6.3.1 Perceptual stability

A salient characteristic of visual perception in a natural environment is that, despite the high frequency of saccadic eye movements performed towards potentially interesting objects in a visual scene, our subjective visual experience is that of a stable and continuous examination of a stationary environment. As previously noted by other authors (Melcher & Colby 2008, Quaia et al. 1998), if latency between the retina and higher order visual areas is anywhere between 60 to 100ms, then our visual system should be expected to be incongruent with the visual world for a significant fraction of the time due to the high frequency of saccadic eye movements. In light of this, there has been a great deal of debate about how the brain integrates visual percepts across saccades (Melcher & Colby 2008). A leading hypothesis dates back to the early work of von Helmholtz & Southall (1924), who suggested that internal monitoring of impending eye movement signals can drive an anticipatory integration mechanism across rapid eye movements. The neurophysiological phenomena of predictive response and trace response truncation have later been suggested as the neural basis for this mechanism (Melcher & Colby 2008) because they have the effect of anticipating the visual consequences of impending eye movements at latencies much lower than in a pure fixation task. A predictive response will reduce the period of incongruence in the activity of a neuron when its classical receptive field is shifted to the location of some salient stimulus. While trace response truncation will do the same for a neuron which has its classical receptive field shifted away from the location of a salient

stimulus.

6.3.2 Spatially accurate shifts in gaze between multiple visual stimuli

A potentially related behavioural challenge is to execute a spatially accurate sequence of saccades between a number of visual stimuli in a scene. In this situation, the intervening saccades creates a discrepancy between the initial sensory representation of a particular target stimulus and the correct motor program required to perform the final saccade to that target. Successfully performing such a motor program is known as *spatially accurate sensory guided action*. The prototypical task involving spatially accurate eye movements is the delayed double-step saccade task, which is contrasted with a simple visually guided saccade task in figure 6.2. In the classical visually guided saccade task shown in Fig. 6.2a, the saccade target is observed from the initial fixation point. This means that the sensory representation of the retinal error RE between the target location T and the initial fixation point FP is identical to the motor error ME that is required to saccade from the fixation point to the target location. In this case, computing the required saccade trajectory is clearly trivial. However, in the delayed double-step saccade task shown in Fig. 6.2b, two visual targets are briefly flashed prior to saccade onset, and the agent has to saccade to the first target and then to the second target in order in the dark. In this case, the retinal error RE_1 and motor error ME_1 for the first target T_1 are identical, hence computing the

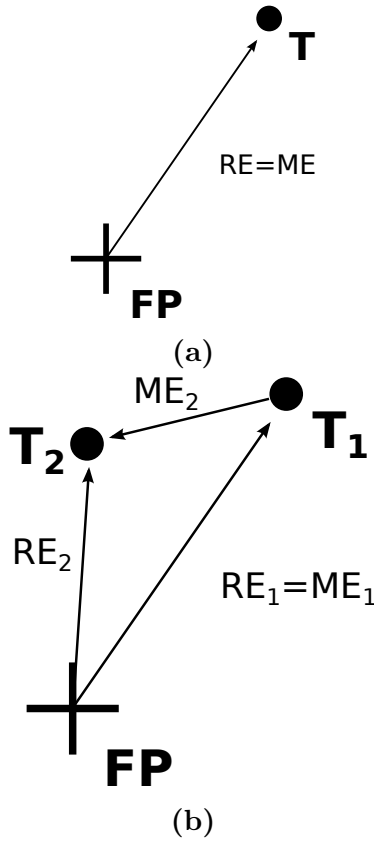


Figure 6.2: Illustrations of a visually guided saccade task (a), and a double-step saccade task (b). Each subfigure shows the spatial structure of the corresponding task, with relevant motor and sensory vectors also included. In both cases, the cross is the initial fixation point (FP), the Ts are the locations of the visual targets, the REs are retinal vectors representing the target positions with respect to the initial fixation point, and the MEs are the motor error (i.e. saccade) vectors with respect to the previous fixation position. For the delayed double-step saccade task shown in (b), it can be seen that there is a discrepancy between the sensory representation RE_2 of the position of target T_2 with respect to the initial fixation point and the motor program ME_2 required for the final saccade to the target T_2.

first saccade is no different from the visually guided saccade task. However, the retinal error RE_2 and motor error ME_2 of the second target T_2 are not the same because of the intervening saccade to the first target. A key challenge for theoreticians is to explain how the visual system is able to bypass this discrepancy in order to perform an accurate saccade to the second target.

It was first demonstrated by Hallett & Lightstone (1976) that human subjects could indeed perform double-step saccade tasks, and later this was confirmed to be the case for non-human primates (Mays & Sparks 1980). Initially, it was suspected that such saccades were programmed by maintaining a head-centered representation of visual space, which was combined with the present eye position to produce spatially accurate saccades (Robinson 1975). This way, the intervening saccade would leave this representation invariant, and the eye position after the first saccade would give the correct motor error for the final saccade. However, perhaps partly due to the apparent absence of explicit head-centered encoding at the single neuron level at the time, an eye-centered visual representation updated by eye-movements was proposed as an alternative hypothesis (Goldberg & Bruce 1990). This way, the representation would be updated internally with the initiation of the first saccade. That is, the neural representation of the retinal location of the second target would be updated during the first saccade by internal mechanisms such that the second target would continue to be represented in the correct eye-centered location. This proposal has the added simplicity of avoiding the need for an explicit eye position signal to program the final saccade, since the second target would already be correctly represented in an eye-centered frame of reference. In Duhamel, Goldberg, Fitzgibbon, Sirigu & Grafman (1992), a human subject with a lateralised fronto-parietal lesion was tested in single and double-step saccade tasks. Despite having no general saccade or perceptual deficits, the subject was unable to accurately execute the second saccade when the first saccade was in

the ipsilesional hemifield, despite being able to execute the first saccade, and being able to perform both saccades when the first saccade was in the contralesional hemifield. The selective undermining of only the second saccade, and only in one hemifield, was interpreted as evidence that an invariant head-centered encoding was not driving the saccades, but rather some neural mechanism for dynamically updating the internal representation based on the first saccade. These results suggest that the delayed double-step saccade task makes use of an eye-centered representation of the retinal locations of visual targets, which may be updated in the absence of visual input by internal saccade signals. This is the approach that is modelled later in this thesis.

6.4 Physiology

Mays & Sparks (1980)

A prominent hypothesis for the function of the superior colliculus (SC) was that the superficial and purely visual layers passed on the retinal error of a visible saccade target to the intermediate saccade layers, which in turn was involved in driving the subsequent saccade. However, the work of Hallett & Lightstone (1976) called this hypothesis into question by demonstrating that humans could perform double-step tasks, thereby dissociating the retinal error and motor error for the second saccade target. Therefore, Mays & Sparks (1980) undertook single unit recordings of both superficial and intermediate layer SC neuron during double-step saccade tasks to reveal whether the second saccade was properly encoded

in the former, and as a result whether there was a discrepancy in the encoding between the two. They found that visual neurons indeed encoded the retinal error of both saccade targets, but did not encode saccades where retinal and motor error was not aligned, i.e. the second saccade in each trial. This showed that saccade related activity in intermediate layers did not simply reflect visual activity in superficial layers, as had been previously thought, and it led the authors to speculate that perhaps a range of different input signals could be driving the intermediate layer, and during double-step tasks it would be some head- or body-centered signal yet to identified. The authors did not consider the possibility that eye-displacement signals, like a saccade signal, could be used to dynamically update the retinal map across saccades.

Goldberg & Bruce (1990)

Shortly after this, the neuronal responses found in the frontal eye fields (FEF) were for the first time precisely characterized by Bruce & Goldberg (1985). Neurons were classified into three primary categories based on their presaccadic response as follows: the visual category, for neurons responding only to visual stimuli in the receptive field, the movement category, responding only to saccades to the movement field without stimuli, and the visuomovement category, responding in both cases. Given these findings, and the general understanding that FEF was thought to be involved in producing eye movements, Goldberg & Bruce (1990) attempted to understand how neurons from the different categories

responded during single and double-step saccades that systematically varied the retinal error and motor error independently w.r.t. the receptive field (Fig. 6.3). The retinal error referred to the retinal location of the last saccade target stimuli as observed from initial fixation, and for a given neuron it was either inside or outside the receptive field. The motor error referred to final eye-centered eye movement vector, which for a given neuron was either inside or outside the movement field. Therefore, there were in total four versions of the saccade task per neuron. The primary findings for the double-step saccade tasks were as follows. For the two tasks shown in Fig.6.3b,d, the motor error but not the retinal error associated with the second saccade extended into the receptive field of the neuron. In this case, most neurons from all categories responded to the motor error. However, for the saccade task shown in Fig.6.3c, the retinal error but not the motor error associated with the second saccade extended into the neuron's receptive field. In this case, most neurons did not respond to the retinal error. This work showed that FEF also maintained a spatially accurate saccade signal along with SC. Given the previous findings that movement neurons could more effectively drive saccades via stimulation, and that they were the only category to project to SC, it was proposed that movement neuron activity was dominant in FEF and that visual input driving visual neurons was only a minor component in the output of FEF. They suggested that the postsaccadic neurons reported by Bizzi (1968), encoding the last saccade, combined with the visual signal of the saccade targets could be used to update activity across saccades, without the need for a head-centered encoding of

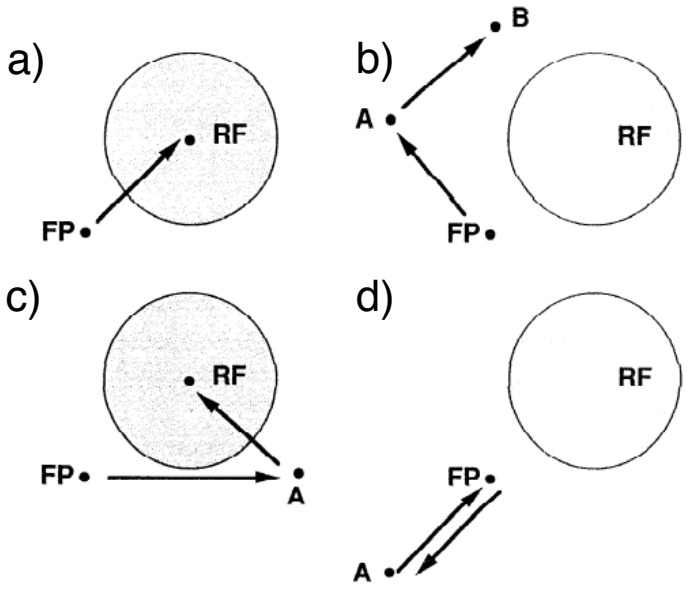


Figure 6.3: Each sub figure shows a type of saccade task used by Goldberg & Bruce (1990); the figure is adapted from the same paper. FP is the initial fixation point, RF is the receptive field of a neuron that has been isolated for the task trial, and A and B are saccade targets. Figure (a) shows a single step saccade task and has both the retinal and motor error inside the receptive field. The remaining subplots (b)-(d) show double-step saccade tasks, where the retinal error and motor error for the second saccade are shown. Figure (b) has the retinal error outside and motor error inside the receptive field. Figure (c) has the retinal error inside and motor error outside the receptive field. Figure (d) has the retinal error outside and motor error inside the receptive field.

visual space.

Duhamel, Colby & Goldberg (1992)

A third parietal area, the lateral intraparietal area (LIP), was known to be reciprocally connected with both the FEF (Blatt et al. 1990) and SC (Lynch et al. 1985), and saccade related activity had also been identified in the area (Gnadt & Andersen 1988). Researchers therefore turned their attention to LIP, and the landmark paper of Duhamel, Colby & Goldberg (1992) examined the responses of neurons to stimuli in their pre and postsaccadic

receptive field locations. A baseline fixation task established that neurons had a response latency to the onset of a visual stimulus in their receptive field, and also that neurons had a tonic discharge for multiple hundreds of milliseconds after stimulus offset, a so called visual trace. For 44% of neurons it was found that they would respond as much as 80ms prior to a saccade that would bring an eccentric stimulus into the receptive field, also known as presaccadic remapping. This was a response which was not only early w.r.t. the saccade itself, but also the latency observed in the fixation task as measured from stimulus onset in the receptive field. Such a response did not occur if the saccade did not bring the stimulus into the receptive field, or if the saccade was not performed while the stimulus remained in the same position. For some neurons it was also found that saccades removing the receptive field from a visual stimulus led to a truncated discharge, or so called trace response truncation. For 96% of neurons it was found that a saccade bringing the receptive field to the location of a presaccadically extinguished stimulus would cause a discharge some time after saccade completion, a so called trace remapping. Even a stimulus presented as long as 1s prior to the saccade could elicit a postsaccadic trace response. The trace remapping and trace response truncation had both been identified in FEF before (Goldberg & Bruce 1990), however these had now been shown to exist in LIP for the first time. The predictive responses described by Duhamel, Colby & Goldberg (1992) had not been observed in any brain area other than LIP. The researchers speculated that this anticipatory response could potentially subserve perceptual visual stability across

saccades.

Walker et al. (1995)

The discovery of predictive remapping in LIP prompted researchers to ask whether such activity also could exist in SC, in particular given the projections from the former to the deep layers of the latter (Lynch et al. 1985). This was investigated by Walker et al. (1995). To identify predictive responses, monkeys were required to make saccades that would bring the location of a stimulus, or recently flashed stimulus, into the classically defined receptive field of collicular neurons. Responses were classified as predictive if the response latency, with respect to saccade onset, during such a saccade task was less than the response latency in a visual fixation task with respect to stimulus onset. Thirty percent of visuomovement neurons, that is neurons responding to both stimuli and saccade onset, identified in the intermediate SC layers had predictive responses. All predictive responses were found to be presaccadic, and the latencies were found to be in the range -192ms to -36ms w.r.t saccade onset. For many neurons there was a brief postsaccadic supression of activity before a reafferent visual response caused by the stimulus in the postsaccadic receptive field. The stimulus onset response, as measured in the 100ms interval after activity onset, was always larger than the predictive response. It was also found that when the stimulus was rendered behaviourally relevant, by making it the target of a second saccade from the second fixation point, then the predictive response was elevated. Two control tasks were

conducted to confirm that the response was indeed predictive. Firstly, the same saccade was made, but in the absence of any stimulus, and in this case there was no response. Hence it was required that the impending saccade would bring a stimulus into the receptive field; it was not just pure saccade planning activity. Secondly, a range of trial types, with different stimulus and saccade parameters, were intermixed to control for the possibility that the results were not a result of the predictability of trial blocks. This showed that under the same protocol as in LIP, neurons in SC were also predictively remapping. The comparable prevalence of predictiveness and visual onset latency in the two areas were taken as evidence for LIP being the driver for such activity in SC.

Umeno & Goldberg (1997)

Predictive remapping was then investigated in FEF by Umeno & Goldberg (1997), and they were in particular interested in predictiveness in both single and double-step saccade tasks. Although previous single-step saccade studies had shown predictive remapping in LIP and SC, these areas also show saccade related activity at the single neuron level. Therefore, it was suggested that the observed predictive remapping in LIP and SC may have been due to a subsequent latent saccade plan, i.e. a second saccade to a peripheral stimulus that was planned but not executed. To rule out the possibility that predictive remapping was always simply due to latent saccade plans, Umeno & Goldberg (1997) proposed to investigate whether purely visual neurons in FEF showed predictive remapping, which

could not then be due to latent saccade activity.

Monkeys were required to perform four different tasks during single neuron recording in the FEF. The first task was a delayed saccade task to an eccentric visual stimulus. This was used to classify each neuron as being either visual, visuomovement or movement. The first responded only to stimulus onset, the second to both stimulus onset and saccade onset, and the third only to saccade onset. The second task was a future receptive field task, where a first saccade brought a peripheral stimulus into the receptive field of the neuron, and then a second saccade was performed to a second fixation location. The purpose of the second saccade was to eliminate any latent saccade plan towards the stimulus, and was thus a control. The third task was a null receptive field task, which was a single saccade while the eccentric stimulus was outside the receptive field both before and after the saccade. The fourth task was a no stimulus task, which was just the same first saccade as in the future receptive field task. The last two tasks were controls used to confirm that remapping only occurred when a saccade brought the receptive field onto a stimulus location.

A neuron was classified as predictive if the response latency during the delayed saccade task w.r.t stimulus onset was greater than the response latency in the future receptive field task w.r.t. saccade onset. Umeno & Goldberg (1997) recorded from 32 visual, 48 visuomovement and 20 movement neurons. Movement neurons had no predictive response, while 31% had predictive latencies across both categories of visually responsive neurons. The remapping latencies ranged approximately from -50ms to 250ms, and the delayed

saccade latencies ranged approximately from 60ms to 110ms. The discharge was, just like in SC, found to be stronger in the delayed saccade task compared to the future receptive field task. The fact that there was predictive remapping in visual neurons, and not in movement neurons, and that such activity was also present in future receptive field tasks with a second saccade away from the receptive field, was together taken as evidence that the remapping was related to visual activity, and did not reflect a latent saccade plan. Reciprocal connectivity between LIP and FEF (Barbas & Mesulam 1981) opened the question of whether predictive remapping originated in only one of these areas. However, since the primary sources of efference from FEF to SC were neurons with strong movement activity (Segraves & Goldberg 1987), which did not show predictiveness, it was concluded that FEF was an unlikely source of predictiveness in SC.

Umeno & Goldberg (2001)

The identification of the trace remapping in LIP neurons (Duhamel, Colby & Goldberg 1992) was really a confirmation of the same phenomenon first found in FEF neurons by Goldberg & Bruce (1990), where this had been revealed in the responses of visuomovement neurons during a double-step saccade task. However, the final saccade was in the direction of the receptive field, by definition. So in theory the neuronal response could have been saccade related. Hence Umeno & Goldberg (2001) attempted to explicitly isolate a potential trace remapping in FEF by instead using the flashed stimulus task in Umeno

& Goldberg (1997), where the final saccade was outside the receptive field of the neuron. They also employed the same three other complementary tasks from Umeno & Goldberg (1997) to perform neuronal classification, and control for saccade and visual stimuli independently. Two different forms of visual memory response where identified, one short term within-trial spatial response and one long term inter-trial response, the former being the familiar trace remapping response found in LIP and the latter being novel. Among visually responsive neurons, that is visual and visuomovement neurons, 61% were found to respond in the flashed stimulus task with a memory response. Control tasks confirmed that it was a memory response to the past presence of visual stimuli. A range of response latencies were found, from −48ms to 272ms w.r.t. saccade onset. As for predictive responses in both FEF (Umeno & Goldberg 1997) and SC (Walker et al. 1995), it was found that response magnitude was stronger in the delayed saccade task than in the flashed stimulus task. During recording the different tasks were typically ordered in block. However, the investigators began intermixing trials from different tasks to investigate the presence of long term memory traces. They found that neurons that did not respond in the no stimulus saccade trials would begin to respond if continuous stimulus trials with the same saccade parameters were recently performed. This long term memory trace would last for as long as 30 trials, or 1.5 minutes, and was found in 35% of visually responsive neurons.

Nakamura & Colby (2002)

The primary behavioural significance of predictive remapping was thought to be supporting stability of visuospatial processing across eye movements by reducing the long latency of propagation of reafferent visual signals from retina to higher order visual areas after saccade completion. However, although predictive remapping was thought to contribute to perceptual stability, the areas FEF, LIP and SC discussed above were known to be involved in attention and eye movement rather than visual perception. Therefore, Nakamura & Colby (2002) looked for predictive remapping effects in other brain areas such as V3A, V3, V2 and V1 more associated with visual perception.

Area LIP was known to be reciprocally connected to V3A (Andersen, Asanuma, Essick & Siegel 1990), which was its primary source of visual input, and this area was also known to have eye-movement related activity (Nakamura & Colby 2000). This led Nakamura & Colby (2002) to speculate that V3A may potentially have some of the perisaccadic receptive field properties found in other oculomotor areas, and they attempted to find such responses in area V3A, as well as V3, V2 and V1. A pure fixation task was used to get a baseline latency of response to stimulus onset, and a single step task was used to study remapping. The latter was a saccade task with the following two parameters. The first parameter was whether the stimulus was flashed in the presaccadic or postsaccadic receptive field. The second parameter was which of four different times w.r.t. saccade onset the stimulus was flashed, where the first two times were before the saccade onset and the last two times

were after. The latency of a neuron was defined as the time between stimulus onset and neuronal discharge, and the remapping amplitude was defined as the response following this discharge. A neuron was classified as a remapping neuron when it responded in a version of the single step task where the stimulus onset was presaccadic in the future receptive field, and simultaneously did not respond in classic stimulus and saccade control tasks.

Remapping was identified in visual areas V1, V2, V3 and V3A, at strongly increasing rates of 2%, 11%, 35% and 52% respectively. Latencies were also increasing in terms of the level of the visual hierarchy, with a majority of V3A neurons being predictive, and even presaccadic in some cases. Across all areas, comparing the responses across the different stimulus onset times revealed that responsiveness in the presaccadic receptive field declined with time, and responsiveness in the future receptive field increased. Similar to the SC and FEF, the remapping amplitude was less, in fact 75.3% on average, of the visual onset response in the fixation task. It was also found that the stimulus response was truncated in the presaccadic receptive field, just like had been found in LIP. The authors suggested that the decline in remapping as one descends the visual pathway most likely reflected either that FEF, as the eye-movement driver of remapping, projected with decreasing strength down the hierarchy (Stanton et al. 1995), or that the same applied to LIP backprojections as the potential source of remapping.

Kusunoki & Goldberg (2003)

Shortly after this, the temporal dimension of remapping in LIP was investigated by Kusunoki & Goldberg (2003) using the same task.

A saccade was executed while a stimulus was flashed some time around the saccade cue. The stimulus appeared in one of two screen locations. In so called current receptive field trials, the stimulus location corresponded to the location of the receptive field before the saccade. While for future receptive field trials, the stimulus location corresponded to the location of the receptive field after the saccade. For each trial type, the average population response following stimulus offset as a function of the delay between saccade onset and stimulus offset was investigated.

It was found that in current receptive field trials, the response was greatest when the stimulus was flashed well before saccade onset, and the response declined as it was flashed at later times. In future receptive field trials the response was greatest when the stimulus was flashed well after saccade onset, and the response declined as it was flashed at earlier times.

These results where interpreted to mean that the responsiveness in the current receptive field declined before an impending saccade while the responsiveness to the future receptive field simultaneously increased, and as such was in agreement with Nakamura & Colby (2002). These findings are discussed further in chapter 8.

Heiser & Colby (2006)

A complementary study by Heiser & Colby (2006) investigated the spatial dimension of remapping in LIP, in particular if single neurons could remap stimulus traces from multiple directions. The primary task was a single step saccade task where the receptive field of the given neuron was brought to the location of a recently flashed stimulus, and it was done in both directions both horizontally and vertically. The primary finding was that 67% of single neurons and 79% of remapping neurons could remap from multiple directions. Moreover, the strength of remapping could depend on direction. The directional selectivity of single neurons was quantified, and it was found that neurons ranged from being highly selective ($\sim$ 50%) to completely non-selective. The preferred direction of single neurons was also quantified. At the population level, the directional preference distribution was uniform, showing that remapping was not directionally biased across the population. Neurons remapping in multiple directions had a larger maximal strength, and finally there was no relationship between selectivity or strength of remapping and receptive field location. These were the key findings relevant for subsequent work in chapter 8, which has more discussion. The access to trace stimuli from multiple directions and the lack of directional bias at the population level were consistent with the possibility that LIP was involved in perceptual stability.

6.5 Models

Droulez & Berthoz (1991)

Droulez & Berthoz (1991) presented, for the first time, a neural network model that demonstrated how a dynamically updated eye-centered visual representation could provide the neural substrate for the findings in Mays & Sparks (1980). They showed how a short-term memory map updated dynamically by an eye velocity signal could maintain a retinotopically accurate representation of a visual scene.

The neural network model had three separate, but aligned, two-dimensional retinotopic populations of neurons, as seen in figure 6.4. A visual input population was driven by an external visual signal representing the retinotopic locations of visual targets in a scene. The input population was connected to the intermediate main population, which represented the actual memory map of target locations to be updated. The input population and main population were connected by bidirectional synaptic connections, which enabled the neurons in these two layers to stably support memory activity packets representing the retinotopic locations of targets in the absence of external visual signals. The main population was also connected to an interneuron population. The interneuron population had two neurons per retinal location, which corresponded to two opposing directions of eye movement. The population of interneurons also received inputs from neurons representing eye movement velocity. Each interneuron was wired to detect the coincidence of activity in the corresponding location in the main map and an eye-movement signal in its preferred

direction, and then stimulate visual input neurons representing a corresponding shift in the retinotopic location of the visual target. Efferent signals from the visual input neurons then updated the location of the activity packet in the main population. In this way, the network was able to store a memory of the retinotopic location of a visual target, which was then updated during a saccade by the internal eye velocity signal.

The authors found that this neural network was able to maintain a packet of neural activity representing the retinotopic location of a visual target in the absence of visual input. Moreover, the model was able to maintain multiple activity packets, given that their distance was sufficiently large w.r.t. the receptive field size of the visual input signal. They also found that a single packet, or multiple packets, could be shifted along the map correctly with respect to an external eye velocity signal. The model generalized well over various saccade durations for the same saccade amplitude, although increasing the amplitude led to larger errors. Thus, the model provided an effective demonstration of how an activity packet representing the retinal location of a target could be updated continuously within a neural population by an external eye velocity signal.

However, the model also has some shortcomings. Two experimental findings appear incompatible with the model. Firstly, predictive remapping, in which activity is updated in presaccadic anticipation of impending eye movements, is prevalent within all areas with remapping. This immediately disqualifies eye-velocity as a candidate for guiding the updating process because, by definition, the predictive remapping begins before any eye

movement has occurred giving rise to an eye velocity signal. So it is not immediately clear how predictive remapping could be simulated by a model in which activity packets are shifted continuously by eye velocity signals. Second, remapping has also been observed to occur over a wide range of neuronal response latencies. However, if the representation of a retinotopic target location was updated during saccades by a continuous shifting of a corresponding activity packet, then this would produce a very narrow range of response latencies because the activities of neurons within the moving packet would be bound to each other and approximately synchronised in time.

The anatomical and physiological properties of the model also appear incompatible with basic principles of neural information processing in cortex. First, the interneurons individually compute arithmetic products of the eye movement signal and afferent signals from corresponding neurons in the main population. However, there is very little evidence for arithmetic multiplication responses in the brain. Second, in the model the effect an interneuron on a given postsynaptic visual input neuron can be both excitatory or inhibitory, depending on the direction of the eye-velocity signal it receives. This is clearly not possible in the brain. Third, at any given time the postsynaptic influence of a given interneuron is both excitatory and inhibitory across all its synapses with visual input neurons. This violates Dale's law (O'Donohue et al. 1985), which states that a neuron must have the same postsynaptic effect across all of its synapses. Lastly, it is also an open question as to whether there actually exists an eye velocity signal in the brain that is precise enough to

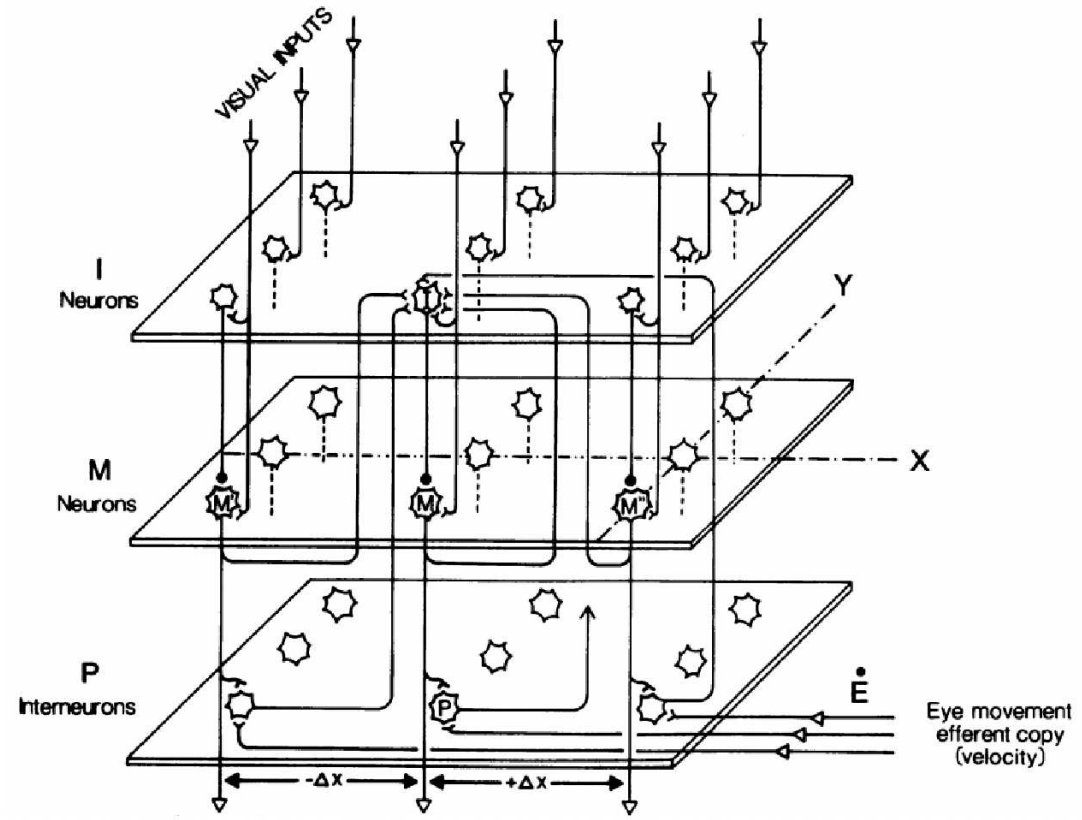


Figure 6.4: Network architecture in figure 1 from Droulez & Berthoz (1991).

facilitate accurate integration.

Krommenhoek et al. (1993)

Sparks & Mays (1983) found that the motor error was accurately encoded in motor error neurons in the SC during a stimulation trial, in which an initial stimulation-induced saccade required an additional adjustment saccade so that a target could be reached. Krommenhoek et al. (1993) investigated through modelling how the intermediate motor layers of the SC, which they referred to as the collicular motor map, could incorporate eye position

signals to produce the spatially accurate motor error signal.

These authors presented two different neural network architectures. Both were feed forward models with an output layer that learned to encode the motor error. The models were trained by a backpropagation of error algorithm. Both models had the following inputs: retinal error at time of target selection (**RE**), eye position at time of target selection ($\mathbf{E}_{\mathrm{T}}$), and present eye position ($\mathbf{E}_{\mathrm{A}}$).

The first model was a three layer model, with all inputs feeding into a fully connected hidden layer, which again projected to the fully connected output layer. The intention was for the model to learn to use the difference between $\mathbf{E}_{\mathrm{T}}$ and $\mathbf{E}_{\mathrm{A}}$, effectively constituting a change-in-gaze-since-target-selection vector, and combine this with **RE** to update the motor error represented in the output layer. This model was, therefore, in some ways similar to the updating model of Droulez & Berthoz (1991).

The second model had an input layer, two hidden layers, and an output layer. First, the inputs **RE** and $\mathbf{E}_{\mathrm{T}}$ were selectively combined in the first hidden layer. Then the output of the first hidden layer was combined with $\mathbf{E}_{\mathrm{A}}$ in the second hidden layer. Finally, the output of the second hidden layer was fed into the output layer. It was expected that the model would develop head-centered representations in the hidden layers during training due to the explicit way that **RE** and $\mathbf{E}_{\mathrm{T}}$ were combined before subtracting $\mathbf{E}_{\mathrm{A}}$ to compute the correct motor error in the output layer. This second model was, therefore, meant to directly correspond to Robinson (1975).

It was found that the first model was able to compute the correct motor error in response to changes in retinal error or actual eye position. They also found that population activity among the hidden units did not encode a head-centered output representation, as had been expected. Moreover, individual output neurons did not purely encode motor error, and hence the code was distributed. The second model was also successful, but it was found to encode a distributed head-centered representation, based on retinal location and present eye position, in the first hidden layer, unlike the first model. This representation also had eye position gain fields.

The finding of this paper, namely that motor error could be computed in two separate ways from the same input signals, was novel. However, the work did not fundamentally overcome these authors' main criticism of the Zipser & Andersen (1988) model, which was that this model used head centered representations to guide supervised learning when in fact no head-centered encoding had yet been found in the brain. The problem was that the two new models of Krommenhoek et al. (1993) had to introduce a novel input signal representing the eye position at the time of target selection, which had also not been found. In fact, the criticism of Zipser & Andersen (1988) was of little significance anyway since these authors never actually claimed that backpropagation of error based on a head-centered supervisory signal was the basis for the self-organization in the brain.

Quaia et al. (1998)

Quaia et al. (1998) presented a model attempting to reproduce the most prominent features of the perisaccadic remapping neurophysiology in LIP, FEF and SC. They also proposed that the ability to remap the activity associated with an extinguished visual saccade target was a possible solution to guiding spatially accurate saccades in the classic double-step saccade paradigm, thereby alleviating the need for a head-centered map. Their model had several broad similarities to the model proposed later in this thesis, and is therefore presented and examined in some detail.

The model attempted to account for the following four observations. First, the phasic onset response of LIP and FEF neurons to stimuli in their visual receptive field. Second, the tonic trace discharge in the time following stimulus offset in the visual receptive field. Third, the presaccadic responses to stimuli that will enter the receptive field with an impending saccade, and how a range of such presaccadic latencies are produced. Fourth, the premature truncation of the tonic trace discharge when a stimulus is removed from the visual receptive field due to a saccade. While these phenomena are present in both LIP and FEF, and some in SC, the authors hypothesized that the LIP was the locus of the signal, and that the other areas merely reflect this activity.

LIP was responsible for the activity remapping with its internal circuitry, and for being able to generate the responses to visual onset and offset described above. An external damping signal caused the truncation of the tonic trace discharge. The part of the model

corresponding to LIP had three primary types of neurons. The phasic neurons (LIP_P) were the only neurons that received a direct visual input signal. All other neurons in the model ultimately had access to the visual signal through these neurons. They had a phasic onset response to visual stimuli in their receptive field, but were then quickly silenced by inhibitory circuitry depressing their visual input at the dendritic level. These neurons projected to phasic neurons in FEF, described later, and also to tonic neurons in LIP (LIP_T) which had a large time constant (3s) making them discharge for a long time after the tonic neurons had gone silent. Both LIP_P and LIP_T neurons projected to phasic-tonic neurons in LIP (LIP_{PT}). These neurons, therefore, inherited both a phasic and tonic response to visual stimuli, and they represented the commonly found LIP neurons in remapping experiments. This basic circuitry is shown in figure 6.5.

FEF was responsible for encoding the impending saccade in an eye-centered reference frame during a brief presaccadic interval, and for propagating the initial remapping activity in LIP to new LIP neurons through the recurrent circuitry, thereby producing the remapping latency distribution desired. The part of the model corresponding to FEF had two primary types of neurons. The FEF movement neurons, or FEF_{MOV}, fired up to 100ms in advance of saccades into their receptive field. The output of FEF_{MOV} and LIP_{PT} neurons was combined in an AND-like operation at a synaptic level on a LIP_P neuron. Critically, this neuron had its receptive field in the retinal location at which remapping activity would appear after the saccade by the FEF_{MOV} neuron was performed while a visual target was

presaccadically located in the receptive field location of the LIP_{PT} neuron. This basic circuitry is shown in figure 6.6. Only some LIP_P neurons had such afferent circuitry, and this allowed them to remap visual signals at very early latencies. Remapping activity was spread to other LIP_P neurons, lacking such circuitry, through a recurrent connectivity with tonic neurons in the FEF (FEF_P). It was this recurrent connectivity, resulting in a cascade of activity back and forth between the LIP and FEF, which produced the range of latencies in remapping. This basic circuitry is shown in figure 6.7. To prevent this cascade from taking place during ordinary visual fixation they gave the visual signal a dendritic inhibitory shunting effect on the synapse between FEF_P to LIP_P neurons. With this architecture they are able to reproduce all four receptive fields effects.

The authors admitted that the circuitry in their model was perhaps implausibly precise, but they were able to somewhat ameliorate this by suggesting a plausible self-organization hypothesis for the key circuitry responsible for the remapping, namely the synaptic level integration of appropriate combinations of FEF_{MOV} and LIP_{PT} neurons onto LIP_P neurons. They proposed that initially, each LIP_P neuron synaptically integrated all combinations of FEF_{MOV} and LIP_{PT} neurons. However, a Hebbian correlation based learning rule adjusting the strength of the LIP_{PT} to LIP_P connections while the subject saccaded around a visual scene would potentiate only connections from the appropriate LIP_{PT} neurons. The authors only sketched the workings of such a mechanism, and even anticipated the need for some timing mechanism. The subtleties of the need for such a mechanism are discussed in further

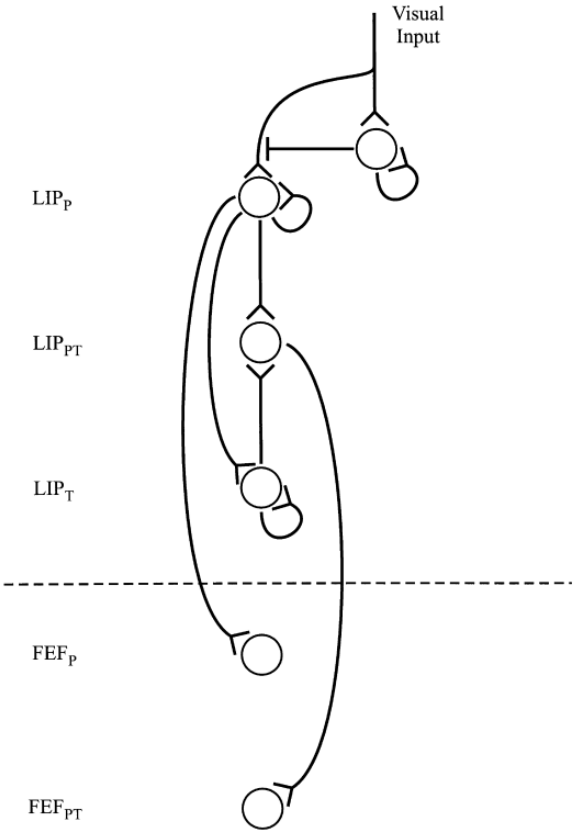


Figure 6.5: LIP visual circuitry (figure 10 from Quaia et al. (1998)).

detail in chapters 7 and 8.

Xing & Andersen (2000*a*)

Mazzoni et al. (1996) found that neurons in area LIP had responses tuned to saccades in the direction of their receptive field, and that these receptive fields would keep a memory trace of a planned saccade if and only if it was the impending saccade. During delayed double-step saccades, which was a task where there was a long delay and a subsequent cue before saccades could be initiated, a neuron would only encode a planned saccade when

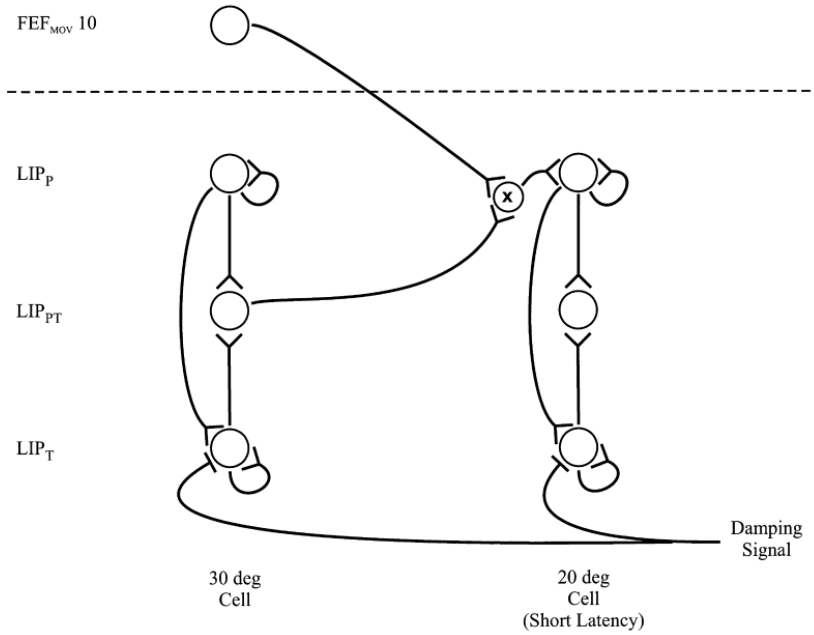


Figure 6.6: FEF-LIP remapping circuitry I (figure 11 from Quaia et al. (1998)).

the intervening saccade was completed. Xing & Andersen (2000*a*) proposed a hypothesis for how LIP potentially could produce these observations. When the first saccade location target is flashed, it is loaded into a primary neural population with internal recurrent circuitry which makes it resist changing its state to future visual input. When the second saccade location is flashed, it is loaded into a secondary neural population along with the present eye position, which only had its input lines open after the first target was flashed. The first saccade would be programmed according to the retinal encoding in the primary population, which would have its activity reset upon saccade completion. This would cause the primary population to compute its new state based on the output of the secondary population, and the present eye position of the agent, in a way so as to encode the motor error for the second saccade.

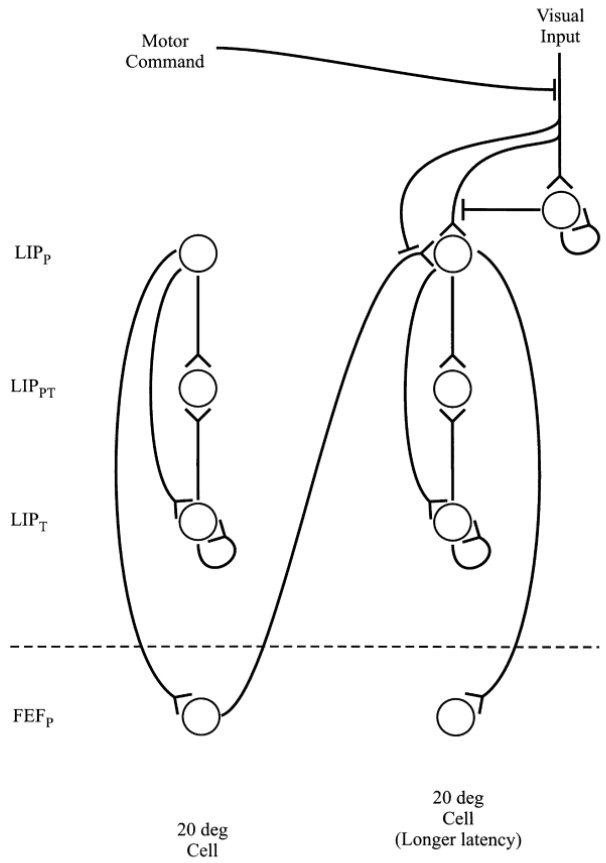


Figure 6.7: FEF-LIP remapping circuitry II (figure 12 from Quaia et al. (1998)).

The hypothesis was instantiated as a neural network model with corresponding architecture. The input layer had disjoint populations encoding the retinal location of visual targets and the present eye position. The input layer was fully connected to a hidden layer. The hidden layer consisted of two neuronal populations, referred to as RN-I and RN-II, which respectively corresponded to the primary and secondary populations in the above hypothesis. The secondary population RN-II projected fully to the primary population RN-I, while each population was also recurrently connected internally. The RN-I population also received an exogenous post-saccadic reset signal. The neurons in RN-I were

fully connected with an output layer encoding the motor error of the impending saccade. All synapses, including the recurrent ones in the hidden layer populations, where adjusted by a dynamical version of the back-propagation of error learning rule, during which the correct motor error was supplied to the output layer as a supervisory training signal.

A reduced model with only the RN-I component of the hidden layer was first trained and tested in a simple delayed saccade trial. After training, the correct motor error was encoded in the output layer, and importantly it was invariant to distractor targets flashed during the delay period. This invariance was produced by recurrent short range excitatory and long range inhibitory circuitry within the hidden layer, an architectural feature that the authors referred to as the 'push-pull structure'.

Then the full model, in which the hidden layer contained both RN-I and RN-II components, was trained and tested in the delayed double-step trial. The correct motor error was again encoded in the output layer after training. Hidden layer units were found to have monotonic eye position gain fields in their visual receptive fields. In particular, the RN-I units had gain fields and retinal receptive fields in the opposite direction, while the RN-II units had gain fields and retinal receptive fields in the same direction. This finding, together with the fact that opposite and equal gain field alignment corresponds to subtraction and addition respectively, pointed to a clear functional role of each population. RN-I would at the beginning of the trial subtract retinal location from eye position to encode the motor error of the first target. RN-II would then, when the second target was flashed,

subtract the eye position from the retinal location, thereby encoding the second target in a distributed head-centered reference frame. When the first saccade was completed, RN-I would be suppressed, and RN-II would be its only input, whereupon it would subtract the new eye position from the head-centered encoding of the second target, thereby producing a motor error signal for the second target. The authors also attempted to explain the remapping findings of Duhamel, Colby & Goldberg (1992) by claiming that the activity really reflected a latent saccade plan to the peripheral stimulus, and that this model would in that case produce the same response.

The main problem for this model was that it was intimately related to reproducing double-step saccade trial response results. Small changes to the task, where for example a cue is used to instruct the agent in which order to make the saccades, rather than the fixed order presumed by the model, would clearly be behaviourally possible, but the model could not handle this. There would be no way of knowing what target to load into what neuronal population. More generally, given the experimentally contrived nature of double-step saccade tasks, it seems unlikely that there would be explicit neural populations in the brain to plan each saccade in such tasks, and that one particular population (RN-II) had a gating signal for selection of a second saccade target that will soon be extinguished. Their claim that remapping activity in Duhamel, Colby & Goldberg (1992) could reflect saccade plans was controlled for in other remapping experiments by introducing a secondary distractor saccade, and remapping still persisted (Umeno & Goldberg 1997, 2001).

Mitchell & Zipser (2001)

Mitchell & Zipser (2001) investigated whether a single three layer neural network model could simultaneously produce both head-centered and eye-centered dynamical output representations for spatially accurate eye movements. A head-centered representation achieves this by coding the final saccade target in a supraretinal reference frame, making it invariant to intervening saccades, and an eye-centered representation does this by shifting activity in correspondence with the intervening saccade vector. Given the difference, it was not clear that a single dynamical model could do both simultaneously.

The network was a three layer model. The input layer encoded the retinal location of targets, as well as eye position and eye velocity. The hidden layer had recurrent excitatory connections, which facilitated a memory trace of extinguished visual input. The hidden layer projected to two output layers. The aim was that one output layer would learn to represent the target location in a head-centered reference frame, while the other output layer would develop an eye-centered representation of the target location. The network was trained to produce the appropriate representation in each output layer using backpropagation of error learning while performing a double-step saccade trial.

During testing, the output representations were decoded during a memory guided saccade task to confirm correct updating. Both representations were found to perform correct updating. That is, the eye-centered output shifted memory packets by an opposite vector to the intervening saccades, while the head-centered output packet remained stationary.

Hidden units were found to have overlapping visual and memory period preferences, and also planar eye position gain fields, both of which are found in LIP. They were also found to participate in both head-centered and eye-centered output representations simultaneously, and have recurrent excitatory connectivity used to sustain memory packets.

While the paper did study a classical problem in an original manner by combining the two output representations, it was not clear why this was a useful analysis. The two mechanisms have nothing inherently in common beyond being candidate mechanisms for the same remapping behaviour and physiology. Hence, insights gained about the interaction between neurons participating in head-centered and eye-centered representations in the model may not be of much use in understanding the neurobiological mechanisms actually present in cortex.

6.6 Perisaccadic Dynamics and Head-Centered Coordinate Transformation

In the previous chapters, the development of head-centered visual representations was explored, and in the following chapters the development of a set of perisaccadic responses is considered. Both phenomena depend on the integration of oculomotor and visual signals, and both have been proposed as candidate mechanisms for spatially accurate eye movements and perceptual stability, as covered in section 6.3. This leads to the question of whether they are related mechanisms, and in particular if the latter could be implemented

using the former. This has been proposed in the work of Krommenhoek et al. (1993), covered in section 6.5. A head-centered visual signal, which is able to maintain a memory of extinguished visual stimuli, does in combination with an eye position signal determine the postsaccadic retinal location of the stimuli. This is precisely what is needed to perform remapping.

There are however a number of reasons why it appears unlikely that perisaccadic responses depend on head-centered representations.

First, remapping responses have been found in a very large number of primate visual areas, such as LIP (Duhamel, Colby & Goldberg 1992), SC (Walker et al. 1995), FEF (Umeno & Goldberg 1997, 2001) and visual areas V3,V3A,V2 and V1 (Nakamura & Colby 2002). Moreover, when found, they are often prevalent. Head-centered responses have however only been found in a few areas, primarily LIP (Mullette-Gillman et al. 2005), V6 (Galletti et al. 1993) and VIP (Duhamel et al. 1997). Moreover, when found, they are always very rare, in particular in a pure head-centered frame of reference. These differences are not easily reconciled with how the former could depend on the latter.

Second, remapping is always observed to occur over a range of latencies, and often even prior to saccade onset, and this is incompatible with a computation which depends on present eye position.

Third, in the only case where it has been examined, namely FEF Sommer & Wurtz (2006), it has been found that remapping activity does not shift continuously from the

initial to the final retinal location of the stimuli in the eye-centered map, but rather jumps ballistically. Eye position based remapping would generate a shift, rather than a jump, and is hence incompatible with this finding.

For these reasons it seems appears unlikely that a head-centered visual representation combined with eye position is unlikely to be the mechanism underlying remapping.

Chapter 7

Perisaccadic Receptive Field Dynamics: Methods

7.1 Introduction

This chapter presents the methodology of the model for perisaccadic receptive field dynamics. The hypothesized functioning of the model is presented, along with the neural network architecture, training stimuli, testing stimuli, model dynamics and data analysis routines.

7.2 Experimental Observations from Single Cell Recording Studies

The model developed here aims to explain the visually-guided development of different kinds of receptive field phenomena observed around the time of saccades in a number of visual and saccade related brain areas, in particular LIP (Duhamel, Colby & Goldberg

1992), SC (Walker et al. 1995), FEF (Umeno & Goldberg 1997, 2001) and V3 (Nakamura & Colby 2002). The primary phenomena to be modelled were as follows:

Predictive Remapping: The saccade aligned latency of the response of a visual neuron to a saccade bringing a visual stimulus into its classical receptive field is less than the latency in a visual onset fixation task (Duhamel, Colby & Goldberg 1992, Nakamura & Colby 2002, Umeno & Goldberg 1997, Walker et al. 1995).

Presaccadic Remapping: Predictive remapping which begins prior to saccade onset (Duhamel, Colby & Goldberg 1992, Nakamura & Colby 2002, Umeno & Goldberg 1997, Walker et al. 1995).

Trace Remapping: The response of a visual neuron to a saccade that brings the site of a recently extinguished stimulus into the neuron's classical receptive field (Duhamel, Colby & Goldberg 1992, Goldberg & Bruce 1990, Umeno & Goldberg 2001).

Responsiveness Shift: Around the time of a saccade, the responsiveness of a neuron gradually declines to stimuli flashed in the presaccadic location of its visual receptive field, and gradually increases to stimuli flashed in the postsaccadic location of its visual receptive field (Kusunoki & Goldberg 2003).

Spatially Independent Remapping: Individual neurons can remap activity from multiple different spatial locations and directions in terms of both strength and latency (Heiser et al. 2005).

7.3 Hypothesis

It was hypothesized that the following core model components, as shown in figure 7.1, would permit the development of the receptive field dynamics described above:

- *Visual population.* There is a population of visual input neurons that represent the retinotopic locations of visual targets. These neurons continue to fire for a short period after the saccade, and thus encode a memory trace of the presaccadic visual scene for some time after saccade onset and stimulus offset. These neurons represent a visual map of the stimuli in a scene.

- *Saccade population.* There is a population of eye-centered saccade neurons, which in some perisaccadic time interval encode the target location of the impending saccade. These neurons also continue to fire for a short period after the saccade, thus implementing a memory trace of the target location of the last saccade.

- *Remapping population.* There is a population of visual neurons that represent the retinotopic locations of visual targets, but where each individual neuron is driven by a combination of three different input signals. Specifically, there is a phasic input component that gives rise to a localised burst of activity peaking after a few tens of milliseconds, an underlying tonic input component that keeps the neuron active while the stimulus is present, and a memory trace component that lasts for up to a few hundred milliseconds after the stimulus is removed. There is a range of latencies

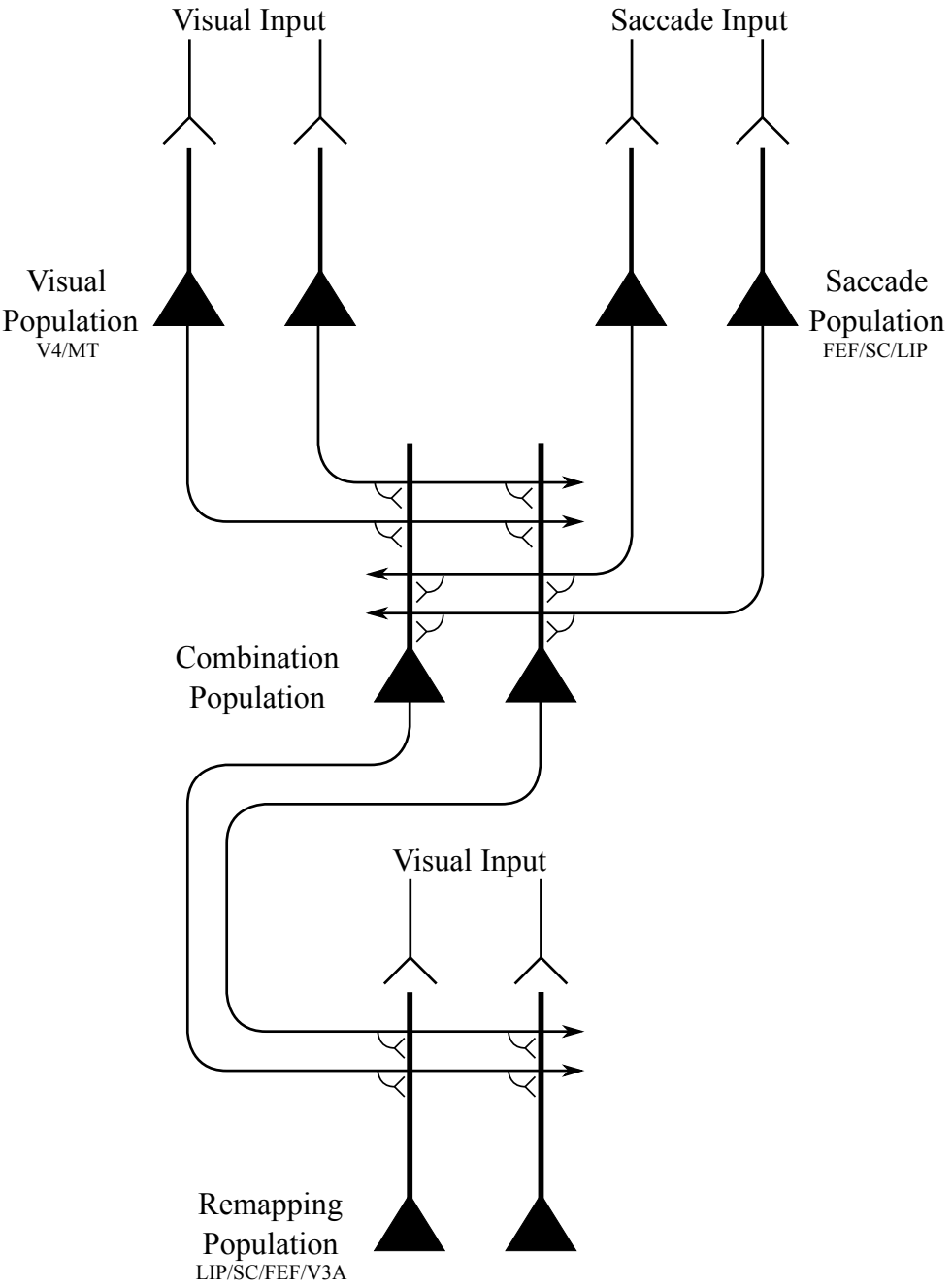


Figure 7.1: Network architecture of model.

in the onset time of the phasic burst. These three input signals are combined for each neuron. Upon saccade onset all three input signals representing the presaccadic stimulus location are truncated.

- *Combination population.* There is an intermediate competitive population of neurons which receives diluted synaptic connections from the visual population and saccade population, and then sends efferent connections to all neurons in the remapping population.

- The synapses from the combination population to the remapping population are dynamically adjusted through associative learning with a Hebbian synaptic learning rule. This layer of synapses effectively implements supervised pattern association learning from the combination population to the remapping population.

- Head-centered visual space is stable across single saccades for some part of the time.

7.3.1 Details of Model Components

The visual input population functions somewhat like a standard eye-centered visual map of the visual stimuli present within a scene. However, neurons in this population have a delayed update time of a few hundred milliseconds in their response to a saccade, which alters the visual representation in eye-centered space. This response inertia is then propagated to the combination population. The effect of this is to allow an inter-temporal association to occur between the combination population, which represents the scene before the saccade, to the remapping population, which represents the scene after the saccade. It is proposed that a population of visual neurons with long postsaccadic response delays of the order of hundreds of milliseconds could arise due to the accumulation of axonal con-

duction delays (Girard et al. 2001) and neuronal response delays of the order of tens of milliseconds as visual signals are propagated between a number of subcortical and cortical brain regions, including being propagated across recurrent connections within individual regions. Furthermore, the memory trace response of a visual neuron after a stimulus has been removed from its receptive field, either due to the extinction of the stimulus or a saccade, may be enhanced by biological factors such as long synaptic time constants that keep the postsynaptic neuron active for tens of milliseconds after the presynaptic signals have been extinguished (Evans & Stringer 2012), and local recurrent circuits that may help to maintain loops of activity after the stimulus is removed (Elliffe et al. 2000). Such trace responses representing a presaccadic stimulus location may also contribute to the delayed response of the visual population to the new location of the stimulus after a saccade. It is proposed that a small subset of visual neurons, intermixed within various stages of processing, will have the required long postsaccadic response latencies. Consequently, the proposed population of visual neurons may be distributed across a number of linked brain areas, including visual areas V1 to V4 and MT, all of which project directly or indirectly to the areas within which remapping has been found (Blatt et al. 1990). However, in the modelling study described below, it was decided to focus on the remapping dynamics, themselves, rather than attempt to simulate how the visual population develops long postsaccadic response latencies over a series of subcortical and cortical stages. Therefore, the postsaccadic response latencies are imposed directly on the visual population in the model

simulations presented here.

The saccade population functions like an eye-centered motor encoding of impending saccades. Critically, this population represents the eye-centered target location of a saccade, both in advance of the saccade and for a few hundred milliseconds after the saccade has been executed. Many examples of neurons encoding impending saccades have been identified, for example, FEF movement cells (Bruce & Goldberg 1985), LIP neurons (Ipata et al. 2006) and SC neurons (Wurtz & Goldberg 1972). Neurons in FEF have also been identified with explicit postsaccadic responses (Bizzi 1968, Bruce & Goldberg 1985).

The remapping population functions like a standard eye-centered visual map, but which is endowed with some extra characteristics to imitate the neuronal responses observed in LIP. These extra response characteristics include a specific response latency to a stimulus that is flashed in the neuron's receptive field, a trace response to a recently extinguished stimulus, and the truncation of this trace response as a result of a saccade onset (Duhamel, Colby & Goldberg 1992). The neurons in this population have different response latencies that result in different remapping latencies. The neuronal trace response and the truncation of the trace response as a result of saccade onset were both included to replicate the findings of Kusunoki & Goldberg (2003).

The remapping population also receives associatively modifiable synaptic connections from the combination population, which represents a combination of the time-delayed visual stimulus location and saccade target location. The time-delayed responses of the

input populations, leading to a corresponding delay in the representation carried by the combination population, leads to a time difference of a few hundred milliseconds between the representations in the combination population and remapping population. This allows the network to learn associations between particular combinations of presaccadic stimulus location and saccade target location represented by the combination population and the resulting postsaccadic stimulus location represented by the remapping population. Thus, the goal was for the model to self-organise its synaptic connections by visually-guided learning in order that the remapping population is able to predict the postsaccadic stimulus location, as well as replicate various other experimentally observed remapping dynamics.

The combination population received diluted **non-modifiable** synaptic connections from the visual population and saccade population. They encode unique combinations of the impending saccade vector and the present presaccadic retinal location of the visual stimulus. Neurons in this population had their firing rate threshold set so that they would only respond if they received simultaneous input from both the visual population and saccade population. The combination population could potentially exist in the same cortical region as the remapping population itself.

It is assumed that the head and visual stimuli often remain stationary during the time course of a saccade. As discussed earlier in section 2.1, this is a reasonable assumption given the known statistics of how primates move their head and eyes. Specifically, a primate adjusts its gaze more frequently by moving its eyes rather than its head (Freedman & Sparks

1997). Evidence for this during exploration of natural environments with free eye, head and body movements has been reported by (Einhäuser et al. 2007). The utilization of the statistics of natural eye and head movements to drive learning in the model represents a fundamental unifying principle that is shared between the simulations of head-centered visual neurons reported in the first part of this thesis and the simulations of perisaccadic receptive field dynamics presented below.

7.3.2 Learning

The updating of the synaptic connections in the model happens as follows. During a training trial, the stimulus remains stationary in head-centered space while a single saccade is performed. Before the planned saccade is initiated, both the visual population and saccade population will be active, encoding the location of a visual stimulus in the scene and the target location of the impending saccade, respectively. This will cause a subset of neurons in the combination population, which are tuned to this particular combination of presaccadic stimulus location and saccade target location, to fire.

When each saccade is initiated, the visual population and saccade population will have delayed responses to the saccade of a few hundred milliseconds. This allows the corresponding set of combination neurons to continue to fire for this period. However, the responses of the remapping population are not subject to the same time delays as the visual population and saccade population. Therefore, immediately after the saccade, the remapping

population will reflect the postsaccadic location of the stimulus. In this situation, the active combination neurons representing the presaccadic activity will be associated onto the remapping neurons that reflect the postsaccadic stimulus location.

A few hundred milliseconds after the saccade, the visual population will begin to represent the postsaccadic location of the stimulus. However, the saccade population will cease responding. This will cause the combination population to cease responding as well due to the firing rate threshold. Thus, further associative learning in the connections from the combination population to the remapping population will cease at this point for the current saccade.

After training, a particular combination of presaccadic stimulus location and saccade target location represented in the visual population and saccade population will stimulate the corresponding subset of combination neurons. These combination neurons, in turn, will stimulate the remapping neurons that represent the corresponding postsaccadic stimulus location.

7.3.3 Predicted Model Behaviour

It was hypothesized that after learning, the model should be able to replicate the experimental observations described above in the following ways:

- Predictive remapping should happen when activity in the visual population representing the presaccadic stimulus location and activity in the saccade population

representing the saccade target location stimulate the corresponding combination neurons, which then stimulate the remapping neurons that represent the corresponding postsaccadic stimulus location.

- Presaccadic remapping should occur by the same means as predictive remapping described above. Although not simulated in this thesis, the broad distribution of response latencies in both predictive and presaccadic remapping could naturally emerge as a result of variability in axonal delays or neuronal and synaptic time constants.

- Trace remapping would occur in the remapping population because the visual input population would continue to represent a stimulus a few hundred milliseconds after the stimulus offset. This would allow memory activity within the visual population reflecting a recently extinguished stimulus, combined with activity from the saccade population representing the intended saccade target location, to stimulate remapping neurons that reflect what would have been the corresponding postsaccadic stimulus location.

- Responsiveness shift around the time of a saccade may also be accounted for with this model. First, consider a stimulus flashed in the spatial location where the receptive field is located before the saccade. When stimulus is flashed at increasing times before saccade onset, trace response truncation causes a declining response as measured after stimulus onset. When stimulus is flashed after saccade onset, there is no response. Second, consider a stimulus flashed in the spatial location where the receptive field

is located after the saccade. When stimulus is flashed at increasing times before saccade onset, remapping causes an increasing response as measured after stimulus onset. When stimulus is flashed after saccade onset, there is pure visual response.

- Spatially independent remapping would occur as long as the model was extensively trained on many different combinations of presaccadic stimulus location and saccade target location. After training, this would enable individual neurons in the remapping population to remap activity from multiple different spatial locations and directions.

Hence, it is hypothesized that the model architecture shown in figure 7.1 incorporates mechanisms that are sufficient for explaining a wide range of observed perisaccadic receptive field dynamics.

7.4 Network Model Architecture

The network model consisted of four populations of neurons, the names and interconnectivity of which is seen in figure 7.1.

There were two input populations that represented corresponding one-dimensional retinotopic spaces. The first input population, the visual population, was purely visual and represented the retinal location of visual stimuli. While the second input population, the saccade population, encoded saccade plans by representing the retinal target locations of impending saccades. The visual population, the size of which was denoted by N^{V}, consisted of visual neurons representing $[-45°, 45°]$ of eye-centered visual space. The saccade

population, the size of which was denoted by N^{S}, consisted of saccade planning neurons representing $[-30°, 30°]$ of eye-centered saccade space. Each population represented the corresponding space by having neurons with a preference for each integer location in the space. Hence $N^{\mathrm{V}} = 91$ and $N^{\mathrm{S}} = 61$. Neither population had any topographic organization. However, simulation results will be presented topographically according to neuron preferences in order to facilitate inspection of the model behaviour.

The combination population consisted of N^{C} combination neurons that each received synaptic connections from a randomly assigned subpopulation of neurons from the visual and saccade populations. Specifically, each combination neuron received inputs from a subpopulation consisting of ϕ^{V} and ϕ^{S} percent of the visual and saccade populations, respectively. Hence the total number of afferents per combination neuron was $(N^{\mathrm{V}}\phi^{\mathrm{V}} + N^{\mathrm{S}}\phi^{\mathrm{S}})/100$.

The remapping population, the size of which was denoted by N^{R}, also represented the retinotopic locations of visual stimuli. This population had the same encoding and structure as the visual input population. It consisted of visual neurons representing $[-45°, 45°]$ of eye-centered visual space, where each integer position in the space was represented by a corresponding neuron with a preference for that location. Hence, the number of remapping neurons was $N^{\mathrm{R}} = 91$. This map was also not topographically organized. However, simulation results will present the remapping population topographically in terms of neuronal preference in order to aid analysis of the model performance. Each remapping neuron re-

ceived synaptic connections from a randomly assigned subpopulation of neurons from the combination population. Specifically, each remapping neuron received inputs from a subpopulation consisting of ϕ^{C} percent of the combination neurons. Hence the total number of afferents per remapping neuron was $(N^{\mathrm{C}}\phi^{\mathrm{C}})/100$.

All synaptic connections between neurons were initially set to a random weight in the interval $[0, 1]$ and subsequently normalized as described in section 7.6.

7.5 Stimuli

The relationship between the eye-centered location r of a visual stimulus, the head-centered eye position e, and the head-centered location h of the same visual stimulus is given by the equation $h = e + r$. The space of head-centered visual stimulus locations, h, covered the interval $[-45°, 45°]$. This meant that visual targets could only be located within this given region of head-centered space. Visual targets were always stationary in head-centered space within a given trial, both in training and testing. The space of eye-centered visual stimulus locations represented by both the visual and remapping populations also covered $[-45°, 45°]$. All trials started with the eye fixating straight ahead at $e = 0°$ in head-centered space. The space of retinotopic saccade target locations represented by the saccade population covered $[-30°, 30°]$, which meant that the saccade population could only represent this range of saccades. Since all trials only had one saccade at most, this implied that the final head-centered eye position e was also confined within the given

saccade plan space $[-30°, 30°]$. However, for any given head-centered stimulus location, the retinal saccade target location was in practice further bounded to ensure that the postsaccadic stimulus location remained within $[-45°, 45°]$. The model was first trained, and then subsequently tested on a range of different tasks, described below, to characterise the various receptive field properties of all neurons in the model. All saccades are performed at a constant velocity of $300°/s$.

7.5.1 Training the Network

A training epoch consisted of M different trials, each of which involved the model performing a saccade while a visual stimulus was present at some location in the head-centered visual space. The head-centered stimulus location and saccade was varied randomly between trials. Each trial thus trained the network on a particular remapping.

In more detail, every trial i began with fixation straight ahead, with initial head-centered eye position $e_i^0 = 0$. For each trial, a stimulus was placed in a head-centered location h_i for the duration of the trial. This corresponded to an initial retinotopic stimulus location $r_i^0 = h_i - e_i^0 = h_i$. Then the model performed a saccade to a retinal target location s_i that brought the stimulus onto a final retinal location $r_i = r_i^0 - s_i$ after the saccade. For each trial, the head-centered stimulus location h_i and eye-centered saccade s_i were picked randomly, but with the constraint that the resulting saccade was no less than $10°$ in magnitude, and that r_i remained within the retinal space.

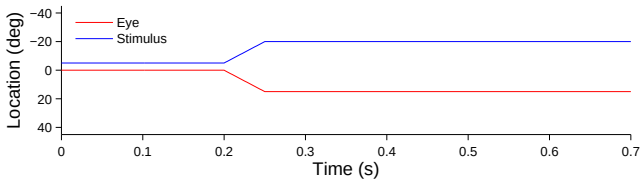


Figure 7.2: The figure shows a typical training trial. The head centered eye position (red) and eye-centered visual stimulus location (blue) are plotted against time. The trial began while fixation was straight ahead (eye position was at $0°$) and the stimulus was located at $-5°$. A saccade of $15°$ was initiated at 200ms with a constant velocity of $300°/s$, and was completed in 50ms. After the saccade, another 450ms fixation period at the new eye position $15°$ was performed, while the stimulus had moved to $-5° - 15° = -20°$ on the retina.

The saccade was initiated after 200ms of initial fixation. Then a post-saccadic fixation period of 300ms combined with some time for the saccade itself brought the total trial length to approximately 700ms. Figure 7.2 shows the typical time course of a training trial.

7.5.2 Testing the Network

Stimulus Control Task

In stimulus control tasks, a visual target was briefly flashed while fixation was maintained throughout the trial. A given trial presented the stimulus in a unique retinal location. Since there was one such trial for each integer location in the retinal space $[-45°, 45°]$ there were 91 trials in total. For each remapping neuron, two particular trials were of interest. Firstly, the trial corresponding to the retinal preference of the given neuron was used to assess the visual response latency of the given neuron. This was later compared to the latency of the neuron when testing remapping, so as to reveal potential predictive remapping. Secondly,

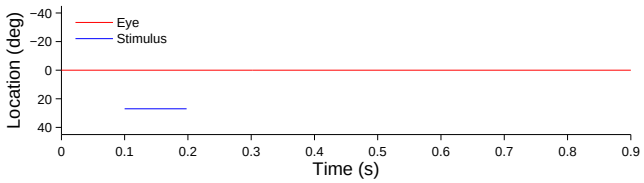


Figure 7.3: The figure shows a typical stimulus control task trial. The task began while fixation was straight ahead. The stimulus was presented at retinal location 32° during the period $100-200$ms. After stimulus offset, there was a 700ms period of maintained fixation.

the trial corresponding to the retinal location in which, during training, the stimulus had been presented presaccadically before remapping into the receptive field of the given neuron. This trial was necessary in order to confirm that the response of the neuron, when testing for remapping, was not simply due to the presence of the stimulus in the visual field. This control was also used by Berman et al. (2007), Dunn et al. (2010), Heiser et al. (2005), Heiser & Colby (2006). Figure 7.3 shows the time course of a typical stimulus control task trial.

Saccade Control Task

In saccade control tasks, a single saccade was executed while no stimulus was present in the visual field. A given trial required performing a unique saccade. Since there was one such trial for each integer location in the space of retinotopic saccade target locations $[-30°, 30°]$, there were 61 trials in total. For each remapping neuron, the task corresponded to the saccade performed during training that set up remapping into the receptive field of the given neuron. This was necessary to confirm that the response of the neuron, when testing for remapping, was not simply due to the saccade execution. This control was also

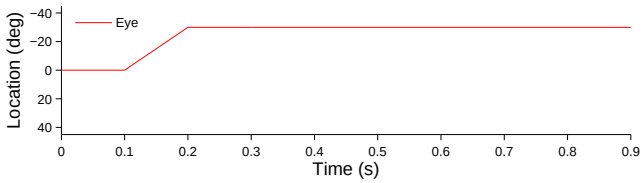


Figure 7.4: The figure shows a typical saccade control task trial. The task began while fixation was straight ahead. A saccade of $-30°$ was initiated at 100ms, and with a constant velocity of $-300°/s$ it was completed 100ms later. After saccade completion, fixation was maintained at the new eye position $-30°$ for another 700ms.

used by Berman et al. (2007), Dunn et al. (2010), Heiser et al. (2005), Heiser & Colby (2006). Figure 7.4 shows the time course of a typical stimulus control task trial.

Probe Task

In probe tasks, a single saccade was executed while a stimulus was presented for the full duration of the trial. A given trial involved a unique combination of stimulus location and saccade target location. Since there was one such trial for each combination of integer positions in each of the two spaces, the total number of trials was $61 \times 91 = 5551$. For each combination neuron, this full set of trials allowed the decoding of what combination of stimulus location and saccade target location the neuron was responsive to. This information was further used to analyze the functional connectivity of the weight vectors of neurons in both the combination and remapping populations. Figure 7.5 shows the time course of a typical probe task trial.

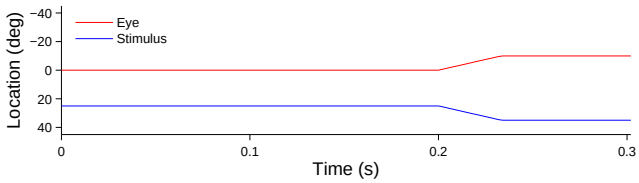


Figure 7.5: The figure shows a typical probe task trial. The trial began while fixation was straight ahead and a stimulus was presented at 25°. A saccade of $-10°$ began at 200ms, and with a constant velocity of $-300°/s$ it was completed 100/3ms later.

Single Step Task

In single step tasks, a stimulus was briefly flashed and a saccade was subsequently performed. A given single step task trial corresponded to a specific training trial in the sense that it involved the same combination of stimulus location and saccade target location. The purpose of this task was to measure the remapping that the corresponding training trial was meant to have taught the network. The difference between the single step task trials and the training trials was that in single step trials the stimulus was flashed for only a short period, as in Berman et al. (2007), Duhamel, Colby & Goldberg (1992), Dunn et al. (2010), Heiser et al. (2005), Heiser & Colby (2006), while in training trials it remained visible for the full duration of the trial. Figure 7.5 shows the time course of a typical single step task trial.

Delayed Stimulus Flash Task

The ith training trial involved a saccade s_i that brought the receptive field of a remapping neuron to the head-centered location of the visual stimulus. For each training trial, there were two families of delayed stimulus flash tasks conducted during testing. In each delayed

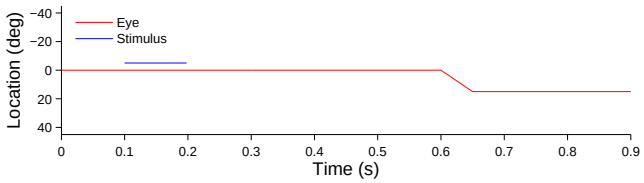
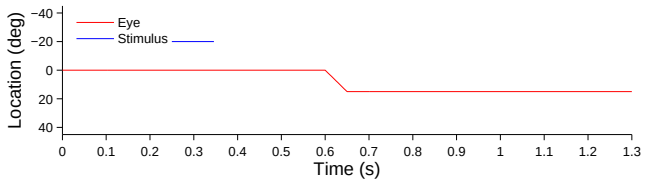
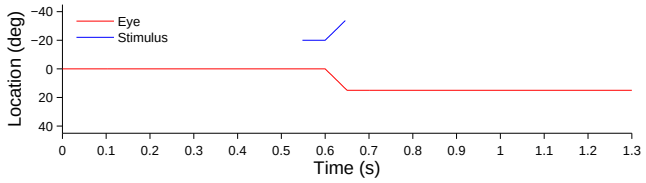


Figure 7.6: The figure shows a typical single step task trial. The trial began while fixation was straight ahead. At 100ms a stimulus was presented at $-5°$ for 100ms. Then at 600ms a 15° saccade was initiated. The constant saccade velocity of $300°/s$ meant the saccade ended after 50ms. After saccade completion, the fixation was maintained for another 250ms at the new eye position of 15°.

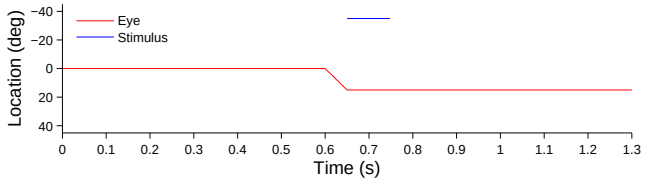
stimulus flash task, a stimulus was briefly flashed at some time w.r.t. the same saccade s_i that was performed during the ith training trial. In one family of task trials, the stimulus was flashed in the head-centered location corresponding to where the receptive field of the remapping neuron was located in the training trial before the saccade. Each of these trials was referred to as a *current receptive field trial*. In the other family of task trials, the stimulus was flashed in the head-centered location corresponding to where the receptive field of the remapping neuron was located in the training trial after the saccade. Each of these trials was referred to as a *future receptive field trial*. Both families had the same number of trials, each of which corresponded to a given stimulus onset time ranging from 100ms to 700ms in increments of 50ms. This resulted in 13 trials in each family. In total, therefore, there were $M \times 2 \times 13 = 26M$ delayed stimulus flash task trials. The purpose of this task was to measure the time course of the responsiveness of the neuron to a stimulus presented in the current and future receptive fields around the time of a saccade, exactly as done by Kusunoki & Goldberg (2003) and Nakamura & Colby (2002). Figure 7.7 shows the time course of a typical delayed stimulus flash task trial.



(a) 250ms



(b) 550ms



(c) 650ms

Figure 7.7: The figure shows three typical delayed stimulus flash task trials with the given stimulus onset times of 250ms (a), 550ms (b) and 650ms (c). All plots are current receptive field trials for a neuron with retinal receptive field $-20°$. In all three cases the stimulus is located in head-centered location $-20°$. The saccade is $15°$ in all three cases, and is initiated at 600ms and lasts for 50ms.

7.6 Neuronal and Synaptic Dynamics

For all simulation experiments, there was at most one saccade performed during each training or testing trial, with the saccade performed from fixation straight ahead to retinal location s. Moreover, there was never more than a single visual stimulus present during a trial, which was kept at a fixed head-centered location h. For those trials with a stimulus present, there was only a single time of onset for the stimulus during the trial. The times of the saccade onset, visual stimulus onset and visual stimulus offset are denoted by t^{SACC}, $t^{\text{STIM}}_{\text{ON}}$ and $t^{\text{STIM}}_{\text{OFF}}$, respectively. Let $e(t)$ denote the eye position at time t during a trial, then

$\Psi(t) = h - e(t)$ is the retinal location of the visual stimulus at time t. However, $\Psi(t)$ is set to ∞ if there is no visible visual stimulus present at time t. Likewise, if there is no saccade in the trial then s is set to ∞.

Visual Population

Each neuron $1 \leq i \leq N^{\mathrm{V}}$ in the visual population was assigned a unique retinal preference $\alpha_i^{\mathrm{V}} \in [-45°, 45°]$. The firing rate $v_i^{\mathrm{V}}(t)$ of visual neuron i was governed by

$$\begin{aligned} \tau_v^{\mathrm{V}} \frac{dv_i^{\mathrm{V}}}{dt} &= \delta(t - t_{\mathrm{ON}}^{\mathrm{STIM}}) \exp\left(-\frac{(\alpha_i^{\mathrm{V}} - \Psi(t))^2}{2\sigma_{\mathrm{V}}^2}\right) \\ &\quad - \delta(t - t^{\mathrm{SACC}} + \Delta^V) v_i^{\mathrm{V}} \\ &\quad + \delta(t - t^{\mathrm{SACC}} + \Delta^V) \exp\left(-\frac{(\alpha_i^{\mathrm{V}} - \Psi(t))^2}{2\sigma_{\mathrm{V}}^2}\right) \end{aligned} \tag{7.1}$$

The time constant τ_v^{V} was uniform for all neurons in the visual input population. The terms on the right hand side of equation 7.1 were as follows.

The first term on the right hand side of equation 7.1 caused an instantaneous rise in the neuronal firing rate due to the onset of a stimulus. The rise occurred at the time of the stimulus onset, and the magnitude of the rise was a Gaussian function of the distance between the neuron's preferred retinal location α_i^{V} and the location of the stimulus $\Psi(t)$. Hence the magnitude of the rise fell off radially as the stimulus location shifted away from the receptive field center. The standard deviation of the Gaussian tuning curve was uniform across all neurons, and was given by σ_{V}.

The second term on the right hand side of equation 7.1 caused an instantaneous decline in the neuron's firing rate due to the initiation of a saccade, as in Duhamel, Colby & Goldberg (1992). The decline occured at a time Δ^V after the saccade onset. The magnitude of the decline was equal to the current firing rate in order to eliminate the activity of the neuron due to previous stimulation by the presaccadic location of the target.

The third term on the right hand side of equation 7.1 caused an instantaneous rise in firing rate due to the the new retinal location of a visual stimulus after a saccade. The rise occurred at a time Δ^V after the saccade onset, and the magnitude of the rise was a Gaussian function of the distance between the neuron's preferred retinal location α_i^{V} and the postsaccadic location of the stimulus $\Psi(t)$. The firing rate dynamics of the visual neurons described above were purely phenomenological, and should be interpreted as being exogenous to the model.

Saccade Population

Each neuron $1 \leq i \leq N^{\mathrm{S}}$ in the saccade population was assigned a unique saccade preference $\beta_i^{\mathrm{S}} \in [-30°, 30°]$. The firing rate $v_i^{\mathrm{S}}(t)$ of saccade neuron i was governed by

$$\tau_v^{\mathrm{S}} \frac{dv_i^{\mathrm{S}}}{dt} = -v_i^{\mathrm{S}} + I(t - t^{\mathrm{SACC}}, \Delta_{\mathrm{PRE}}^{\mathrm{S}}, \Delta_{\mathrm{POST}}^{\mathrm{S}}) \exp\left(-\frac{(\beta_i^{\mathrm{S}} - s)^2}{2\sigma_{\mathrm{S}}^2}\right) \tag{7.2}$$

where

$$I(x,y,z) = \begin{cases} 1 & \text{if } -y \le x \le z \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

The time constant τ_v^{S} was uniform for all neurons in the saccade population. The indicator function I governed when the saccade neurons were activated with respect to the time of saccade onset t^{SACC}. Specifically, I had a value of 1 whenever t was in the interval $[t^{\mathrm{SACC}} - \Delta_{\mathrm{PRE}}^{\mathrm{S}}, t^{\mathrm{SACC}} + \Delta_{\mathrm{POST}}^{\mathrm{S}}]$, where $\Delta_{\mathrm{PRE}}^{\mathrm{S}}$ and $\Delta_{\mathrm{POST}}^{\mathrm{S}}$ were positive time delays representing how soon in advance and how long after saccade onset the saccade neurons started and stopped responding. During this time interval, the firing rate of the saccade neuron was driven up by a Gaussian function of the difference between the saccade preference of the neuron β_i^{S} and the actual saccade plan s, with a standard deviation of σ_{S}.

Combination Population

The combination neurons were driven dynamically by synaptic inputs from presynaptic visual neurons and saccade neurons. Consequently, for each neuron $1 \le i \le N^{\mathrm{C}}$ in the combination population there were two dynamical quantities defined: an internal activation $h_i^{\mathrm{C}}(t)$ and a firing rate $v_i^{\mathrm{C}}(t)$.

The activation h_i^{C} of combination neuron i was governed by

$$\begin{aligned}\tau_h^{\mathrm{C}} \frac{dh_i^{\mathrm{C}}}{dt} &= -h_i^{\mathrm{C}} \\ &+ \psi^{\mathrm{V}\to\mathrm{C}} \sum_{j=1}^{N^{\mathrm{V}}} w_{ij}^{\mathrm{V}\to\mathrm{C}} v_j^{\mathrm{V}} \\ &+ \psi^{\mathrm{S}\to\mathrm{C}} \sum_{j=1}^{N^{\mathrm{S}}} w_{ij}^{\mathrm{S}\to\mathrm{C}} v_j^{\mathrm{S}} \\ &+ w_{\mathrm{INHB}}^{\mathrm{C}} \sum_{j=1}^{N^{\mathrm{C}}} v_j^{\mathrm{C}} \end{aligned} \tag{7.4}$$

The time constant τ_h^{C} was uniform for all neurons in the combination population. The terms on the right hand side of equation 7.4 were as follows. The first term represented the constant leak in the activation. The second term represented the excitatory synaptic input to the combination neuron from the visual population, where $w_{ij}^{\mathrm{V}\to\mathrm{C}}$ was the weight of the synapse from visual neuron j to combination neuron i. The third term represented the excitatory synaptic input to the combination neuron from the saccade population, where $w_{ij}^{\mathrm{S}\to\mathrm{C}}$ was the weight of the synapse from saccade neuron j to combination neuron i. The fourth term represents inhibitory feedback from the population of combination neurons, where $w_{\mathrm{INHB}}^{\mathrm{C}}$ is a constant scaling parameter. This implements competition between combination neurons and helps to control the overall level of activity within the combination population.

The firing rate v_i^{C} of combination neuron i was a function of the activation and was

given by the sigmoid function

$$v_i^{\mathrm{C}} = \frac{1}{1 + \exp\left(-2\varphi^{\mathrm{C}}(h_i^{\mathrm{C}} - \theta^{\mathrm{C}})\right)} \tag{7.5}$$

where φ^{C} was the sigmoid slope, and θ^{C} was the threshold.

After initially being randomized, synaptic weights were then normalized by ensuring that

$$\sum_j \left(w_{ij}^{\mathrm{V}\to\mathrm{C}}\right)^2 = 1 \tag{7.6}$$

and

$$\sum_j \left(w_{ij}^{\mathrm{S}\to\mathrm{C}}\right)^2 = 1 \tag{7.7}$$

for each postsynaptic combination neuron i (Dayan & Abbott 2001).

Remapping Population

Each neuron $1 \leq i \leq N^{\mathrm{R}}$ in the remapping population was assigned a unique retinal preference $\alpha_i^{\mathrm{R}} \in [-45°, 45°]$. The remapping neurons were driven dynamically by synaptic inputs from presynaptic combination neurons. Consequently, for each remapping neuron i there were two dynamical quantities defined: an internal activation $h_i^{\mathrm{R}}(t)$ and a firing rate $v_i^{\mathrm{R}}(t)$.

The activation h_i^{R} of remapping neuron i was governed by

$$\begin{aligned}\tau_h^{\mathrm{R}} \frac{dh_i^{\mathrm{R}}}{dt} &= -h_i^{\mathrm{R}} \\ &\quad + \psi^{\mathrm{C}\to\mathrm{R}} \sum_{j=1}^{N^{\mathrm{C}}} w_{ij}^{\mathrm{C}\to\mathrm{R}} v_j^{\mathrm{C}} \\ &\quad - w_{\mathrm{INHB}}^{\mathrm{R}} \sum_{j=1}^{N^{\mathrm{R}}} v_j^{\mathrm{R}} \\ &\quad + K_i \end{aligned} \tag{7.8}$$

The time constant τ_h^{R} was uniform for all neurons in the remapping population. The terms on the right hand side of equation 7.8 were as follows. The first term represented the constant leak in the activation. The second term represented the excitatory synaptic input to the remapping neuron from the combination population, where $w_{ij}^{\mathrm{C}\to\mathrm{R}}$ was the weight of the synapse from combination neuron j to the remapping neuron i. The third term represents inhibitory feedback from the population of remapping neurons, where $w_{\mathrm{INHB}}^{\mathrm{R}}$ is a constant scaling parameter. The fourth term K_i represented the external visual input to

remapping neuron i. This was governed by

$$\begin{aligned}\tau^{\mathrm{K}}\frac{dK_i}{dt} &= -K_i \\ &+ \psi^{\mathrm{K}} I(t - t^{\mathrm{STIM}}_{\mathrm{ON}}, -\Gamma^{\mathrm{K}}_i, t^{\mathrm{STIM}}_{\mathrm{OFF}} - t^{\mathrm{STIM}}_{\mathrm{ON}}) \exp\left(-\frac{(\alpha^{\mathrm{R}}_i - \Psi(t))^2}{2\sigma^2_{\mathrm{K}}}\right) \\ &- \delta(t - t^{\mathrm{SACC}} + \Delta^K) K_i \\ &+ P_i \end{aligned} \tag{7.9}$$

The time constant τ^{K} was uniform for all neurons in the remapping population, and was relatively short in all experiments, e.g. 20ms. The terms on the right hand side of equation 7.9 were as follows.

The first term on the right hand side of equation 7.9 represented a constant leak.

The second term on the right hand side of equation 7.9 represented the visual drive due to the presence of a visual stimulus in the visual receptive field of the neuron. The indicator function I, defined by equation 7.3, governed when the remapping neurons were activated with respect to the times of stimulus onset and stimulus offset. Specifically, I had a value of 1 whenever t was in the interval $[t^{\mathrm{STIM}}_{\mathrm{ON}} + \Gamma^{\mathrm{K}}_i, t^{\mathrm{STIM}}_{\mathrm{OFF}}]$, where Γ^{K}_i represented a positive onset delay for remapping neuron i. The values of Γ^{K}_i were randomly drawn from $\mathcal{N}(0ms, 50ms)$, with negative values flipped to positive and all values above 80ms clipped to this limit. During this time interval, K_i was driven up by a Gaussian function of the distance between the remapping neuron's preferred retinal location α^{R}_i and the location of

the stimulus $\Psi(t)$, with a standard deviation of σ_{V}.

The third term on the right hand side of equation 7.9 implemented an instantaneous truncation due to the initiation of a saccade, which was effected a short time Δ^K after the saccade onset, as in Duhamel, Colby & Goldberg (1992). Such saccade aligned activity truncation has also been implemented in other models such as those of Quaia et al. (1998) and Xing & Andersen (2000*b*). The magnitude of the decline in K_i was equal to the present value of K_i in order to reduce this variable to a baseline value of zero.

The fourth term, P_i, on the right hand side of equation 7.9 was a tonic driving input specific to each remapping neuron i, and its dynamics were governed by

$$\begin{aligned} \tau^{\mathrm{P}} \frac{dP_i}{dt} = & -P_i \\ & + \delta(t - t^{\mathrm{STIM}}_{\mathrm{OFF}}) K_i \\ & - \delta(t - t^{\mathrm{SACC}} + \Delta^K) P_i \end{aligned} \tag{7.10}$$

The time constant τ^{P} was relatively long in all experiments, e.g. 300ms. P_i provided the driving input to K_i upon stimulus offset, giving the i^{th} remapping neuron a slow prolonged trace response when the stimulus was removed as in Duhamel, Colby & Goldberg (1992). The terms on the right hand side of equation 7.10 were as follows. The first term represented a constant leak. The second term caused an instantaneous rise in P_i upon stimulus offset. The magnitude of this rise was equal to K_i. This resulted in K_i, which was driven by P_i, having a long trace response to a visual stimulus after its removal. The third term effected

an instantaneous truncation of P_i a short time Δ^K after the saccade onset, as for K_i.

The dynamics of the driving input signals K_i and P_i described by equations 7.9 and 7.10 were purely phenomenological, designed through a process of experimentation to produce neuronal responses in the remapping population that matched the observed neuronal responses in LIP.

The firing rate v_i^{R} of remapping neuron i was a function of the activation and was given by the sigmoid function

$$v_i^{\mathrm{R}} = \frac{1}{1 + \exp\left(-2\varphi^{\mathrm{R}}(h_i^{\mathrm{R}} - \theta^{\mathrm{R}})\right)} \tag{7.11}$$

where φ^{R} was the sigmoid slope, and θ^{R} was the threshold.

During training, each synaptic weight $w_{ij}^{\mathrm{C}\to\mathrm{R}}$ was modified according to a Hebbian learning rule

$$\frac{dw_{ij}^{\mathrm{C}\to\mathrm{R}}}{dt} = \varrho^{\mathrm{C}\to\mathrm{R}} v_i^{\mathrm{R}} v_j^{\mathrm{C}} \tag{7.12}$$

where $\varrho^{\mathrm{C}\to\mathrm{R}}$ was the learning rate. Unbounded growth of the synaptic weights during training was prevented by imposing the constraint

$$\sum_j \left(w_{ij}^{\mathrm{C}\to\mathrm{R}}\right)^2 = 1 \tag{7.13}$$

for each postsynaptic remapping neuron i after each weight update (Dayan & Abbott 2001). Experimental evidence for renormalisation of synaptic weights in the brain has

been provided by Royer & Paré (2003).

7.7 Numerical Simulation

The system of differential equations were integrated numerically using the Forward-Euler scheme, where the numerical time step was set to one tenth of the smallest neuronal time constant among $\tau^{\mathrm{V}}, \tau^{\mathrm{S}}, \tau^{\mathrm{C}}, \tau^{\mathrm{R}}, \tau^{\mathrm{K}}$ and τ^{P}. All input stimuli was dynamically simulated and sampled at the same frequency as the time step.

7.8 Analysis

Neuronal Period Response

The *period response* of a neuron was an analogue to the average spike count rate over some time interval as measured in single unit recording neurophysiology studies. Specifically, let $v(t)$ be the firing rate of a neuron for $t \in [0, T]$ during a task trial. During the time period $[t_1, t_2]$, the period response of the neuron was defined as

$$\bar{v} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} v(t) dt \tag{7.14}$$

This integral was numerically integrated using the trapezoidal method.

Neuronal Response Latency

The *response latency* of a neuron, that is the earliest time during the trial at which point the neuron is considered to have a response in its discharge, was defined as the time of the start of the first 30ms window where the slope of the response curve was consistently above a minimal threshold (0.002). If no such time was found, then the neuron was considered unresponsive in the given trial.

Remapping Index

The *remapping index* of a remapping neuron was a measure of the strength of the activity of the neuron which could be attributed to remapping. For each single step task trial i, in which activity should be remapped by a saccade s_i into the retinal receptive field location r_i, a remapping index is computed by first computing a visual index and a saccade index. These two indices measure how much of the neuronal activity in the single step task trial can be attributed to remapping when controlling for either purely visual or purely saccadic activity, respectively, as in Berman et al. (2007), Dunn et al. (2010), Heiser et al. (2005), Heiser & Colby (2006). For both indices, the remapping activity was defined as the period response in a 300ms saccade onset aligned epoch in the single step task.

The visual index was defined as the remapping activity in the single step task trial minus the period response from a corresponding stimulus control task trial, in which a visual stimulus was briefly flashed while fixation was maintained throughout the trial in

the absence of a saccade. In particular, the visual stimulus was flashed at the presaccadic retinal location of the stimulus in the training trial that set up remapping into the retinal receptive field location r_i. The period response from the stimulus control trial was computed over a 300ms interval starting 200ms after stimulus onset.

The saccade index was defined as the remapping activity in the single step task trial minus the period response from a corresponding saccade control task trial, in which the saccade s_i was performed with no visual stimulus present. The saccade control task performed the saccade s_i of the training trial that set up remapping into receptive field location r_i. The period response from the saccade control trial was computed over a $300ms$ interval aligned on the saccade onset.

Both of these indices were confined to $[-1, 1]$.If either index was negative, then the remapping index was set to zero, otherwise the remapping index was the norm of a vector of the two indices, confining it to $[0, \sqrt{2}]$. A value of 0 indicated that no remapping activity can be attributed beyond either purely visual or purely saccadic activity. While, conversely, a value of $\sqrt{2}$ indicated that there was a maximal remapping response and that nothing could be attributed to purely visual or purely saccadic activity.

Probe Task Decoding

For the purpose of understanding the functional structure of synaptic connectivity between the visual and saccade populations and the combination population, and also the connec-

tivity between the combination population and the remapping population, it was necessary to decode the combination of retinal stimulus location and saccade to which a combination neuron was selective. To do this, we analysed the period response of a combination neuron to all of the probe task trials. For the i^{th} probe task trial, let r_i be the retinal stimulus location at the start of the trial and let s_i be the saccade. Also, let $\bar{v}_i$ be the period response of the given combination neuron over the 50ms epoch aligned at saccade initiation time for the i^{th} probe task trial. The selectivity of the given neuron in each of the two input spaces, encoded by the visual population and saccade population, was decoded by performing a center-of-mass calculation in the given space across all task trials. Hence the retinal selectivity of a given combination neuron was decoded as

$$\frac{\sum_i r_i \bar{v}_i}{\sum_i \bar{v}_i} \tag{7.15}$$

where the summation is carried out over all task trials i. Likewise, the saccade selectivity was decoded as

$$\frac{\sum_i s_i \bar{v}_i}{\sum_i \bar{v}_i} \tag{7.16}$$

Chapter 8

Perisaccadic Receptive Field Dynamics: Results

8.1 Introduction

This chapter presents simulation results from the neural network model, described in the previous chapter, of a set of perisaccadic receptive field dynamics. The basic operation of the model is established first in a manually prewired model, demonstrating the ideal computational mechanisms which the learning of the synaptic connectivity is intended to develop. This experiment also serves as a baseline to validate the output data decoding procedures. Next, the performance of the model with learning is investigated, and it is found to account for the major experimental findings in area LIP in Duhamel, Colby & Goldberg (1992). The model is then retested to investigate the temporal characteristics of remapping, and is found to be in agreement with Kusunoki & Goldberg (2003) and Nakamura & Colby (2002). The spatial characteristics of remapping are investigated by retraining the model to be able to remap multiple retinal locations into a given receptive

field, and the model performance is found to be in agreement with Heiser & Colby (2006).

8.2 Results

8.2.1 Hardwired Model: Replicating the observations of Duhamel, Colby & Goldberg (1992)

This experiment explored how a network with hardwired synaptic connections could exhibit the same perisaccadic receptive field dynamics as those listed in section 7.3, and found in areas LIP (Duhamel, Colby & Goldberg 1992), SC (Walker et al. 1995), FEF (Umeno & Goldberg 1997, 2001) and visual areas V3,V3A,V2 and V1 (Nakamura & Colby 2002). The experiment with the prewired model had two purposes as follows. First, the experiment was used test the hypothesis for how the model with learning should function after learning. Secondly, the experiment was used to test and validate the procedures for decoding neuronal responses described in section 7.8 by comparing the values of the decoded responses recorded during the operation of the model with the desired neuronal responses used to guide the prewiring of the model. The parameters for the model can be found in table 8.1.

The population of combination neurons corresponded exactly to the set of all possible combinations of visual input neuron preference and saccade input neuron preference. Hence $N^{\mathrm{C}} = N^{\mathrm{V}} \times N^{\mathrm{S}} = 91 \times 61 = 5551$. To elaborate, the i^{th} combination neuron was assigned

the same retinal location α_i^{C} as its corresponding visual neuron, and the same saccade β_i^{C} as its corresponding saccade neuron. The synaptic weight vectors between this combination neuron and the visual and saccade input populations were then constructed to make the combination neuron selective to its assigned retinal location and saccade. Specifically, the weight vector of combination neuron i with the visual input population was given by

$$w_{ij}^{\mathrm{V}\to\mathrm{C}} = \exp\left(-\frac{\left\|\alpha_j^{\mathrm{V}} - \alpha_i^{\mathrm{C}}\right\|^2}{2\sigma_{\mathrm{W}}^2}\right) \tag{8.1}$$

where σ_{W} governs the spread of the prewired weights. Likewise the weight vector with the saccade input population was given by

$$w_{ij}^{\mathrm{S}\to\mathrm{C}} = \exp\left(-\frac{\left\|\beta_j^{\mathrm{S}} - \beta_i^{\mathrm{C}}\right\|^2}{2\sigma_{\mathrm{W}}^2}\right) \tag{8.2}$$

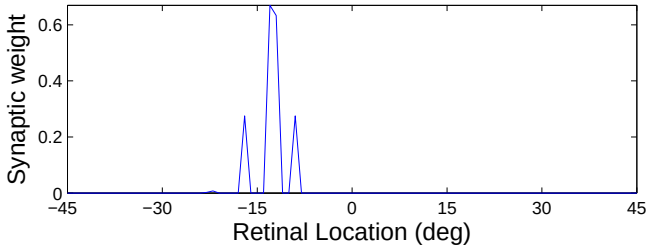
The weight vectors of the remapping neurons with the combination population were set to encode all possible remappings, whereby each saccade leads to a shift in the retinal location of a visual stimulus. Thus, the weight vector of remapping neuron i with the combination population was given by

$$w_{ij}^{\mathrm{C}\to\mathrm{R}} = \exp\left(-\frac{\left\|\alpha_j^{\mathrm{R}} - \left(\alpha_i^{\mathrm{C}} - \beta_i^{\mathrm{C}}\right)\right\|^2}{2\sigma_{\mathrm{W}}^2}\right) \tag{8.3}$$

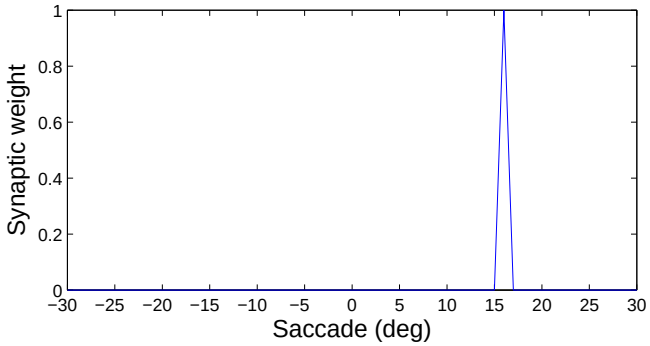
Lastly, each weight vector was subsequently diluted to the correct level as prescribed by ϕ^{V}, ϕ^{S} and ϕ^{C} respectively, and finally normalized according to equations 7.6, 7.7 and 7.13.

Figure 8.2 shows the synaptic connectivity in the prewired network model. Fig. 8.2a and Fig. 8.2b show the weight vector of combination neuron #2000 with the visual and saccade input populations, respectively. This combination neuron was tuned approximately to retinal stimulus location $-16°$ and saccade $16°$. Fig. 8.2c shows the synaptic weights from the same combination neuron to all remapping neurons. The combination neuron had strongest connections to remapping neurons with retinal preferences that were approximately $-31°$. These feedforward synaptic connections between successive layers of the network were appropriate to produce approximately the correct remapping for a presaccadic combination of retinal location $-16°$ and saccade $16°$, which would bring the stimulus to retinal location $-32°$. Fig. 8.2d examines the combinations of retinal and saccade inputs that drive the remapping neuron with preferred retinal location $-30°$. It is clear that this remapping neuron has the strongest connections from the subset of combination neurons that represent combinations of presaccadic retinal location and saccade which will bring the stimulus into the remapping neuron's receptive field. These combinations of visual and saccade inputs should, therefore, produce remapping activity in the given remapping neuron, as desired.

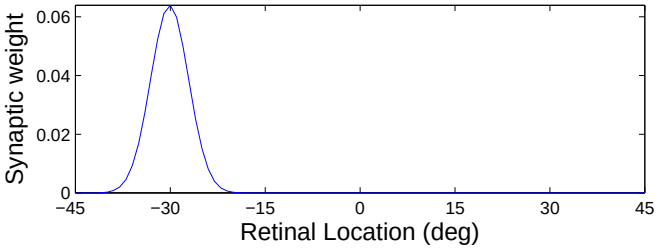
The prewired network was tested on the stimulus control task, saccade control task, probe task and single step task. For comparison, baseline simulations were also run with the same model architecture and parameters, but with randomised synaptic weights throughout the network. The inputs used for the single step task are shown in figure 8.3.



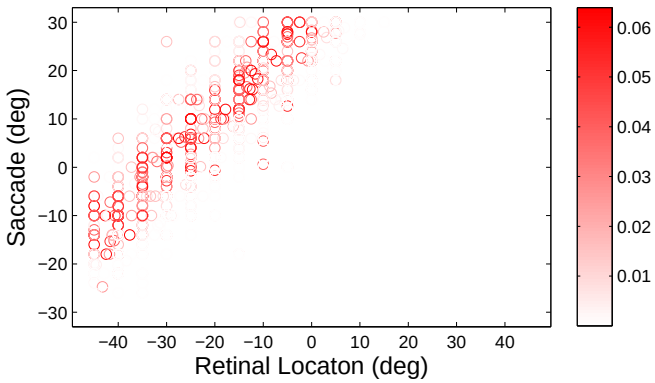
(a) $w_{i*}^{\mathrm{V}\to\mathrm{C}}$ vector of combination neuron #2000



(b) $w_{i*}^{\mathrm{S}\to\mathrm{C}}$ vector of combination neuron #2000



(c) $w_{*j}^{\mathrm{C}\to\mathrm{R}}$ vector of combination neuron #2000



(d) $w_{i*}^{\mathrm{C}\to\mathrm{R}}$ vector of remapping neuron $-30°$

Figure 8.2: Synaptic connectivity in prewired neural network model. Plots (a) and (b) show the afferent synaptic weights to combination neuron #2000 from the visual and saccade input neurons, respectively. The visual and saccade neurons are ordered along the abscissa of these two plots according to their respective retinal and saccade preferences. Plot (c) shows the synaptic weights of all efferent synaptic connections from the same combination neuron onto the remapping population, where the remapping neurons are ordered along the abscissa according to their retinal response preferences. Observe that the latter plot is smoother than the former two plots because there is no dilution in the connectivity from combination neurons to remapping neurons. Plot (d) shows the visual and saccade inputs that drive the remapping neuron with retinal preference $-30°$, which has maximum weight in plot (c). Specifically, plot (d) is created by tracing back the weighted connections from this remapping neuron to the preceding layer of combination neurons, and then computing the retinal and saccade preferences of each combination neuron using the probe task decoding procedure described in section 7.8. A scatterplot of these decoded preferences is shown for all combination neurons, where the boldness of each data point indicates the strength of the connection from that particular combination neuron to the remapping neuron. It can be seen that the remapping neuron receives strong inputs from combinations of retinal stimulus location and saccade that lie along the diagonal region which would be mapped onto the retinal preference $-30°$ of the remapping neuron.

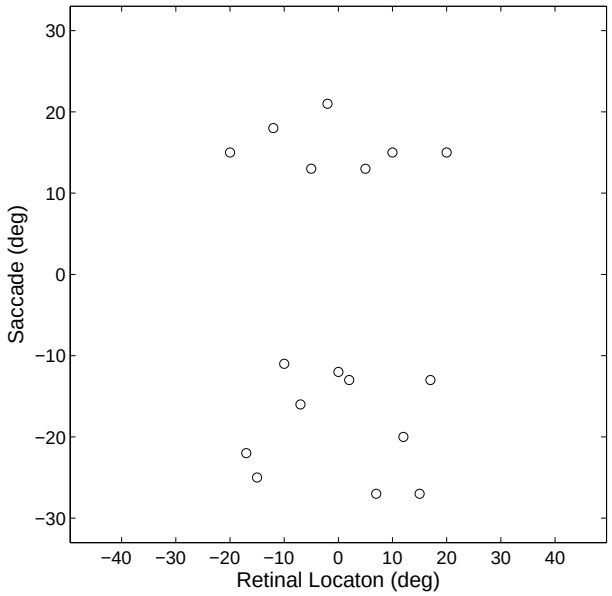
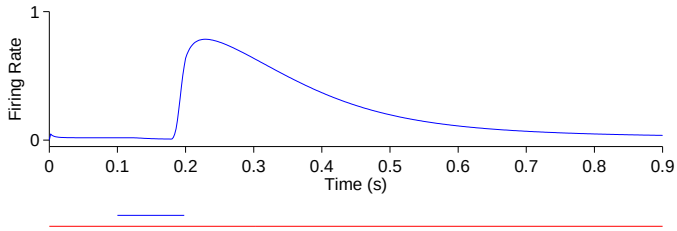


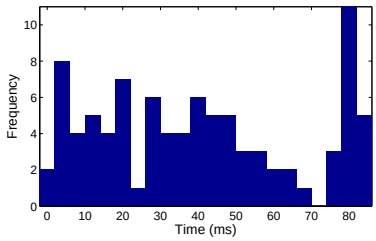
Figure 8.3: Plot representing the inputs used for all of the single step task trials. Each data point corresponds to the presaccadic retinal location of the flashed stimuli (abscissa) and the saccade performed (ordinate) in one of the trials. The $20°$ central portion of the saccade space, for which there are no trials, reflects the fact that all saccades were at least $10°$ in magnitude.

Figure 8.4 shows the responses of a typical remapping neuron, with receptive field at $-20°$, from the prewired network model. Fig. 8.4a shows the response of the remapping neuron during a stimulus control task trial, in which a stimulus is flashed in the neuron's receptive field while constant fixation is maintained in the absence of any saccade. The neuronal response exhibited two prototypical LIP characteristics (Duhamel, Colby & Goldberg 1992). First, the response is delayed significantly with respect to the stimulus onset, with a latency of approximately 80ms. Across different remapping neurons, this response latency varied considerably, and in fig. 8.4b one can observe the distribution of response latencies found in the remapping population. Secondly, after stimulus offset at 200ms, figure 8.4a shows that there was a long trace of decaying response for several hundred milliseconds. Fig. 8.4c shows the response of the same remapping neuron in a delayed stimulus task trial, where the stimulus was flashed from 300ms to 400ms, and a saccade was subsequently initiated at 600ms. This response exhibited a third important phenomenon, namely that the onset of a saccade after the brief flashing of a stimulus in the neuron's receptive field caused a truncation in the trace of neuronal activity, as has been observed in LIP (Duhamel, Colby & Goldberg 1992, Kusunoki & Goldberg 2003). In summary, the remapping neurons exhibited visual onset variability, offset trace and saccade onset truncation, as found among LIP neurons.

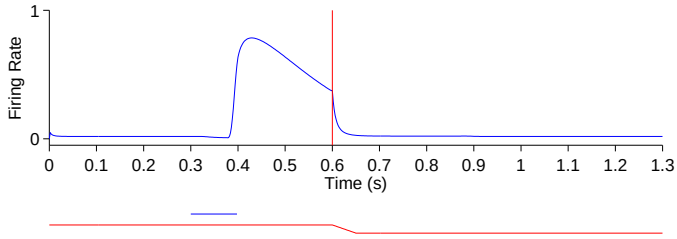
Figure 8.5 shows the remapping behaviour of a remapping neuron with retinal preference $-20°$, in the prewired network. The remapping behaviour was tested in a single



(a) Stimulus control trial



(b) Stimulus onset latency distribution



(c) Delayed stimulus trial

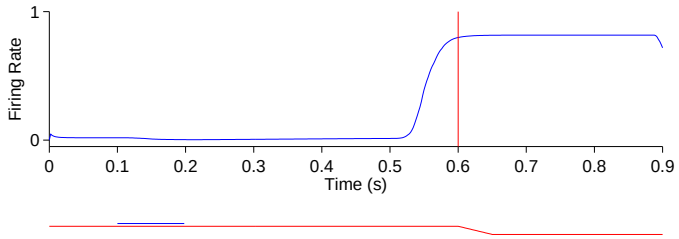
Figure 8.4: The response properties of a remapping neuron, with assigned retinal preference $-20°$, in the prewired network. Results are shown for both a stimulus control task in which a stimulus is flashed in the neuron's receptive field and no saccade is performed, and a delayed stimulus task in which a saccade is performed a short time after the stimulus has been flashed in the neuron's receptive field. Plot (a) shows the response of the remapping neuron to the stimulus control task trial, in which the stimulus was presented directly in the neuron's receptive field. The two traces below the abscissa indicate the eye position (red) and stimulus presence (blue). Hence it can be seen that fixation was kept throughout the trial and the stimulus was flashed from 100ms to 200ms. The remapping neuron does not begin to respond to the stimulus until about 80ms after the stimulus onset. Plot (b) shows the distribution of the response latencies of all remapping neurons as determined from their corresponding stimulus control task trial. Plot (c) shows the response of the same remapping neuron that was plotted in (a) in a delayed stimulus task trial, in which a stimulus was presaccadically flashed in the neuron's receptive field in advance of a saccade. The vertical red line shows the moment of saccade onset. The response of the remapping neuron is sharply truncated from this moment onwards.

step task trial (Fig. 8.5a), where the stimulus was located presaccadically at retinal location $-5°$, and a saccade of $15°$ was initiated at 600ms into the trial. Two important observations can be made by looking at the response of the given remapping neuron in the single step task trial. Firstly, despite the fact that the stimulus was never flashed in the neuron's classically defined receptive field, the neuron still responded around the time of (i.e. a few hundred milliseconds before) the saccade. However, the remapping neuron did not respond during a stimulus control task trial, in which the stimulus was flashed in a retinal location distant from the neuron's receptive field in the absence of a saccade (Fig. 8.5b). Nor did the remapping neuron respond during a saccade control task trial, in which the saccade was executed in the absence of a visual stimulus (Fig. 8.5c). These three task trials confirmed that both the stimulus and the saccade are required to trigger a response in the remapping neuron. In particular, in order for the remapping neuron to respond, the saccade had to bring the location of the recently extinguished stimulus into the receptive field of the neuron. Such a saccade could trigger a response even if the stimulus had been extinguished 400ms prior to saccade initiation. Secondly, the remapping neuron began responding well in advance of the initiation of the saccade itself (Fig. 8.5a), and the response latency with respect to saccade onset does not even take into account that the saccade must be completed before the stimulus is brought into the receptive field of the neuron. This was despite the fact that in a stimulus control task the same neuron had a reponse latency of approximately 80ms with respect to pure fixation (Fig. 8.4a).

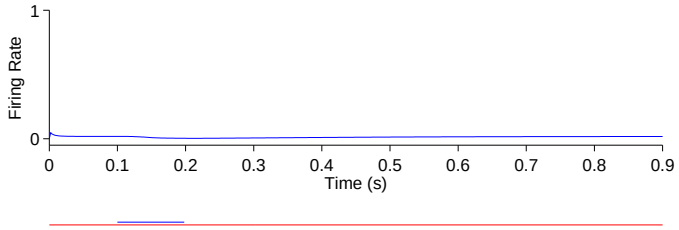
Both the presaccadic and trace responses have been observed in multiple oculomotor areas such as LIP (Duhamel, Colby & Goldberg 1992), SC (Walker et al. 1995), FEF (Umeno & Goldberg 1997, 2001) and visual areas V3,V3A,V2 and V1 (Nakamura & Colby 2002). In summary, the model successfully reproduces both presaccadic and trace responses among remapping neurons.

Figure 8.6 shows the population responses of neurons in the prewired model during the same flashed stimulus single step task trial as before. Activity in all neuronal populations was absent prior to stimulus onset at 100ms. However, at this point activity begins immediately in the visual input population corresponding to the retinal location of the visual target. Remapping neurons with similar retinal preferences started responding with various delays after this time. At 500ms the saccade population started encoding the 15° impending saccade. It was not long after this that activity in the remapping population corresponding to −20° began to increase, and this happened before saccade onset. Activity among the remapping neurons encoding the old retinal location was also truncated quickly, i.e. well before the saccade onset. Activity in all three populations was in equilibrium until around 900ms into the trial when both input populations began to go silent, and as a result so did the remapping population.

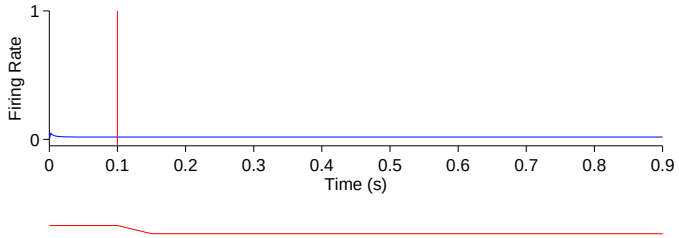
The remapping behaviour of 17 randomly selected remapping neurons in single step task trials was analysed for both the manually prewired network and randomly connected network. The results are shown in figure 8.7 and key population metrics are summa-



(a) Single step task trial



(b) Stimulus control trial



(c) Saccade control trial

Figure 8.5: The remapping behaviour of a remapping neuron with retinal preference $-20°$, in the prewired network. Plot (a) shows a single step task trial in which a stimulus is flashed from 100ms to 200ms, and a saccade is subsequently performed with onset at 600ms. The saccade brings the stimulus into the receptive field of the remapping neuron, which shows a remapping response that begins before the saccade onset - a presaccadic remapping response. Plot (b) shows the response of the same remapping neuron in a stimulus control task trial, in which a stimulus is flashed from 100ms to 200ms, but there is no saccade. The retinal location of the stimulus is the same as the presaccadic location of the stimulus in the single step task trial. In this case the remapping neuron does not respond. Plot (c) shows the response of the same remapping neuron in a saccade control task trial, in which there is no stimulus present and a saccade is performed with onset at 100ms. The saccade performed is the same as that performed in the single step task trial. Again, the remapping neuron shows no response. These three different kinds of task trial confirm that both the presence of the stimulus and the saccade are needed to trigger a response in the remapping neuron.

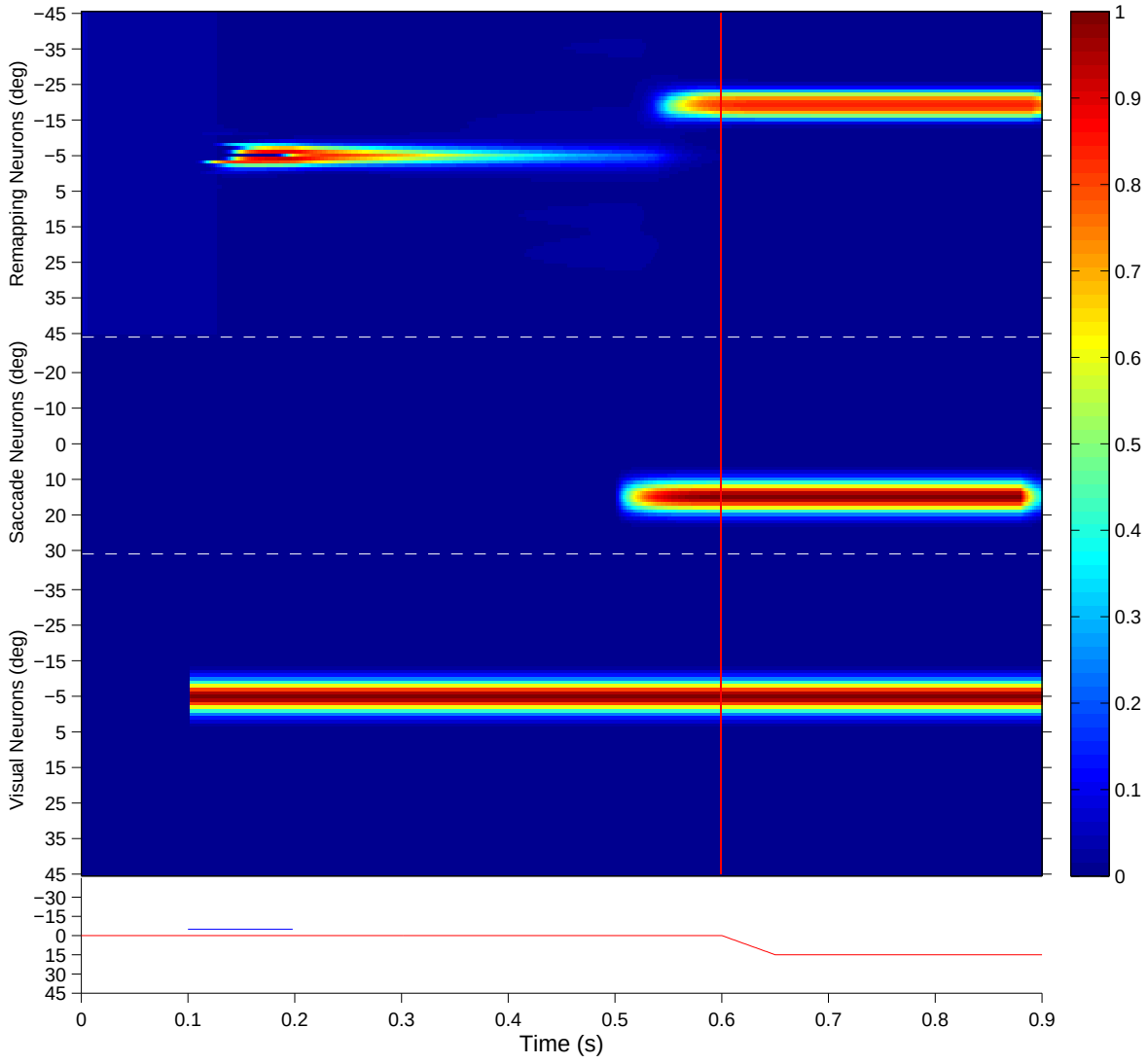
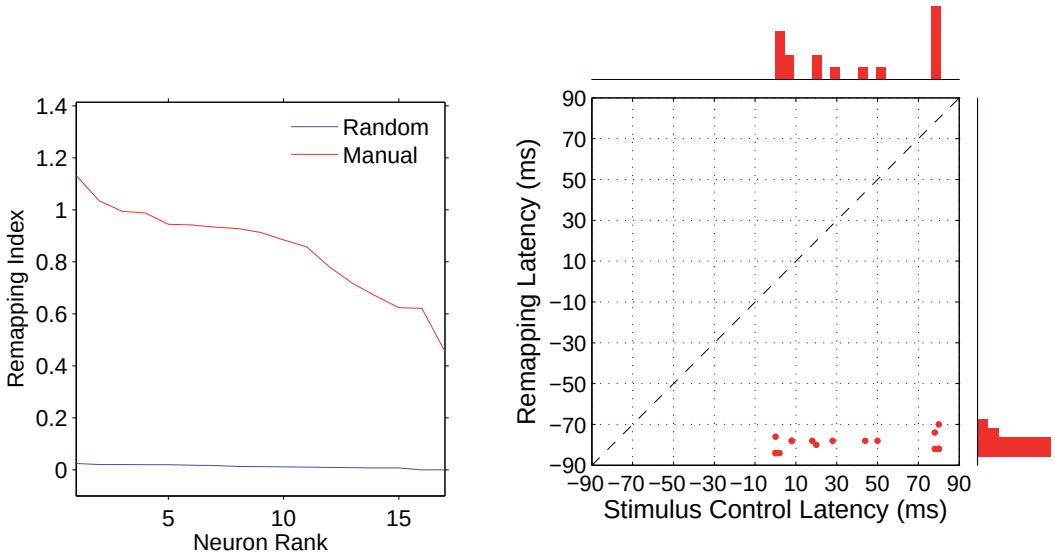


Figure 8.6: The population responses of neurons in the prewired network during a single step task trial for the remapping neuron with retinal preference $-20°$. The figure shows the responses of all neurons in the remapping population (top), saccade population (middle), and visual population (bottom). Each population subplot arranges the neurons topographically along the ordinate in terms of their assigned retinal or saccade preference, α_i and β_i, respectively. The traces at the bottom show the eye position (red) and retinal location of the stimulus (blue) for the trial, and the red vertical bar designates the saccade onset. It is evident that remapping neurons with preferences for retinal locations near $-20°$ show presaccadic remapping in anticipation of the visual stimulus shifting into their receptive fields after the saccade.

rized in table 8.2. Two key response properties were analysed in Fig. 8.7, the strength of remapping and the latency of remapping. Fig. 8.7a shows the remapping index of the remapping neurons in both the manually prewired and randomly connected networks, with the remapping neurons ranked according to the value of their remapping index. It was clear that all remapping indices were lower in the random, compared to the manually prewired model. The remapping indices were near zero for all remapping neurons in the randomly connected network, with an average value of 0.0131. However, the average remapping index increased dramatically to 0.848 for the manually prewired network. Fig.8.7b compared the remapping latency with the stimulus control latency for remapping neurons in the manually prewired model. The remapping latency was assessed for each remapping neuron by computing the neuronal response latency with respect to saccade onset in a single step task trial. While the stimulus control latency was assessed by computing the neuronal response latency with respect to stimulus onset in a stimulus control task trial. No remapping latencies could be decoded for remapping neurons in the randomly wired model. However, in the manually prewired model all neurons were found to remap both predictively and presaccadically. In summary, remapping activity was present only in the manually prewired model, and a range of response latencies were found with presaccadic remapping being represented with high frequency.

Lastly, it was investigated whether the decoded retinal and saccade preferences for the neurons in the combination population corresponded to the preassigned values α_i^{C} and β_i^{C}



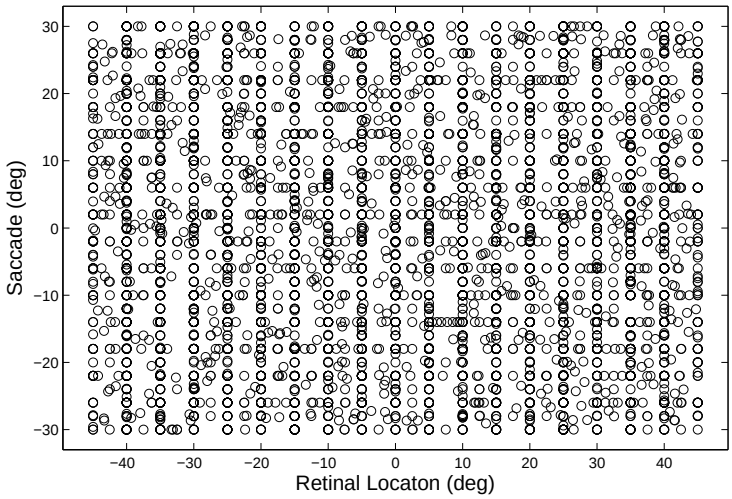
(a) **(b)**

Figure 8.7: Population analysis of the remapping responses of 17 randomly selected remapping neurons from the manually prewired network model. Each remapping neuron is individually tested on a single step task trial that brings the visual stimulus into the receptive field of the neuron, as well as a corresponding stimulus control task trial and saccade control task trial. The results of these three different task trials are then used to compute the remapping index for that neuron as described in section 7.8. The left plot shows the remapping index for each remapping neuron, with the neurons plotted in rank order according to the value of their remapping index. Results are shown for the manually prewired network (red) and the network with randomised synaptic weights (blue). It can be seen that remapping neurons in the manually prewired network have much larger remapping indices than remapping neurons in the network with a random synaptic connectivity. The right plot is a scatterplot in which each data point corresponds to one of the remapping neurons in the manually prewired network, with each data point showing the remapping latency and stimulus control latency for that neuron. The remapping latency is measured as the neuronal response latency, the calculation of which is described in section 7.8, between the saccade onset and the neuron's response in the single step task trial for that neuron. The stimulus control latency is measured as the neuronal response latency between the stimulus onset and the neuron's response in the stimulus control task trial for that neuron. Each data point below the unit diagonal (dashed line) had a longer stimulus control latency than remapping latency, implying that the neuron performed predictive remapping. Furthermore, since all neurons had a negative remapping latency, they were in fact all performing presaccadic remapping as well.

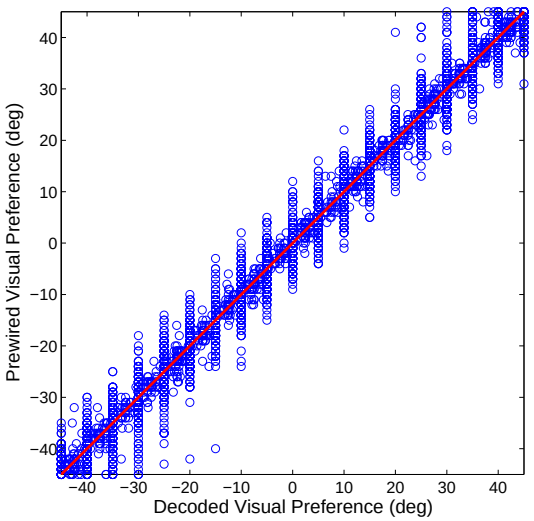
used to prewire the afferent synaptic connections to these neurons. For each combination neuron, the responses from the probe task were decoded as described in section 7.8. This yielded a decoded value for the retinal and saccade preference of each combination neuron. Figure 8.8a shows the set of decoded retinal and saccade preferences of all neurons from the combination population. The grid like structure in the input space is in broad agreement with the prewiring of the network. To verify the decoding at the single neuron level, a correlation analysis was performed in which the decoded retinal and saccade preferences for each combination neuron were compared to the preassigned preferences α_i^{C} and β_i^{C} respectively. Figure 8.8a,b shows two scatter plots, one for each preference, which confirms that there was substantial agreement, indeed near perfect correlation in both cases, between the decoded and the preassigned preferences. This confirmed that the decoding method used was operating properly.

8.2.2 Model with Learning: Replicating the observations of Duhamel, Colby & Goldberg (1992)

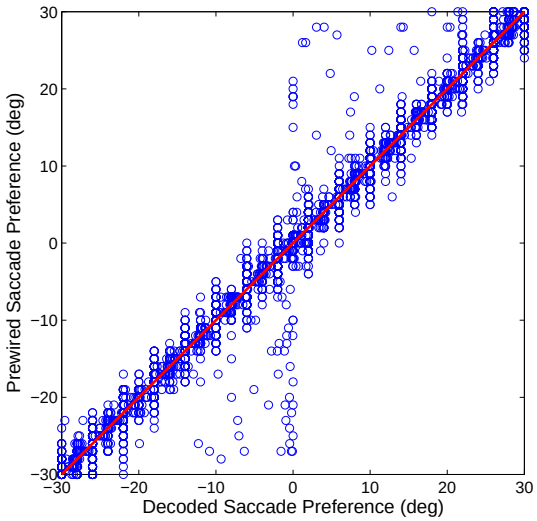
Unlike the previous model, the model explored here modified the weights from the combination to the remapping population through the unsupervised learning mechanism described in the previous chapter. The model was trained on the same set of single step task trials that were used in the previous experiment (Fig. 8.3), and hence was also tested on the same set of task trials. The new parameter values for this model can be found in table 8.3,



(a) Decoded preferences in combination population



(b) Retinal preference decoding



(c) Saccade preference decoding

Figure 8.8: Decoded response preferences of the entire population of 5551 combination neurons in the manually prewired network. Plot (a) shows the retinal and saccade preferences of each combination neuron computed using the probe task decoding procedure described in section 7.8. Each data point in the scatterplot gives the decoded preferences for one of the combination neurons. The scatterplot (b) shows the relationship between the decoded retinal preferences of the combination neurons and the preassigned retinal preferences α_i^{C} used to prewire the afferent synaptic connections to these neurons. It can be seen that there is a very good correspondence between the preassigned and decoded retinal preferences, with the data points clustered along the unit diagonal (red line). Similarly, plot (c) shows the relationship between the decoded saccade preferences of the combination neurons and the preassigned saccade preferences β_i^{C} used to prewire the synaptic connections. Again, there is a very good correspondence between the preassigned and decoded saccade preferences.

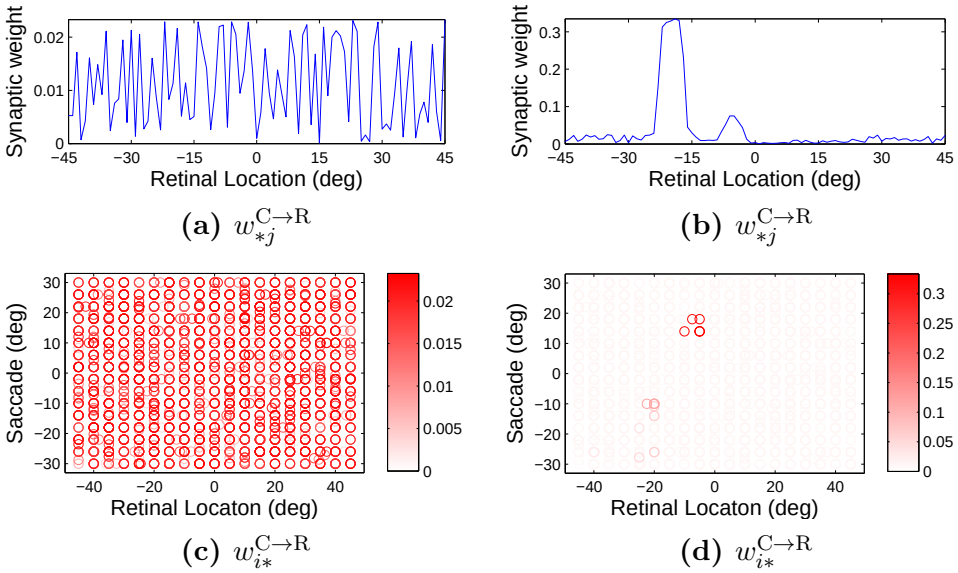
with the remaining parameters as given in Table 8.1.

Figure 8.9 shows examples of synaptic connectivity before (left column) and after (right column) training. The combination neuron #181 projected to all neurons in the remapping population, i.e. there was full connectivity, and prior to training there was no structure in these efferents (Fig. 8.9a). However, after training there was a clear topographic structure in which connections were potentiated to remapping neurons with a retinal preference close to either $-5°$ or $-20°$ (Fig.8.9b). The presence of strong connections to both of these retinal locations in the remapping population could be explained by the time course of the training trials and the preferences of this combination neuron. During training, the given combination neuron would have responded to its preferred combination of pre-saccadic stimulus retinal location and saccade. Moreover, the combination neuron would have started to respond in advance of the saccade due to the pre-saccadic response of the saccade population. Consequently, the efferent connections from the given combination neuron to remapping neurons representing the pre-saccadic stimulus location $-5°$ would be strengthened by associative learning during the period of the pre-saccadic response of the saccade neurons, which was $\Delta^{\mathrm{S}}_{\mathrm{PRE}} = 70\mathrm{ms}$ in this experiment. After the completion of the saccade, the stimulus is relocated to a new post-saccadic retinal location, that is $-20°$. Meanwhile, the post-saccadic latency $\Delta^{\mathrm{V}} = 280ms$ in the visual population and the post-saccadic latency $\Delta^{\mathrm{S}}_{\mathrm{POST}} = 300\mathrm{ms}$ in the saccade population will keep the given combination neuron active while both of these inputs remain simultaneously active af-

ter the saccade. Consequently, the efferent connections from this combination neuron to remapping neurons representing the post-saccadic stimulus location $-20°$ will be potentiated through associative learning up to $\Delta^{\mathrm{V}} = 280ms$ after saccade onset. Notice that this relatively long duration of post-saccadic associative learning, relative to the period of pre-saccadic associative learning, explains why significantly more synaptic weight is devoted to remapping neurons representing the post-saccadic location.

The weight vector of a remapping neuron with retinal preference $-20°$ shows that before training it receives connections from combination neurons with a wide range of preferences (Fig.8.9c). However, after training, the presynaptic combination neurons with strong connections to the given remapping neuron all have very similar preferences to combination neuron #181 (Fig. 8.9d). Also notice that, unlike in the manually prewired model, the given remapping neuron does not receive connections from a full unit diagonal in the combined space of saccade and retinal location, just a localised region. This is because the connections in the prewired model were set up to facilitate remapping from all configurations of retinal location and saccade which brought a stimulus into the remapping neuron's receptive field, while this model was only trained on a small subset of possible configurations.

The model displayed the same onset and offset characteristics of remapping neurons in stimulus control tasks and delayed stimulus tasks as presented for the manually prewired model in Fig. 8.4.



(a) $w_{*j}^{\mathrm{C}\to\mathrm{R}}$ **(b)** $w_{*j}^{\mathrm{C}\to\mathrm{R}}$

(c) $w_{i*}^{\mathrm{C}\to\mathrm{R}}$ **(d)** $w_{i*}^{\mathrm{C}\to\mathrm{R}}$

Figure 8.9: Synaptic connectivity in the model before (left column) and after training (right column). Conventions as for Fig. 8.2. First row shows the synaptic weights of all efferent synaptic connections from the same combination neuron onto the remapping population. Second row shows the visual and saccade inputs that drive the remapping neuron with retinal preference $-20°$, which was computed as described for Fig. 8.2.

Figure 8.10 shows the remapping behaviour of a remapping neuron with retinal preference $-20°$ after training. As for the manually prewired model, it is informative to analyse the remapping behaviour of the remapping neuron by inspecting its corresponding single step task trial (Fig. 8.10a), stimulus control task trial (Fig. 8.10b) and saccade control task trial (Fig. 8.10c). Recall that the latter two task trials isolate either the visual or saccadic component of the single step task trial to which they correspond, and that discharge in either of these two task trials indicates that discharge in the single step task trial cannot be solely attributed to remapping.

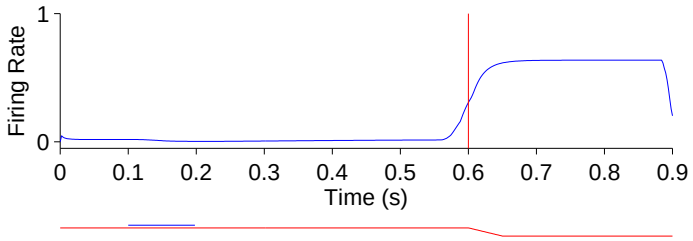
In the single step task trial, the visual stimulus was briefly flashed presaccadically for 100ms at head-centered, and therefore also retinal, location $-5°$. Exactly 400ms after

stimulus offset a saccade of 15° was initiated, bringing the given remapping neuron, with a retinal receptive field at −20°, over the head-centered location of the extinguished stimulus. Notice that the stimulus was never presented close to the classical receptive field location of the given remapping neuron.

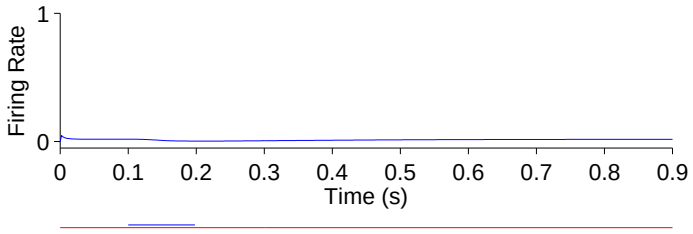
Fig. 8.10a shows that the given remapping neuron did indeed respond during the single step task trial despite the visual stimulus never being present in the receptive field of the neuron. However, when looking at the corresponding stimulus control task trial where the stimulus was flashed in retinal location −5° (Fig. 8.10b), and the saccade control task trial where the saccade of 15° was executed (Fig. 8.10c), there was no discharge. These results confirm that the given remapping neuron displayed genuine remapping activity during the single step task trial. Moreover, the response in the remapping neuron began well in advance of the saccade onset. Hence this negative onset latency shows that the neuron performed predictive remapping.

Figure 8.11 shows the population responses of neurons during the same flashed stimulus single step task trial. The results are very similar to those shown above for the prewired network model.

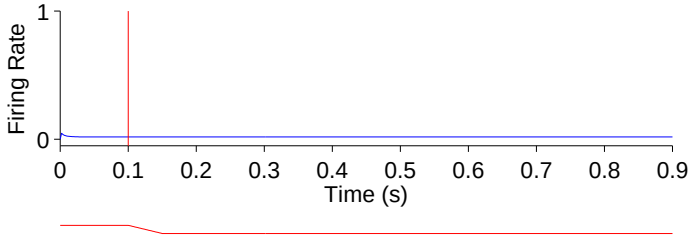
Prior to 100ms, the visual, saccade and remapping populations were all quiescent. At 100ms, a visual stimulus was introduced and activity immediately developed in the visual population representing the presaccadic retinal location of the stimulus at −5°. Next, a subset of remapping neurons with similar retinal location preferences began to respond



(a) Single step task trial



(b) Stimulus control trial



(c) Saccade control trial

Figure 8.10: The remapping behaviour of a remapping neuron with retinal preference $-20°$ after training. Plot (a) shows a single step task trial in which a stimulus is flashed from 100ms to 200ms, and a saccade is subsequently performed with onset at 600ms. The saccade brings the stimulus into the receptive field of the remapping neuron, which shows a remapping response that begins before the saccade onset - a presaccadic remapping response. Plot (b) shows the response of the same remapping neuron in a stimulus control task trial, in which a stimulus is flashed from 100ms to 200ms, but there is no saccade. The retinal location of the stimulus is the same as the presaccadic location of the stimulus in the single step task trial. In this case the remapping neuron does not respond. Plot (c) shows the response of the same remapping neuron in a saccade control task trial, in which there is no stimulus present and a saccade is performed with onset at 100ms. The saccade performed is the same as that performed in the single step task trial. Again, the remapping neuron shows no response. These three different kinds of task trial confirm that both the presence of the stimulus and the saccade are needed to trigger a response in the remapping neuron.

with varying delays after the stimulus onset. At 200ms the visual stimulus was removed and activity among the remapping neurons representing the presaccadic stimulus location started to decay rapidly. At 500ms a subset of neurons in the saccade population began to represent the impending saccade of 15°. The combined activity in the visual and saccade populations then activated the corresponding combination neurons (not shown), which then stimulated a subset of neurons in the remapping population representing the upcoming postsaccadic stimulus location of −20°. These remapping neurons were activated before the saccade onset at 600ms, and thus demonstrated presaccadic remapping.

Interestingly, the onset of activity in the saccade population, corresponding to the impending saccade, also caused a small level of activity to restart at the presaccadic retinal location of the extinguished stimulus in the remapping population. In the manually prewired model there was no such reignition of activity at the presaccadic stimulus location in the remapping population. However, in the self-organizing model this was due to corresponding projections from combination neurons representing a combination of the presaccadic stimulus location and impending saccade to remapping neurons representing the same presaccadic stimulus location, which had been strengthened by a brief period of associative learning during training. This subtle effect is thus a specific prediction of the hypothesis explored here.

Activity in the visual, saccade and remapping populations remained in equilibrium until the end of the single step task trial at 900ms. At the end of the trial, the visual

and saccade input populations began to go silent, and consequently so did the remapping population.

The latency and strength of remapping activity in the 17 single step task trial configurations the model was trained on was analysed before training and after training. The results are shown in figure 8.12 and key population metrics are summarized in table 8.4. Fig. 8.12a shows the remapping index of the remapping neurons before training and after training. The remapping indices of neurons in the trained model were mostly much larger than for the untrained model, indicating that there was only significant remapping activity in the trained model (Fig.8.12a). The average remapping index in the trained model was 0.54939, while in the untrained model it was 0.007686.

Fig.8.12b compared the remapping latency with the stimulus control latency for remapping neurons in the trained model. The remapping latencies and stimulus control latencies were computed as described above for the manually prewired model in Fig. 8.7 of section 8.2.1. Due to the absence of sufficient remapping activity, no remapping latency could be decoded for the untrained model. However, for the trained model the average remapping latency among the fourteen neurons for which it could be decoded was -49ms, and all neurons were found to remap both predictively and presaccadically. A remapping latency could not be decoded for the three remaining neurons because they had extremely weak remapping activity, as also seen in the last three ranks for Fig. 8.4a. In summary, remapping activity was found only in the model after training, and a range of response latencies

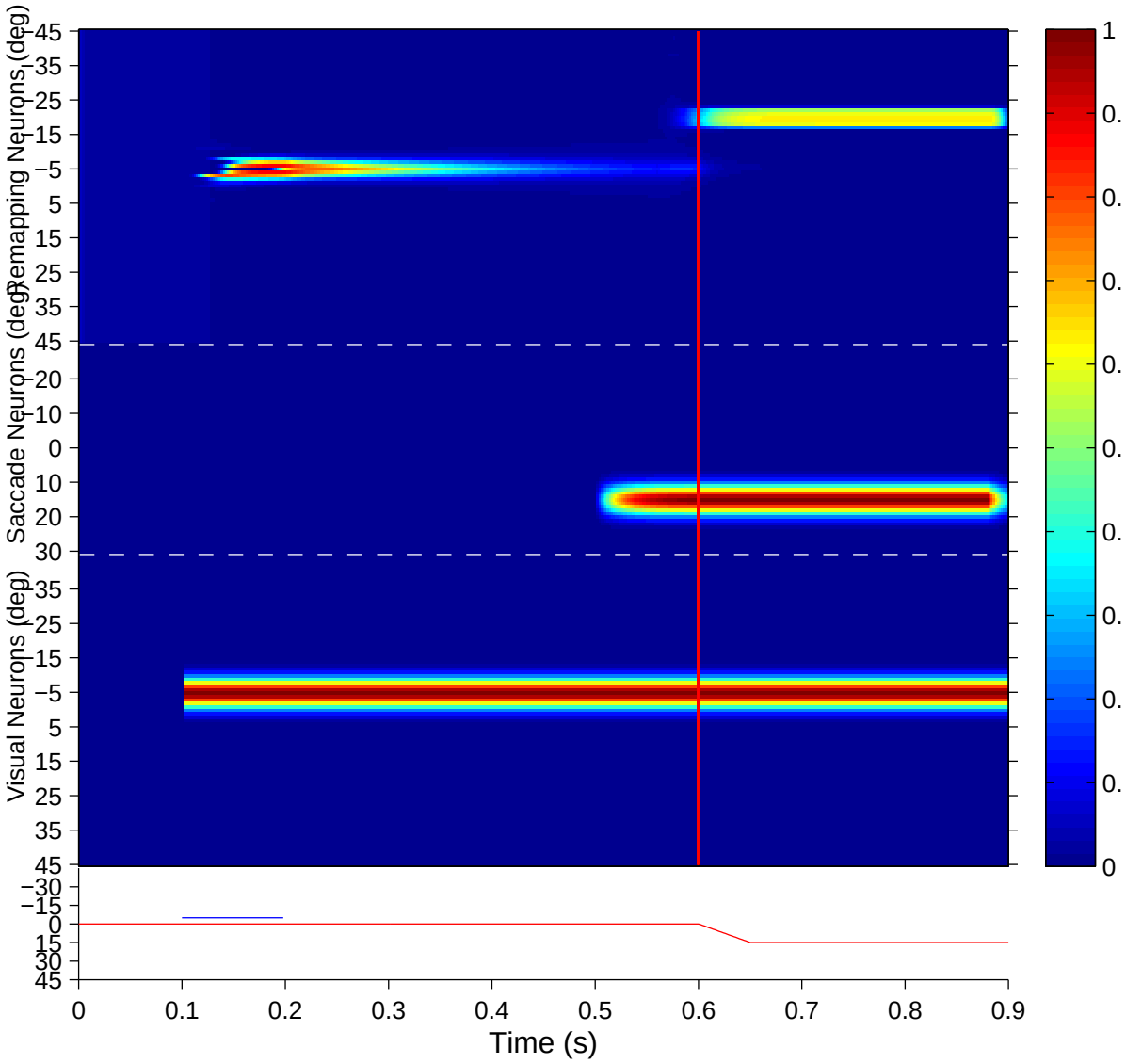


Figure 8.11: The population responses of neurons during a single step task trial for the remapping neuron with retinal preference $-20°$, where the stimulus is flashed at retinal location $-5°$ and the saccade of $15°$ is executed 400ms later. The figure shows the responses of all neurons in the remapping population (top), saccade population (middle), and visual population (bottom). Each population subplot arranges the neurons topographically along the ordinate in terms of their assigned retinal or saccade preference, α_i and β_i, respectively. The traces at the bottom show the eye position (red) and retinal location of the stimulus (blue) for the trial, and the red vertical bar designates the saccade onset. It is evident that remapping neurons with preferences for retinal locations near $-20°$ show presaccadic remapping in anticipation of the visual stimulus shifting into their receptive fields after the saccade.

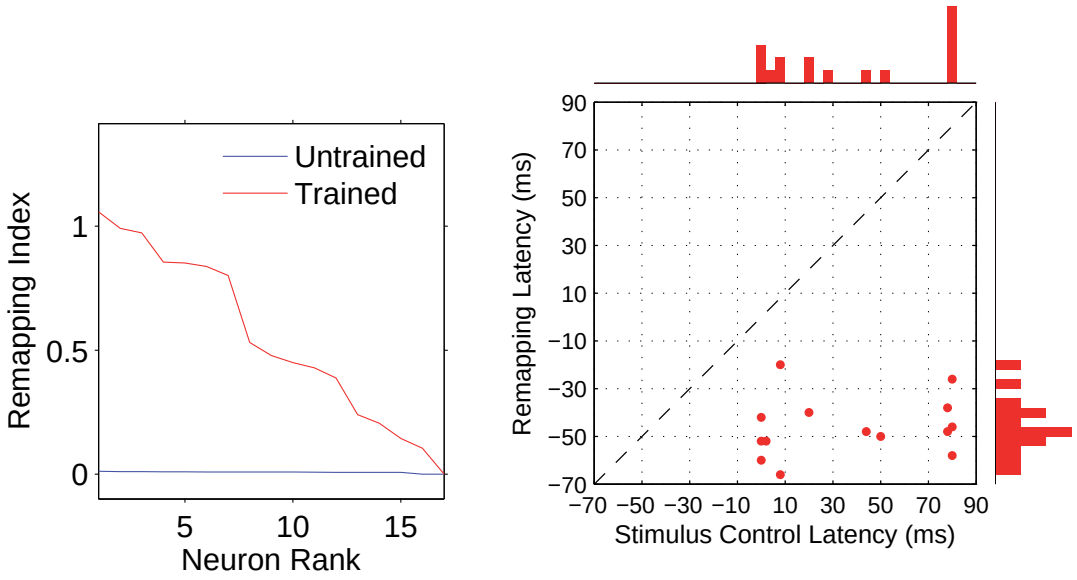


Figure 8.12: Population analysis of the remapping responses of 17 randomly selected remapping neurons. Conventions and analytical procedures as for Fig. 8.7.

were found that were presaccadic.

8.2.3 Minimal post-saccadic latency required for input representations: Varying Δ^{V} and Δ^{S}_{POST}

One of the key assumptions in the hypothesis for how the model operated was that the visual input population required some postsaccadic latency before it updated its representation from the presaccadic stimulus location to the new postsaccadic location of the stimulus, and that the saccade input population also needed a similar postsaccadic latency so that it continued to represent the saccade for some time after the saccade was completed. This, in turn, ensured that the combination neurons, which are driven by the visual and saccade populations, also contined to represent a combination of the presaccadic

stimulus location and saccade for some time after the saccade was performed. This was necessary to allow the combination neurons to associate the combination of presaccadic stimulus location and saccade, as encoded by the two delayed input populations, onto the new postsaccadic stimulus location represented by the remapping neurons.

In this experiment it was investigated how short the latency in the two input populations could be while still facilitating successful learning. The latency in the visual population was controlled by the parameter Δ^{V}, which determined how long after saccade onset the visual representation was updated. The latency in the saccade population was controlled by the parameter $\Delta^{\mathrm{S}}_{\mathrm{POST}}$, which determined how long after saccade onset the last saccade stopped being represented. For simplicity, it was always the case that $\Delta^{\mathrm{V}} - 20\mathrm{ms} = \Delta^{\mathrm{S}}_{\mathrm{POST}}$, and this was chosen because of intrinsic differences in the dynamics of the equations for the two input populations that had to be compensated. Six simulations were conducted where Δ^{V} was varied over values of 50ms, 100ms, 150ms, 200ms, 250ms, 300ms, All stimulus and model parameters were otherwise as in experiment 8.2.2.

Figure 8.14 shows the effect of varying the response latencies of the visual and saccade input populations on the remapping indices. Key population summary statistics are given in table 8.5. It is clear that with small response latencies, there was a complete collapse in the development remapping dynamics in the model. In particular, for $\Delta^{\mathrm{V}} = 50ms$ the remapping indices of all neurons was close to zero for both the trained and untrained network models. This can be explained by the following. Since most saccades took at

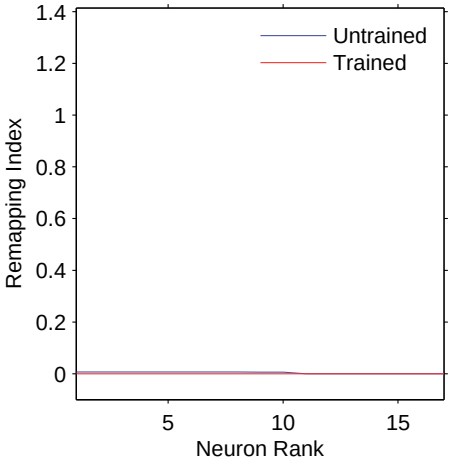
least 50ms to be completed during training, as soon as the remapping population began to represent the new postsaccadic stimulus location, there was no more input to sustain the combination layer representing the combination of presaccadic stimulus location and saccade. This then prevented the network learning an association between the combination of presaccadic stimulus location and saccade represented by the combination neurons and the postsaccadic stimulus location represented by the remapping neurons. Significant differences between the remapping indices in the trained and untrained networks is first seen for much larger delays (Fig.8.14c,d). The general trend is that remapping performance in the trained network improves as the response latencies get larger, with the best performance at $\Delta^{\mathrm{V}} = 300$ms.

However, one must take into account that the remapping indices were computed based on the activities of remapping neurons within a fixed analysis window of 300ms across all of the simulations shown in Fig. 8.14. Hence, even if there were effective remapping dynamics, at least some decline in the remapping indices as the response latencies of the visual and saccade neurons is reduced should still be expected. This is because during testing the remapping neurons are driven solely by the visual and saccade inputs. In this case, the remapping neurons also respond for shorter periods as the response latencies of the visual and saccade input neurons are reduced. Consequently, the activity of the remapping neurons across ever larger portions of the analysis window will remain silent due to the shortened duration of any possible remapping activity, which will lead to a

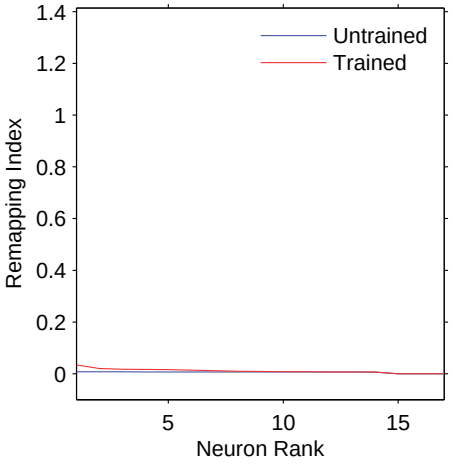
spurious additional reduction in the computed remapping indices. Because of this, the observed reduction in the remapping indices as the response latencies of the visual and saccade neurons are reduced, as shown in Fig. 8.14, will be due to a combination of two possible factors: (i) a reduction in the remapping performance of the model due to failure to self-organise properly during training, (ii) a measurement bias due to the use of a fixed 300ms analysis window over all simulations with different response durations among the remapping population. So, how much of the decline in the remapping indices shown in Fig. 8.14 is due to each of these factors? One way to answer this question is to run comparison simulations with a manually prewired network, which eliminates the first of these factors. That is, the advantage of the manually prewired model is that it can be prewired to have the correct synaptic weights for remapping even in simulations with relatively low response latencies for the visual and saccade input neurons. Therefore, as a control, the manually prewired model was similarly tested with the same set of values for Δ^{V}, as seen in figure 8.16. The trend in model performance is the same for the manually prewired network; the remapping indices declines as Δ^{V} is reduced. This confirms that the measurement bias due to the fixed analysis window of 300ms is contributing to the observed reduction in the remapping indices as the response latencies of the visual and saccadic populations are reduced. However, now there is also a substantive difference between the manually prewired and randomly wired networks at much lower latencies, such as $\Delta^{\mathrm{V}} = 50$ms as shown in Fig. 8.16a. In particular, the manually prewired model has much larger remapping indices

than both the randomly wired model or the model with learning, for low response latencies. This confirms that the reduction in the remapping indices due to the second factor, i.e. the measurement bias due to the use of a fixed 300ms analysis window over all simulations, had only a limited effect. This in turn shows that the reduction in the remapping indices in the model with learning, shown in Fig. 8.14, was at least in large part due to the first factor, that is, a reduction in the remapping performance of the model due to failure to self-organise properly during training.

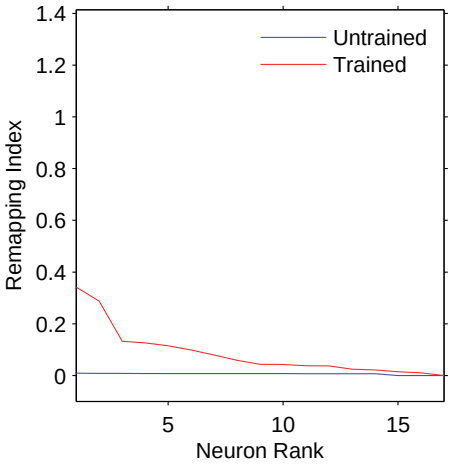
Taken together, the above results confirm the hypothesis that a postsaccadic response latency in the visual and saccade input populations is required during training in the self-organizing model, and that longer response latencies facilitate the development of stronger remapping activity.



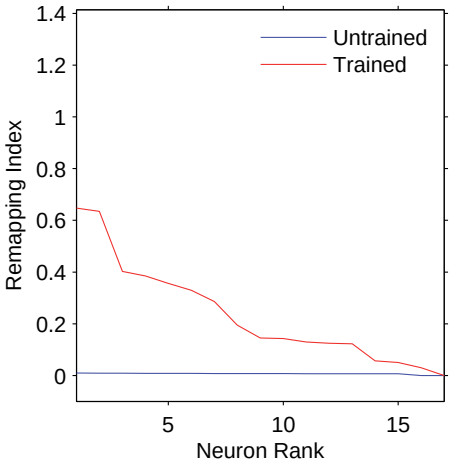
(a) 50ms



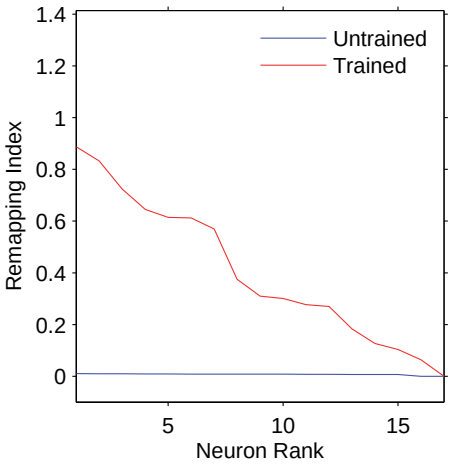
(b) 100ms



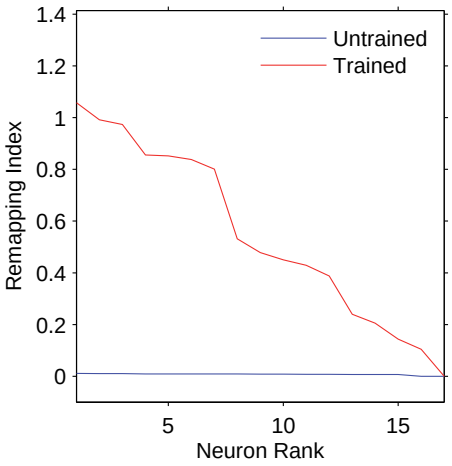
(c) 150ms



(d) 200ms



(e) 250ms



(f) 300ms

Figure 8.14: The effect of setting the response latencies of the visual and saccade input populations, Δ^{V} and $\Delta^{\mathrm{S}}_{\mathrm{POST}}$. Each plot corresponds to the given value of Δ^{V} and $\Delta^{\mathrm{S}}_{\mathrm{POST}} = \Delta^{\mathrm{V}} - 20$.

Parameter	Symbol	Value
Stimuli		
Number of remappings in training & testing	M	17
Visual Input Population		
Population size	N^{V}	91
Firing rate time constant	τ_v^{V}	1s
Visual receptive field size	σ_{V}	3°
Saccade supression delay	Δ^{V}	300°
Saccade Input Population		
Population size	N^{S}	61
Firing rate time constant	τ_v^{S}	20ms
Saccade receptive field size	σ_{S}	3°
Presaccadic onset time	$\Delta^{\mathrm{S}}_{\mathrm{PRE}}$	100ms
Postsaccadic offset delay	$\Delta^{\mathrm{S}}_{\mathrm{POST}}$	280ms
Combination Population		
Population size	N^{C}	5551
Activation time constant	τ_h^{C}	20ms
Visual input coefficient	$\psi^{\mathrm{V}\to\mathrm{C}}$	10
Saccade input coefficient	$\psi^{\mathrm{S}\to\mathrm{C}}$	8
Lateral competition coefficient	$w^{\mathrm{C}}_{\mathrm{INHB}}$	0.1
Activation function slope	φ^{C}	100
Activation function threshold	θ^{C}	15
Visual input connectivity rate	ϕ^{V}	20%
Saccade input connectivity rate	ϕ^{S}	40%
Remapping Population		
Population size	N^{R}	5000
Activation time constant	τ_h^{R}	20ms
Combination population input coefficient	$\psi^{\mathrm{C}\to\mathrm{R}}$	7
Lateal competition coefficient	$w^{\mathrm{R}}_{\mathrm{INHB}}$	0.6
Activation function slope	φ^{R}	0.5
Activation function threshold	θ^{R}	3
Combination population connectivity	ϕ^{C}	100%
K time constant	τ^{K}	20ms
K coefficient	ψ^{K}	8ms
K truncation delay	Δ^{K}	0ms
P time constant	τ^{P}	300ms
Prewired connection weights		
Spread of connection weights	σ_{W}	3°

Table 8.1: Parameters of prewired model.

Experiment 8.2.1		
	Random	**Manual**
Average remapping latency	N/A	−79ms
Average remapping index	0.0131	0.848
Predictive remapping	0 (0%)	17 (100%)
Presaccadic remapping	0 (0%)	17 (100%)

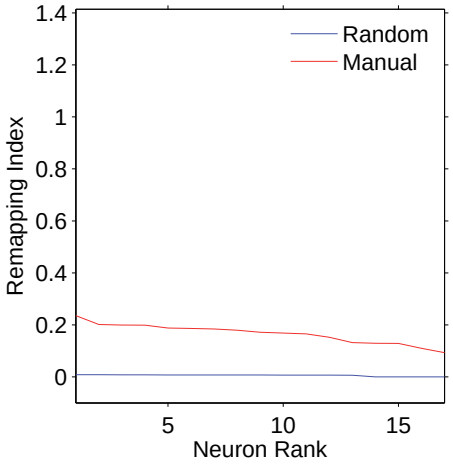
Table 8.2: Population summary statistics of the response properties of remapping neurons in the randomly and manually prewired models during single step task trials. Results for the randomly wired model are shown in the left column, while results for the manually prewired model are shown in the right column. Each row corresponds to a different performance metric: Average remapping latency, average remapping index, predictive remapping, presaccadic remapping. The average remapping latency is given to the closest millisecond, and the number of neurons for which latency could be computed is given.

Parameter	**Symbol**	**Value**
Learning rate: combination to remapping neurons	$\varrho^{\mathrm{C}\to\mathrm{R}}$	0.1
Visual input connectivity rate	ϕ^{V}	2%
Saccade input connectivty rate	ϕ^{S}	2%
Presaccadic onset time	$\Delta^{\mathrm{S}}_{\mathrm{PRE}}$	70ms
Combination population input coefficient	$\psi^{\mathrm{C}\to\mathrm{R}}$	3
Saccade input coefficient	$\psi^{\mathrm{S}\to\mathrm{C}}$	10
Activation function threshold: combination neurons	θ^{C}	18

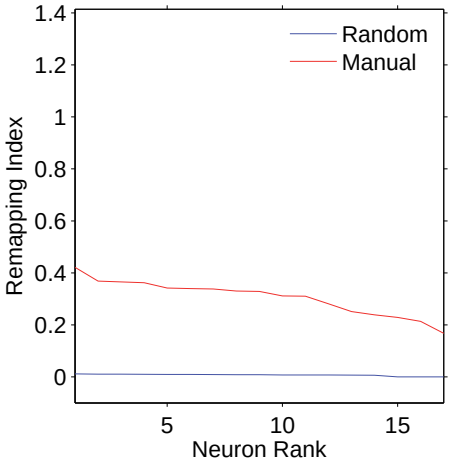
Table 8.3: Parameters of self-organizing model.

Experiment 8.2.2		
	Untrained	**Trained**
Average remapping latency	N/A (0/17)	−46ms (14/17)
Average remapping index	0.007686	0.54939
Predictive remapping	0 (0%)	14 (82%)
Presaccadc remapping	0 (0%)	14 (82%)

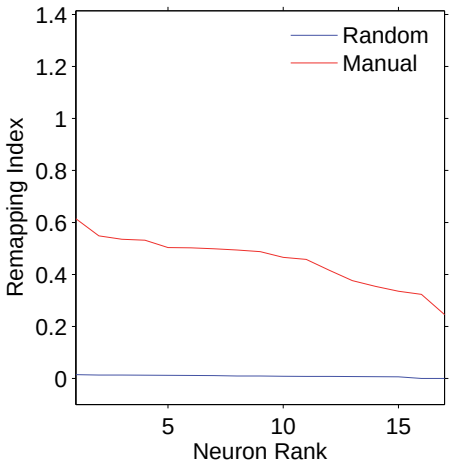
Table 8.4: Population summary statistics of the response properties of remapping neurons before and after training during single step task trials. Results for the untrained model are shown in the left column, while results for the trained model are shown in the right column. Each row corresponds to a different performance metric: Average remapping latency, average remapping index, predictive remapping, presaccadic remapping. The average remapping latency is given to the closest millisecond, and the number of neurons for which latency could be computed is given.



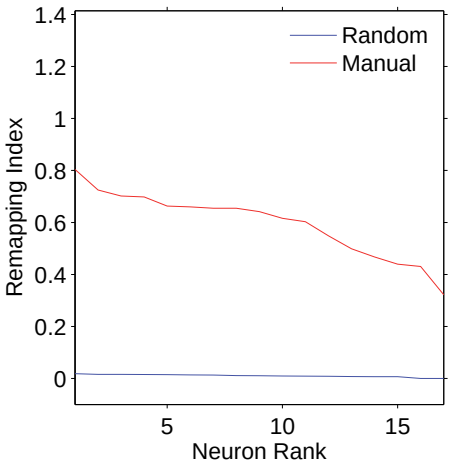
(a) 50ms



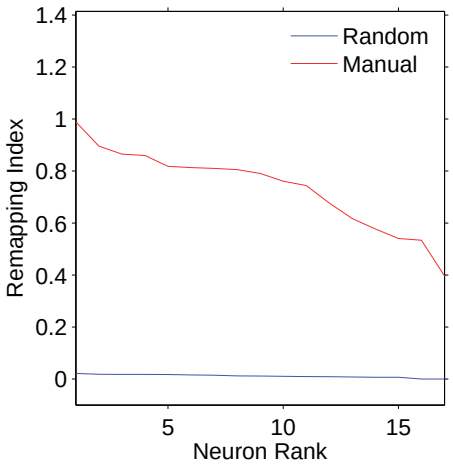
(b) 100ms



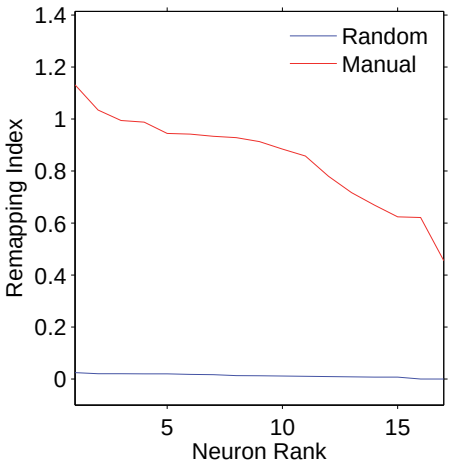
(c) 150ms



(d) 200ms



(e) 250ms



(f) 300ms

Figure 8.16: The effect of varying the response latencies of the visual and saccade input populations, Δ^{V} and $\Delta^{\mathrm{S}}_{\mathrm{POST}}$, on remapping in the manually prewired and randomly wired network models.

Experiment 8.2.3						
	50ms	**100ms**	**150ms**	**200ms**	**250ms**	**300ms**
Average remapping latency	N/A (0/17)	N/A (0/17)	-25ms (7/17)	−43ms (13/17)	−46ms (14/17)	−46ms (14/17)
Average remapping index	0	0.010891	0.0819	0.23761	0.40556	0.54939
Predictive remapping	0 (0%)	0 (0%)	7 (41%)	13 (76%)	14 (82%)	14 (82%)
Presaccadic remapping	0 (0%)	0 (0%)	6 (35%)	13 (76%)	13 (76%)	14 (82%)

Table 8.5: Population summary statistics showing the effect of varying the response latencies of the visual and saccade input populations, Δ^{V} and $\Delta^{\mathrm{S}}_{\mathrm{POST}}$, respectively. Each column is for a simulation with corresponding value of Δ^{V}. Each row corresponds to a different performance metric: Average remapping latency, average remapping index, predictive remapping, presaccadic remapping. The average remapping latency was computed over those neurons for which this metric could be defined.

8.2.4 Replicating the observations of Kusunoki & Goldberg (2003)

The classic experimental study of Kusunoki & Goldberg (2003) investigated the time course of the responsiveness of LIP neurons when visual dot stimuli were presented in the receptive fields of these neurons around the time of a saccade. In particular, they sought to characterize precisely how much an LIP neuron responded as a function of when the dot stimulus was briefly flashed with respect to a given saccade. The stimulus was flashed in either the presaccadic or postsaccadic head-centered location of the receptive field, and trials of each kind were referred to as *current receptive field* and *future receptive field* trials, respectively. They found that the response of LIP neurons to a stimulus in the current receptive field *decreased* as the stimulus was flashed later with respect to the saccade, and conversely that response in the future receptive field *increased* as the stimulus was flashed later with respect to the saccade.

In the experiment described next it was attempted to confirm that the model presented in section 8.2.2 also displayed the above experimentally observed behaviour, and to investigate the mechanisms by which the model achieved this. The trained model was tested on a set of delayed stimulus flash task trials, as described in section 7.5.2, corresponding to the single step saccade trials the model had been trained on. The responses of the remapping neurons were analysed using a 300ms window aligned at 50ms after stimulus onset, just as in Kusunoki & Goldberg (2003).

Looking at the responses of the remapping neuron with a receptive field at -20° in both

current receptive field trials (Fig. 8.17) and future receptive field trials (Fig. 8.18) for four different stimulus onset times provides a good example for understanding the peri-saccadic shift in receptive field sensitivity of LIP neurons.

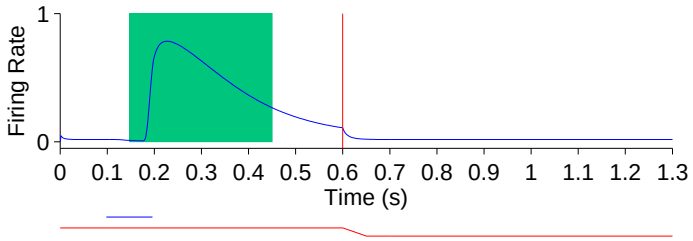
First, consider the current receptive field trials shown in Fig. 8.17, where the visual stimulus is presented in the presaccadic head-centered receptive field location of the remapping neuron. When the stimulus onset is well before the saccade (Fig. 8.17a), then the stimulus aligned window captures the majority of the response. Furthermore, the response captured by the window is not truncated by the impending saccade, which is beyond the window. So the response in the stimulus aligned window is maximal at this point. As the stimulus onset approaches saccade onset (Fig. 8.17b,c), a longer interval of the response of the neuron is truncated by the saccade. In particular, larger portions ot the stimulus onset window capture the response truncation effect of the saccade, thus leading to a reduced response within the window. When the stimulus onset is after saccade onset (Fig. 8.17c) there is of course no response by the neuron at all, given that the neuron's receptive field has been removed from the presaccadic head-centered location where the stimulus is presented. These simulation results show that the neuronal response in the stimulus onset aligned window decreases as the stimulus is flashed later with respect to the saccade.

Secondly, consider the future receptive field trials. So long as the stimulus onset is prior to saccade onset, the neuron responds more or less invariantly in terms of the magnitude and time course of the response around the time of saccade onset (Fig. 8.18a,b,c). So if the

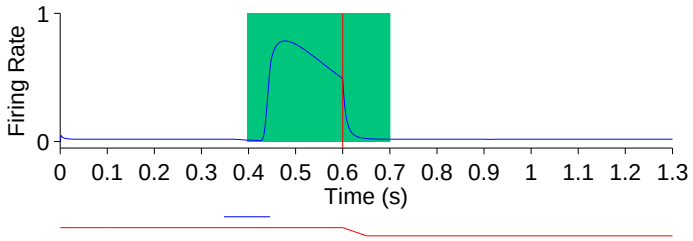
stimulus onset is well in advance of the saccade onset (Fig. 8.18a), then the stimulus onset aligned analysis window does not capture any of the response, and thus registers a minimal response. With later stimulus onset times the stimulus aligned analysis window captures ever larger portions of the remapping activity, thus registering a larger response. As the stimulus onset occurs after saccade onset (Fig. 8.18d), the response of the neuron now becomes a pure visual onset response with accompanying decay. This causes the stimulus onset aligned window to capture the majority of the strong neuronal response regardless of the stimulus onset time, thus registering a maximal response here. These simulation results show that the response in the stimulus onset aligned window increases as the stimulus is flashed later with respect to the saccade.

To compare these findings with the population level analysis in Kusunoki & Goldberg (2003), all task trials of a given type (current or future receptive field) and stimulus onset time were grouped. For each such group, the average response within a stimulus onset aligned analysis window was computed. Figure 8.19 shows these results before and after training. The current and future receptive field trials, both before and after training, were in agreement with the experimental observations of Kusunoki & Goldberg (2003). That is, the response of the remapping neurons to a stimulus in the current receptive field decreased as the stimulus was flashed later with respect to the saccade, while response in the future receptive field increased as the stimulus was flashed later with respect to the saccade.

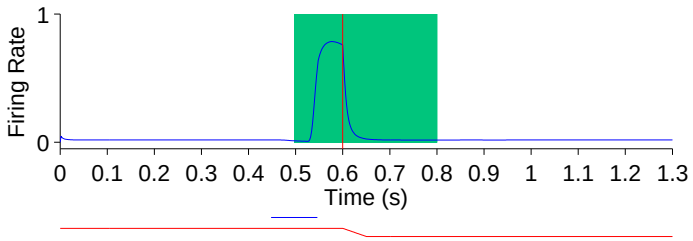
Interestingly, training had negligible influence on the current receptive field trial curve.



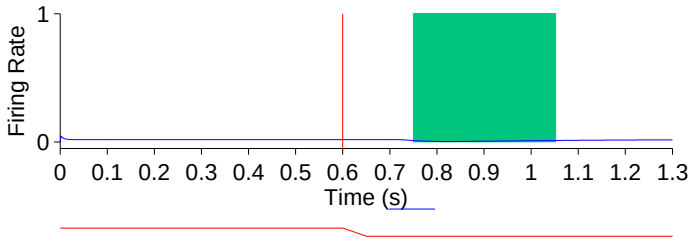
(a) 100ms



(b) 350ms

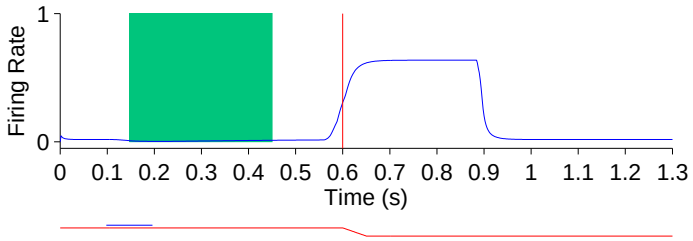


(c) 450ms

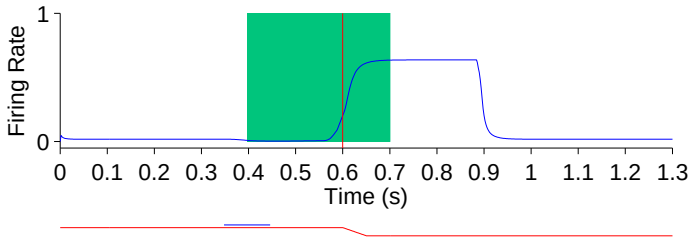


(d) 700ms

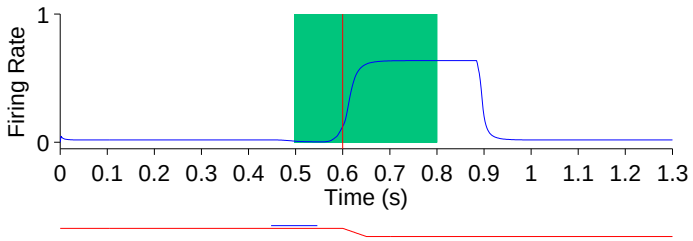
Figure 8.17: Peri-saccadic shift in the receptive field sensitivity of a remapping neuron as the stimulus onset time is varied during current receptive field trials. In current receptive field trials the stimulus is presented in the presaccadic head-centered receptive field location. Each plot shows the response of the remapping neuron with a receptive field at $-20°$ during a current receptive field trial where the stimulus onset is at one of the the following times: 100ms, 350ms, 450ms and 700ms. In each trial the saccade onset is at 600ms. The two traces below the abscissa indicate the eye position (red) and stimulus presence (blue). The green rectangle shows the stimulus onset-aligned response analysis window. When analysing the neuronal response in the stimulus onset-aligned window, it is evident that the response of this remapping neuron to a stimulus presented in its presaccadic receptive field decreases as the stimulus is flashed later with respect to the saccade.



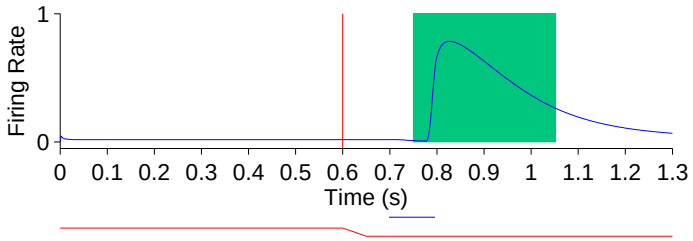
(a) 100ms



(b) 350ms



(c) 450ms



(d) 700ms

Figure 8.18: Peri-saccadic shift in the receptive field sensitivity of a remapping neuron as the stimulus onset time is varied during future receptive field trials. In future receptive field trials the stimulus is presented in the postsaccadic head-centered receptive field location. Conventions as in Fig. 8.17. When analysing the neuronal response in the stimulus onset-aligned window, it is evident that the response of this remapping neuron to a stimulus presented in its postsaccadic receptive field increases as the stimulus is flashed later with respect to the saccade.

This should be expected because the remapping neurons representing the presaccadic receptive field were being driven directly by the external visual signal during the time course of the current receptive field trials. These neurons initially responded to the stimulus, which was followed by a truncation of their responses at the saccade. Because these neuronal dynamics were driven directly by the external visual signal, they were not affected by training of the network.

Training did have a clear effect on the future receptive field trial curve. This was because the remapping neurons representing the postsaccadic stimulus location were driven by both the feedforward synaptic connections within the network, which are modified during training, as well as the direct external visual signal during the timecourse of future receptive field trials. The average response increased with increasing stimulus onset time. The initial responses corresponded to onset times where the analysis window was too early to capture any remapping activity among the remapping neurons. However, the monotonic increase reflected the fact that larger and larger portions of the analysis window were being filled by remapping activity in the future receptive field around the time of the saccade. The responses at the latest stimulus onset times reached a saturated maximum, where the analysis window simply captured the visual onset activity of the remapping neurons. These findings were in argreement with the experimental observations reported in Kusunoki & Goldberg (2003) (Fig. 7).

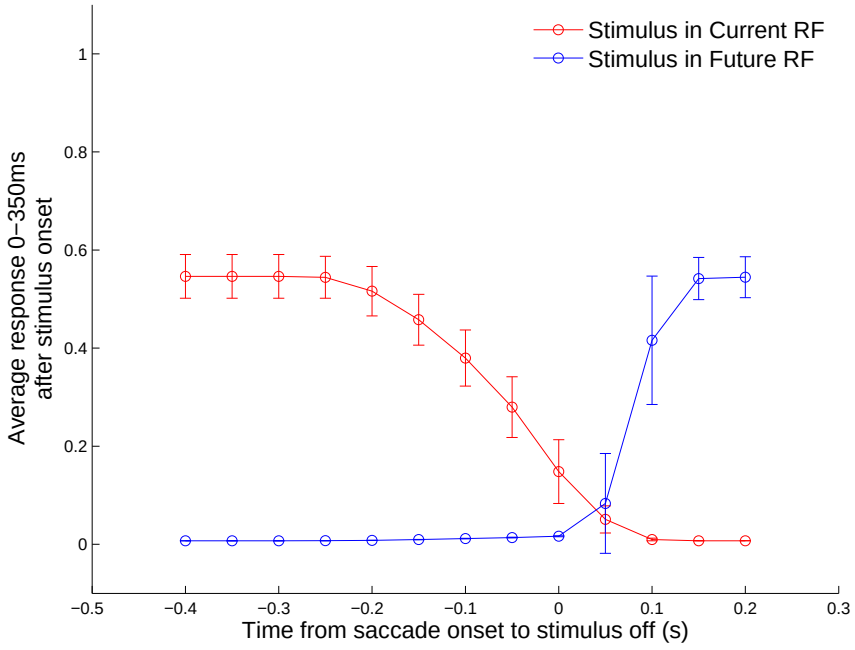
In summary, the decreasing responsiveness of remapping neurons as the stimulus onset

time occurs later with respect to the saaccade in current receptive field trials can be attributed to saccade onset aligned activity truncation, and not to any change in the feedforward visual sensitivtiy of remapping neurons after training. Likewise, the increasing responsiveness of remapping neurons in future receptive field trials can be attributed to the remapping activity displayed by these neurons.

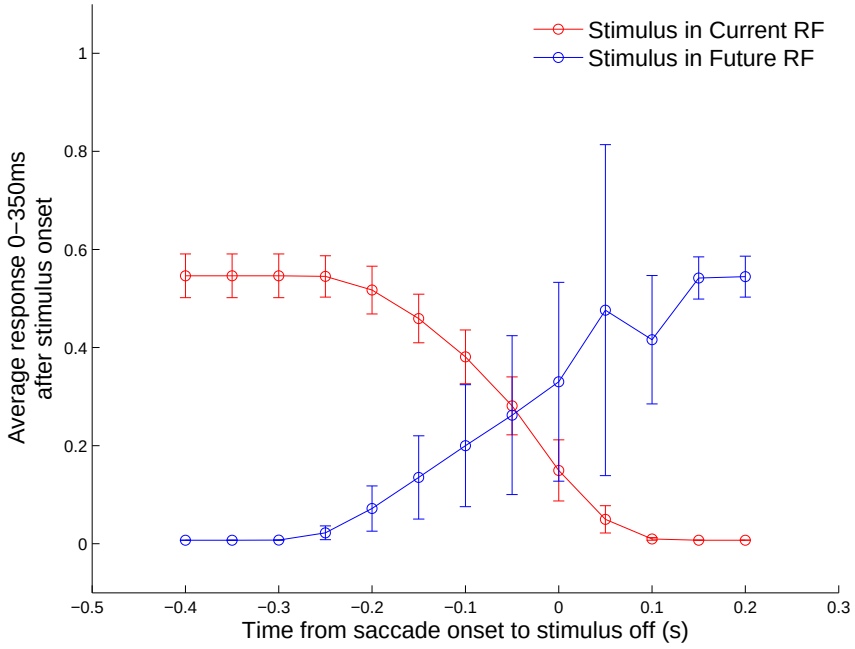
8.2.5 Replicating observations of Heiser & Colby (2006)

In Heiser & Colby (2006) the authors attempted, for the first time, to comprehensively investigate the spatial characteristics of remapping in LIP. They investigated whether and how individual LIP neurons remapped activity into their receptive field from multiple directions. This was done by measuring the remapping activity of a single neuron in single step tasks requiring a saccade in each of the four cardinal directions with respect to the neuron's receptive field, where each trial relocated the receptive field to the head-centered location of a recently extinguished visual stimulus. Determining the spatial characteristics of remapping was particularly important to resolve, as it had a significant bearing on whether LIP, and other similar areas, could be supporting perceptual spatial constancy as many authors had suggested.

Importantly, the experiment did not dissociate the saccade direction from retinal location since only a single location in a given cardinal direction was examined for remapping activity. In particular, the saccades in the four cardinal directions were all of the same reti-



(a) Before training



(b) After training

Figure 8.19: Population analysis of peri-saccadic shift in the receptive field sensitivity of remapping neurons as the stimulus onset time is varied during both current and future receptive field trials. Results are given before the network is trained (a) and after the network is trained (b). The plots show the average response, as measured by stimulus onset aligned window, across the 17 remapping neurons as a function of the time from the saccade onset to the stimulus extinction. The error bars represent the standard deviations.

nal distance. However, it seems unlikely that remapping activity in LIP neurons should be restricted to saccades in cardinal directions that are all of equal distance from the neuron's receptive field. Therefore, the model simulations described below investigated remapping activity when saccades were performed over different retinal distances. Though the saccades were peformed in only two cardinal directions because the retinal space simulated in the model was one dimensional.

In the previous experiment 8.2.2 there where 17 training trials, where each trial corresponded to one particular combination of presaccadic stimulus location, saccade, and resulting postsaccadic stimulus location. This set up the synaptic connections to enable the remapping neurons representing the postsaccadic stimulus location to exhibit remapping activity in response to a stimulus presented in the trained presaccadic location followed by the corresponding saccade. In the experiment described next, there were four different training trials associated with each of the same original 17 postsaccadic stimulus locations, each with a different combination of random presaccadic stimulus location and corresponding saccade. This made the total number of training trials $17 \times 4 = 68$. Figure 8.20 shows all training trials in terms of the presaccadic retinal location of the stimulus and the corresponding saccade. The model was trained for 20 epochs with the same parameters as before, except that ϕ^{C} was varied because this parameter influenced the likelihood of a given combination neuron driving a particular remapping neuron. The following subsections describe simulation results addressing specific issues from Heiser &

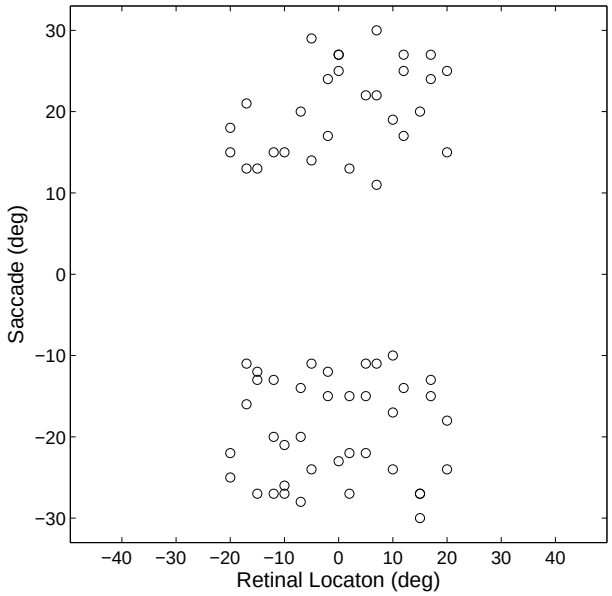


Figure 8.20: Simulations in which individual remapping neurons are trained to remap activity from multiple (four) presaccadic stimulus locations. The scatterplot shows all of the training trials in terms of the presaccadic retinal location of the stimulus and the corresponding saccade.

Colby (2006).

Individual remapping neurons can remap activity from multiple presaccadic retinal stimulus locations

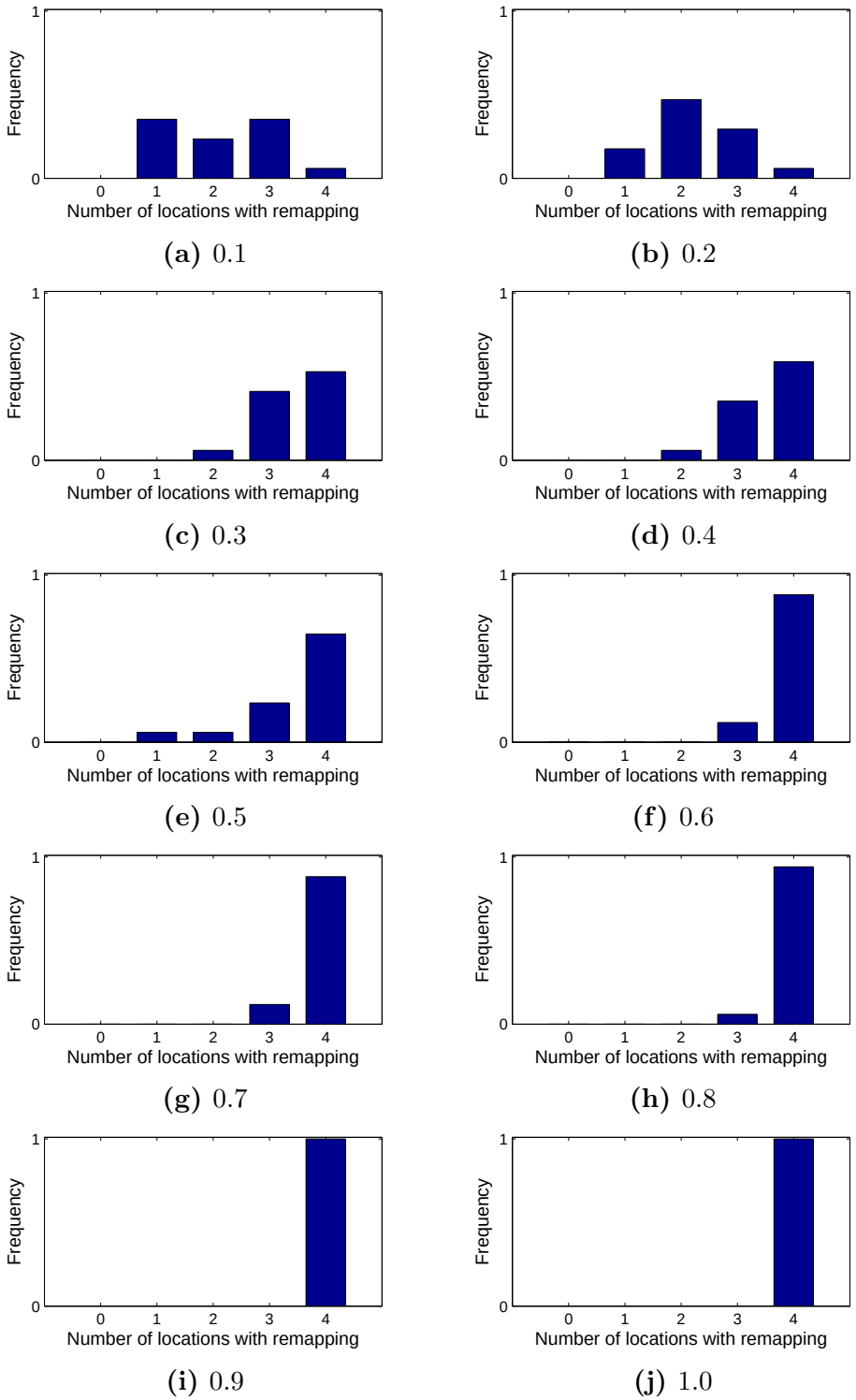
Given the training and testing trials described above, a given remapping neuron could remap activity anywhere from 0 to 4 presaccadic stimulus locations. A neuron was classified as being able to remap a given presaccadic stimulus location if the remapping index from the single step trial was greater than zero and larger than the remapping index from the same trial prior to training. Frequency distributions showing the proportion of the 17 selected remapping neurons that learned to remap activity from 0, 1, 2, 3 or 4 presaccadic

stimulus locations are shown in figure 8.21. It can be seen that for all values of ϕ^{C} there were at least some remapping neurons that were able to remap activity from all four presaccadic stimulus locations. However, as ϕ^{C} was increased from 0.1 to 1.0, the latter of which being the standard value, the neurons tended to remap from an increasing number of presaccadic retinal locations. The parameter ϕ^{C} controlled how well connected a neuron in the remapping population was with neurons in the combination population. Hence, an increase in this parameter made it more likely that a given remapping neuron trained on a particular single step trial would be connected with combination neurons representing the corresponding presaccadic stimulus location and saccade. So increasing ϕ^{C} made it possible for a remapping neuron to remap from more presacadic stimulus locations.

In summary, these results show that remapping neurons in the model were able to remap from multiple presaccadic stimulus locations, which was in agreement with the findings of Heiser & Colby (2006). Furthermore, the population distribution of remapping neurons that could remap activity from a number of different presaccadic stimulus locations became more skewed towards a larger number of locations as the connectivity rate between the remapping and combination populations was increased.

Remapping strength of individual remapping neurons varies with the presaccadic stimulus location

This experiment investigated how the remapping strength of individual remapping neurons depended upon the presaccadic stimulus location. To quantify this, a selectivity index was



(a) 0.1 **(b)** 0.2

(c) 0.3 **(d)** 0.4

(e) 0.5 **(f)** 0.6

(g) 0.7 **(h)** 0.8

(i) 0.9 **(j)** 1.0

Figure 8.21: Frequency distributions showing the proportion of the 17 selected remapping neurons that learned to remap activity from 0, 1, 2, 3 or 4 presaccadic stimulus locations. Each distribution corresponds to a simulation with a particular value of ϕ^{C} ranging from 0.1 to 1.0.

used to measure how selective an individual neuron was to a single presaccadic stimulus location, versus being completly unbiased. It is assumed that a particular remapping neuron is trained on $L = 4$ different single step trials, each of which involves a different combination of presaccadic stimulus location and saccade, but with the same postsaccadic stimulus location corresponding to the retinal preference of the given remapping neuron. Each such trial has a stimulus located in presaccadic retinal location r_i, and a saccade s_i is performed. Let the remapping index of the given remapping neuron in the i^{th} single step trial, among $L = 4$, be denoted by RI_i. Then the preferred presaccadic stimulus location of the given neuron for remapping was decoded as

$$\frac{\sum_{i=1}^{L} r_i \text{RI}_i}{\sum_{i=1}^{L} \text{RI}_i} \tag{8.4}$$

A selectivity index, SI, for the remapping neuron was computed by first normalizing each remapping index

$$\text{RI}_i^{\text{norm}} = \frac{\text{RI}_i}{\sum_{i=1}^{L} \text{RI}_i} \tag{8.5}$$

and then using the entropy of the remapping index distribution across the directions

$$\text{SI} = 1 + \frac{1}{\log_2 L} \sum_{i=1}^{L} \text{RI}_i^{\text{norm}} \log_2 \text{RI}_i^{\text{norm}} \tag{8.6}$$

This metric gave a maximum value of 0 for a completely uniform remapping index distribution across the $L = 4$ different presaccadic stimulus locations, and 1 for a perfectly selective distribution where the neuron remapped from only one presaccadic stimulus location. Figure 8.22 shows the remapping index distributions of eight different remapping neurons from simulations of with $\phi^{\mathrm{C}} = 1.0$. Each subplot corresponding to one of the remapping neurons also shows the selectivity index for that neuron, as well as its preferred presaccadic stimulus location for remapping. The first thing which is clear is that many remapping neurons have a significant variability in the strength of remapping across the four presaccadic stimulus locations. However, these neurons nearly always remap more strongly in the trained compared to the untrained condition for all presaccadic stimulus locations. Some remapping neurons clearly respond very strongly to a single presaccadic stimulus location (Fig. 8.22a,c,g), while other neurons respond more evenly across all locations (Fig. 8.22d,e,f,h). Notice that the larger values of the selectivity index are associated with single peaked distributions.

Going from the single neuron to the population level, it can be confirmed that across all values of ϕ^{C}, a range of selectivity indices were present in the remapping population (Fig. 8.23). Also, increasing values of ϕ^{C} caused the selectivity indices to decrease, as one would expect. At the extreme value of $\phi^{\mathrm{C}} = 0.1$ (Fig. 8.23a), there was a large number of perfectly selective neurons, and at the extreme value $\phi^{\mathrm{C}} = 1.0$ (Fig. 8.23j) there were no perfectly selective neurons. Thus, the population of remapping neurons displayed variability in their

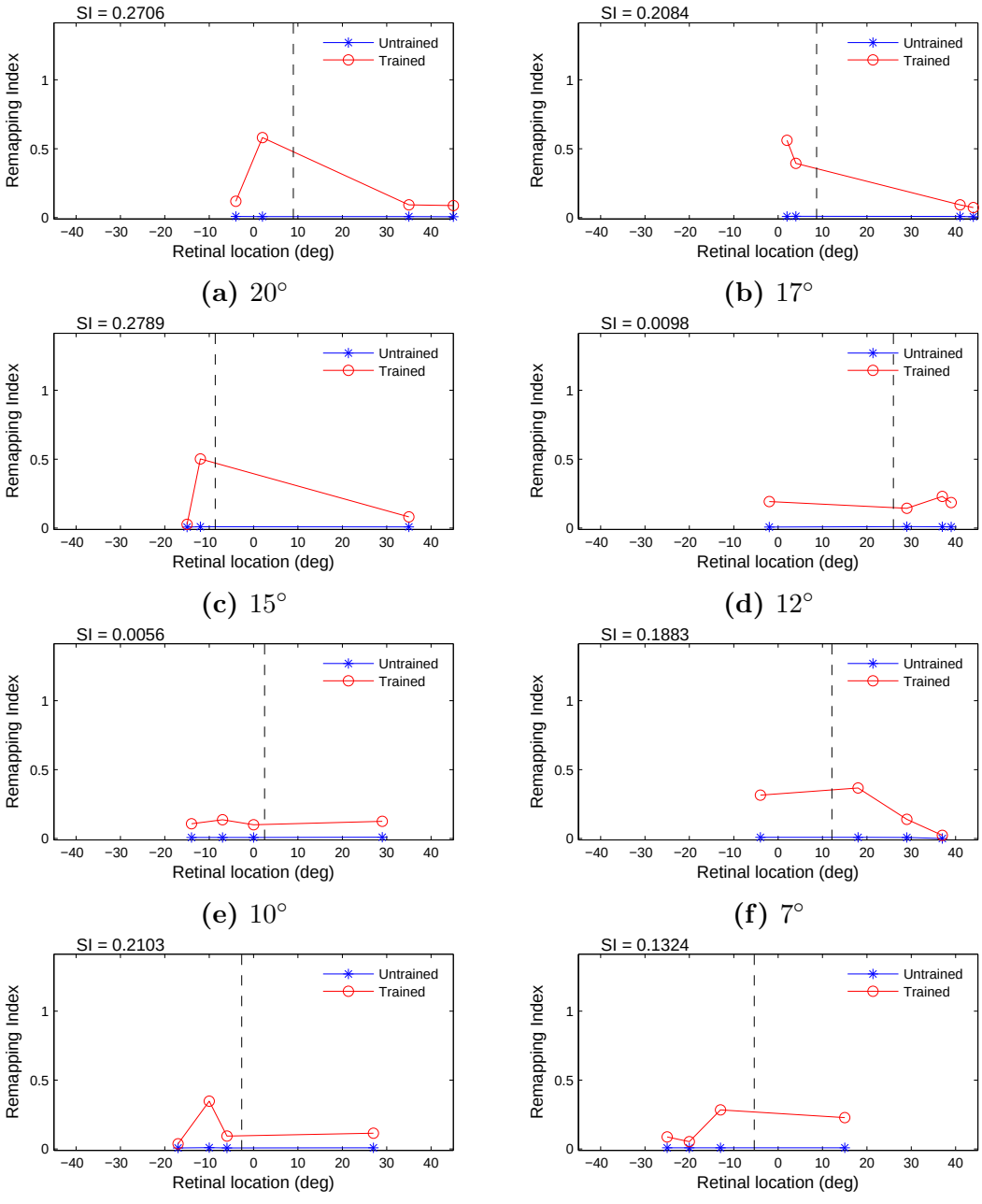


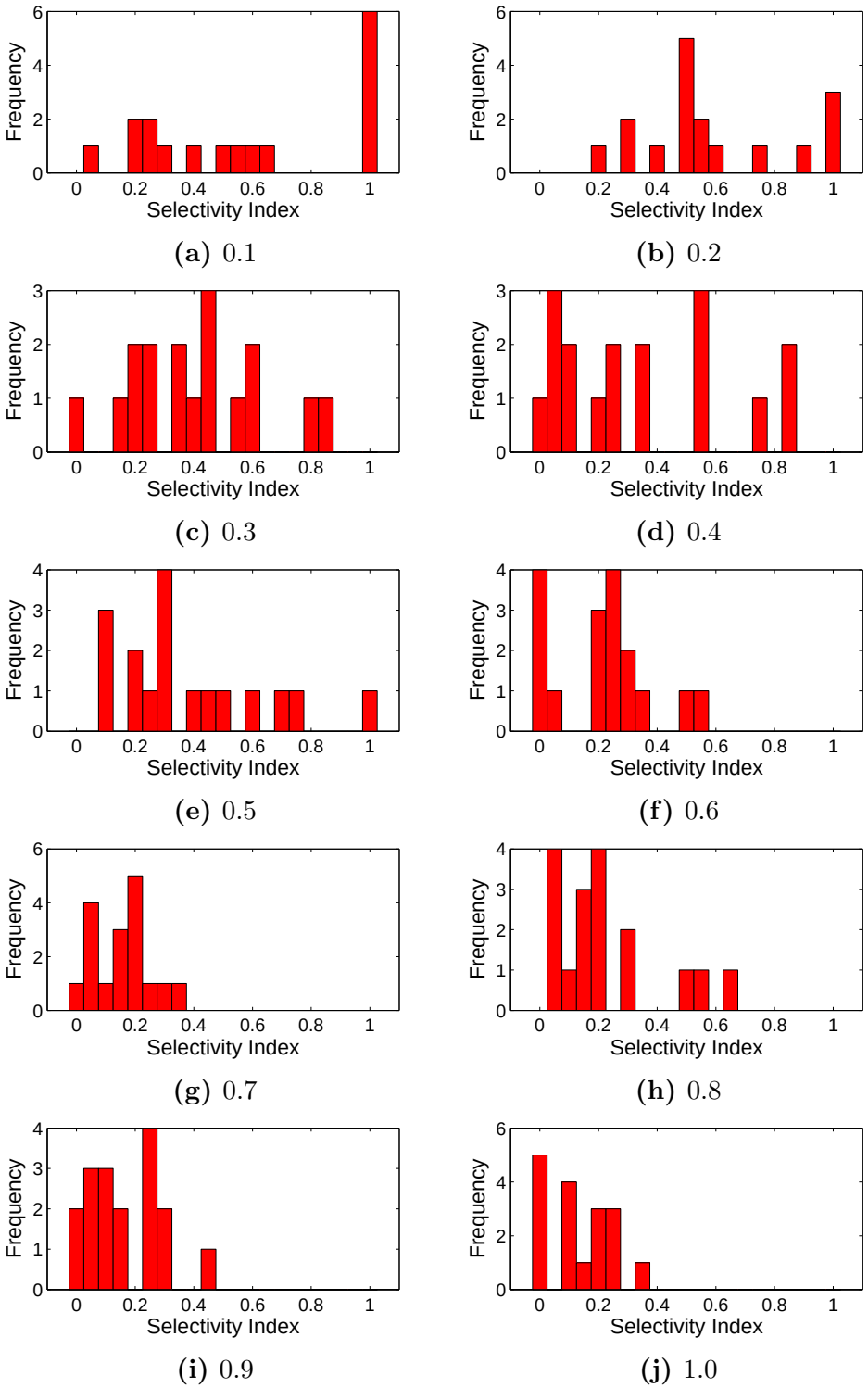
Figure 8.22: The remapping index distributions of eight different remapping neurons with retinal preferences 20°, 17°, 15°, 12°, 10°, 7°, 5° and 2° from simulations with $\phi^{\mathrm{C}} = 1.0$. Each of the subplots (a-g) shows the remapping index as a function of presaccadic stimulus location for one of the neurons in the remapping population. These results are presented before training (blue) and after training (red). The selectivity index, SI, of each remapping neuron after training is shown at the top left of the corresponding subplot. The vertical dashed line in each subplot represents that neuron's decoded preferred presaccadic stimulus location for remapping.

selectivities for remapping from different presaccadic stimulus locations.

Furthermore, the distribution of preferences for remapping from different presaccadic stimulus locations across the population of remapping neurons was also relatively uniform across most values of ϕ^{C} (Fig. 8.24). This effectively meant that stimulus traces could be remapped from the entire retina. These simulation results were thus in agreement with the experimental findings of Heiser & Colby (2006).

Strength and selectivity of remapping are related

Given the broad range of selectivity indices observed above, it was natural to investigate whether there was any relationship between the selectivity index of a remapping neuron and the neuron's strength of remapping from the different presaccadic stimulus locations as measured by the remapping index. In particular, for $\phi^{\mathrm{C}} = 0.1$ there was a strong bias towards remapping neurons with high selectivity indices close to 1 (Fig. 8.23a), which was what was found by Heiser & Colby (2006). This raised the question of whether such neurons with high selectivity indices were especially responsive to remapping from a particular presaccadic stimulus location, while neurons with low selectivity indices remapped weakly across all presaccadic stimulus locations, as Fig. 8.22 also indicated. This was analysed by looking at the average maximal remapping index among remapping neurons responding to the same number of presaccadic stimulus locations (Fig.8.25). There was no clear relationship between the average maximal remapping index and number of presaccadic stimulus locations remapped. However, in the three cases (Fig. 8.25a,b,e) where there



(a) 0.1 **(b)** 0.2

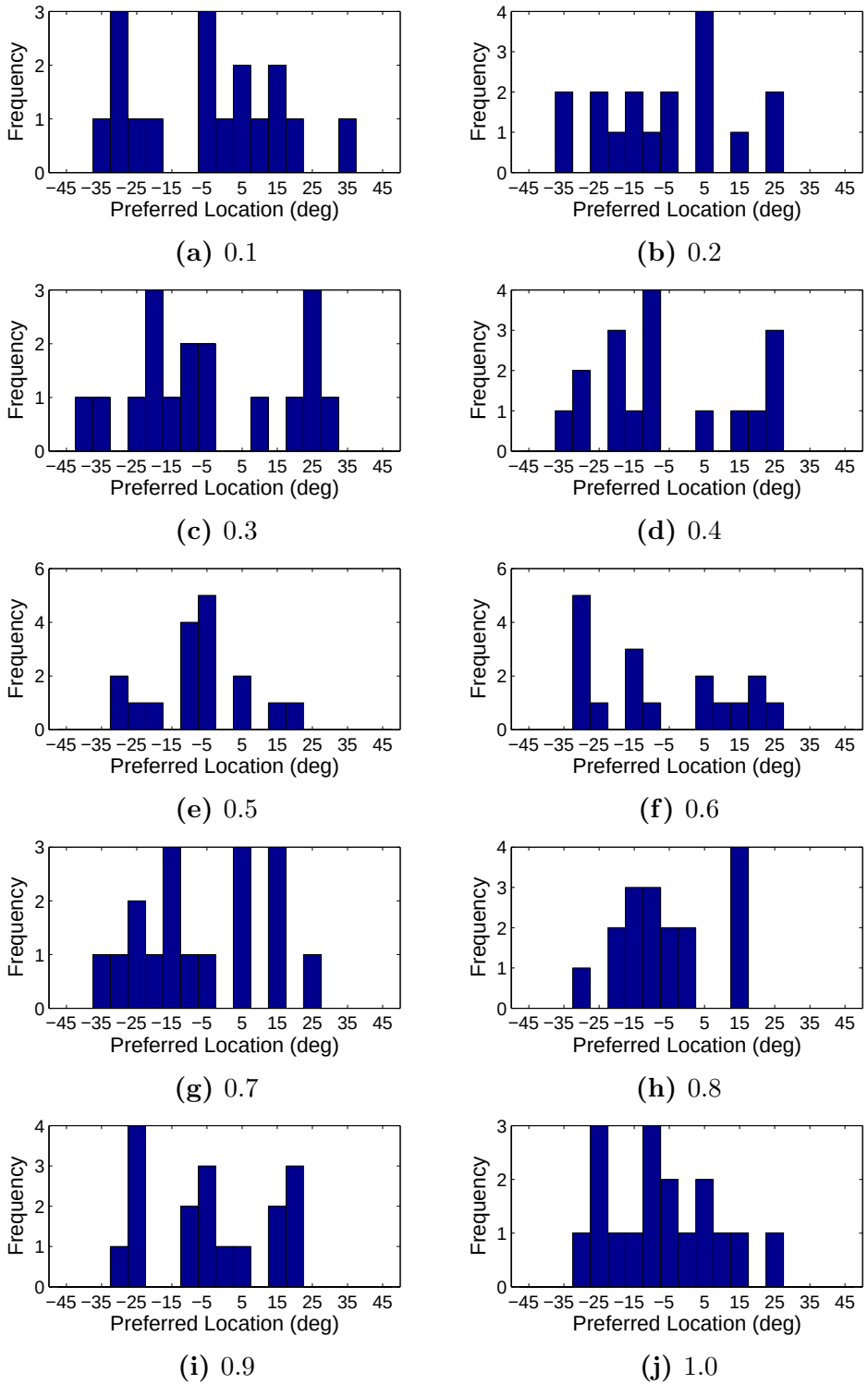
(c) 0.3 **(d)** 0.4

(e) 0.5 **(f)** 0.6

(g) 0.7 **(h)** 0.8

(i) 0.9 **(j)** 1.0

Figure 8.23: Frequency distributions showing the number of remapping neurons with different selectivity indices after training. Each distribution corresponds to a simulation with a particular value of ϕ^{C}.



(a) 0.1 **(b)** 0.2

(c) 0.3 **(d)** 0.4

(e) 0.5 **(f)** 0.6

(g) 0.7 **(h)** 0.8

(i) 0.9 **(j)** 1.0

Figure 8.24: Frequency distributions showing the number of remapping neurons with preferences for remapping from different presaccadic stimulus locations after training. Each distribution corresponds to a simulation with a particular value of ϕ^{C}.

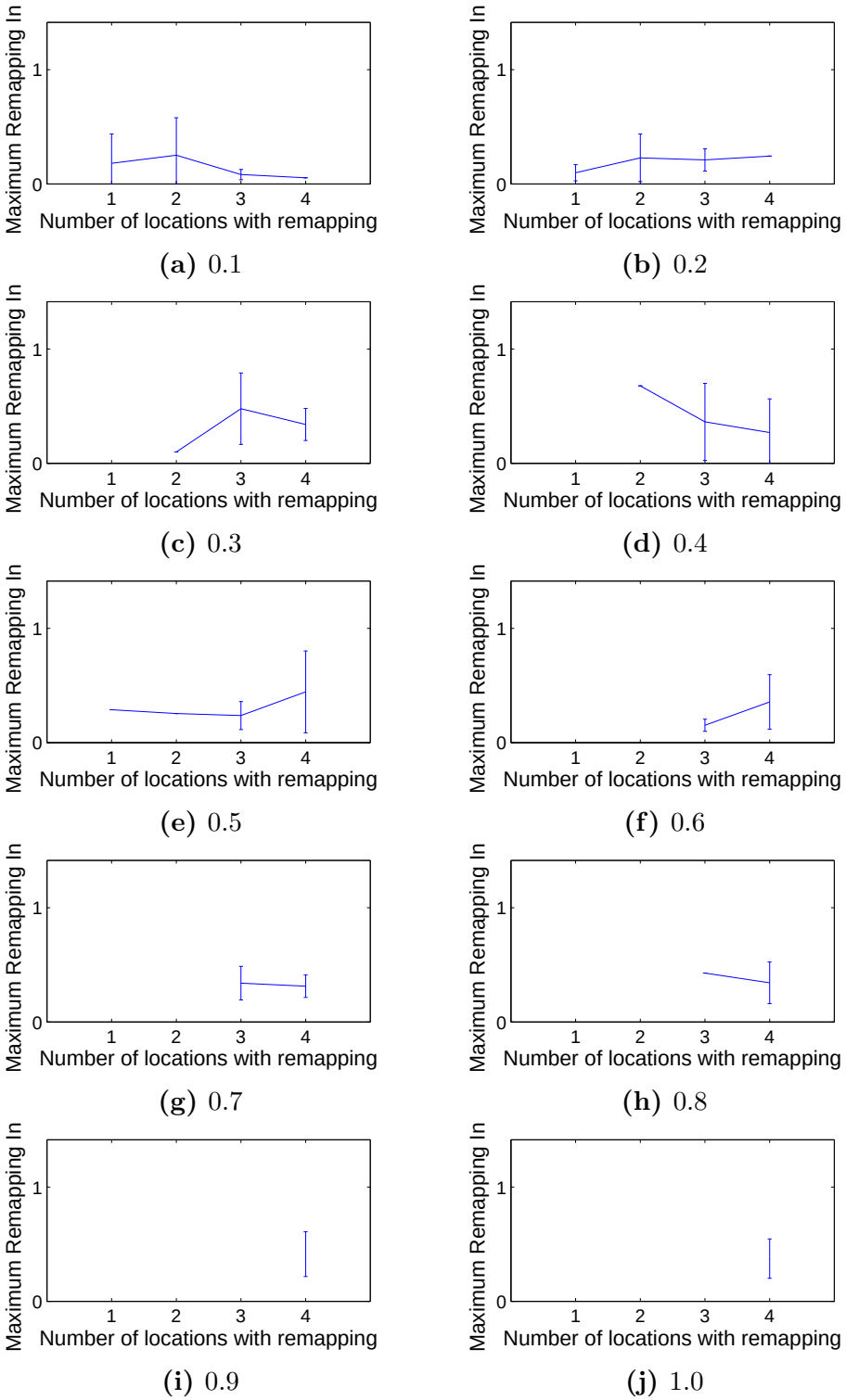
existed neurons that remapped only a single presaccadic stimulus location, the average maximal remapping index was less than the average across neurons remapping multiple locations. This shows that neurons remapping only a single presaccadic stimulus location were not more responsive than those remapping multiple locations, precisely as found in the experimental study of Heiser & Colby (2006).

Remapping is independent of receptive field location

The final question considered was whether remapping neurons with receptive fields distributed throughout the retina had equally selective and strong remapping. Figure 8.26 shows, for simulations with different values of ϕ^{C}, the selectivity index and average remapping index of remapping neurons as a function of the neuron's receptive field location. It is clear that in all cases, no correlation exists between the retinal location of the neuron's receptive field and either of the remapping indices. This shows that remapping neurons with receptive fields distributed across the visual field had comparable selectivity and remapping strength, which was in agreement with the experimental findings of Heiser & Colby (2006).

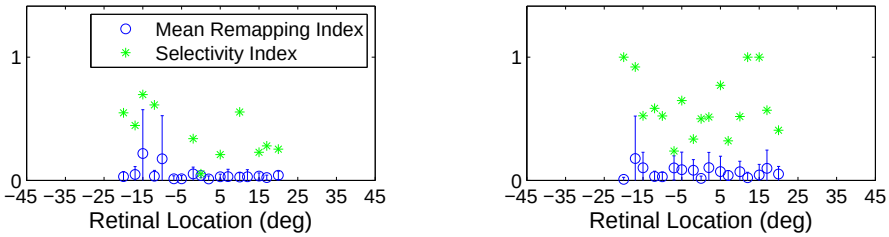
8.3 Discussion

This chapter investigated the feasibility of the hypothesis, described in section 7.3, that the presented neural network model can exhibit a range of experimentally observed perisaccadic



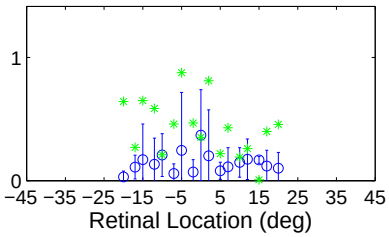
(a) 0.1 **(b)** 0.2 **(c)** 0.3 **(d)** 0.4 **(e)** 0.5 **(f)** 0.6 **(g)** 0.7 **(h)** 0.8 **(i)** 0.9 **(j)** 1.0

Figure 8.25: The average maximal remapping index among remapping neurons responding to the same number of presaccadic stimulus locations. Each plot corresponds to a simulation with a particular value of ϕ^{C}. Within each plot, each remapping neuron is categorised according to the number of presaccadic stimulus locations that it remaps (abscissa), and then the maximal remapping index of neurons was averaged within each category (ordinate). The error bars represent the standard deviations. Missing values indicate that no neurons remapped the given number of presaccadic stimulus locations.

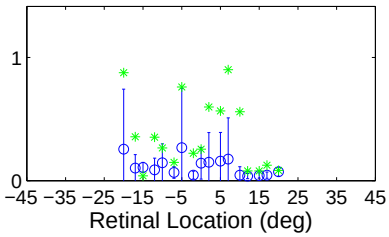


(a) 0.1

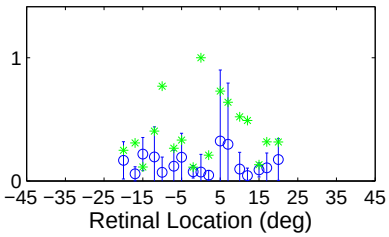
(b) 0.2



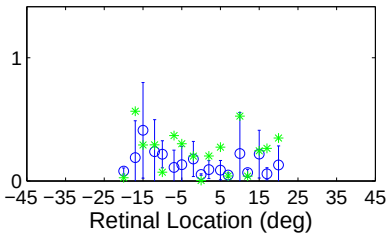
(c) 0.3



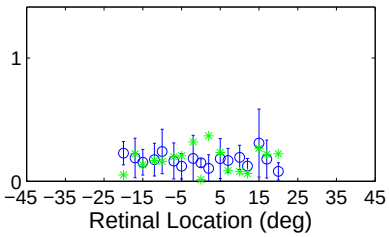
(d) 0.4



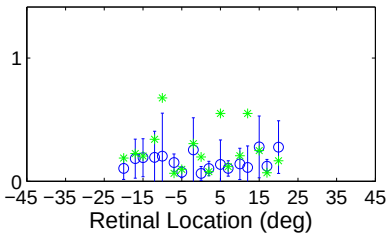
(e) 0.5



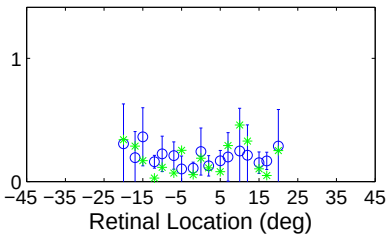
(f) 0.6



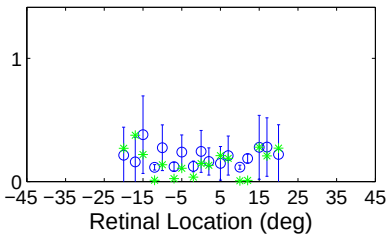
(g) 0.7



(h) 0.8



(i) 0.9



(j) 1.0

Figure 8.26: The selectivity index and average remapping index of remapping neurons as a function of the neuron's receptive field location. Each plot corresponds to a simulation with a particular value of ϕ^{C}. Each data point within a plot corresponds to a different remapping neuron, where the selectivity index (green asterisk) and average remapping index across all remapping trials for the neuron (blue circle) are plotted against the retinal receptive field location of the neuron. The error bars represent the standard deviations of the remapping indices computed across all remapping trials for the given neuron.

receptive field properties.

The model was constructed to have three intrinsic, experimentally observed, receptive field behaviours among visual input neurons that were not products of learning, but which were needed for the successful operation of the model. First, visual input neurons had a latency in their response to the onset of a visual stimulus in their receptive field, and when they did respond it took the form of an initial phasic discharge. Second, this was followed by a longer period of decaying tonic response over several hundred milliseconds. Third, this tonic response was abruptly truncated if the receptive field was removed from the location of a stimulus due to an eye-movement. All three response properties have been observed among visual neurons in LIP (Duhamel, Colby & Goldberg 1992). In the model, these behavours were driven by two external signals, K_i and P_i. It is known that neuronal responses in LIP are modulated by a wide range of experimental and behavioural parameters (Bisley & Goldberg 2010, Kusunoki et al. 2000), and it is not known whether these are produced in LIP itself, or in afferent areas. Hence the causes of these dynamics were excluded from consideration. They were, however, relevant for producing key model behaviours. For example, visual onset latency was important for defining remapping latency, while stimulus offset trace responses and saccade onset response truncation were important for reproducing the results in Kusunoki & Goldberg (2003).

A manually prewired model was constructed and tested to check that the remapping mechanism suggested in the hypothesis could work in principle. In this construction,

combination neurons were set up to have afferent weight vectors corresponding to unique combinations of the presaccadic retinal location of a stimulus and impending saccade, and efferent weight vectors corresponding to the post-saccadic retinal location of the stimulus after the completion of the given saccade. This was found to function properly, and therefore demonstrated the feasibility of the computational mechanism that it was hoped would emerge with learning.

A model with learning, initially set up with random synaptic connectivity and weighting, was exposed to a series of visual stimuli that were stationary in head-centered space across saccades. Around the time of saccade onset the model learned to associate the presaccadic retinal location of the stimulus and the associated saccade with the corresponding postsaccadic retinal location of the stimulus. Through computer simulation, a number of experimentally observed response properties were found to have developed. After training, the model could remap activity into the receptive field of remapping neurons corresponding to the postsaccadic retinal location of the stimulus, if the appropriate combinaton of presaccadic stimulus location and saccade occured. This remapping occurred even though the visual stimulus had been extinguished well in advance of the saccade which would bring the neuron's retinal receptive field over the head-centered location of the stimulus. This was because the remapping was being driven by a transient encoding of the visual scene in a retinal frame of reference. This explains the classic result of Duhamel, Colby & Goldberg (1992) and also Goldberg & Bruce (1990), which found that even a stimulus that had been

extinguished as long as 1s prior to saccade onset would excite a neuron if its receptive field was brought over the head-centered location previously occupied by the stimulus. Moreover, it was found that the remapping neurons in the model would start responding with a much lower latency compared to a visual onset trial. Indeed, most neurons responded even in advance of the saccade onset, resulting in predictive remapping, which was also observed by Duhamel, Colby & Goldberg (1992). This was because the saccade population in the model began encoding the impending saccade in advance. Hence, the coincidence of the encoded saccade and transiently encoded visual scene excited the combination population, which in turn stimulated the remapping neurons, before the saccade had been initiated. A range of response latencies was observed in the model, as has been observed in LIP (Duhamel, Colby & Goldberg 1992),FEF (Umeno & Goldberg 1997) and SC (Walker et al. 1995). However, there was a population bias towards predictiveness which was not reflected in experimental observations. This could, however, be ameliorated by introducing sufficient variability in the $\Delta^{\mathrm{S}}_{\mathrm{PRE}}$ parameter across neurons, which had been kept uniform at 70ms in all experiments.

In the hypothesis for how the model would learn, the latency in the two input populations was thought to be essential in the learning between the combination and remapping population. To explore this, the effect of varying the saccade aligned response latency of the two input populations was investigated. It was found that shorter response latencies during training undermined the network's ability to learn to associate a combination of

presaccadic stimulus location and saccade with the corresponding postsaccadic stimulus location. This was because some time was required for the saccade to be completed, and additional time was required after the saccade for the network to learn the association. This minimal response latency was only required for the learning. When the prewired model was tested, it was able to remap successfully with much shorter response latencies. This showed that the requirement for long response latencies was a constraint imposed by the learning, not the remapping itself. This serves as a good example of how considering the development of a neural circuit may produce constraints on the model beyond what prewired or supervised models may offer.

How might the required postsacadic response latencies in the visual, saccade and combination populations be achieved? Fundamentally, what is required is that the combination neurons are able to maintain their activity until some time after the saccade in order that the network can learn to associate the combination of presaccadic stimulus location and saccade with the corresponding postsaccadic stimulus location. In the model presented, the required response latencies were implemented in the visual and saccade input neurons, which then continued to drive the combination neurons. However, firstly, it is important to emphasise that the visual input population in the model is not explicitly identified with a particular brain area. Instead, it is proposed that as visual signals are propagated between a number of subcortical and cortical brain regions, including being propagated across recurrent connections within individual regions, the axonal transmission delays and

neuronal response delays (in the order of 10ms) add up cumulatively to eventually produce a small subpopulation of visually-driven neurons with suitably long postsaccadic response latencies (in the order of 100ms). This subpopulation of visual neurons may be distributed across a number of linked brain areas. Secondly, neurons may also develop that encode saccades with similarly long response latencies due to the accumulation of axonal transmission delays and neuronal response delays as signals are propagated between and within brain areas. Another possibility is that the response latencies develop after the visual and saccade signals have been brought together by combination neurons. However, regardless of exactly how the latencies accumulate, even if only a small subpopulation of combination neurons distributed across a number of brain regions develop the required postsaccadic response latencies, the model should still be able to learn the statistically regular associations between these combination neurons and remapping neurons, which reflect the correlated activities of these neurons over many thousands of saccades. So, in future modelling work, it is intended to first investigate how a subpopulation of combination neurons may develop these postsaccadic response latencies naturally in a multi-layer neural network model with bi-directional connections between multiple brain regions and recurrent connections within individual regions. Then it is intended to explore whether the network is able to learn associations between these combination neurons representing a combination of presacadic stimulus location and saccade and remapping neurons representing the corresponding postsaccadic stimulus location.

The timecourse of remapping dynamics around the time of a saccade were tested in the same manner as in Kusunoki & Goldberg (2003) and Nakamura & Colby (2000). The population results found in the model agreed qualitatively very well with the experimental results from LIP, and also from visual cortical areas V3A, V3, V2 and V1. These result have been interpreted as a shift in the responsiveness in the current and future receptive field locations around the time of the saccade. Specifically, the responses of LIP neurons to a stimulus in the current receptive field decrease as the stimulus is flashed later with respect to the saccade, while the responses in the future receptive field increase as the stimulus is flashed later with respect to the saccade. The model examined here explains the data with more clarity. First consider the responses of remapping neurons with their receptive field located on the stimulus presaccadically, i.e. the current receptive field trials, and the corresponding population curves shown in figure 8.19. With a stimulus aligned analysis window, the decrease in the neuronal responses emerges as a result of the increasing portions of the analysis window passing the onset time of the saccade, and therefore being void of activity, despite the fact that the timecourse of responses prior to this point is always the same. Second consider the response of remapping neurons with their receptive fields located on the stimulus postsaccadically, i.e. the future receptive field trials, and the corresponding population curves shown in figure 8.19. With a stimulus aligned anlaysis window, the increase in the neuronal responses emerges as a result of the increasing portion of the analysis window passing the onset time of the saccade, and therefore capturing the

remapping activity. However, neither of these scenarios actually requires any change to the timecourse of neuronal responses prior to saccade onset, and hence are perhaps not best described as a change in responsiveness leading up to the saccade. Instead, the apparent change in responsiveness is due to the interaction of saccade onset aligned events such as response truncation with the relatively long analysis windows.

The spatial properties of remapping were tested in a similar manner to Heiser & Colby (2006), where they examined the remapping of stimuli located 20° from the receptive field along one of the four cardinal directions. The experimental results of Heiser & Colby (2006) were presented as being about the effects of saccade direction. However, their findings are perhaps better understood as being about presaccadic stimulus location. When a neuron was found to remap stimuli only in a single cardinal direction, it was classified as unidirectional. However, it is possible that the neuron would have remapped stimuli at other untested distances along the three other cardinal directions, or also along other non-cardinal directions. Likewise, it was quite possible that the neuron would not have remapped stimuli at all other distances along the cardinal direction for which it did display remapping. Hence, their experimental results are best interpreted in a more general way as about remapping from different stimulus locations. Indeed, although the model tested in this chapter could develop some remapping neurons that failed to learn to remap from all four presaccadic locations that they were trained on, this effect was quite stochastic and would be shaped by random factors such as the diluted connectivity and initial strengths of

the afferent connections to the combination neurons. In other words, with regular unbiased training, there was no systematic way in which remapping neurons could learn a regular bias in terms of their selectivity to remapping from four cardinal directions as tested by Heiser & Colby (2006). As a result, the model was examined for remapping from four random directions and distances.

Multiple key results from the original paper of Heiser & Colby (2006) were confirmed in the model. Many remapping neurons remapped in multiple directions. However, there was substantial variability among remapping neurons which depended critically on the rate of connectivity between the combination population and remapping population. Lower connectivity caused remapping neurons to learn to remap in fewer directions, as expected. Individual remapping neurons also had a range of selectivities across their four presaccadic stimulus locations, and low connectivity rates were associated with the high level of selectivity observed in the experiments. A preferred presaccadic stimulus location was decoded for each remapping neuron, and the preference distribution in the population revealed no underlying bias. This showed that stimuli could be remapped from any retinal location.There was in general no relationship between the number of remapping locations and the strength of remapping. These findings, as noted in Heiser & Colby (2006), are consistent with the proposal that LIP is involved in producing visual perceptual constancy.

A range of models have previously been proposed to explain various remapping phenomena that the model presented here has sought to account for. Some of these earlier

modelling studies are described next.

The early work of Droulez & Berthoz (1991) showed that an activity packet could be moved within a topographic population of neurons by an external eye velocity signal. Their model shared some architectual similarities to the work presented here. However, it required topographic organization of the neural population, something which is not prevalent in LIP (Blatt et al. 1990). Their model also predicted that activity would move continuously through the neural population, guided by an eye velocity signal. However, the model is not compatible with predictive and presaccadic remapping, nor with the variability in remapping latency across neurons. Lastly, the circuit elements in the model switch between being excitatory or inhibitory depending on the eye velocity input, something which is not thought to be possible in biological neurons (O'Donohue et al. 1985).

The work of Krommenhoek et al. (1993), Mitchell & Zipser (2001) and White & Snyder (2004) represent classic examples of models organized through biologically implausible processes of supervised error correction learning to produce the correct motor error for double-step saccade tasks. The two models of Krommenhoek et al. (1993) required eye position at the time of target selection as an explicit input signal, and were able to combine this with the current eye position and retinal target location to produce the correct motor error. It does, however, seem unclear why there should be an explicit mechanism for saving eye position at the initation of double-step saccades given the highly unrealistic nature of the task. The model of Mitchell & Zipser (2001) was a shifting packet model, much in the

spirit of Droulez & Berthoz (1991), which was found to be able to produce outputs in both eye-centered and head-centered reference frames. The model of White & Snyder (2004) also had the same basic structure. However, it was trained to remap gaze fixed stimuli when a separate reference frame cue was turned on without an adequate explanation for where such a cue might actually derive from in the brain. These three models thus offered no plausible mechanism by which they could self-organize their synaptic connectivity.

The work of Quaia et al. (1998) is most similar to the model developed here. It had a very similar basic architecture, where a module in LIP was responsible for shifting activity between different parts of a retinal population in LIP, based on the activity in a population of FEF neurons encoding impending saccades. The model was hardwired to produce the correct remap, in a similar manner to the first experiment in this chapter. Like the model developed here, it also had a saccade aligned damping signal that produced response truncation. However, in contrast, it produced a range of remapping latencies by having remapping activity flow back and forth between phasic neurons in FEF and LIP. While the authors do suggest that some of the primary circuitry could self-organize in a similar manner to the model presented here, it still has some important shortcomings. First, its function depends on very detailed dendritic operations and wiring that could not be accounted for by self-organization. Second, the model depends on topographically aligned nesting between LIP and FEF, which is not likely to be possible given that LIP has very weak topography (Blatt et al. 1990).

A potentially fruitful further line of inquiry based on the initial work presented in this chapter could be to attempt to interpret the range of physiological and behavioral results associated with split brain macaques performing interhemispheric stimulus remapping. It has been found that both behavioural performance during double-step saccades (Berman et al. 2005*a*) and interhemispheric remapping in such trials (Heiser et al. 2005) is influenced by commissurotomizing the forebrain commissures, and that the two are related (Berman et al. 2007). Moreover, improvement in the performance of these tasks after initial impairment was found, and neural plasticity was thought to be a primary candidate for this improvement (Berman et al. 2005*a*). This could potentially be modeled by commissurotomizing the model and attempting to retrain it in accordance with the behavioural protocol. Another issue this model could help shed light on is how multiple interconnected areas, such as LIP, FEF and SC, interact in producing their remapping responses. It is not known whether each area independently generates its own remapping, or whether it is a distributed process.

There are also a number of important shortcomings of the work in this chapter which would benefit from future investigation. First, the model was only trained on visual scenes with a single visual target visible at any time, while natural visual scenes most often have multiple salient visual targets. Under the assumption of visual scenes with uncorrelated visual target locations, as used in section 4.2.6, one would hope and expect that remapping neurons would learn connections to the correct combination neurons over time, as

any given incorrect combination neurons would only drive them spuriously. Second, the inputs to the combination population were not plastic, and future work should attempt to study how learning in these may be important. Third, the neurons in the remapping population did not learn their visual responses, and future work should attempt to learn these simultaneously with the remapping responses.

8.4 Summary

The following is a summary of the main results and insights in this chapter.

- It was possible to manually hardwire the feedforward synaptic connections in a model with two separate populations of input neurons encoding the presaccadic retinal stimulus location and impending saccade, combination neurons detecting particular combinations of the two inputs, and remapping neurons that remapped this input activity to the corresponding postsaccadic location of the visual stimulus.

- The model can develop stimulus remapping by implementing the visually-guided learning mechanism described in section 7.3.

- If the saccade onset latency in the visual and saccade input populations is not sufficiently long, then the network is unable to learn an association from the presaccadic stimulus location and saccade encoded by the combinbation neurons to the corresponding postsaccadic stimulus location encoded by the remapping neurons.

- When the model was tested in a delayed stimulus flash task, the results agreed well with the experimental observations of Kusunoki & Goldberg (2003). The decreasing responsiveness of remapping neurons in current receptive field trials as the stimulus was flashed later with respect to saccade could be attributed to saccade onset aligned activity truncation, and not any change in the feedforward visual sensitivtiy of remapping neurons. Likewise, the increasing responsiveness of remapping neurons in the future receptive field trials as the stimulus was flashed later with respect to the saccade could be attributed to the stimulus aligned analysis window capturing more of the remapping activity, thus registering a larger response.

- When the model was trained to remap multiple presaccadic retinal stimulus locations into each postsaccadic retinal stimulus location, the results agreed well with the experimental observations of Heiser & Colby (2006). Individual remapping neurons were able to remap from multiple locations. The population distribution of the number of presaccadic locations that individual remapping neurons could remap was skewed towards a larger number of locations as the connectivity rate between the combination and remapping populations was increased. There was also variability among remapping neurons in their selectivity for remapping from particular presaccadic locations, and for low connectivity rates there was a bias towards more selective neurons. However, at the population level, activity could be remapped from the entire retina. Neurons remapping only a single location were not more responsive than

those remapping multiple locations, and neurons with receptive fields distributed across the visual field had comparable selectivity and remapping strengths.

Chapter 9

Conclusion

This chapter provides a unified summary of the findings in the thesis and discusses potential avenues for future work.

The work in this thesis was an investigation into a set of non-classical visual receptive field properties observed in the primate brain. This investigation was conducted by inspecting the behaviour of mathematical neural network models, instantiated as computer simulations, when placed in conditions similar to experimental tasks used to interrogate the physiology of primate neural circuits. The models were investigated both before and after learning, during which they where subjected to external stimuli corresponding to the learning hypothesis for the given model. The difference in the behaviour of the model, as a result of training, and in particular the degree to which the post-training behaviour was in agreement with empirical observations in primates, was used as a metric for the explanatory power of the models.

The two main phenomena investigated were the head-centered visual receptive fields, in chapters 1 to 5, and perisaccadic receptive field dynamics, in chapters 6 to 8. The former refers to a visual receptive field which responds maximally to a visual stimulus in the same head-centered location across all eye positions. The latter refers to a set of observed neuronal response patterns associated with the advent of a saccade which relocates an eye-centered receptive field. These two phenomena were chosen because they are among the most well known and well studied, both empirically and theoretically, examples of how

visual and motor information is integrated in a behaviourally relevant way. This is in general known as sensorimotor integration, in this case in the oculomotor system. This system has been studied in great detail in the context of such integration because of its relative simplicity compared to other motor systems, while still preserving the fundamental problem of how sensory stimuli are integrated with motor information, so as to build neural representations suitable for ultimately guiding motor action.

The development of head-centered receptive fields was investigated by proposing, testing and examining a hypothesis for how such receptive fields could emerge. The proposed hypothesis was that a model with the following core features was sufficient: (i) the input neurons encode the retinal target location and eye position in a coupled manner, (ii) the output neurons compete with each other through inhibitory interactions, (iii) the synaptic weights are modified by a learning rule that incorporates a memory trace of recent neural activity in order to effect temporal binding of temporally proximal input patterns, and (iv) there are periods during training when the eyes are moving while the head and visual target remain stationary.

When considering input neurons with peaked eye position gain fields and Gaussian tuning curves, as investigated in chapters 3 and 4, then this learning hypothesis implicitly suggested a specific functional circuit. This functional structure was tested directly in a manually hardwired model, and was shown to succeed. This confirmed the existence of a suitable network synaptic weight structure for the given network architecture, and also that the conjectured result of the hypothesis was indeed one such sufficient weight structure. The hypothesis was directly tested in an experiment where the model was required to make sequences of saccades while visual targets remained stationary in head-centered space for periods of time. It was found that this resulted in the development of neurons with head-centered visual receptive fields in the output population, as hypothesized.

A set of follow up experiments were conducted to confirm the necessity of each of the components in the hypothesis. First, the visual and eye position components of the recep-

tive fields of the input neurons were decoupled, resulting in an input population where the eye position and retinal location of visual stimuli were encoded disjointly. This lead to a collapse of the model performance, and no head-centered neurons emerged. This was because each input neuron would participate equally in the representation of large portions of head-centered space, hence it was not useful in allowing output neurons to develop a preference for a specific location in head-centered space. Second, the competition was effectively eliminated in the output layer by setting the sparseness percentile to $\pi = 0\%$. This lead to the development of output neurons with a very low selectivity in terms of receptive field size in head-centered space. Third, the synaptic learning rule was modified in two ways to explore the consequences of decreasing its efficacy in binding together temporally proximal input patterns. First, the original trace rule was modified by decreasing the time constant τ_q governing the time course of the trace term q_i. Second, a standard Hebbian learning rule based on the pre and postsynaptic firing rate was also explored across a range of values for the time constant τ_h governing the time course of the activation term h_i. In both cases, reducing the time constant led to reduction, and eventual absence, of neurons developing head-centered receptive fields during training. Moreover, when the time constants were the same, the model with a trace learning rule was was more efficacious in developing head-centered receptive fields compared to the Hebbian learning rule. Fourth, the relative frequency of eye movements compared to relocations of stimuli in head-centered space was reduced. This also lead to an initial decline, and eventual absence, of head-centered receptive fields in the output population. In total, these four findings confirmed the individual role and necessity of each core model feature in the hypothesis.

Various model parameters, such as learning rate, sparsity and competition model, were altered to confirm the robustness of the model behaviour.

After these basic experiments, which provided initial confirmation of the viability of the hypothesis, a series of experiments were conducted in order to explore if, and how well, the model would behave when the training stimuli and the model were altered to better

approximate ecological validity. The number of training locations within which the head-centered visual stimuli could be located during training was increased significantly, so as to approximate the continuous sampling of head-centered space during development. This lead to an increase in the number of neurons developing head-centered receptive fields, and the size of these receptive fields was becoming smaller, and therefore more selective. The full range of head-centered space was also evenly represented.

The receptive field size of input neurons was increased in both the eye position and retinal space. This lead to the collapse of the even representation of head-centered space and decline of prevalence of head-centered receptive fields. This failure was hypothesized to be due to increasing interaction between the edges of the neural input representation and the output representation of head-centered spatial locations. The failure was ameliorated by altering the synaptic and neuronal dynamics so as to impose a non-linearity on the influence of presynaptic firing on postsynaptic learning and activation. This alteration reduced the effective size of presynaptic receptive fields, and it re-established successful model performance after training.

The prevalence of random stimulus dynamics in the training stimuli was varied to explore the robustness of the model to more ecological stimulus dynamics. In all other experiments the model was subjected to perfectly stationary head-centered visual stimuli inspected by sequences of saccadic eye movements. In this experiment, such sequences were interleaved with sequences of random eye position and retinal locations of stimuli of varying length, corresponding to self-movement and stimulus movement. It was found that even for the longest sequences of such non-ideal training stimuli, the model was able to develop head-centered neurons at significant rates.

The duration of fixation was varied over a wide range, and for sufficiently long duration periods, the model failed to develop head-centered neurons. This was attributable to a forgetting effect where the last fixation dominated the weight vectors of neurons at the expense of previously accommodated input patterns. Altering the learning rule slightly, by

introducing an extra explicit upper bound on the weight of individual synapses, overcame this problem, and also allowed the model to deal with random variability in the duration of fixation over a wide range of standard deviations of fixation duration.

The eye position component of the receptive field of input neurons, or gain field, was altered to have two peaks of varying random size in eye position space, as has been observed in experimental studies. The model was still able to learn properly with this more ambiguous and ecological input representation.

The training stimuli were altered so that multiple visual targets were always visible in the designated training locations, so as to approximate natural training conditions which will include multiple salient stimuli frequently, if not at all times. Ensuring that all combinations of locations were observed equally lead to successful model performance after training, just as in the single stimulus training case. However, when this, and the original model, were tested with multiple visual stimuli present at any given time, it was found that the competitive interaction between the output representation of each stimulus lead to the suppression of the response of otherwise well behaved output neurons for some eye positions. In total, these findings show that the model is robust to many changes approximating more ecological, and non-ideal, conditions of development.

When considering input neurons with monotonic eye position gain fields, as investigated in chapter 5, the representation in the input layer had to be changed. In previous experiments, the gain fields took the form of a Gaussian peaked response function in the eye position space, but for these experiments it was substituted with a sigmoid curve in the eye position space. At the single neuron level, the gain field was characterized by two parameters, the inflection point and the slope sign. The inflection point determined the location in the space at which the sigmoid was centered, and the slope sign determined whether the gain field increased or decreased in terms of eye position. This meant that the size of the input population had to be doubled, since for each combination of retinal location and eye position there were two distinct input units, each with a different sign

gain field. This meant that input units now effectively encoded whether a stimulus was in a given retinal location and whether the eyes were to the left, or right, of a given reference eye position.

With this new input representation, the functional circuit suggested by the learning hypothesis was a bit different. This functional structure was tested directly in a manually hardwired model, and was shown to succeed. This confirmed the existence of a suitable network synaptic weight structure for the given network architecture. However, when plasticity was introduced and the model was trained in the exact same manner as the first model with peaked eye position gain fields, the outcome was very different. There was a dramatic decline in head-centeredness and dramatic increase in eye-centeredness, as a result of training, that is the opposite of the desired and conjectured outcome. The weight vectors in this trained model did not correspond to input patterns associated with a given head-centered location in space, but rather the structure in the response correlations between neurons in the input population. Hence it seemed that only with peaked gain fields was the model able to achieve the proper temporal association, and with monotonic gain fields this failed. Critically, even if the manually prewired model with monotonic gain fields, which had successfully produced head-centered output neurons, was subsequently subjected to visual stimuli combined with synaptic plasticity, then it would fail in the same manner. This showed that, independently of whatever neurodevelopmental process established the working neural circuitry for the head-centered receptive fields, any remaining level of synaptic plasticity would undermine the functioning of the circuit. These findings showed the problems associated with such input representations, so it was explored how well the model could perform when peaked and monotonic gain fields were mixed at various rates in the input population. It was found that the fraction of head-centered neurons after training remained above roughly 15% so long as less than one in five neurons had monotonic gain fields. To better understand how other models of coordinate transformation coped with the same input gain fields, some well known neural network models were replicated

and studied. The primary way in which they overcame the problem was either by relying on error-correction, which also had other undesirable consequences for the resulting neural circuit, or by relying on an external hypothetical supervisory signal. However, in this last case the problem of remaining synaptic plasticity was still a problem.

This work presented and examined a model instantiating a hypothesis for how head-centered visual receptive fields in primates could operate. The focus on development revealed that monotonic gain fields were an obstacle for learning, and even mere compatibility with synaptic plasticity. This may shed some light on the functional significance of the peaked and non planar gain fields which have not received as much attention in the past. It also highlights the need to distinguish between the possibility that some processing can be implemented as a neural circuit, and whether such a circuit actually could develop. Considerations about the latter may serve as a useful constraint when considering candidate neural mechanisms behind some behaviour or observation.

While this work established the baseline result, and also explored issues of ecological validity, a meaningful next step could be to train the model on natural eye and head movements recorded from human subjects. That is likely to reveal further important insights about the viability of this hypothesis, beyond what can be revealed by a one dimensional model with artificially generated stimuli. Another important avenue of future research would be to develop new ideas about what the role of monotonic eye position gain fields could be, and in what alternative model could they support the development of head-centered receptive fields in a biologically plausible way. The fundamental idea of using temporal association for performing coordinate transformation could also possibly be extended at least one more step by recognizing that visual stimuli in the world are likely to be even more stable in a body-centered reference frame, compared to head-centered and eye-centered space.

A set of perisaccadic phenomena were investigated by proposing, testing and examining a neural network model of such receptive field dynamics. There were seven main observa-

tions which the hypothesis was intended to account for: predictive remapping, presaccadic remapping, trace remapping, stimulus offset trace, saccade onset trace truncation, responsiveness shifts around the time of saccades, and spatially independent remapping. All of these are precisely defined in chapter 6.

The model had two eye-centered input populations, encoding the locations of visual stimuli and the impending saccade, projected onto a competitive population, called the combination population, which again projected onto an eye-centered visual population encoding the locations of visual stimuli, called the remapping population.

The learning between the combination and remapping population was hypothesized to work as follows. The combination of a visual target and an impending saccade would cause some neurons in the combination population to start firing. The saccade would relocate the stimulus, assumed to be stationary in a head-centered reference frame, to a new retinal location. This location would be encoded by the remapping population after saccade completion. Due to the postsaccadic latency in both input populations, the combination population would remain invariant for some period after saccade completion. This would cause the projections from the combination population to the remapping population to be modified by supervised pattern association learning. After this learning process, the combination population would always excite the neurons in the remapping population corresponding to the postsaccadic retinal location of the stimulus. In this way, a given subset of combination neurons could effectively bind the presynaptic retinal stimulus location and saccade with the corresponding new retinal location of the same stimulus after completion of the saccade. This architecture would account for predictive, presaccadic and trace remapping. The two trace phenomena were not endogenously explained by the model, but rather attributed to an external signal, which caused a tonic response and a saccade aligned truncation respectively. The responsiveness shift and spatially independent remapping was hypothesized to follow from the combination of the remapping and trace phenomena.

In an initial experiment, a model was manually hardwired to perform remapping by assigning a unique combination of a presynaptic retinal location and saccade to each combination neuron, and setting up its afferent and efferent weight vectors to reflect this. This led to successful predictive and presaccadic remapping as required, and it was also possible to confirm that the exogenous trace phenomena, that is decay and truncation, were both functioning properly. Subsequently, the model was trained as described, that is performing saccades while a stationary visual stimulus was located in different head-centered locations. After this, the same model was tested by exposing it to task trials with the same combinations of saccades and stimulus locations. This also led to the development of predictive and presaccadic remapping, and due to the choices of model parameters, all inspected neurons in the remapping population were in fact remapping presaccadically. Inspecting the structure of the resulting weight vectors, and also population activity during testing trials, revealed that the model operated as hypothesized.

A control experiment was conducted to investigate the impact of reducing the saccade aligned latency in the two input populations on remapping. The self-organizaiton hypothesis states that some delay is required to facilitate associative learning between the combination and remapping populations in some postsaccadic time interval. Reducing the period of latency in the input populations was found to reduce, and eventually eliminate, remapping. To show this clearly and conclusively, the manually wired model was tested with shorter latencies as well, but in this case remapping remained. This showed that successful self-organization depended on a sufficiently large latency.

The model was then tested in trials where the stimulus was flashed at various times with respect to saccade onset, and also in either the presaccadic or postsaccadic retinal locations of the remapping neuron's receptive field. The response of the remapping neuron was found to decrease as the stimulus onset time occurred later with respect to the saccade when the stimulus was flashed in the presaccadic retinal location of the receptive field. This could be explained by the truncation of the trace response of the neuron when the stimulus

was flashed before saccade onset, and when flashed after saccade onset it was simply never flashed in the receptive field of the neuron. The response of the remapping neuron was found to increase as the stimulus onset time occurred later with respect to the saccade when the stimulus was flashed in the postsaccadic retinal location of the receptive field. This could be explained by the remapping activity of these neurons when the stimulus was flashed before saccade onset, and when flashed after saccade onset the stimulus was simply flashed in the receptive field of the neuron. These insights accounted for the observed changes in the responsiveness of remapping neurons around the time of saccades in terms of already well known perisaccadic phenomena.

The model was finally trained to remap multiple presynaptic stimulus locations into the receptive field of a given remapping neuron in order to investigate the spacial properties of remapping. The model was tested on all of the remapping configurations that it had been trained on, and the testing and training were done across a range of connectivity rates between the combination population and remapping population. This connectivity parameter was inherently relevant to the question of spatial remapping since it controlled the likelihood of successful associative learning between the two populations. A number of observations were made that were in strong agreement with experimental results. Single remapping neurons would remap from multiple locations, and this increased with increasing connectivity between the combination and remapping populations. There was varying selectivity in the remapping strength of individual remapping neurons across the multiple presynaptic stimulus locations, and neurons became more selective with decreasing connectivity. Activity could be remapped from the entire retina at the population level. Neurons remapping only a single location were not more responsive than those remapping multiple locations. Neurons with receptive fields across the visual field had comparable selectivity and remapping strength.

The majority of previous modelling work has been with neural network models that were either hardwired or organized through an implausible process of supervised error-correction

learning, and which did not account for such a wide range of experimentally observed phenomena. The model investigated here thereby constitutes the first unsupervised model to account for such a wide range of empirical observations. For the learning process to work effectively, the visual input neurons needed to keep responding for a short time after the saccade onset. The focus on learning in this study has thus led to a prediction about the existence of visual neurons in the relevant areas of the primate brain with relatively long response latencies of the order of a 100ms.

There are a number of avenues for future modelling research. First, an open question in the literature is how the remapping taking place across multiple brain areas is related, and if so, which signals in which brain areas are the true origin of the responses observed in other areas. The model investigated here could be extended to more closely correspond anatomically to the connectivity between the relevant areas of the primate brain. Such a unified model could be used to examine the division of labour among these areas in terms of driving the perisaccadic remapping phenomena. Secondly, recent experimental work has examined the effect of commissurotomy of the forebrain commissures on interhemispheric remapping and double-step saccades (Berman et al. 2007, 2005*b*, Heiser et al. 2005). It has been found that there are changes in neurophysiology and behaviour over time after such procedures, indicating a possible role of neuroplasticity in organizing and reorganizing such neural circuits. The model developed here would be well suited for studying and explaining such observations.

Appendices

Appendix A

Appendix: Eye-Centeredness Reference Frame Analysis

During testing, the visual target was located in head-centered locations

$$t_j = t_1 + \Delta h(j-1) \tag{A.1}$$

for $j = 1, \ldots, T$, and while in each location it was observed from eye positions

$$e_i = e_1 + \Delta e(i-1) \tag{A.2}$$

for $i = 1, \ldots, E$. The eye position shift Δe was set to a multiple of the head centered target location shift Δh to cause resampling of the neuron's response at the same retinal location for different eye positions, thereby providing a resampling of the response in both head-centered and eye-centered space across multiple eye positions.

The set of head-centered locations $\{t_1, \ldots, t_T\}$ corresponded to retinal locations $R_i = \{t_1 - e_i, \ldots, t_T - e_i\}$ when the model was fixating eye position e_i, and from this it is clear that among retinal locations common to all eye positions, $t_1 - e_1$ was the first and $t_T - e_E$ was the last, that is

$$\begin{aligned} \min \bigcap_i R_i &= t_1 - e_1 \\ \max \bigcap_i R_i &= t_T - e_E \end{aligned} \tag{A.3}$$

Therefore f_i and l_i, denoting the first and last position included from the i^{th} response vector respectively, had to correspond to these two retinal locations respectively

$$t_{f_i} - e_i = t_1 - e_1 \tag{A.4}$$

$$t_{l_i} - e_i = t_T - e_E \tag{A.5}$$

We can resolve each equation to find an explicit formula for f_i and l_i in terms of i as follows.

By substituting equations A.1 and A.2 into equation A.4 we obtain

$$(t_1 + \Delta h(f_i - 1)) - (e_1 + \Delta e(i - 1)) = t_1 - e_1$$

Rearranging this gives

$$\Delta h(f_i - 1) - \Delta e(i - 1) = 0$$

$$f_i = \frac{\Delta e}{\Delta h}(i - 1) + 1 \tag{A.6}$$

By substituting equations A.1 and A.2 into equation A.5 we obtain

$$(t_1 + \Delta h(l_i - 1)) - (e_1 + \Delta e(i - 1)) = (t_1 + \Delta h(T - 1)) - (e_1 + \Delta e(E - 1))$$

Rearranging this gives

$$\Delta h(l_i - 1) - (e_1 + \Delta e(i - 1)) = \Delta h(T - 1) - (e_1 + \Delta e(E - 1))$$

$$\Delta h(l_i - 1) - \Delta e(i - 1) = \Delta h(T - 1) - \Delta e(E - 1)$$

$$\Delta h(l_i - 1 - (T - 1)) = -\Delta e(E - 1 + (i - 1))$$

$$\Delta h(l_i - T) = -\Delta e(E + i)$$

$$l_i = T - \frac{\Delta e}{\Delta h}(E + i) \tag{A.7}$$

We can also deduce the length V of the portion of each response vector that is used in the eye centered correlation analysis as follows. By definition, for each response vector

$$V = l_i - f_i + 1$$

Substituting in equations A.6 and A.7 gives

$$V = T - \frac{\Delta e}{\Delta h}(E + i) - \left(\frac{\Delta e}{\Delta h}(i - 1) + 1\right) + 1$$

Rearranging gives

$$\begin{aligned} V &= T - \frac{\Delta e}{\Delta h}(E + i) - \frac{\Delta e}{\Delta h}(i - 1) \\ &= T - \frac{\Delta e}{\Delta h}(E + i - (i - 1)) \\ &= T - \frac{\Delta e}{\Delta h}(E + 1) \end{aligned} \tag{A.8}$$

Appendix B

Appendix: Calculation of Head-Centered Receptive Field Location

For each output neuron, the head-centered receptive field location $\hat{h}_i$ for the i^{th} eye position was calculated using the center of mass of the head-centered response vector at this eye position, that is

$$\hat{h}_i = \frac{\sum_{j=1}^{T} t_j \mathbf{R}[i,j]}{\sum_{j=1}^{T} \mathbf{R}[i,j]} \tag{B.1}$$

The values $\hat{h}_1, \ldots, \hat{h}_E$ may be regarded as estimates of a single global head-centered receptive field location h across all eye positions. In this case, the error $E(h)$ between the set of estimates $\hat{h}_1, \ldots, \hat{h}_E$ and the head-centered receptive field location h was defined as

$$E(h) = \sum_{i=1}^{E} \left(\hat{h}_i - h\right)^2 \tag{B.2}$$

This error was minimized over $h \in \mathbb{R}$ by first finding the derivative of this function

$$\begin{aligned}\frac{dE}{dh} &= \sum_{i=1}^{E} \frac{d\left(\left(\hat{h}_i - h\right)^2\right)}{dh} \\ &= -2\sum_{i=1}^{E}\left(\hat{h}_i - h\right) \\ &= -2\left(\sum_{i=1}^{E}\hat{h}_i - Eh\right)\end{aligned} \tag{B.3}$$

and then finding the critical points h^* by solving $\frac{dE}{dh}(h^*) = 0$

$$\begin{aligned}0 &= -2\left(\sum_{i=1}^{E}\hat{h}_i - Eh^*\right) \\ 0 &= \sum_{i=1}^{E}\hat{h}_i - Eh^* \\ Eh^* &= \sum_{i=1}^{E}\hat{h}_i \\ h^* &= \frac{1}{E}\sum_{i=1}^{E}\hat{h}_i\end{aligned} \tag{B.4}$$

The fact that $\frac{dE(h_1)}{dh} < 0 < \frac{dE(h_2)}{dh}$ when $h_1 < h^* < h_2$ shows that h^* minimizes E. Therefore the head centred receptive field location of a given neuron was given by

$$\frac{1}{E}\sum_{i=1}^{E}\frac{\sum_{j=1}^{T} t_j \mathbf{R}[i,j]}{\sum_{j=1}^{T}\mathbf{R}[i,j]} \tag{B.5}$$

Bibliography

Andersen, R. A. (1987), 'Inferior parietal lobule function in spatial perception and visuomotor integration', *Comprehensive Physiology* .

Andersen, R. A. (1989), 'Visual and eye movement functions of the posterior parietal cortex', *Annual Review of Neuroscience* **12**, 377–403.

Andersen, R. A. (1995), 'Encoding of intention and spatial location in the posterior parietal cortex', *Cerebral Cortex* **5**(5), 457–469.
URL: *http://cercor.oxfordjournals.org/content/5/5/457.abstract*

Andersen, R. A., Asanuma, C., Essick, G. & Siegel, R. M. (1990), 'Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule', *The Journal of Comparative Neurology* **296**(1), 65–113.
URL: *http://onlinelibrary.wiley.com/doi/10.1002/cne.902960106/abstract*

Andersen, R. A. & Mountcastle, V. B. (1983), 'The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex', *The Journal of Neuroscience* **3**(3), 532–548.

Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. (1997), 'MULTIMODAL REPRESENTATION OF SPACE IN THE POSTERIOR PARIETAL CORTEX AND ITS USE IN PLANNING MOVEMENTS', *Annual Review of Neuroscience* **20**(1), 303–330.
URL: *http://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.20.1.303*

Andersen, R., Bracewell, R., Barash, S., Gnadt, J. & Fogassi, L. (1990), 'Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque', *The Journal of Neuroscience* **10**(4), 1176–1196.
URL: *http://www.jneurosci.org/content/10/4/1176.abstract*

Andersen, R., Essick, G. & Siegel, R. (1985), 'Encoding of spatial location by posterior parietal neurons', *Science* **230**(4724), 456–458.
URL: *http://www.sciencemag.org/content/230/4724/456.abstract*

Barbas, H. & Mesulam, M.-M. (1981), 'Organization of afferent input to subdivisions of area 8 in the rhesus monkey', *Journal of Comparative Neurology* **200**(3), 407–431.

Barto, A. G. (1985), 'Learning by statistical cooperation of self-interested neuron-like computing elements', *Human Neurobiology* **4**(4), 229–256.

Berman, R. A., Heiser, L. M., Dunn, C. A., Saunders, R. C. & Colby, C. L. (2007), 'Dynamic circuitry for updating spatial representations. III. from neurons to behavior', *Journal of Neurophysiology* **98**(1), 105–121.
URL: *http://jn.physiology.org/content/98/1/105.abstract*

Berman, R. A., Heiser, L. M., Saunders, R. C. & Colby, C. L. (2005*a*), 'Dynamic circuitry for updating spatial representations. i. behavioral evidence for interhemispheric transfer in the Split-Brain macaque', *Journal of Neurophysiology* **94**(5), 3228–3248.
URL: *http://jn.physiology.org/content/94/5/3228.abstract*

Berman, R. A., Heiser, L. M., Saunders, R. C. & Colby, C. L. (2005*b*), 'Dynamic circuitry for updating spatial representations. i. behavioral evidence for interhemispheric transfer in the split-brain macaque', *Journal of neurophysiology* **94**(5), 3228–3248.

Bisley, J. W. & Goldberg, M. E. (2010), 'Attention, intention, and priority in the parietal lobe', *Annual Review of Neuroscience* **33**, 1–21.
URL: *http://www.annualreviews.org/doi/abs/10.1146/annurev-neuro-060909-152823*

Bizzi, E. (1968), 'Discharge of frontal eye field neurons during saccadic and following eye movements in unanesthetized monkeys', *Experimental Brain Research* **6**(1), 69–80.

Blatt, G. J., Andersen, R. A. & Stoner, G. R. (1990), 'Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area lip) in the macaque', *Journal of Comparative Neurology* **299**(4), 421–445.

Breveglieri, R., Bosco, A., Canessa, A., Fattori, P. & Sabatini, S. P. (2009), Evidence for peak-shaped gaze fields in area v6a: implications for sensorimotor transformations in reaching tasks, *in* 'Bioinspired Applications in Artificial and Natural Computation', Springer, pp. 324–333.

Brotchie, P. R., Andersen, R. A., Snyder, L. H. & Goodman, S. J. (1995), 'Head position signals used by parietal neurons to encode locations of visual stimuli', *Nature* **375**(6528), 232–235.
URL: *http://dx.doi.org/10.1038/375232a0*

Bruce, C. J. & Goldberg, M. E. (1985), 'Primate frontal eye fields. i. single neurons discharging before saccades', *Journal of Neurophysiology* **53**(3), 603–635.

Cohen, Y. E. & Andersen, R. A. (2000), 'Reaches to sounds encoded in an Eye-Centered reference frame', *Neuron* **27**(3), 647–652.
URL: *http://www.sciencedirect.com/science/article/pii/S0896627300000738*

Colby, C. L. & Duhamel, J. (1991), 'Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey', *Neuropsychologia* **29**(6), 517–537.
URL: *http://www.sciencedirect.com/science/article/pii/002839329190008V*

Colby, C. L. & Goldberg, M. E. (1999), 'SPACE AND ATTENTION IN PARIETAL CORTEX', *Annual Review of Neuroscience* **22**(1), 319–349.
URL: *http://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.22.1.319*

Dayan, P. & Abbott, L. F. (2001), *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, ISBN 0-262-04199-5.

Droulez, J. & Berthoz, A. (1991), 'A neural network model of sensoritopic maps with predictive short-term memory properties.', *Proceedings of the National Academy of Sciences* **88**(21), 9653–9657.

Duhamel, Colby, C. & Goldberg, M. (1992), 'The updating of the representation of visual space in parietal cortex by intended eye movements', *Science* **255**(5040), 90–92.
URL: *http://www.sciencemag.org/content/255/5040/90.abstract*

Duhamel, J., Goldberg, M., Fitzgibbon, E., Sirigu, A. & Grafman, J. (1992), 'Saccadic dysmetria in a patient with a right frontoparietal lesion the importance of corollary discharge for accurate spatial behaviour', *Brain* **115**(5), 1387–1402.

Duhamel, J.-R., Bremmer, F., BenHamed, S. & Graf, W. (1997), 'Spatial invariance of visual receptive fields in parietal cortex neurons', *Nature* **389**(6653), 845–848.

Dunn, C. A., Hall, N. J. & Colby, C. L. (2010), 'Spatial updating in monkey superior colliculus in the absence of the forebrain commissures: Dissociation between superficial and intermediate layers', *Journal of Neurophysiology* **104**(3), 1267–1285.
URL: *http://jn.physiology.org/content/104/3/1267.abstract*

Eccles, J. C., Fatt, P. & Koketsu, K. (1954), 'Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones', *The Journal of physiology* **126**(3), 524–562.

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E. & König, P. (2007), 'Human eye-head co-ordination in natural exploration', *Network: Computation in Neural Systems* **18**(3), 267–297.

Elliffe, M., Rolls, E. T., Parga, N. & Renart, A. (2000), 'A recurrent model of transformation invariance by association', *Neural Networks* **13**(2), 225–237.

Evans, B. & Stringer, S. (2012), 'Transform-invariant visual representations in self-organizing spiking neural networks', *Frontiers in Computational Neuroscience* **6**(46).

Földiák, P. (1991), 'Learning invariance from transformation sequences', *Neural Computation* **3**(2), 194–200.

Freedman, E. G. & Sparks, D. L. (1997), 'Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys', *Journal of Neurophysiology* **77**(5), 2328–2348.

Galletti, C., Battaglini, P. & Fattori, P. (1991), 'Functional properties of neurons in the anterior bank of the parieto-occipital sulcus of the macaque monkey', *European Journal of Neuroscience* **3**(5), 452–461.

Galletti, C., Battaglini, P. & Fattori, P. (1995), 'Eye position influence on the parieto-occipital area po of the macaque monkey', *European Journal of Neuroscience* **7**(12), 2486–2501.

Galletti, C., Battaglini, P. P. & Fattori, P. (1993), 'Parietal neurons encoding spatial locations in craniotopic coordinates', *Experimental Brain Research* **96**(2), 221–229.

Girard, P., Hupe, J. & Bullier, J. (2001), 'Feedforward and feedback connections between areas v1 and v2 of the monkey have similar rapid conduction velocities', *Journal of neurophysiology* **85**(3), 1328–1331.

Gnadt, J. W. & Andersen, R. A. (1988), 'Memory related motor planning activity in posterior parietal cortex of macaque', *Experimental Brain Research. Experimentelle*

Hirnforschung. Exprimentation Crbrale **70**(1), 216–220. PMID: 3402565.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3402565*

Goldberg, M. E. & Bruce, C. J. (1990), 'Primate frontal eye fields. III. maintenance of a spatially accurate saccade signal', *Journal of Neurophysiology* **64**(2), 489–508.
URL: *http://jn.physiology.org/content/64/2/489.abstract*

Gottlieb, J. P., Kusunoki, M. & Goldberg, M. E. (1998), 'The representation of visual salience in monkey parietal cortex', *Nature* **391**(6666), 481–484.

Hallett, P. E. & Lightstone, A. (1976), 'Saccadic eye movements to flashed targets', *Vision Research* **16**(1), 107–114.

Heiser, L. M., Berman, R. A., Saunders, R. C. & Colby, C. L. (2005), 'Dynamic circuitry for updating spatial representations. II. physiological evidence for interhemispheric transfer in area LIP of the Split-Brain macaque', *Journal of Neurophysiology* **94**(5), 3249–3258.
URL: *http://jn.physiology.org/content/94/5/3249.abstract*

Heiser, L. M. & Colby, C. L. (2006), 'Spatial updating in area lip is independent of saccade direction', *Journal of neurophysiology* **95**(5), 2751–2767.

Hertz, J. A., Krogh, A. S. & Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Vol. 1, Basic Books.

Ipata, A. E., Gee, A. L., Goldberg, M. E. & Bisley, J. W. (2006), 'Activity in the lateral intraparietal area predicts the goal and latency of saccades in a Free-Viewing visual search task', *The Journal of Neuroscience* **26**(14), 3656–3661.
URL: *http://www.jneurosci.org/content/26/14/3656.abstract*

Kandel, E. R., Schwartz, J. H., Jessell, T. M. et al. (2000), *Principles of neural science*, Vol. 4, McGraw-Hill New York.

Krommenhoek, K., Van Opstal, A., Gielen, C. & Van Gisbergen, J. (1993), 'Remapping of neural activity in the motor colliculus: a neural network study', *Vision research* **33**(9), 1287–1298.

Kullmann, D. M., Moreau, A. W., Bakiri, Y. & Nicholson, E. (2012), 'Plasticity of inhibition', *Neuron* **75**(6), 951–962.

Kusunoki, M. & Goldberg, M. E. (2003), 'The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey', *Journal of Neurophysiology* **89**(3), 1519–1527.

Kusunoki, M., Gottlieb, J. & Goldberg, M. E. (2000), 'The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance', *Vision research* **40**(10), 1459–1468.

Lynch, J., Graybiel, A. & Lobeck, L. (1985), 'The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus', *Journal of Comparative Neurology* **235**(2), 241–254.

Mays, L. E. & Sparks, D. L. (1980), 'Dissociation of visual and saccade-related responses in superior colliculus neurons', *Journal of Neurophysiology* **43**(1), 207–232.
URL: *http://jn.physiology.org/content/43/1/207.abstract*

Mazzoni, P., Andersen, R. A. & Jordan, M. I. (1991), 'A more biologically plausible learning rule for neural networks', *Proceedings of the National Academy of Sciences* **88**(10), 4433–4437.

Mazzoni, P., Bracewell, R. M., Barash, S. & Andersen, R. A. (1996), 'Motor intention activity in the macaque's lateral intraparietal area. i. dissociation of motor plan from sensory memory', *Journal of Neurophysiology* **76**(3), 1439–1456.

Melcher, D. & Colby, C. L. (2008), 'Trans-saccadic perception', *Trends in cognitive sciences* **12**(12), 466–473.

Mender, B. M. & Stringer, S. M. (2013), 'A model of self-organizing head-centered visual responses in primate parietal areas', *PloS one* **8**(12), e81406.

Mender, B. M. & Stringer, S. M. (2014), 'Self-organization of head-centered visual responses under ecological training conditions', *Network: Computation in Neural Systems* **25**(3), 116–136.

Mender, B. M. W. & Stringer, S. M. (2015), 'A self-organizing model of perisaccadic visual receptive field dynamics in primate visual and oculomotor system', *Frontiers in Computational Neuroscience* **9**(17).

Mitchell, J. & Zipser, D. (2001), 'A model of visual–spatial memory across saccades', *Vision research* **41**(12), 1575–1592.

Molotchnikoff, S. & Ono, T. (1993), *The visually responsive neuron: From basic neurophysiology to behavior*, Vol. 95, Elsevier.

Mullette-Gillman, O. A., Cohen, Y. E. & Groh, J. M. (2005), 'Eye-Centered, Head-Centered, and complex coding of visual and auditory targets in the intraparietal sulcus', *Journal of Neurophysiology* **94**(4), 2331–2352.
URL: *http://jn.physiology.org/content/94/4/2331.abstract*

Nakamura, K. & Colby, C. L. (2000), 'Visual, Saccade-Related, and cognitive activation of single neurons in monkey extrastriate area V3A', *Journal of Neurophysiology* **84**(2), 677–692.
URL: *http://jn.physiology.org/content/84/2/677.abstract*

Nakamura, K. & Colby, C. L. (2002), 'Updating of the visual representation in monkey striate and extrastriate cortex during saccades', *Proceedings of the National Academy*

of Sciences **99**(6), 4026–4031.
URL: *http://www.pnas.org/content/99/6/4026.abstract*

O'Donohue, T. L., Millington, W. R., Handelmann, G. E., Contreras, P. C. & Chronwall, B. M. (1985), 'On the 50th anniversary of dale's law: multiple neurotransmitter neurons', *Trends in Pharmacological Sciences* **6**(0), 305 – 308.
URL: *http://www.sciencedirect.com/science/article/pii/0165614785901415*

Pouget, A. & Sejnowski, T. J. (1997), 'Spatial transformations in the parietal cortex using basis functions', *Journal of Cognitive Neuroscience* **9**, 222–237.
URL: *http://dl.acm.org/citation.cfm?id=1331235*

Quaia, C., Optican, L. M. & Goldberg, M. E. (1998), 'The maintenance of spatial accuracy by the perisaccadic remapping of visual receptive fields', *Neural Networks* **11**(7), 1229–1240.

Robinson, D. (1975), 'Oculomotor control signals', *Basic mechanisms of ocular motility and their clinical implications* pp. 337–374.

Rolls, E. T. (2012), 'Invariant visual object and face recognition: neural and computational bases, and a model, visnet', *Frontiers in Computational Neuroscience* **6**(35).

Rolls, E. T. & Milward, T. (2000), 'A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures', *Neural Computation* **12**(11), 2547–2572.

Rolls, E. T. & Treves, A. (1998), *Neural Networks and Brain Function*, Oxford University Press Oxford.

Royer, S. & Paré, D. (2003), 'Conservation of total synaptic weight through balanced synaptic depression and potentiation', *Nature* **422**(6931), 518–522.

Salinas, E. & Abbott, L. (1995), 'Transfer of coded information from sensory to motor networks', *The Journal of Neuroscience* **15**(10), 6461–6474.
URL: *http://www.jneurosci.org/content/15/10/6461.abstract*

Salinas, E. & Thier, P. (2000), 'Gain modulation-a major computational principle of the central nervous system', *Neuron* **27**(1), 15–21.

Segraves, M. A. & Goldberg, M. E. (1987), 'Functional properties of corticotectal neurons in the monkey's frontal eye field', *Journal of Neurophysiology* **58**(6), 1387–1419.

Snyder, L. H., Grieve, K. L., Brotchie, P. & Andersen, R. A. (1998), 'Separate body- and world-referenced representations of visual space in parietal cortex', *Nature* **394**(6696), 887–891.
URL: *http://dx.doi.org/10.1038/29777*

Soechting, J. & Flanders, M. (1992), 'Moving in three-dimensional space: frames of reference, vectors, and coordinate systems', *Annual review of neuroscience* **15**(1), 167–191.

Sommer, M. A. & Wurtz, R. H. (2006), 'Influence of the thalamus on spatial visual processing in frontal cortex', *Nature* **444**(7117), 374–377.

Sparks, D. L. & Mays, L. E. (1983), 'Spatial localization of saccade targets. i. compensation for stimulation-induced perturbations in eye position', *J Neurophysiol* **49**(1), 45–63.

Spratling, M. W. (2009), 'Learning posture invariant spatial representations through temporal correlations', *Autonomous Mental Development, IEEE Transactions on* **1**(4), 253–263.

Stanton, G. B., Bruce, C. J. & Goldberg, M. E. (1995), 'Topography of projections to posterior cortical areas from the macaque frontal eye fields', *The Journal of*

Comparative Neurology **353**(2), 291–305.
URL: *http://dx.doi.org/10.1002/cne.903530210*

Stringer, S. M. & Rolls, E. T. (2002), 'Invariant object recognition in the visual system with novel views of 3D objects', *Neural Computation* **14**(11), 2585–2596.

Stringer, S. M. & Rolls, E. T. (2008), 'Learning transform invariant object recognition in the visual system with multiple stimuli present during training', *Neural Networks* **21**(7), 888–903.

Stringer, S., Rolls, E. & Tromans, J. (2007), 'Invariant object recognition with trace learning and multiple stimuli present during training', *Network: Computation in Neural Systems* **18**(2), 161–187.

Umeno, M. M. & Goldberg, M. E. (1997), 'Spatial processing in the monkey frontal eye field. i. predictive visual responses', *Journal of Neurophysiology* **78**(3), 1373–1383.

Umeno, M. M. & Goldberg, M. E. (2001), 'Spatial processing in the monkey frontal eye field. ii. memory responses', *Journal of Neurophysiology* **86**(5), 2344–2352.

von Helmholtz, H. & Southall, J. P. (1924), 'Helmholtz's treatise on physiological optics, vol 1 (trans'.

Walker, M. F., Fitzgibbon, E. J. & Goldberg, M. E. (1995), 'Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements', *Journal of Neurophysiology* **73**(5), 1988–2003.

Wallis, G. & Rolls, E. T. (1997), 'Invariant face and object recognition in the visual system', *Progress in neurobiology* **51**(2), 167–194.

White, R. L. & Snyder, L. H. (2004), 'A neural network model of flexible spatial updating', *Journal of Neurophysiology* **91**(4), 1608–1619.
URL: *http://jn.physiology.org/content/91/4/1608.abstract*

Widrow, B. & Hoff, M. (1960), Adaptive switching circuits, *in* 'IRE WESCON convention record', Vol. 4, pp. 96–104.

Wurtz, R. H. & Goldberg, M. E. (1972), 'Activity of superior colliculus in behaving monkey. iii. cells discharging before eye movements', *J Neurophysiol* **35**(4), 575–586.

Xing, J. & Andersen, R. A. (2000*a*), 'Memory activity of LIP neurons for sequential eye movements simulated with neural networks', *Journal of Neurophysiology* **84**(2), 651–665.
URL: *http://jn.physiology.org/content/84/2/651.abstract*

Xing, J. & Andersen, R. A. (2000*b*), 'Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames', *Journal of Cognitive Neuroscience* **12**, 601–614.
URL: *http://dl.acm.org/citation.cfm?id=1163508*

Zipser, D. & Andersen, R. A. (1988), 'A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons', *Nature* **331**(6158), 679–684. PMID: 3344044.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3344044*

Zucker, R. S. & Regehr, W. G. (2002), 'Short-term synaptic plasticity', *Annual review of physiology* **64**(1), 355–405.