The Development of Neuronal Responses in the Primate Lateral Intraparietal Area



Daniel de Moraes Navarro
The Queen's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2019

This thesis is dedicated to

My beloved parents Glaucio Martines Navarro and Elisabete Aparecida de Moraes Navarro, thank you for your endless love, pray, support, sacrifice, guidance, patience and everything that you have done since I was born.

My beloved brothers Emerson de Moraes Navarro and Allan de Moraes Navarro, thank you for your care, pray and support.

Acknowledgements

At last this *"long and winding road"*, paraphrasing the famous song by The Beatles, has come to a successful end. It has been a much rougher journey than I believe one could predict due to a series of unexpected events and unforeseen circumstances. For instance, the unfortunate and sudden death of my co-supervisor and Head of Department, Professor Glyn Humphreys, was followed the year after by the historical closure of our departmental building and in the subsequent year by some serious funding problems concerning my original scholarship. For this reason, at the end of this journey I would like to express my warmest gratitude to everyone whose comments, criticism, support and encouragement, academic and personal, have left a mark on this work.

Firstly I would like to thank my supervisor and friend, Dr Simon Stringer, for his advice, guidance, contributions, enthusiasm and continued encouragement even in the face of the greatest challenges. Secondly, I would like to thank my co-supervisor, Dr Hannah Smithson, for her invaluable comments, stimulating discussions, constructive and encouraging feedback and for always being available when I needed the most.

I would like to express my sincere gratitute to the Oxford Foundation for Theoretical Neuroscience and Artificial Intelligence (OFTNAI) for their financial

support of my doctoral studies at Oxford University.

I also would like to thank my labmates: Daniel Walters, Hector Page, Irina Higgins, Juan Galeazzi-González, Akihiro Eguchi, Harry Jordan, Hanna Dury, Nasir Ahmed, James Isbister, Elliot Smith and Niels Leadholm. I am also deeply thankful to my co-supervisor's doctoral students and postdocs: Laura Young, Manuel Spitschan, Joshua Harvey, Takuma Morimoto, Tsvetomira Dumbalska, Alexandra Hibble, and in particular Sarah Regan and Anna-Katharina Hauperich, with whom I had the pleasure of collaborating with. I could not have asked to be part of a more inspiring and talented group of researchers. It was an absolute honour to work with you all.

I must also give special thanks to Paul Madden, the Provost at The Queen's College, and his wife Alison, for their friendship and support throughout my doctoral studies. Likewise, I would like to thank Joyce Millar, Marie Bracey, Melanie Harris, Emily Downing, Adrian Faulkner, John Maguire, Hayley Fitzhugh, Sarah Daley, Dawn Grimshaw, Amanda Saville, Tessa Shaw, Sarah Arkle and all members of staff at The Queen's College who have always being remarkably supportive, friendly and caring. Thanks to each of you I will always treasure the countless hours spent in College and the friendships that were built during this time. In particular, I would like to thank Alexander Vervuurt, Sachintha Dias, Alexander Mortimore, Kevin Schlegel for being such a positive influence in my life and for always challenging me to be a better me.

Finally, I would like to thank my family, to whom this thesis is dedicated. They have given me everything, and their support and encouragment, even if

most times from the other side of the Atlantic, were crucial not only to the completion of the work presented in this thesis, but along every step of the journey that has led me to this point. My gratitude to them is beyond words.

Abstract

The Development of Neuronal Responses in the Primate Lateral Intraparietal Area
Daniel de Moraes Navarro
The Queen's College
Submitted for the degree of Doctor of Philosophy in Experimental Psychology
Trinity Term 2019

This thesis develops biologically plausible neural network models of how the response properties of certain classes of visual neurons in the primate parietal cortex may emerge through visually-guided learning. We start with an investigation of how different functional forms of gain modulation of neuronal responses, in particular gain modulation by the position of the eye, may develop in the brain. Gain modutalion refers to the modulatory effect that a particular bodily state or posture has on the firing rate of some visual neurons responding in a given reference frame. The remainder of the thesis focuses on the development of neurons with head-centred receptive fields. Such neurons respond maximally to a visual stimulus in the same head-centred location across all eye positions. In this thesis we show that different functional forms of gain modulation deeply affects the self-organisation of head-centred responses using simulated training data. We then recorded simultanous eye and head movements from human participants during a visual search task and used these to train our self-organising model of coordinate transformation. A further contribution of this thesis is the development of an open source scientific tool for Android

devices to record rotation movements. Our neural network models were trained using a biologically plausible visually-guided learning process and then tested in simulated conditions similar to the experimental tasks used when the investigated responses were observed in the primate brain. The trained responses of our models showed a broad agreement with neurophysiological observations and our simulations help advance understanding of how the neuronal response properties investigated in this thesis may develop in the visual brain.

Contents

1 Introduction **1**

1.1 Head-Centred reference frames 1

1.2 Eye position gain modulation 4

1.3 Computer Models 8

1.4 Research presented in this thesis 15

2 The emergence of eye position gain fields modulating the responses of visual neurons **18**

2.1 Introduction 18

2.2 Hypothesis 20

2.3 Hardwired model 21

2.3.1 Methods 21

2.3.2 Simulation results 25

2.4 Self-organising model 34

2.4.1 Methods 35

2.4.2 Simulation results 39

2.5 Conclusion 46

3 Self-organisation of head-centred responses peaked and monotonic eye position gain fields **50**

3.1 Introduction 50

3.2 Hypothesis 51

3.3 Self-organising Neural Network Model of Coordinate Transformation 53

3.4 Network Components 55

3.5 Network Architecture 58

3.6 Training the Network 59

3.7 Testing the Network 60

3.8 Network Dynamics 61

3.8.1 Input Layer 61

3.8.2 Output Layer 64

3.8.3 Modification of Synaptic Weights by Trace Learning 64

3.8.4 Simulation of the Differential Model 65

3.9 Analysis 66

3.10 Results 68

3.10.1 Self-organisation with peaked and monotonic gain fields 68

3.10.2 Covariance Analysis of the Effects of Gain Modulation 77

3.10.3 Introducing Plasticity into the Prewired Model 83

3.10.4 Standard Trace Learning Rule with Mixed Peaked and Sigmoidal eye-position Modulation of Input Neurons 91

3.10.5 Modified Learning Rule: Delayed Postsynaptic Trace with Anti-Hebbian Learning 94

3.10.6 Modified Learning Rule: Delayed Postsynaptic Firing Rate with Anti-Hebbian Learning . . . 99
3.10.7 Modified Learning Rule: Current Postsynaptic Trace with Anti-Hebbian Learning . . . 104
3.10.8 Modified Learning Rule: Delayed Postsynaptic Trace Learning Rule 107
3.10.9 Delayed Postsynaptic Trace Learning Rule with Mixed Peaked and Monotonic eye-position Modulation of Input Neurons . . . 112
3.11 Conclusion . . . 116

4 DN Mobile Rotation Tracking System: An Android application for recording head-rotation movements 120

4.1 Introduction . . . 120
4.2 Motivation and significance . . . 122
4.3 Software description . . . 124
4.3.1 Operational states of the application . . . 125
4.3.2 User interactive components and software functionalities . . . 127
4.3.2.1 Home activity . . . 127
4.3.2.2 Recording activity . . . 130
4.3.3 Recording rotation movements . . . 132
4.3.3.1 Defining rotation movements . . . 132
4.3.3.2 The challenge of continuously recording real-time sensor data 134
4.4 Conclusion . . . 136

5 Self-organisation of head-centred responses using eye and head movements recorded from human participants **139**

5.1 Introduction . . . 139

5.2 Materials and Methods . . . 141

5.2.1 Experimental task . . . 141

5.2.1.1 Participants . . . 141

5.2.1.2 Eye and head-movement recording . . . 141

5.2.1.3 Stimuli and task . . . 143

5.2.1.4 Procedure . . . 143

5.2.1.5 Physical properties of the recorded eye and head movement data . . . 145

5.3 Network Architecture . . . 149

5.3.0.1 Input Layer . . . 150

5.3.0.2 Output Layer . . . 151

5.3.0.3 Modification of the Synaptic Weights . . . 151

5.4 Training the Network . . . 152

5.5 Testing the Network . . . 153

5.5.1 Analysis of Network Performance . . . 154

5.6 Results . . . 154

5.7 Conclusion . . . 163

6 Conclusions **167**

A **181**

Bibliography **184**

List of Figures

2.1 Architecture of the hardwired neural network model 22

2.2 Firing rate responses in the hardwired model. 27

2.3 Influence of varying the width ρ in the hardwired model 28

2.4 Neural modulation in the hardwired model 30

2.5 Neural modulation in the hardwired model (sigmoidal) 32

2.6 Architecture of the self-organising neural network model 36

2.7 Firing rate responses in the self-organising model 41

2.8 Influence of varying the width ρ in the self-organising model 43

2.9 Modulation in the self-organising model 45

3.1 Architecture of the coordinate transformation neural network model 59

3.2 Examples of the two alternative response functions of input neurons 62

3.3 Comparison of model performance with peaked and monotonic modulation 72

3.4 Strengths of afferent synapses during testing 75

3.5 Scatter plot of eye-centredness and head-centredness 76

3.6 Covariance between input neurons 78

3.7 Weight vectors of two typical output neurons 82

3.8 Synaptic weight structure of a head-centred neuron 85

3.9 Performance of the model with monotonic modulation 87
3.10 Population analysis with plasticity 88
3.11 Effects of peaked and monotonic input neurons 93
3.12 Firing rate responses with equation 3.14 97
3.13 Synaptic weight vector with equation 3.14 98
3.14 Reference frame response with equation 3.14 100
3.15 Firing rate responses with equation 3.15 101
3.16 Synaptic weight vector with equation 3.15 102
3.17 Reference frame response with equation 3.15 104
3.18 Firing rate responses with equation 3.16 106
3.19 Synaptic weight vector with equation 3.16 107
3.20 Reference frame response with equation 3.16 109
3.21 Firing rate responses with equation 3.17 111
3.22 Synaptic weight vector with equation 3.17 112
3.23 Reference frame response with equation 3.17 114
3.24 Firing rate responses of mixed modulation with equation 3.17 115
3.25 Synaptic weight vector of mixed modulation with equation 3.17 116
3.26 Reference frame response of mixed modulation with equation 3.17 118

4.1 Operational State Diagram of the application 126
4.2 Examples of the Home screen 128
4.3 Examples of the Recording screen 131
4.4 Orientation of a mobile device 133
4.5 Example of an output file created by the application 135

5.1 Interpretation of HREF coordinates 142

5.2 Experimental setup of the visual search task 144

5.3 Recorded saccades sequence 147

5.4 Heat maps of exploration patterns 148

5.5 Box plot of saccades frequencies and head movements 149

5.6 Firing rate responses before and after training 156

5.7 Synaptic weight vectors from monotonic neurons 157

5.9 Synaptic weight vectors from peaked neurons 159

5.10 Population analysis for each model 162

List of Tables

2.1 Parameters of hardwired model of eye position gain fields 26

2.2 Parameters of self-organising model of eye position gain fields 40

3.1 Parameters of self-organising model of coordinate transformation 69

3.2 Results for experiments with either peaked or monotonic input modulation. 73

3.3 Parameters of hardwired model of coordinate transformation 86

3.4 Results for experiments with modified learning rule 3.14. 99

3.5 Results for experiments with modified learning rule 3.15. 103

3.6 Results for experiments with modified learning rule 3.16. 108

3.7 Results for experiments with modified learning rule 3.17. 113

3.8 Results for experiments with modified learning rule 3.17 and mixed input modulation. 117

5.1 Parameters of self-organising with recorded data 155

5.2 Experiments using experimentally recorded data with the modified learning rule 3.17 and the standard Hebbian learning rule (equation 2.9). 160

Chapter 1

Introduction

1.1 Head-Centred reference frames

Within the primate dorsal visual pathway, neurons encode the locations of visual objects in different body-centred frames of reference. For example, neurons at an early stage of processing encode the locations of objects within an eye-centred reference frame, while neurons in later stages may encode the positions of objects with respect to the head or hand, which are more relevant for guiding motor actions (Andersen et al., 1985, Bremner and Andersen, 2012, Brotchie et al., 1995, Buneo et al., 2002, Duhamel et al., 1997, Galletti et al., 1993, Mullette-Gillman et al., 2005, Pesaran et al., 2006). A key question that this thesis aims to address is how the visual system learns to perform such coordinate transformations between different body-centered reference frames. In particular, a number of experimental studies have found visual neurons in the monkey brain that encode the locations of visual stimuli with respect to the head, i.e. within a head-centred frame of reference (Duhamel et al., 1997, Galletti et al., 1993, Mullette-Gillman et al., 2005). The model simulations presented in this thesis seek to explain how such head-centred neurons may develop in the brain.

Galletti et al. (1993) was the first experimental study to provide evidence of head-centred visual representations in the preoptic area (PO) of the macaque brain. A head-centred representation of visual space is assumed to be important for visually guided reaching and perceptual stability. In fact, Galletti et al. (1991) had previously predicted the existence of head-centred visual neurons in PO due to the presence of retinotpic visual neurons with eye-position gain modulation in areas 7a (Andersen et al., 1985), LIP (Andersen et al., 1990) and V3A (Galletti et al., 1991), which are a key precursor to the development of head-centred visual representations (Mender and Stringer, 2014, 2013). Eye position gain modulation is discussed in section 1.2. In the experimental study of Galletti et al. (1993), a head-restrained monkey performed a series of fixations to different locations on a screen whilst a visual target was presented in various other screen locations for each fixation. This allowed the authors to assess how individual neuron's activity depended on either the eye-centred or head-centred location of the visual target. It was reported that the receptive fields of 11% of recorded neurons were not completely eye-centred, with six of these neurons presenting clear head-centred receptive fields. The authors concluded that the previous claim by Andersen et al. (1990) that there was no explicit head-centred neuronal representation of visual space was incorrect. Furthermore, head-centred neurons were suggested to originate from pooling the output of preceding eye-position gain-modulated retinotopic visual neurons.

In Duhamel et al. (1997) the authors studied the visual reference frames found in the ventral intraparietal area (VIP). In particular, they were interested in dissociating head-centred and eye-centred reference frames. First the authors mapped the visual receptive field of each neuron at 9 different fixation locations organised into a 3×3 grid. Then, for

each fixation location, a visual stimulus was flashed in 49 different screen locations in a 7×7 grid. A cross-correlation analysis of the response matrices for different fixation locations allowed the authors to compute the receptive field shifts caused by different fixation locations. For instance, a correlation shift smaller than the fixation shift was interpreted as a head-centred receptive field, with zero indicating perfect head-centredness. Similarly, a correlation shift closer or equal to the fixation shift was interpreted as a eye-centred receptive field. The authors reported a continuum of reference frames amongst neurons in area VIP, with perfect reference frames at both ends of the spectrum. In particular, the authors used a clustering routine that revealed two populations of neurons that could be classified as eye-centred and head-centred. Moreover, eye postion gain modulation was observed in approximately 53% of recorded neurons, where the modulation of 37% of these neurons was classified as planar.

The work of Mullette-Gillman et al. (2005) was motivated by the behavioural requirement for the brain to be able to integrate visual and auditory signals, such as the sight of moving lips and the corresponding speech sound, within a common body-centred frame of reference. The authors' investigation of the reference frames used to encode visual and auditory responses in area LIP revealed the first head-centred visual representations in this area. This contradicted the suggestion made by Andersen et al. (1990) that head-centred responses would not exist in area LIP. The proportion of neurons sensitive to visual target locations was 72%, whilst 51.4% of neurons were sensitive to auditory target locations. Moreover, 5% to 43% of neurons were simultaneously responsive to both visual and auditory target locations, depending on how responsiveness was defined. For visual neurons, 33% had eye-centred responses and 18% had head-centred responses. Regarding auditory

neurons, 10% had eye-centred responses while 23% had head-centred responses. Neither the remaining 49% of visual neurons nor the remaining 67% of auditory neurons could have their responses classified as eye-centred or head-centred. In summary, it was found that both auditory and visual neurons had responses compatible with either eye-centred or head-centred frames of reference, although most neurons had complex responses that could not be classified into either of these categories. This was the first time that head-centred neuronal responses were indentified in area LIP.

1.2 Eye position gain modulation

A neural phenomenon which is thought to play a key role in coordinate transformations in the dorsal visual pathway is *gain modulation*. Gain modulation arises when one input signal affects only the sensitivity of the neuron to another signal without modifying its selectivity. In other words, gain modulation refers to the modulatory effect that a particular bodily state or posture, like eye or head position, has on the firing rate of some visual neurons responding in a given reference frame. In this thesis we focus on the eye-position modulation of visual neuronal firing responses first discovered by Andersen and Mountcastle (1983) in area 7a in parietal cortex (discussed below) and later confirmed in multiple stages of the primate dorsal visual pathway. Moreover, computer simulations carried out within our laboratory have demonstrated that an important precursor to the emergence of head-centred visual neurons in the parietal cortex is likely to be the presence of retinotopic visual neurons with responses that are gain modulated by the position of the eye in the orbit (Mender and Stringer, 2014, 2013, Navarro et al., 2018). A number of experimental studies have previously confirmed the existence of such retinotopic visual neurons with

eye-position gain fields in the monkey brain.

The work by Andersen and Mountcastle (1983) was the earliest demonstration of the influence of the position of the eye in the orbit on the responses of retinotopic light sensitive neurons in area 7a of the monkey parietal cortex. The authors studied the responses of light sensitive neurons using both an attentive task and an inattentive task. In both tasks, the position of the head was kept fixed while the animal was able to move its eyes. During the attentive task the monkey had a peripheral stimulus flashed in a particular eye-centred location whilst maintaining fixation at some gaze angle. Conversely, during the inattentive task targets were flashed on random locations of the screen whilst the monkey freely oriented its gaze. The authors reported that the position of the eyes in the socket affected the responses of neurons in area 7a of the inferior parietal lobule during both attentive fixation and under the inattentive condition. Although it was also found that the responses of a much larger proportion of neurons were modulated by eye position during the attentive task than the inattentive task. The first task, the attentive task, revealed that 61% of neurons had their responses to visual stimuli significantly changed by the eye position. The neuronal responses could be more than three times stronger when the eye position shifted by 20° towards the preferred direction (optimal eye position). The second task, the inattentive task, revealed that a significantly smaller proportion of examined neurons (10%) had their responses to visual stimuli significantly changed by the eye position. The precise interaction between the visual signal encoding the retinotopic location of the target and the eye position signal was later characterised by Andersen et al. (1985). In Andersen et al. (1985) the authors mapped the receptive field of isolated neurons with a task that required head-restrained monkey to fixate straight ahead while

a stimulus was presented systematically across a tangent screen. The responses revealed the prefered eye-centred location of the recorded neurons. A following task then presented a visual stimulus in the preferred eye-centred location whilst the animal fixated at nine locations organised into a 3×3 grid. This task allowed the authors to systematically analyse the influence of the eye position on the responses of the neuron. The authors reported that the eye position modulation of 39% of recorded neurons was planar, whilst the eye position modulation of another 38% had a significant planar component, but with a bump in the fit. The eye position modulation of the remaining neurons was peaked, without a planar component. Moreover, these effects were also observed in the lateral intraparietal area (LIP) by Andersen et al. (1990). In fact, it was in this work that the role of LIP in sensorimotor integration was suggested for the first time. In Andersen et al. (1990) the authors first determined the preferred saccade direction of each neuron by using a memory guided saccade task. The memory guided saccade task consisted of the following four periods. The first period, the *baseline period*, consisted of fixating at the centre of the screen for either $300ms$ or $800ms$. During the second period, the *light sensitive period*, a peripherical saccade stimulus was presented for $300ms$. During the third period, the *memory period*, the saccade target was extinguished, but fixation had still to be kept for either $400ms$ or $700ms$. Finally, the fourth period, the *saccade period* started when the fixation light was extinguished and a saccade to the remembered target location was made. After determining the preferred saccade direction, the gain field of each neuron was mapped during a task that required sacccades to be performed from nine different gaze angles organised in a 3×3 grid, whilst the saccade target was always positioned in the neuron's preferred direction. The authors reported that the majority of recorded cells

had significant eye position gain modulation that was planar or had a planar component. The authors also reported that the observed gain fields in LIP had larger gradients and intercepts than gain fields in 7a.

The presence of more *peaked* eye position gain modulation has been observed in the parietal occipital area (PO) by Galletti et al. (1995) and Breveglieri et al. (2009). In Galletti et al. (1995) the authors first recorded neural responses when a monkey fixated in the dark at nine screen locations organised in a 3×3 grid. Then during a second task performed both in light and dark, the authors mapped the visual receptive fields of these neurons with the same nine screen fixation locations. It was reported that 61% of visual neurons had responses modulated by the position of the eye. Moreoever, approximatelly 32% of non-visual neurons and 48% of visual neurons had eye-position gain modulation in the task performed in the dark. The authors discussed that the observed gain modulation in the dark showed that the effect was caused by the position of the eye and not by changes in the background when the eyes were moving. In the more recent study of Breveglieri et al. (2009), an experimental task was designed to investigate the proportion of retinotopic visual neurons in area V6A of the primate brain that are modulated by eye position with either peaked (non-monotonic) or planar (monotonic) gain fields. During the experimental task the monkey had nine equally spaced fixation locations organised as a 3×3 grid, with the visual stimulus always presented at the fixation point. These authors found that approximately 56% of recorded neurons had their responses modulated by the position of the eye. Furthermore, 27% of the neurons with responses that were modulated by eye position had planar gain fields, whilst the remaining 73% had peaked gain fields.

Furthermore, the recent work of Morris and Krekelberg (2019) showed for the first time

the existence of neurons in the primary visual cortex (V1), the earliest stage of visual information processing in the cortex, that are modulated by the position of the eye. In Morris and Krekelberg (2019) the authors implanted multi-electrode arrays to record the activity of V1 neurons of macaque monkeys during tasks that required either fast or slow eye movements. The authors reported that V1 neurons encoded the real location of objects in the world via a population code, i.e. the location of the objects derived from the recorded V1 retinal map was identical to the real location of the visual stimulus on the screen (i.e. the real location of the object in the world). This finding contrasted with the common belief that V1 neurons would only encode the retinal location of visual stimuli. Moreover, the authors reported that 55% of single neurons recorded (127/232) had significant eye position modulation. In particular, the author reported that the form of eye position gain fields of V1 neurons was planar.

1.3 Computer Models

Previous modelling studies have attempted to explain how gain modulation may develop in the brain. The influential work of Salinas and Abbott (1996) investigated the development of gain modulated responses using recurrent connections within a network model of cortical neurons. The recurrent connectivity was defined according to the distance between the preferred retinal locations of the neuronal responses. Although the model was based on ordinary properties of cortical circuits, the authors used a linear function to model the response profiles of eye position input neurons. This is an important issue in terms of biological plausibility because the response profiles of eye position neurons have been reported to be *Gaussian* rather than *linear* (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja,

1980, Wang et al., 2007). Furthermore, the recurrent connections in their model violated the widely accepted Dale's law, which establishes that a presynaptic neuron cannot make both excitatory and inhibitory synaptic connections on postsynaptic neurons (O'Donohue et al., 1985). In addition, the model of Salinas and Abbott (1996) was not trained to *learn* visual responses that were multiplicatively modulated by eye position. Instead, the synaptic connections were hardwired to ensure that the visual responses of each neuron within the network were gain modulated by eye position. A more recent simulation study by De Meyer and Spratling (2011) showed that gain modulation can appear in a predictive coding model using unsupervised learning. In the context used by the authors, predictive coding is defined as neural theory based on the principle of minimising error in the stimulus-driven activity between bottom-up and top-down predictions of the internal representation of the stimulus. Although predictive coding has been suggested to be implemented by cortical feedback connections calculating residual errors that would be then propagated by cortical feedforward connections, currently there is no physiological evidence of predictive coding being implemented on the neuronal level in the cortex (Bastos et al., 2012, Kwisthout and Van Rooij, 2013, M and S, 2015). In this thesis we address the biological issues in Salinas and Abbott (1996) with both a hardwired neural network model and a self-organising neural network model of the development of retinotopic visual responses modulated by eye position (chapter 2). Understanding how gain modulation may develop in the brain provides an insight into the way the brain encodes information that is relevant to high-level visual functions, such as coordinate transformation, eye movement planning, visual-motor behaviours and stability of spatial awareness.

Zipser and Andersen (1988) developed an early influential model that learned to transform an input representation consisting of the position of the eye in the socket and the retinotopic position of the visual target to an output representation consisting of the position of the visual target with respect to the head. Specifically, two-dimensional representations of the eye-position $\mathbf{e}$ and of the retinal location of the visual target $\mathbf{r}$ were used as input for the neural network model. The target output used to guide the error-based update of the synaptic weights during training was the head-centred location $\mathbf{h}$ of the visual target represented by $\mathbf{r} + \mathbf{e}$. The model thus utilised a supervised, multi-layer backpropagation of error network architecture, in which the output layer of the network was provided with an explicit training signal representing the current head-centred location of the visual target. Such a network architecture is not biologically plausible for a couple of reasons. Firstly, it is not clear where such a training signal representing the current head-centred location of the visual target might originate from in the brain. Secondly, a multi-layer backropagation network architecture requires that the afferent weights to the hidden layer be updated using an error signal based on the efferent synaptic weights from these hidden neurons to the output layer scaled by the corresponding errors in the output layer. Such a network architecture in itself is not biologically plausible. Nevertheless, Zipser and Andersen (1988) found that training such a network to transform independent eye position and retinal target position signals to head-centred coordinates led to hidden units in the intermediate layer developing retinocentric receptive fields with planar eye-position gain fields, very similar to those found in the posterior parietal cortex by Andersen and Mountcastle (1983).

A number of later models that simulate the development of head-centred visual neurons

have also relied on some form of error correction learning. For example, although Mazzoni et al. (1991) recognised that the backpropagation of error architecture used by Zipser and Andersen (1988) was biologically implausible as a learning process to be implemented by the cortex, the model used in Mazzoni et al. (1991) relied on a form of supervised global error correction learning that utilised an error term which is unlikely to be present in the cortex. Mazzoni et al. (1991) suggested that the global error correction signal in their model could be implemented in the brain by neuromodulator release. However, this would require the cortex to compute a desired head-centred output with the current neuronal response, then compare the desired head-centred output with the actual output to finally calculate a global error term at population level that would be used to adjust the synaptic weights through an exceptionally complicated learning rule. There is no evidence of how such a mechanism could be implemented in the cortex. An alternative approach to reference frame transformation based on nonlinear function approximation theory is proposed by Pouget and Sejnowski (1997). However, the model of sensorimotor transformation by Pouget and Sejnowski (1997) also used a supervised error correction learning rule to modify the synaptic connectivity within the network. So, although supervised error correction learning does not appear to offer a biologically plausible way of modelling the development of head-centred visual neurons in the brain, it has nevertheless remained a popular approach.

In contrast to the above error correction network models, Salinas and Abbott (1995) developed a model that utilised associative learning. These authors investigated how retinotopic visual representations could be transformed into head-centred representations that are relevant to reaching tasks. It was hypothesised that the observation of our own motor movements during a reaching task would be used to develop a map between the eye-

centred visual representation of the target and the head-centred motor representation of the movement required to reach the target. In other words, the input layer of the neural network model represented the visual signals generated by observing the reaching movements, whilst the output layer represented the corresponding motor movements that were performed. As with Zipser and Andersen (1988), the input representation consisted of the position of the eye in the socket and the retinotopic position of the visual target, while the output representation consisted of the head-centred target position. Hebbian learning, in which the strength of each synaptic connection is increased in proportion to the product of the firing rates of the pre- and post-synaptic neurons, was then used during training to successfully associate each eye-centred visual input representation of the target to the corresponding head-centred output motor representation. In contrast to Zipser and Andersen (1988), the model developed by Salinas and Abbott (1995) did not use a backpropagation of error network architecture or any other form of error correction. Instead, they used a more biologically plausible associative learning rule to modify the synaptic connectivity within their model. However, the model was still trained in a supervised manner. That is, the network still made use of an explicit training signal in the output layer representing the head-centred location of the visual target in order to guide learning without an adequate explanation of where such a signal might originate from in the brain.

In Breveglieri et al. (2009) the authors explored the influence of each type of gain modulation on the traditional neural network model of sensorimotor transformation proposed by Salinas and Abbott (1995). Their main motivation was to understand the implications of the functional form of gain fields, i.e. peaked (non-monotonic) *vs* planar (monotonic), for sensorimotor transformations in reaching tasks. In particular, they investigated how the

functional form of the gain fields affected the performance of the model proposed by Salinas and Abbott (1995) to transform eye-centred visual representations into head-centred representations. The authors found that the incorporation of planar rather than peaked eye-position gain fields led to reduced model performance, with the population of output neurons providing a less accurate representation of the location of the visual stimulus with respect to the head. In contrast, Spratling (2009) proposed a hierarchical neural network model of coordinate transformation that used associative learning to set up the synaptic connectivity but did not require a biologically implausible supervisory training signal as used by Salinas and Abbott (1995). The model developed by Spratling (2009) had the following two main processing stages. The first processing stage used signals representing the retinotopic location of the visual target and signals representing the position of the eye to learn head-centred visual representations. The second processing stage used head position signals coupled with the head-centred visual representations developed by the first processing stage to learn body-centred visual representations. Training consisted of continously shifting a visual target presented to the network whilst randomly varying the position of the eyes and the head. Both processing stages used competitive learning with an associative learning rule to self-organise conjunctive representations of its respective inputs, and then used competitive learning with a form of temporally associative learning rule to bind representations that occurred close together in time. Learning in the second processing stage only started after the first processing stage had finished learning head-centred representations. This allowed each processing stage of the model to self-organise either head-centred or body-centred representations. Most importantly, in contrast to the models developed by Zipser and Andersen (1988) and Salinas and Abbott (1995), learning

in Spratling (2009) did not require the use of a biologically implausible supervisory training signal to guide learning of the coordinate transformations. The model of Spratling (2009), however, did not investigate how the functional form of eye position gain modulation, i.e. monotonic *vs* peaked, may affect the development of head-centred visual representations.

The more recent work by Mender and Stringer (2013) also showed simulation results of how head-centered visual representations in primate parietal areas may self-organise through unsupervised visually-guided learning. The model in Mender and Stringer (2013) consists of an input population of neurons with eye position gain modulated retinal receptive fields that send efferent synaptic connections to a competitive output layer of neurons. An associative trace learning rule was used to update the synaptic connections during training. The associative learning rule had the effect of successfully encouraging output neurons to learn to respond to subsets of input patterns that tend to occur close together in time. The simulation results in Mender and Stringer (2013) showed that the model was able to self-organise head-centred output responses after training. In Mender and Stringer (2014) the authors explored a range of more ecological conditions to train the model, such as, for example, the duration of fixations and the time duration that the visual target is presented. The simulation results in Mender and Stringer (2014) showed that the underlying learning mechanism used in their model was robust to many forms of stimulus variability under ecological training conditions. An outstanding issue with their model, however, was the fact that all neurons in the input population were modulated only by peaked eye-position gain fields, although a variety of forms of eye-position gain modulation has been observed in the brain.

1.4 Research presented in this thesis

This thesis develops biologically plausible neural network models of how the response properties of certain classes of visual neurons in the primate parietal cortex may emerge through visually-guided learning. Chapter 2 presents neural network simulations of the emergence of eye position gain fields modulating the responses of visual neurons. In particular, the simulations discussed in chapter 2 investigates how multiplicative eye position gain modulation may develop using inputs from eye position neurons with Gaussian response profiles. Previous modelling studies used a linear function to model the dependence of the visual responses on eye position (Salinas and Abbott, 1996). This is an important issue as the response profiles of eye position neurons in parietal areas have been reported to be a better fit to Gaussian rather than to linear functions (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007). The remaining chapters concern the development of visual neurons representing the locations of visual targets in a head-centred frame of reference. The forms of eye position gain modulation studied in chapter 2 are relevant to the development of the head-centred visual responses studied in chapters 3 and 5. Chapter 3 presents neural network simulations of how head-centred visual neurons may develop through a biologically plausible process of visually-guided self-organisation. In particular, we explore how different forms of gain modulation of visual input neurons by eye-position, e.g. peaked and monotonic, influence the self-organisation of output neurons with head-centred visual responses. We show in chapter 3 how a biologically plausible neural network model successfuly self-organises head-centred output responses when the input neurons have peaked modulation by eye position, but fails when the modulation is changed from peaked to monotonic. We then present a mathematical analysis of the influ-

ence of monotonic modulation on the self-organisation of the model. This analysis leads us to investigate a modified version of the associative trace learning rule used by Mender and Stringer (2013) to self-organise head-centred output responses with input neurons modulated by monotonic gain fields. Although the simulation results presented in chapter 3 show how head-centred visual responses may be successfully self-organised in the presence of input neurons with a mixture of different forms of eye-position gain modulation, the neural network model was trained with artificially created visual input patterns. In contrast, chapter 5 presents neural network simulations of how head-centred visual neurons may develop through a biologically plausible process of visually-guided self-organisation using recorded data from human participants during a visual search task. Chapter 4 describes a smartphone application developed by the author to record head movements from human participants that was used to train the model in chapter 5. The smartphone application described in chapter 4 has been published as an open source scientific tool and provides the user with the contact details of the developer to facilitate the request of new features and to instigate the establishment of future collaboration with other researchers within the scientific community. This is an additional important contribution of the work presented in this thesis. Chapter 5 shows for the first time simulation results of our neural network model of coordinate transformation trained with recorded eye and head-movements from human participants, instead of artificially created eye and head-movements. By showing how the neural network model self-organised head-centred visual responses using realistic eye and head-movements, the simulation results in this thesis make an important contribution to understanding how head-centred visual responses may develop in the brain through an unsupervised process of visually-guided learning given visual precursor neu-

rons with peaked and monotonic eye-position gain modulation. Moreover, although we investigated one particular type of coordinate transformation (i.e., from eye-centred to head-centred visual representations), our findings may also apply to a range of other coordinate transformations with sensorimotor integration of monotonically encoded motor variables (Bremner and Andersen, 2012, Duhamel et al., 1997, Pesaran et al., 2006).

Chapter 2

The emergence of eye position gain fields modulating the responses of visual neurons

2.1 Introduction

Visual neuronal responses are often characterised in terms of the selectivity of the responses to the location of a given visual stimulus. The portion of the visual space in which a visual stimulus changes the responses of a neuron is conventionally referred to as the *receptive field*. For instance, the receptive field of neurons in the primary visual cortex (V1) is fixed to a visual space defined in a retinocentric frame of reference (Guo and Li, 1997), whilst the receptive field of neurons within other areas of the primate dorsal visual pathway is fixed to a visual space defined in different body-centred reference frames such as head (Galletti et al., 1993, O'dhaniel et al., 2005) or hand (Bremner and Andersen, 2012, Buneo and Andersen, 2006), which are relevant for guiding motor actions. Visually responsive neurons in the cortex have been observed to fire maximally when visual stimuli are positioned in

a particular *preferred location* within the receptive field. A key experimental observation that we investigate in the models presented in this chapter is that such visual responses are commonly reported to be modulated by bodily state or posture, e.g. position of the eyes, head or hand (Andersen et al., 1985, Bremner and Andersen, 2012, Brotchie et al., 1995, Buneo et al., 2002, Pesaran et al., 2006). This *gain modulation* effect, however, does not influence the preferred location or the general response profile of visual neurons, but only the amplitude of the responses. In particular, different forms of eye position gain modulation have been observed in many cortical areas (Andersen and Mountcastle, 1983, Breveglieri et al., 2009, Brotchie et al., 1995, Galletti et al., 1995, Galletti and Battaglini, 1989, Hoffmann, 1998, Lal and Friedlander, 1990, Merriam et al., 2013, Trotter and Celebrini, 1999). For example, gain modulation by eye position in parietal neurons has been described as a multiplicative modulation of the underlying Gaussian retinotopic receptive field by a linear or saturating function of the position of the eye (Andersen and Mountcastle, 1983, Brotchie et al., 1995). Previous modelling studies investigating the development of multiplicative gain modulation used a linear function to model the response profiles of eye position input neurons (Salinas and Abbott, 1996). Linear functions, however, do not accurately represent the Gaussian response profile of eye position neurons in the cortex (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007).

2.2 Hypothesis

In this chapter we hypothese that a variety of different functional forms of gain modulation, including multiplicative gain modulation, may develop when a Gaussian function is used to model the responses of both eye position input neurons and retinocentric input neurons. Moreover, we investigate how the width of the Gaussian response profile of the eye position input neurons affects the linearity of the gain modulation by eye position of the retinotopic visual output neurons. In the first part of this chapter we study a hardwired model similar to Salinas and Abbott (1996). The visual input to our hardwired model consists of the sum of a term encoding the retinal location of the visual stimulus and a term encoding the position of the eye (i.e. the gaze direction). In contrast to Salinas and Abbott (1996), our model does not violate the Dale's Law (O'Donohue et al., 1985) and response profiles of the eye position input neurons and retinotopic visual input neurons were both modelled by a more biologically realistic (Gaussian) function. In the second part of this chapter we study a self-organising model of the development of multiplicative gain modulation by eye position. Our simulation results again show the emergence of a variety of different functional forms of gain modulation. Moreover, the linearity of the gain modulation by eye position of the retinotopic visual output neurons is affected by the width of the Gaussian response profile of the eye position input neurons.

2.3 Hardwired model

In this section we present simulation results investigating the development of multiplicative gain modulation by eye position. In particular, we address the following issues of a similar model by Salinas and Abbott (1996). First we use Gaussian functions to model both retinal stimulus position and eye position input components of each visual output neuron. In Salinas and Abbott (1996) the authors used a linear function to model the influence of the position of the eyes on the visual responses. Gaussian functions, however, have been reported to be a better fit to the response profile of parietal eye position neurons (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007). Secondly we show that multiplicative gain modulation can be obtained without violating the Dale's Law (O'Donohue et al., 1985). Finally, the firing rate responses were modelled with a sigmoid transfer function rather than the linear function used by Salinas and Abbott (1996). A sigmoid transfer function represents more accurately the firing rate profile of individual neurons observed in the cortex (Dayan and Abbott, 2001, Marĉelja, 1980). Furthermore, we show how the width of the eye position tuning curve affects the overall linearity of gain modulation by eye position of the retinotopic visual output neurons.

2.3.1 Methods

The architecture of the hardwired neural network model consists of a single layer of visual output neurons (figure 2.1). The internal activation h of each visual neuron (equation 2.1)

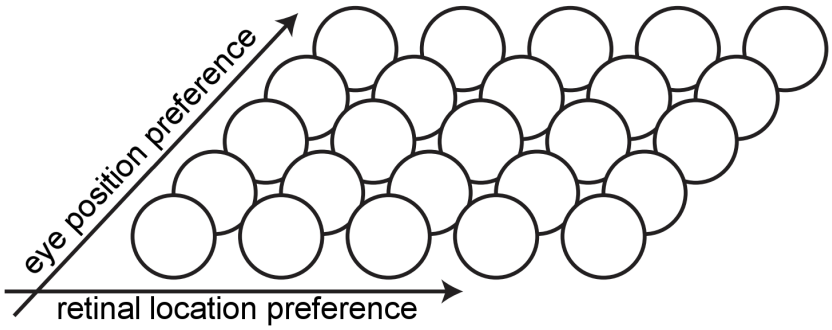


Figure 2.1: Architecture of the hardwired neural network model. The architecture of the hardwired neural network model consists of a single layer of visual output neurons. Each neuron is set to respond maximally to a unique combination of eye position and retinal location of the visual stimulus.

consists of the sum of a component h_r encoding the retinal location x of the visual stimulus and a component h_e encoding the position y of the eye in the orbit. Each neuron is set to respond maximally to a unique combination of retinal location of the visual stimulus and eye position. The range of eye positions and retinal locations were defined within $[-35°, 35°]$ and $[-10°, 10°]$, respectively. The entire two dimensional space consisting of all possible combinations of eye position and retinal location of the visual stimulus is covered in integer steps of one degree in each dimension by the population of visual neurons. This resulted in a total of $71 \times 21 = 1,491$ neurons.

$$h(x, y) = h_r(x) + h_e(y) \tag{2.1}$$

The component of the internal activation h_r to each visual neuron that depends on the retinal location x of the visual stimulus is given by the Gaussian function

$$h_r(x) = h_{max_r} \times \exp\left(-\frac{(x-\alpha)^2}{2\sigma^2}\right) \tag{2.2}$$

where x is the current retinal location of the visual stimulus, α is the neuron's preferred retinal location and σ is the width of the Gaussian tuning curve. The maximum value h_{max_r} is achieved only when the visual stimulus is located at the neuron's preferred retinal location, i.e. when $x = \alpha$. If the stimulus is located at any other retinal location, then the value of h_r corresponds to a Gaussian function of the difference between the stimulus location on the retina x and the neuron's preferred location α.

Similarly, the component of the internal activation h_e to each visual neuron that depends on the position y of the eye in the orbit is described by the Gaussian function

$$h_e(y) = h_{max_e} \times \exp\left(-\frac{(y-\beta)^2}{2\rho^2}\right) \tag{2.3}$$

where y is the current position of the eye in the orbit, β is the neuron's preferred eye position and ρ is the width of the Gaussian tuning curve. The response of a given visual neuron is maximally amplified (i.e. $h_e(y) = h_{max_e}$) when the eye shifts to the neuron's preferred eye position β. For all other eye positions the neuron's responses are modulated by a Gaussian function of the difference between the neuron's preferred eye position and the current position of the eye.

The instantaneous firing rate v_i of each visual neuron i in the hardwired model is given

by

$$v_i = \frac{1}{\exp\left(-2\varphi(h-\theta)\right)} \tag{2.4}$$

where the activation h is defined by equation 2.1 and the slope and threshold of the activation function are denoted, respectively, by φ and θ.

Rearranging equation 2.4 will result in

$$v_i = \frac{1}{\exp\left(-2\varphi(h_r + h_e - \theta)\right)}$$

$$v_i = \frac{1}{\exp\left(-2\varphi h_r - 2\varphi h_e + 2\varphi\theta\right)}$$

$$v_i = \frac{1}{\exp\left(-2\varphi h_r\right) \times \exp\left(-2\varphi h_e\right) \times \exp\left(2\varphi\theta\right)}$$

$$v_i = \frac{1}{\exp\left(-2\varphi h_r\right)} \times \frac{1}{\exp\left(-2\varphi h_e\right)} \times \frac{1}{\exp\left(2\varphi\theta\right)}$$

$$v_i = \exp\left(2\varphi h_r\right) \times \exp\left(2\varphi h_e\right) \times \frac{1}{\exp\left(2\varphi\theta\right)}$$

$$v_i = \exp\left(2\varphi h_e\right) \times \left(\exp\left(2\varphi h_r\right) \times \frac{1}{\exp\left(2\varphi\theta\right)}\right) \tag{2.5}$$

Equation 2.5 shows that a multiplicative effect by eye-position may be obtained when the firing rate of visual neurons are defined by equation 2.4. Nevertheless, some later simulations discussed in this chapter were run with the firing rate given by

$$v_i = \frac{1}{1 + \exp\left(-2\varphi(h-\theta)\right)} \tag{2.6}$$

where the activation h is defined by equation 2.1 and the sigmoidal slope and threshold are denoted, respectively, by φ and θ.

Although equation 2.6 cannot be expressed in an exactly multiplicative form like equation 2.4, the use of equation 2.6 assures the bounding of firing rates between 0 and 1.

2.3.2 Simulation results

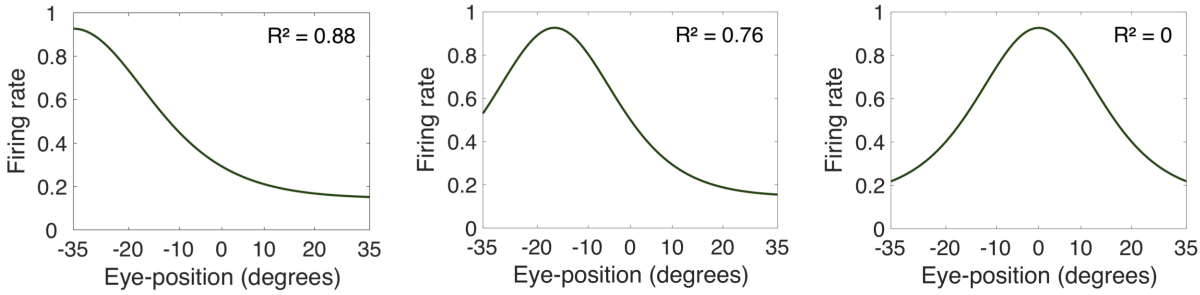
This experiment investigates the existence and nature of multiplicative gain modulation in retinotopic visual output neurons when the response profiles of both the retinotopic visual input component and eye position input component are modelled by Gaussian functions. The former modelling work of Salinas and Abbott (1996) used a linear function to model the influence of the position of the eye on visual responses. This is a particularly important issue because the responses of eye position neurons in the cortex have been reported to be a better fit to a Gaussian function (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007). In the simulations presented here we show how multiplicative gain modulation of the visual neurons can occur even if a Gaussian function is used to model the term $h_e(y)$ denoting the component of the visual neuron activation that depends on eye position. The simulation parameters for the model are given in table 2.1.

Figure 2.2 shows the visual responses of neurons in the hardwired model across different eye positions. For each subplot in figure 2.2 the visual target was kept at the neuron's preferred retinal location whilst the eyes shifted in steps of one degree to all positions

Parameter	Symbol	Value
Activation function slope	φ	1.9
Activation function threshold	θ	0.99
Population size	–	1491
Maximum eye position activation	h_{max_e}	0.485
Maximum retinal location activation	h_{max_r}	0.485
Width of eye position tuning curve	ρ	20°
Width of retinal location tuning curve	σ	6°

Table 2.1: Simulation parameters of hardwired model.

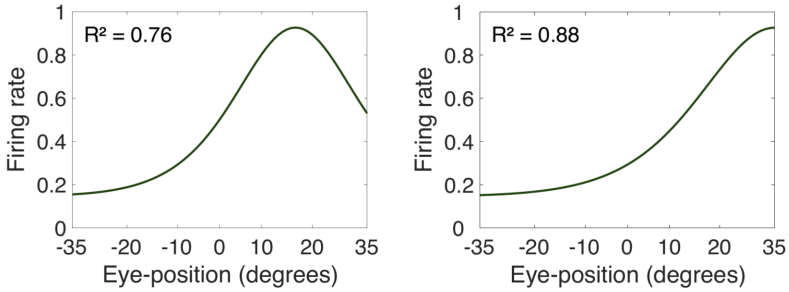
within $[-35°, 35°]$. This range of eye positions has been observed as the functional limit on orbital position during natural visual exploration (Freedman and Sparks, 1997, Navarro et al., 2019). Previous modelling studies of brain function have also used this range of eye positions (Mender and Stringer, 2014, 2013, Navarro et al., 2018). Figure 2.2 shows the visual responses of neurons #1, #373, #475, #1117 and #1491. The preferred retinal location of the visual target for each of these neurons was respectively set to $-10°$, $-5°$, $0°$, $5°$ and $10°$. The linearity of the visual responses shown in figure 2.2 may be compared by calculating the *coefficient of determination* R^2 of each response curve. The coefficient of determination is traditionally used in regression analysis to measure the success of predicting a dependent variable from independent variables (Rao et al., 1973). In other words, the coefficient of determination R^2 provides statistical information about the *goodness-of-fit* of a model. In this work the coefficient of determination measures the goodness-of-fit of the visual responses across different eye positions to a *linear* model. The values of R^2 are defined within $[0, 1]$, where R^2 equal to 1 indicates that a linear model perfectly predicts the visual responses. Such visual responses are referred to as *linear responses*. If R^2 is



(A) Neuron #1 ($\alpha = -10°, \beta = -34°$).

(B) Neuron #373 ($\alpha = -5°, \beta = -18°$).

(C) Neuron #475 ($\alpha = 0°, \beta = 0°$).



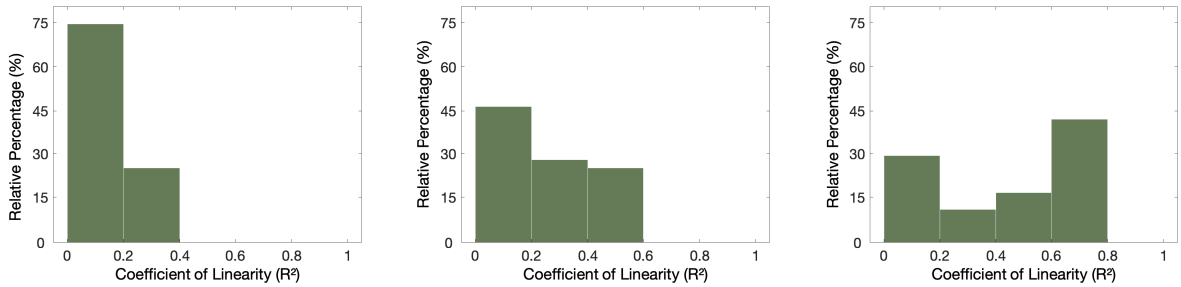
(D) Neuron #1117 ($\alpha = 5°, \beta = 18°$).

(E) Neuron #1491 ($\alpha = 10°, \beta = 34°$).

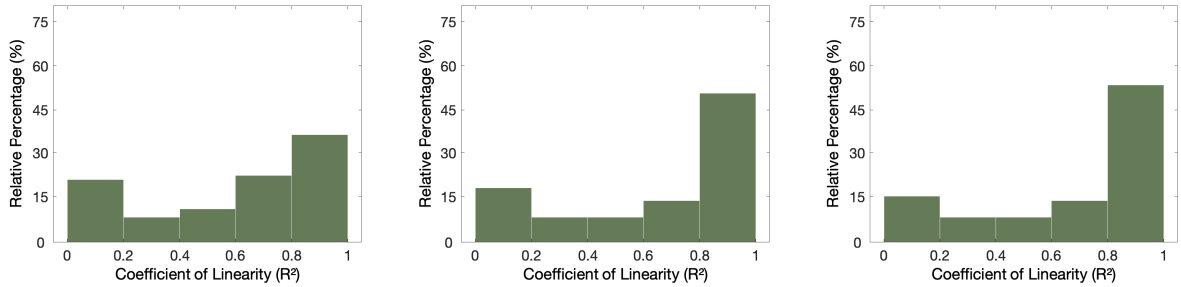
Figure 2.2: Firing rate responses of visual neurons in the hardwired model. Subplots 2.2**A**–2.2**E** show, respectively, the responses of neurons #1, #373, #475, #1117 and #1491 when the visual target was kept at each neuron's preferred retinal location whilst the eyes shifted in integer steps of one degree to all positions within $[-35°, 35°]$. The preferred retinal location of each of these neurons was set to $-10°$, $-5°$, $0°$, $5°$ and $10°$. The preferred eye position of each neuron was $-34°$, $-18°$, $0°$, $18°$ and $34°$, respectively. The coefficient of determination R^2 provides a measure of linearity of the visual responses of each visual neuron. The values of R^2 are defined within $[0, 1]$, where 1 implies a perfect linear fit. In particular, whilst subplot 2.2**C** shows that the response profile of neuron #475 was not linear (i.e. $R^2 = 0$), all other subplots had more linear responses (i.e. values of R^2 closer to 1). Subplots 2.2**A** and 2.2**E** had approximately linear monotonic response profiles covering the whole eye position space. Table 2.1 shows the simulation parameters for the model. These simulation results show that modelling the input components representing both the position of the eye and the retinal location of the visual stimulus using biologically realistic Gaussian functions results in a variety of forms of visual responses similar to parietal responses, including peaked and monotonic.

equal to 0, then none of the visual responses can be predicted by a linear model. In this case the visual responses are referred to as *non-linear responses*.

Subplot 2.2**A**, subplot 2.2**B**, subplot 2.2**D** and subplot 2.2**E**, respectively, show the



(A) $\rho = 2.5°$. **(B)** $\rho = 5°$. **(C)** $\rho = 10°$.



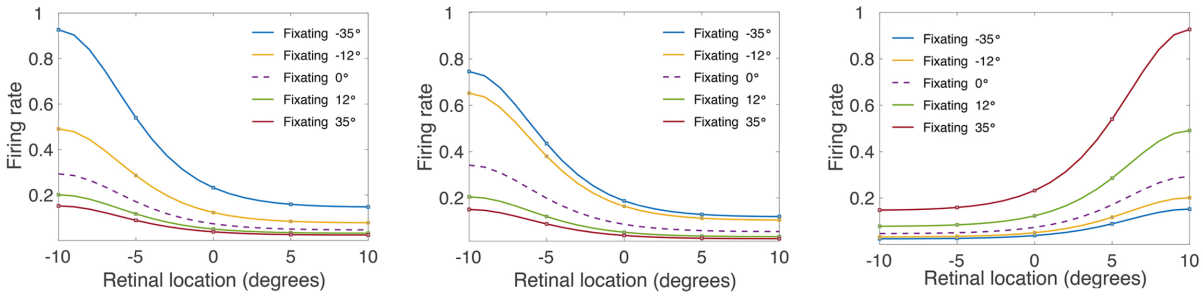
(D) $\rho = 15°$. **(E)** $\rho = 20°$. **(F)** $\rho = 25°$.

Figure 2.3: Influence of varying the width ρ of the eye position tuning curve in the hardwired model. The subplots show how the width ρ of the eye position tuning curve affected the linearity of the eye position gain modulation of the responses of the visual neurons in the hardwired model. Simulation parameters are given in table 2.1. The coefficient of determination R^2 was calculated to assess the linearity of the response curves of each neuron when the visual stimulus was in the neuron's preferred retinal location and the eye position varied in integer steps of one degree within $[-35°, 35°]$. Subplots 2.3**A**–2.3**F** show frequency histograms where individual visual neurons are binned according to their calculated value of R^2. The plotted values are scaled to represent the relative percentages of the entire population of visual neurons. Subplots 2.3**A**–2.3**C** show that the majority of responses were not linear (i.e. values of R^2 were closer to 0) for values of ρ not greater than 10, whilst subplots 2.3**D**–2.3**F** show a higher proportion of more linear responses (i.e. values of R^2 closer to 1) for greater values of ρ. These simulation results show a clear increase in the linearity of the visual responses caused by an increase in the width of the eye position tuning curve.

responses of neurons #1, #373, #1117 and #1491. The response curves of neurons #1 and #1491 had a R^2 value of 0.88, whilst neurons #373 and #1117 had response curves

with R^2 equal to 0.78. The values of R^2 reflect the fact that the visual responses of neurons #1, #373, #1117 and #1491 were linear for the majority of eye positions. Subplots 2.2**A** and 2.2**E** had approximately linear monotonic response profiles covering the whole eye position space. However, the peaked response curve of neuron #475 shown in figure 2.2**C** had a R^2 value equal to 0. Therefore, figure 2.2 shows that modelling the influence of both eye position and retinal location of the visual stimulus using biologically realistic Gaussian functions results in a variety of functional forms of gain modulation of visual responses including both peaked and monotonic. This is a particularly relevant result because experimental studies have reported the existence of these different forms of visual responses in the cortex (Andersen and Mountcastle, 1983, Breveglieri et al., 2009, Brotchie et al., 1995, Galletti et al., 1995, Galletti and Battaglini, 1989, Hoffmann, 1998, Lal and Friedlander, 1990, Merriam et al., 2013, Trotter and Celebrini, 1999).

Next we study the impact of varying the width ρ of the eye position Gaussian tuning curve on the linearity of the eye position gain modulation of the responses of the entire visual population of neurons. We hypothesised that increasing the width ρ of the eye position tuning curve would increase the overall linearity of the gain modulation of visual responses by eye position. Figure 2.3 shows how the linearity of the eye position gain modulation of the responses of visual neurons changed for different values of ρ within $[2.5°, 25°]$. The visual responses were tested for the same range of eye positions in figure 2.2. The changes in the linearity of the responses were assessed using the coefficient



(A) Neuron #1 ($\alpha = -10°, \beta = -34°$).

(B) Neuron #10 ($\alpha = -10, \beta = -25°$).

(E) Neuron #1491 ($\alpha = 10°, \beta = 34°$).

Figure 2.4: Modulation of visual neuron responses by eye position in the hardwired model. The plot shows the gain modulation by eye position of visual responses of neurons #1, #7 and #1491 in the hardwired model. The x-axis (abscissa) indicates the retinal location of the visual stimulus. Table 2.1 gives the simulation parameters for the model. Each curve corresponds to fixations with the following eye positions: $-35°, -12°, 0°, 12°$ and $35°$. The broken line represents fixation straight ahead (i.e. $0°$). The symbols were obtained by multiplying the straight-ahead fixation response curve by a different constant value for each of the other response curves. The precise alignment of the symbols over the response curves shows that the form of modulation of the visual responses by eye position is multiplicative over all the retinal locations of the stimulus.

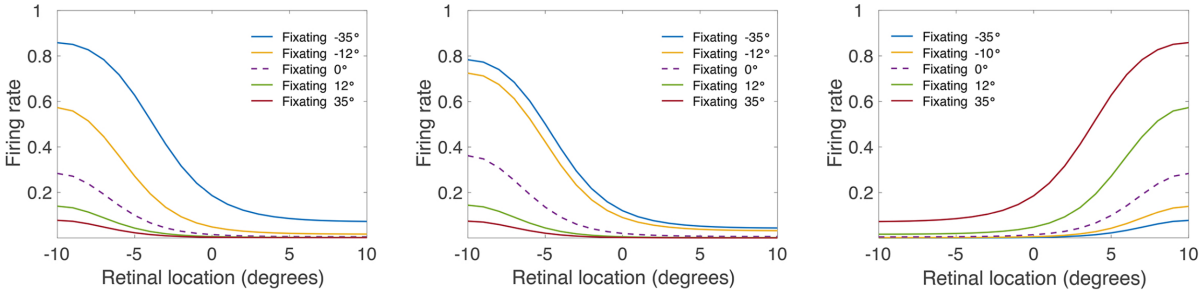
of determination R^2 for each response curve with the visual stimulus positioned in each neuron's preferred retinal location. Table 2.1 shows the simulation parameters for the model. The bars in each subplot represent the proportion of visual neurons in the hardwired model with responses resulting in values of coefficient of determination R^2 from 0 to 1.0 with a bin size of 0.2. Subplot 2.3**A**, subplot 2.3**B** and subplot 2.3**C** show that at least 57% of visual neurons had values of coefficient of determination R^2 smaller than 0.5 when ρ was not greater than $10°$. In other words, more than half of all visual responses were highly non-linear for small values of ρ. However, subplot 2.3**D**, subplot 2.3**E** and subplot 2.3**F** show that the values of coefficient of determination R^2 of the majority of response curves are closer to 1.0 for greater values of ρ. In fact, subplot 2.3**E** and subplot 2.3**F**

show that more than 45% of all visual response curves in the hardwired model had R^2 greater than 0.8. In particular, figure 2.3 shows that the proportion of response curves with coefficient of determination R^2 not greater than 0.2, decreased from approximately 75% to 17% when the value of ρ increased from 2.5° to 25°. These simulation results show that a higher proportion of linear responses is obtained with greater values of ρ.

Finally, we investigate whether the linear responses in the hardwired model (figure 2.2) would produce multiplicative gain modulation by eye position over all retinal locations of the stimulus. Table 2.1 shows the simulation parameters for the model. In particular ρ was set equal to 20°. Figure 2.4 shows the responses of neuron #1, neuron #7 and neuron #1491 in the hardwired model as a function of the retinal location of the visual stimulus and eye position. For each different fixation made to $-35°, 12°, 0°, 12°$ and 35° the responses of each neuron were recorded for all retinal stimulus locations within $[-10°, 10°]$ in steps of one degree. To demonstrate the multiplicative nature of the eye position gain modulation of the neuronal responses, symbols were plotted in figure 2.4 by multiplying the response corresponding to the straight ahead fixation (broken line) by a constant value for each fixation. The precise alignment of the symbols over the response curves shows that the visual responses are modulated multiplicatively by the position of the eye. Thus, figure 2.4 shows that the form of gain modulation by eye position of the visual responses of neuron #1, neuron #7 and neuron #1491 is multiplicative.

Although we carefully controlled the parameter space for the simulation results dis-

cussed above to assure that firing rate values were always within $[0, 1]$, equation 2.4 may produce firing rates greater than 1. Equation 2.6, however, bounds the firing rates between 0 and 1 for all values of the parameter space. Therefore, a further remaining question is whether an approximate to multiplicative effect by eye-position would be empirically observed in simulations with equation 2.6? Figure 2.5 shows the responses of the same neurons presented in figure 2.4 in the hardwired model when the instantaneous firing rate v_i of each visual neuron i was defined by equation 2.6 with $\varphi = 4.5$ and $\theta = 0.77$. The values of other simulation parameters are the same as shown in table 2.1. The responses of each neuron were recorded for the same retinal stimulus locations and for each each different fixation in figure 2.4. The response curves in figure 2.5 show that when a sigmoidal transfer



(A) Neuron #1 ($\alpha = -10°, \beta = -34°$).

(B) Neuron #10 ($\alpha = -10, \beta = -25°$).

(E) Neuron #1491 ($\alpha = 10°, \beta = 34°$).

Figure 2.5: Modulation of visual neuron responses by eye position in the hardwired model with a sigmoidal transfer function The plot shows the gain modulation by eye position of visual responses of neurons #1, #7 and #1491 in the hardwired model with transfer function defined by equation 2.6 with $\varphi = 4.5$ and $\theta = 0.77$. Other simulation parameters are the same as shown in table 2.1. The x-axis (abscissa) indicates the retinal location of the visual stimulus. Each curve corresponds to fixations with the following eye positions: $-35°, -12°, 0°, 12°$ and $35°$. The broken line represents fixation straight ahead (i.e. $0°$). The response curves in each subplot show that there is a form of modulation of the visual responses by eye position that is aproximatelly multiplicative with a sigmoidal transfer function (equation 2.6).

function (equation 2.6) is used to define the instantaneous firing rates of visual neurons in the hardwired model the form of modulation of the visual responses by eye position is aproximatelly multiplicative, similar to the responses shown in figure 2.4.

In this section we presented a hardwired model of multiplicative eye-position gain modulation of the responses of visual neurons in the cortex. In figure 2.2 we showed that modelling both the retinal preference and the eye position preference of visual neurons using a biologically realistic Gaussian function resulted in a variety of forms of eye position gain modulation across the entire population of visual output neurons. This is a particularly important result as different forms of eye position gain modulation of visual neurons have been observed in the cortex (Breveglieri et al., 2009, Galletti et al., 1995, Galletti and Battaglini, 1989, Hoffmann, 1998, Lal and Friedlander, 1990, Merriam et al., 2013, Trotter and Celebrini, 1999). Next we showed that varying the width of eye position tuning curves increased the linearity of eye position gain modulation across the entire population of visual neurons (figure 2.3). In figure 2.4 we showed that the model can produce visual neurons with multiplicative gain modulation by eye position over all retinal locations of the stimulus with firing rates of visual neurons given by equation 2.4. In figure 2.5 we showed that the model with firing rates defined by equation 2.6 can produce visual neurons with a form of gain modulation by eye position over all retinal locations of the stimulus that is approximatelly multiplicative, similar to the responses observed with equation 2.4.

In the next section we study a self-organising model of how multiplicative eye position

gain modulation may develop in the brain. Firing rates in the simulations discussed in the next section are given by equation 2.6. The choice of using a sigmoidal transfer function (equation 2.6) to model the firing rates of visual neurons is justified by the fact that the form of gain modulation by eye-position in the hardwired model with equation 2.6 was found to be approximately multiplicative (figure 2.5) in the region of the parameter space explored whilst also bounding the firing rates between 0 and 1.

2.4 Self-organising model

In this section we present simulation results of a self-organising neural network model of the development of multiplicative gain modulation of visual responses by eye position. That is, we show how the synaptic connectivity could develop through a biologically plausible learning process. In the self-organising model, the visual output neurons are driven by two explicitly modelled populations of input neurons. One input population encodes the retinotopic position of the stimulus, whilst the other input population encodes the position of the eyes. In a manner analogous to the hardwired model, the response profiles of both populations of input neurons are modelled using biologically realistic Gaussian functions. The activations of the visual output neurons were driven by additively combining the inputs from these two populations of input neurons as would occur in the brain. A sigmoid transfer function was used to model the firing rate responses of the visual output neurons in the self-organising model. In this section we investigate the influence of the width of the

eye position input neuron tuning curves on the self-organisation of linear eye position gain modulation at preferred retinal stimulus locations and the development of multiplicative eye position gain modulation across all retinal stimulus locations.

2.4.1 Methods

The architecture of the self-organising neural network model consists of two layers of neurons. The first layer, the *input layer*, had a mixture of neurons with responses depending on either the position of the eye or the retinal location of the visual stimulus. Feedforward synaptic connections from neurons in the input layer to neurons in the second layer, the *output layer*, were updated with a Hebbian learning rule during training. After training, neurons in the output layer were expected to develop the form of retinotopic visual responses with multiplicative eye position gain modulation shown in the previous section (figure 2.2 and figure 2.4). Figure 2.6 shows the architecture of the model.

Neurons in the input population had responses depending on either the position of the eye or the retinal location of the visual stimulus. Neurons with responses depending on the position of the eye had a Gaussian internal activation given by equation 2.3. Each of these neurons is set to fire maximally to a unique eye position within $[-35°, 35°]$ in integer steps of one degree. Neurons with responses depending on the retinal location of the visual stimulus had a Gaussian internal activation described by equation 2.2, where the range of preferred retinal locations is defined within $[-10°, 10°]$ in integer steps of one degree. This

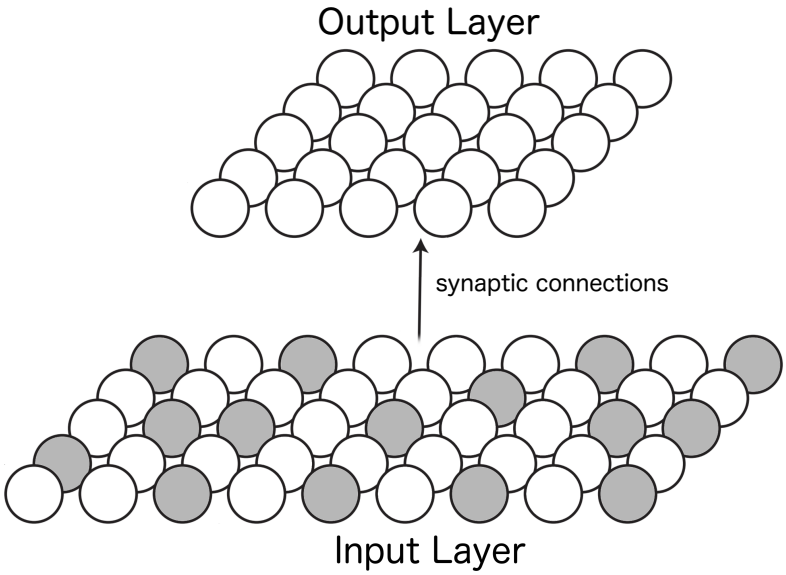


Figure 2.6: Architecture of the self-organising neural network model. Afferent synaptic connections from neurons in the input layer (bottom) are projected to neurons in the output layer (top). The responses of neurons within the input layer depend on either the position of the eye or the retinal location of the visual stimulus. The two colours of neurons in the input layer represent these two different forms of neuronal input responses. A Hebbian learning rule is used to modify the strengths of the feedforward synaptic connections from the input layer to the output layer during training.

resulted in a total of $71 + 21 = 92$ input neurons. The instantaneous firing rate v_j of each input neuron j is given by equation 2.6 where the activation h is defined by either equation 2.2 if the neuron had responses depending on the retinal location of the visual stimulus or equation 2.3 if the neuron had responses depending on the position of the eye.

The output population consists of a total of 100 neurons. The internal activation h_i of each neuron i in the output layer is governed by

$$\tau_h \frac{dh_i}{dt} = -h_i + \sum_j w_{ij} v_j \tag{2.7}$$

where w_{ij} is the strength of the synapse from input neuron j to output neuron i. The

parameter τ_h is a time constant common for all output neurons.

The instantaneous firing rate v_i of each output neuron i is given by the sigmoidal transfer function

$$v_i = \frac{1}{\exp\left(-2\varphi(h_i - p_\pi - \theta)\right)} \tag{2.8}$$

where φ and θ, respectively, denote the slope and threshold of the activation function. The parameter p_π controls the level of competition amongst output neurons by limiting the proportion of output neurons that are active at any time. For instance, p_π is set to the top fifteenth percentile activation value within the output population when the sparseness percentile π is set to 85. This form competition within the output layer has been previously used in neural network models of the primate visual system (Rolls, 2012). In the cortex this form of lateral inhibition is implemented via inhibitory interneurons (Dayan and Abbott, 2001).

Each output neuron receives synaptic connections from a unique randomly assigned subset of input neurons corresponding to ϕ percent of the input population. Before training, the weights of these feedforward synaptic connections are initialised to random values within $[0, 1]$. During training the synaptic weights are updated by a Hebbian learning rule given by

$$\frac{dw_{ij}}{dt} = \varrho v_i v_j \tag{2.9}$$

where ϱ is the learning rate, v_j is the current firing rate of the presynaptic neuron j (equation 2.6) and v_i is the current firing rate of the postsynaptic neuron i (equation 2.8).

To prevent unbounded growth of the synaptic weights during training (Dayan and Abbott, 2001), after each weight update the length of the synaptic weight vector $\mathbf{w}_i$ for each output neuron i is renormalised by setting

$$\mathbf{w}_i := \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \tag{2.10}$$

The experimental work of Royer and Paré (2003) provides evidence for renormalisation of synaptic weights in the brain.

The self-organising model makes no assumptions about the temporal dynamics of eye movements relative to the changing positions of visual stimuli within the environment. Therefore, training involves applying lots of different randomised combinations of eye position and retinal stimulus location to the input layer of the model. The activity is propagated from the input layer to the visual output layer, where a small portion of output neurons win the competition to respond to the current inputs. Next, the feedforward synaptic weights from the active input neurons to the active output neurons are strengthened by Hebbian learning. This training process eventually endows the visual output neurons with their learned response properties.

Training lasts for 10 epochs. In each training epoch, the model is exposed to time-varying combinations of eye position and stimulus retinal location. Specifically, each epoch

consists of periods in which a visual stimulus is presented for $2s$ at each retinal location within $[-10°, 10°]$ in integer steps of one degree. During each of these periods the model performs 14 saccades at a constant velocity of $400°/s$ to a different random eye position within $[-35°, 35°]$. These saccades are interleaved with 15 fixations, each lasting for $300ms$. Therefore, training was completed after approximately $420s$ of simulated time. The differential equations 2.7 and 2.9 were numerically integrated using a Forward-Euler scheme with numerical time step set as one-tenth of the time constant τ_h.

2.4.2 Simulation results

This experiment investigates the development of visual output neurons with multiplicative gain modulation by eye position when the input population consists of a mixture of neurons with Gaussian responses depending on either the position of the eye or the retinal location of the visual stimulus. The influential model by Salinas and Abbott (1996) used a linear function to model how visual responses depended on the position of the eyes. However, the response profiles of eye position neurons in parietal cortex have been observed to be a better fit to a Gaussian function than to a linear function (Dayan and Abbott, 2001, Marĉelja, 1980). Furthermore, the efferent synaptic connections were both excitatory and inhibitory for the same neuron in the model of Salinas and Abbott (1996), which violates the Dale's law (O'Donohue et al., 1985). The synaptic connections within the self-organising model discussed in this section do not violate the Dale's law. Table 2.2 gives the parameters of

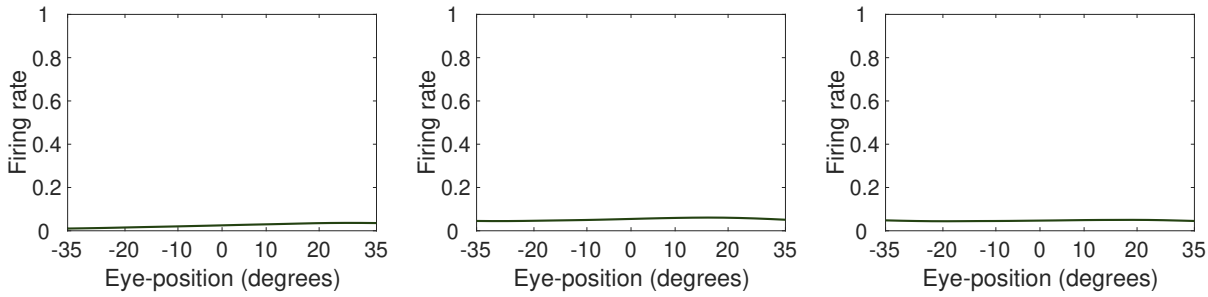
the self-organising model.

Parameter	Symbol	Value
Activation function slope	φ	4.5
Activation function threshold	θ	0.99
Activation time constant	τ_h	$100ms$
Input neuron population size	-	92
Learning rate	ϱ	0.05
Maximum eye position activation	h_{max_e}	2.0
Maximum retinal location activation	h_{max_r}	2.0
Number of training epochs	-	10
Output neuron population size	-	100
Sparseness percentile	π	90%
Synaptic connectivity	ϕ	10%
Width of eye position tuning curve	ρ	$20^{\circ\dagger}$
Width of retinal tuning curve	σ	6°

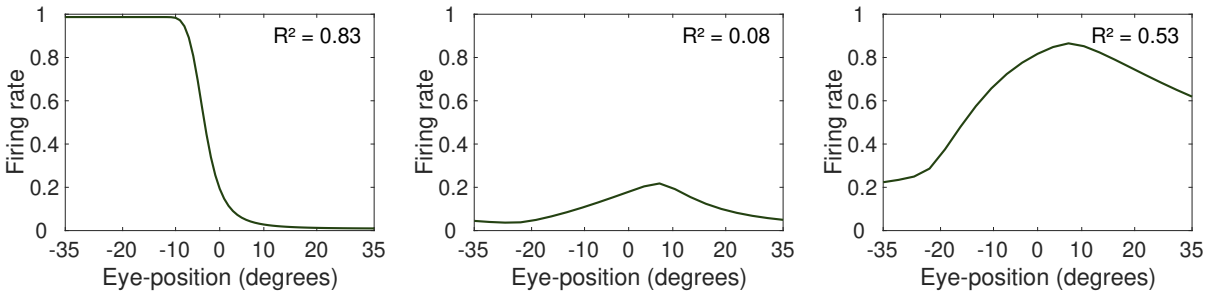
† Except when explictly stated, the value of ρ was set to 20°.

Table 2.2: Simulation parameters of self-organising model.

Figure 2.7 shows the visual responses across different eye positions of selected output neurons before training (top row) and after training (bottom row). Figure 2.7**A** and figure 2.7**D** show the responses of neuron #87 when the visual stimulus was presented at retinal location -10°. For the responses of neuron #36 shown in subplot 2.7**B** and subplot 2.7**E** the visual stimulus was presented at retinal location 0°. Finally, the visual stimulus was presented at retinal location 2° for responses of neuron #97 shown in subplot 2.7**C** and subplot 2.7**F**. The comparison of the responses in the untrained model (subplots 2.7**A**–2.7**C**) with the respective responses in the trained model (subplots 2.7**D**–2.7**F**) shows how training changed the responses of output neurons. For instance, the response curve of output neuron #87 after training resembles the response curve of neuron #1 shown in figure 2.2, with similar coefficients of determination R^2 of 0.83 and 0.88, respectively.



(A) Neuron #87 (untrained). **(B)** Neuron #36 (untrained). **(C)** Neuron #97 (untrained).

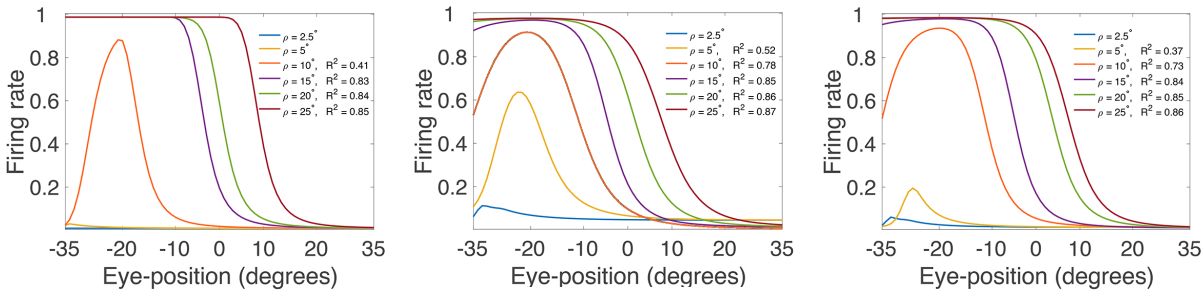


(D) Neuron #87 (trained). **(E)** Neuron #36 (trained). **(F)** Neuron #97 (trained).

Figure 2.7: Firing rate responses of output visual neurons in the self-organising model. Each subplot shows the firing rate responses of output neurons #87, #36 and #97 in the self-organising model either before training (top row) or after training (bottom row). The responses are plotted for the preferred retinal locations of neurons #87, #36 and #97 at $-10°$, $0°$ and $2°$, respectively. These neurons were chosen to show the variety of different forms of visual output responses in the trained model. Table 2.2 shows the simulation parameters for the model. The value of ρ was set to 15. The comparison of responses prior to training (subplots 2.7**A**–2.7**C**) to the respective responses after training (subplots 2.7**D**–2.7**F**) shows that training changed the responses of these output neurons. Particularly, the response curves of neurons #87, #36 and #97 in the trained model (bottom row) resemble the response curves shown in figure 2.2 of neurons #1 (subplot 2.2**A**), #475 (subplot 2.2**C**) and #1117 (subplot 2.2**D**) in the hardwired model, respectively. The values of the coefficient of determination R^2 for each response curve in the trained model show that the form of visual responses in the output layer after training varied from non-linear (e.g. $R^2 = 0.08$ for neuron #36) to more linear responses (e.g. $R^2 = 0.53$ and $R^2 = 0.83$ for neurons #97 and #87, respectively). These simulation results show that the use of more biologically realistic Gaussian functions to model the response profiles of input neurons encoding either eye position or retinal location of the visual stimulus enables the self-organising model to develop a variety of forms of visual responses, including peaked and monotonic, similar to cortical responses.

Likewise, the response curves of neurons #36 (subplot 2.7**E**) and #97 (subplot 2.7**F**) in the trained model resembles the responses of neurons #475 (subplot 2.2**C**) and #1117 (subplot 2.2**D**) in the hardwired model. The values of coefficient of determination R^2 of output neuron #97 in the trained model and neuron #1117 in the hardwired model were 0.53 and 0.78, respectively. The values of coefficient of determination R^2 were very similar for both output neuron #36 in the self-organising model after training ($R^2 = 0.08$) and neuron #475 in the hardwired model ($R^2 = 0$). Therefore figure 2.7, together with the analysis of the coefficient of determination R^2 of the visual response curves, show that modelling the response profiles of eye position and retinotopic input neurons with more biologically realistic Gaussian functions drives the self-organisation of a variety of visual output responses similar to the responses in the hardwired model (figure 2.2).

Next we investigate what would be the impact of varying the width ρ of the eye position tuning curve on the self-organisation of responses in the output population of visual neurons. In particular, we investigate whether greater values of ρ would result in an increase in the linearity of how visual output responses vary with eye position in the self-organising model, similar to the effects presented in previous section. The model was individually trained as described above for the following different values of ρ: 2.5°, 5°, 10°, 15°, 20° and 25°. Simulation parameters are given in table 2.2. Figure 2.8 shows the influence of varying ρ on the self-organisation of linear responses for neuron #87, neuron #92 and neuron #95. The responses of each of these output neurons in the trained model were



(A) Neuron #87. **(B)** Neuron #92. **(C)** Neuron #95.

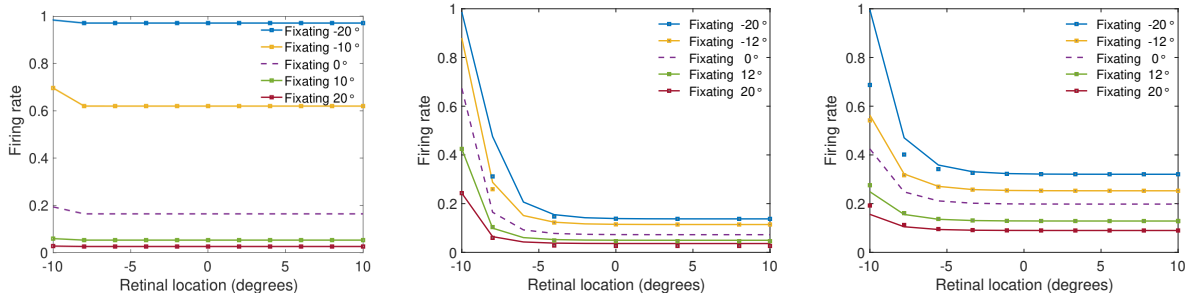
Figure 2.8: Influence of varying the width ρ of the eye position tuning curve in the self-organising model. The subplots show how the width ρ of the eye position tuning curve affected the self-organisation of the visual responses of neuron #87, neuron #92 and neuron #95. Each subplot shows the responses of these output neurons in the trained model when the visual stimulus was presented at retinal location $-10°$ and the eyes shifted from $[-35°, 35°]$ in integer steps of one degree. Table 2.2 gives the simulation parameters for the model. The values of ρ that were used during training and testing are indicated in each subplot. For each of these neurons, the subplots show how increasing the value of ρ changes the firing rate responses from showing little activity to presenting an approximately Gaussian response profile. In particular, each subplot shows that training the model with greater values of ρ resulted in more sigmoidal response profiles for each of these output neurons. These simulations results show that the width of the eye position tuning curve has a large impact on the self-organisation of the functional form of the output responses, with the profiles passing from peaked to monotonic as ρ increases from $10°$ to $25°$. The values of the coefficient of determination R^2 in each subplot show how the width ρ of the tuning curve of the eye position neurons affected the degree of linearity of the self-organised output responses. The values of R^2 were only calculated when the output neuron had significant activity for a range of eye positions after training.

recorded when the visual stimulus was presented at retinal location $-10°$ and the eyes shifted from $[-35°, 35°]$ in integer steps of one degree. Figure 2.8 shows that increasing the value of ρ drives the development of more sigmoidal output visual responses. These simulation results show that the width ρ of the tuning curve of the eye position neurons has an impact on the self-organisation of the degree of linearity of the output responses, as indicated by the values of the coefficient of determination R^2 in each subplot. The values of R^2 were only calculated when the output neuron had significant activity for a range

of eye positions. Subplot 2.8**A** shows that for values of ρ not greater than 5° the firing rate of output neuron #87 in the trained model was close to zero for all eye positions, whilst a more Gaussian response profile was observed when $\rho = 10^\circ$ for the same output neuron. Moreover, subplots 2.8**A** show that for greater values of ρ the response profile of output neuron #87 after training is more sigmoidal across all eye positions. Subplot 2.8**B** and subplot 2.8**C** show that output neuron #92 and output neuron #95, respectively, were mostly unresponsive for $\rho = 2.5^\circ$. However, the response profiles of these neurons were approximately Gaussian for $\rho = 5^\circ$ and $\rho = 10^\circ$. In addition, subplots 2.8**B**–2.8**C** show that the response profiles of output neurons #92 and #95 in the trained model were approximately sigmoidal for all eye positions within $[-35^\circ, 35^\circ]$ when the values of ρ were greater than 10°. Moreover, figure 2.8 shows that the coefficient of determination R^2 increases from values not greater than 0.52 for smaller values of ρ to values greater than 0.7 for larger values of ρ. The effects of varying ρ presented in figure 2.8 were generally observed amongst a large population of output visual neurons after training.

Finally, we studied whether output neurons displayed multiplicative gain modulation by eye position over all retinal locations of the stimulus. Figure 2.9 shows how the output responses of neuron #87, neuron #92 and neuron #95 changed for different fixations when the visual target was shifted across the retina to locations within $[-10^\circ, 10^\circ]$ in integer steps of one degree. Table 2.2 gives the simulation parameters for the model. The value of ρ was set to 17°. The responses of each of these output neurons were tested for

fixations to $-20°, -12°, 0°, 12°$ and $20°$. To demonstrate the multiplicative nature of the eye position gain modulation of the neuronal responses, symbols were plotted in figure 2.9 by multiplying the response corresponding to the straight ahead fixation (broken line) by a constant value for each fixation. The alignment of the symbols over the response curves shows that the visual responses are modulated multiplicatively by the position of the eye. This is the same methodology used in the influential work of Salinas and Abbott (1996). Therefore the simulation results in figure 2.9 show that the position of the eye modulated the responses of output neurons #87, #92 and #95 in a multiplicative manner.



(A) Modulation of responses of neuron #87.

(B) Modulation of responses of neuron #92.

(C) Modulation of responses of neuron #95.

Figure 2.9: Modulation of visual neuron responses by eye position in the self-organising model after training. The plot shows the gain modulation by eye position of the visual responses of neurons #87, #92 and #95 after training in the self-organising model. The x-axis (abscissa) indicates the retinal location of the visual stimulus. The simulation parameters for the model are shown in table 2.2. The value of ρ was set to 17°. Each curve corresponds to fixations with the following eye positions: $-20°, -12°, 0°, 12°$ and $20°$. The broken line represents fixation straight ahead. The symbols were obtained by multiplying the straight-ahead fixation response curve by a different constant value for each of the other response curves. The general alignment of the symbols over the response curves shows that the form of modulation of the visual responses by eye position is approximately multiplicative over all the retinal location of the stimulus. Although subplots 2.9**B**–2.9**C** show that the visual responses were not perfectly multiplicative by eye position for a small part of the response curves corresponding to fixating at $-20°$ (blue curve).

2.5 Conclusion

In this chapter we study both hardwired and self-organised neural network models of how multiplicative gain modulation of visual responses by the position of the eye may develop in the brain. A variety of forms of visual responses, such as peaked, sigmoidal and more linear responses, have been observed in cortical areas (Andersen and Mountcastle, 1983, Breveglieri et al., 2009, Brotchie et al., 1995, Galletti et al., 1995, Galletti and Battaglini, 1989, Hoffmann, 1998, Lal and Friedlander, 1990, Merriam et al., 2013, Trotter and Celebrini, 1999). Previous modelling studies (Salinas and Abbott, 1996) used a linear function to model the dependence of the visual responses on eye position. However, the response profiles of eye position neurons in parietal areas have been reported to be a better fit to Gaussian than to linear functions (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007). The simulation results presented in this chapter showed how multiplicative gain modulation by eye position may develop using inputs from eye position neurons with Gaussian response profiles. We started by studying a hardwired model of how cortical visual neuron responses depend on both the retinal location of the visual target and the position of the eyes. The simulation results in figure 2.2 showed a variety of different functional forms of gain modulation by eye position of visual responses in the hardwired model that were similar to cortical responses. In particular, the responses in subplot 2.2**A** and subplot 2.2**E** in the hardwired model were approximately linear for all eye positions when the visual stimulus was at the neuron's preferred retinal location.

We also investigated how varying the width ρ of the eye position Gaussian tuning curve affected the functional form, including the linearity, of the eye position gain modulation of visual responses. We showed in figure 2.3 that an increase in the width of the eye position tuning curve resulted in more linear responses. These simulation results showed that more biologically realistic Gaussian tuning curves, modelling how visual responses depend on the retinal location of the visual stimulus and on the position of the eye, resulted in different functional forms of gain modulation by eye position similar to those seen in parietal cortex, including responses that were linear across different eye positions. The theory shows that the form of gain modulation by eye-position would be multiplicative with equation 2.4. The simulation results in figure 2.4 showed that this was indeed the case. Furthermore, although the theory does not suggest that the form of gain modulation by eye-position would be perfectly multiplicative with a sigmoidal transfer function (equation 2.6), in figure 2.5 we showed that in practice the form of gain modulation by eye-position was approximately multiplicative for the range of parameters explored.

A further important contribution of the work discussed in this chapter is the investigation of how such a variety of different functional forms of gain modulation by eye position, as well as multiplicatively gain modulated visual responses, would self-organise in a neural network model of cortical neurons. Our self-organising model consisted of a population of input neurons sending feedforward synaptic connections to an output population of visual neurons. The input population consisted of a mixture of neurons with Gaussian responses

depending on either the position of the eye or the retinal location of the visual stimulus. Training consisted of presenting the network with lots of different randomised combinations of retinal stimulus location and eye position. The weights of the synaptic connections were updated during training using a Hebbian learning rule (equation 2.9). Figure 2.7 showed that output neurons in the trained model had responses similar to the ones plotted for the hardwired model in figure 2.2. These simulation results showed that the model successfully self-organised the forms of visual responses observed in the parietal cortex including peaked and monotonic. In particular, subplot 2.7**D** showed that the responses of neuron #87 in the trained model were monotonic and somewhat similar to the approximately linear responses of neuron #1 in the hardwired model (subplot 2.2**A**). The simulation results in figure 2.9 showed that after training the position of the eye modulated the responses of neuron #87 in a multiplicative manner. Additionally, in figure 2.8 we showed how varying the width ρ of the tuning curve of eye position input neurons affected the self-organisation of output neuron responses. Similar to the effects presented in figure 2.4, the simulation results in figure 2.8 showed that increasing the width of the eye position tuning curve resulted in the self-organisation of responses that varied more monotonically with eye position. However, the neuronal response profiles shown in figure 2.8 were only linear over a narrow range of eye positions. This may have been due to the relatively high value of the slope ϕ of the sigmoid transfer function (equation 2.8) used in the simulations. Changing this parameter in isolation made the self-organisation behaviour less robust and the exploration over the

full parameter set is beyond the scope of this paper. A larger slope ϕ produces a more sharp increase in the neuronal firing rate v_i as the input activation h varies.

Gain modulated responses in parietal cortex are critical for coordinate transformation from eye-centred or retinocentric frames of reference to other body-centred frames of reference more suitable to guide motor behaviours, such as head-centered and hand-centered that have been observed in cortex. We have shown in this chapter the emergence of different functional forms of gain modulation, including peaked and monotonic modulation, in both a hardwired and a self-organising neural network model of cortical visual neurons. Chapter 3 and chapter 5 concern the development of visual neurons representing the locations of visual targets in a head-centred frame of reference. In particular, in chapter 3 we show that the forms of eye position gain modulation studied in this chapter are relevant to the development of the head-centred visual responses.

Chapter 3

Self-organisation of head-centred responses peaked and monotonic eye position gain fields

3.1 Introduction

We study a self-organising neural network model of how visual representations in the primate dorsal visual pathway are transformed from an eye-centred to head-centred frame of reference. The model has previously been shown to robustly develop head-centred output neurons with a standard trace learning rule, but only under limited conditions. Specifically it fails when incorporating visual input neurons with *monotonic* gain modulation by eye-position. Since eye-centred neurons with monotonic gain modulation are so common in the dorsal visual pathway, it is an important challenge to show how efferent synaptic connections from these neurons may self-organise to produce head-centred responses in a subpopulation of postsynaptic neurons. In this chapter we show for the first time how a variety of modified, yet still biologically plausible, versions of the standard trace learning rule enable the model to perform a coordinate transformation from eye-centred to head-centred reference frames when the visual input neurons have monotonic gain modulation

by eye-position. The simulation results presented in this chapter have been published in Navarro et al. (2018).

3.2 Hypothesis

Within the primate dorsal visual pathway, neurons encode the locations of visual objects in different body-centred frames of reference. For example, neurons at an early stage of processing encode the locations of objects within an eye-centred reference frame, while neurons in later stages may encode the positions of objects with respect to the head or hand which is more relevant for guiding motor actions. A key question is how the visual system learns to perform such coordinate transformations between different body-centred reference frames. The neural mechanisms supporting this form of transformation, critical for visually guided motor function, have long been studied in primates. A neural phenomenon which is thought to play a key role in coordinate transformations in the dorsal visual pathway is *gain modulation*. This refers to the modulatory effect that a particular bodily state or posture, like eye or head position, has on the firing rate of some visual neurons responding in a given reference frame. Gain modulation of neuronal firing responses has been discovered in multiple stages of the primate dorsal visual pathway.

Parietal area 7a (Andersen and Mountcastle, 1983, Andersen et al., 1985), LIP (Andersen et al., 1990) and PO (Galletti et al., 1995) have visual neurons modulated by the position of the eye in the orbit. This form of gain modulation is thought to be involved in supporting the development of head-centred visual representations, which were later identified in the same areas (Duhamel et al., 1997, Galletti et al., 1993, Mullette-Gillman et al., 2005). The parietal reach region (PRR) has hand-centred visual neurons with eye-

position gain modulation (Chang et al., 2009), which are thought to be involved in guiding reaching to target locations. In certain regions such as parietal area 7a and LIP, the gain modulation is usually a *monotonic* function of the relevant bodily posture, such as eye or head position. For example, eye-position modulation often takes the form of a linear or saturating function of eye-position, which multiplicatively modulates the underlying Gaussian retinotopic receptive field of a visually responsive neuron (Andersen and Mountcastle, 1983). While in other areas, such as V6A, there can be a higher proportion of retinotopic visual neurons with peaked eye-position gain fields (Breveglieri et al., 2009).

Contemporary understanding of the potential role of such gain modulation in coordinate transformation has been informed by the highly influential coordinate transformation model of Zipser and Andersen (1988), reviewed in chapter 1. The majority of models studying coordinate transformation since that time (Chang et al., 2009, Mazzoni et al., 1991, Pouget and Sejnowski, 1997, Pouget and Snyder, 2000, Xing and Andersen, 2000) have been heavily inspired by this early work. These models set up the synaptic weight matrix by some form of supervised, error correction, learning algorithm which cannot be implemented in the cortex (Rolls and Treves, 1998). Such algorithms provide no biologically plausible learning hypothesis, and also produce circuits which violate Dale's law, the widely accepted neuroanatomical fact that a given presynaptic neuron cannot be both excitatory and inhibitory across its efferent synapses (O'Donohue et al., 1985).

Our laboratory has developed a biologically plausible neural network model that self-organises its synaptic connectivity during visual experience such that the model learns to transform eye-centred visual representations to head-centred representations (Mender and Stringer, 2014, 2013). The model can achieve this using purely associative local synaptic

learning rules with no supervisory training signal - i.e. unsupervised learning. The model is able to self-organise its synaptic connectivity by exploiting the natural eye and head movements of primates. However, a limitation of previous studies with this model has been their reliance on incorporating retinotopic visual input neurons with responses that are modulated by only *peaked* eye-position gain fields. This is not biologically realistic because many retinotopic visual neurons in the monkey brain are found to have *monotonic* eye-position gain fields. In the new simulations presented below, we first show that the incorporation of retinotopic visual input neurons with monotonic gain fields leads to a failure of the model to develop head-centred output neurons. We then show how model performance can be substantially improved by employing a range of more sophisticated, yet still biologically plausible, learning rules. A review of relevant physiological and behavioural data is presented in chapter 1 along with a discussion of some previous computational models in order to provide the context for the new simulation results discussed in this paper.

3.3 Self-organising Neural Network Model of Coordinate Transformation

Our laboratory has previously published a biologically plausible neural network model of the visually-guided development of head-centred visual neurons, which relies on associative learning rules and does not require a supervisory training signal to guide learning (Mender and Stringer, 2013). Instead, the model utilises an unsupervised form of competitive learning that exploits the natural statistics of how primates move their eyes and head as they shift their gaze around their visual environment. Specifically, the model employs a

form of *trace learning*, which encourages neurons in higher network layers to bind together input patterns that tend to occur close together in time. If a primate tends to move its eyes more frequently than adjusting the position of its head, then retinal images that occur close together in time will tend to correspond to different positions of the eyes but the same position of the head. In this case, trace learning will encourage neurons in higher layers to learn to respond to the position of a visual target in the same head-centred location across different retinal positions. Such neurons will have thus learned to represent the position of visual targets within a head-centred reference frame. The inputs to the model are eye-centred visual neurons that represent the locations of visual targets on the retina, but which have responses that are also gain modulated by the position of the eyes in the socket. In this paper, we investigate how the learning process depends on the functional form of this gain modulation by eye-position. Two forms of eye-position gain modulation are explored: monotonic gain modulation, which is dominant in most primate parietal areas (LIP, 7a, PRR), and peaked gain modulation which is primarily found in area PO.

In the simulations that we have run in this chapter we show that the model can robustly develop head-centred output neurons with a standard trace learning rule when incorporating visual input neurons with *peaked* eye-position gain modulation, which has also been shown by Mender and Stringer (2013). However, our simulation results in this chapter show that this is not the case when the input neurons are modulated with *monotonic* eye-position gain fields. Moreover, our simulation results in this chapter revealed that even if the model has its synaptic connectivity perfectly prewired to perform a coordinate transformation from eye-centred input neurons with monotonic gain modulation to head-centred output neurons, subsequently introducing plasticity into the synaptic connections with

the standard trace learning rule quickly degrades the synaptic connectivity and eventually leads to elimination of the head-centred output responses. This problem has also been raised in Mender (2014). Since eye-centred visual neurons with monotonic eye-position gain modulation are so common in the dorsal visual pathway (Andersen et al., 1985, Daddaoua et al., 2014, Merriam et al., 2013), it is an important challenge to show how efferent synaptic connections from these neurons may self-organise to produce head-centred visual responses in a subpopulation of postsynaptic receiving neurons. A subsequent analysis that we have carried out in this chapter of the nature of the failure of the self-organisation of the synaptic connectivities led us to explore the performance of the model with a variety of modified, yet still biologically plausible, more powerful versions of the standard trace learning rule. The choice of the modified versions of the trace learning rule used in this paper was motivated by the superior performance of these learning rules reported by Rolls and Stringer (2001). Here we show for the first time how these modified learning rules enable the model to learn to perform a coordinate transformation from eye-centred to head-centred reference frames even when the visual input neurons have monotonic gain modulation by eye-position.

3.4 Network Components

The model uses four core components to self-organise head-centred visual representations through a biologically plausible process of visually guided learning. The first component is a population of input neurons that encode both the position of the eyes in the orbit and the retinotopic location of visual targets. Such retinotopic visual neurons with eye-position gain modulation have been identified in multiple primate cortical areas (Andersen

et al., 1990, 1985, Daddaoua et al., 2014, Galletti et al., 1995, Merriam et al., 2013). The second component is a population of output neurons that compete with each other through mutual inhibitory interactions, which is a standard feature of cortical architecture (Rolls and Treves, 1998). The third component is a local synaptic *trace learning* rule to update the feedforward synaptic connections between the input and output neurons. The trace learning rule is a local associative learning rule that incorporates an exponentially decaying temporal trace of past neuronal activity, and it has been widely used in the context of developing invariant visual object recognition (Földiák, 1991, Wallis and Rolls, 1997). The trace learning rule has the effect of encouraging individual postsynaptic neurons to learn to respond to subsets of input patterns that tend to occur close together in time. Finally, the fourth component comes under the assumption that visual stimuli are relatively static in a world reference frame during natural self-motion. This assumption is justified by experimental findings in which primates adjust their gaze more often by moving their eyes rather than the head itself (Freedman and Sparks, 1997). This behavioural strategy to adjust gaze is preferable because it reduces the frequency of making more energetically costly and slow head movements. In fact, it has been found that during exploration of natural environments with free head, eye and body movements, at any time when there is movement, isolated head and isolated eye movements occurred 13.3% and 33.1% of the time respectively, whilst a mixture of movements was observed in the remaining time. That is, during natural movement there are periods when the eyes are moving whilst the head remains stationary with respect to the visual environment and visual objects also remain stationary within the environment.

These four model components allow the model to self-organise its synaptic connectivity

during visually guided training in the following way. If the eyes move around a scene containing a stationary visual target while the head also remains stationary, then the visual system will receive a sequence of input patterns with the visual target in different retinal locations but the same head-centred location. That is, the visual target remains stationary in the head-centred space, but changes its position in the eye-centred space. The sequence of eye-positions and resulting retinal locations of the visual target are represented by retinotopic visual input neurons with responses that are gain modulated by eye-position. The synaptic trace learning rule is able to bind subsets of input patterns corresponding to a visual target in the same head-centred location, albeit with the visual target situated in different eye-centred locations, onto the same output neurons. This is because input patterns corresponding to a visual target situated in the same head-centred location tend to occur close together in time due to the statistics of natural eye and head movements, in which the eyes tend to saccade about a static visual scene while the head remains stationary. Moreover, the naturally rapid movements of the eyes may expose the visual system to many such input pattern sequences, where each such sequence has the visual target situated in the same head-centred location but different randomised subsets of retinal locations. This process will ensure that all possible input patterns corresponding to the same head-centred location but different retinal locations are eventually brought into temporal proximity with each other as training progresses. Hence, all input patterns corresponding to the same head-centred location but different retinal locations would tend to occur clustered together in time. This process continues with the visual target seen in a different position within the head-centred space every time the position of the head is shifted. That is, the natural head movements that occur between sequences of rapid eye movements would

shift the location of the visual target to new head-centred locations. In this manner, the learning process is repeated with the visual target presented in many different head-centred locations. New subsets of output cells would learn to respond to the visual target in each different head-centred location due to the competitive interactions between the output cells. Consequently, the output layer would eventually develop neurons that cover the entire space of head-centred target locations.

3.5 Network Architecture

The neural network architecture of the model is shown in figure 3.1. The network consists of the following two layers of neurons. The first layer is a population of input neurons that simultaneously encode the eye-position of the agent and the retinal location of the visual target. These visual input neurons are modelled as retinotopic neurons with eye-position gain modulation. The eye and retinal position spaces, representing the range of eye-positions in orbit and retinal target locations, covered $[-30°, 30°]$ and $[-90°, 90°]$, respectively. Feedforward synaptic connections project from neurons in the input layer to neurons in the second layer.

The second layer is a population of N output neurons that compete to represent patterns in the input layer (Rolls and Treves, 1998). Neurons in the second layer, the *output layer*, all receive the same number of afferent connections from neurons in the input layer, that is ϕ percent of the input population, but each output neuron receives connections from its own randomly assigned subset of the input neurons. Neither the output layer nor the input layer is topographically organised.

The strengths of the feedforward synaptic connections from the input layer to the

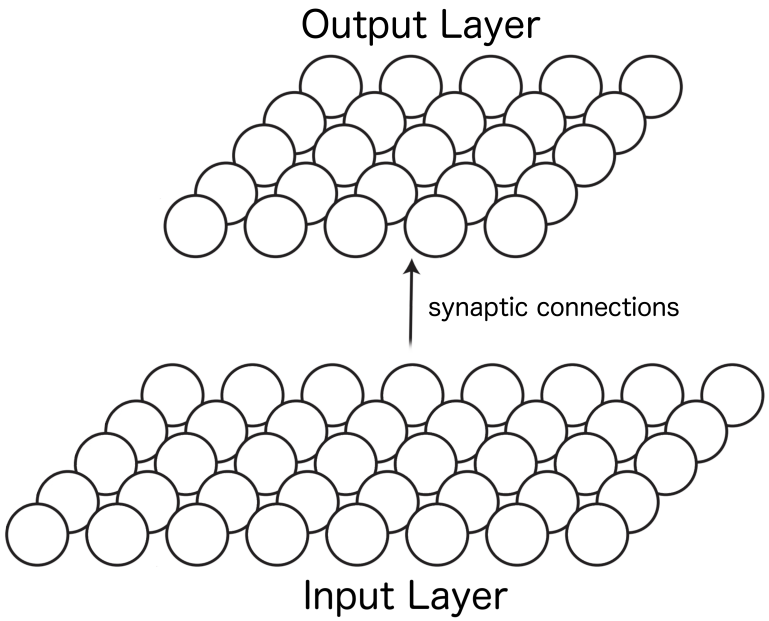


Figure 3.1: Architecture of the neural network model. The competitive output layer on the top receives afferent synaptic connections from neurons in the input layer on the bottom. A trace learning rule is used to modify the strengths of the feedforward synaptic connections from the input layer to the output layer during learning.

output layer are initialised to random weights in the interval $[0, 1]$ at the start of each simulation. The synaptic weight vector of each output neuron is then renormalised as is typical in competitive networks (Rolls and Treves, 1998).

3.6 Training the Network

The neural network is trained on input patterns that simultaneously encode both the position of the eyes in orbit and the retinotopic position of the visual target. The position of the eyes is kept within $[-24°, 24°]$, whilst the retinal locations of visual targets are kept within the interval $[-63°, 63°]$. Keeping the position of the eyes and the retinal locations of visual targets within these respective intervals reduces edge effects due to clipping of the input representations.

Similarly, all M evenly spaced head-centred locations chosen for each experiment are confined within $[-63°, 63°]$ to ensure that visual targets always remain in view as the eyes move. Each training epoch consists of M periods, where each period individually corresponds to one of the M chosen head-centred locations. During training, for each period a visual target is fixed in a given head-centred location whilst the eyes perform a series of saccades to P different and uniformly sampled eye-positions within $[-24°, 24°]$. The saccades between successive eye-positions are performed at a constant velocity of $400°/s$. The duration of each fixation is set to $300ms$.

Thus, the training process consisted of presenting to the network sequences of combined visual and eye-position input signals, which represent the visual target in fixed head-centred locations, whilst the eyes randomly shifted to different positions in the orbit.

3.7 Testing the Network

The responses of the output units for all combinations of T head-centred visual target locations and E eye fixation positions are recorded after training to test the model. In order to test the ability of the trained model to generalise to new input patterns, the model is tested with a set of novel combinations of eye-position and visual target location not encountered during training. Specifically, the following $E = 4$ eye-positions are used during testing: $-18°, -6°, 6°$ and $18°$. For each of these eye-positions, the visual target is shifted in increments of $2°$ every $330ms$ to the next one of the $T = 80$ head-centred target locations within $[-79°, 79°]$. The firing rates of all output neurons are saved at the end of each fixation period to analyse the receptive field properties of the neurons, including the receptive field size, receptive field location and reference frame of response.

3.8 Network Dynamics

3.8.1 Input Layer

The firing rates of neurons in the input layer were modelled by a response function that encodes the retinotopic location of a visual target, where the responses were modulated by the position of the eyes in the orbit. The response function of each input neuron j maps the current retinal location r of the visual target and the eye-position e onto the neuron's instantaneous firing rate v_j^I. The instantaneous firing rate values are defined within the range $[0, 1]$. We investigated the performance of the model when the response functions of the input neurons were modulated by two different functional forms of eye positon gain field as described next.

The first response function has a peaked eye-position gain field as shown in figure 3.2**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

$$v_j^I = \exp\left(-\frac{\|e-\beta_j\|^2}{2\rho^2}\right) \times \exp\left(-\frac{\|r-\alpha_j\|^2}{2\sigma^2}\right) \tag{3.1}$$

This response function is formed from a product of two components: the first component represents the eye-position signal, whilst the second component encodes the retinotopic position r of the visual target. In equation 3.1, the neuronal response is modulated by a peaked Gaussian function of eye-position. The parameter β_j denotes the preferred eye-position for each input neuron j. The width of the corresponding Gaussian eye-position tuning curve is determined by the standard deviation ρ. The preferred retinal location of a target stimulus for each input neuron j is denoted by the parameter α_j. The width of the

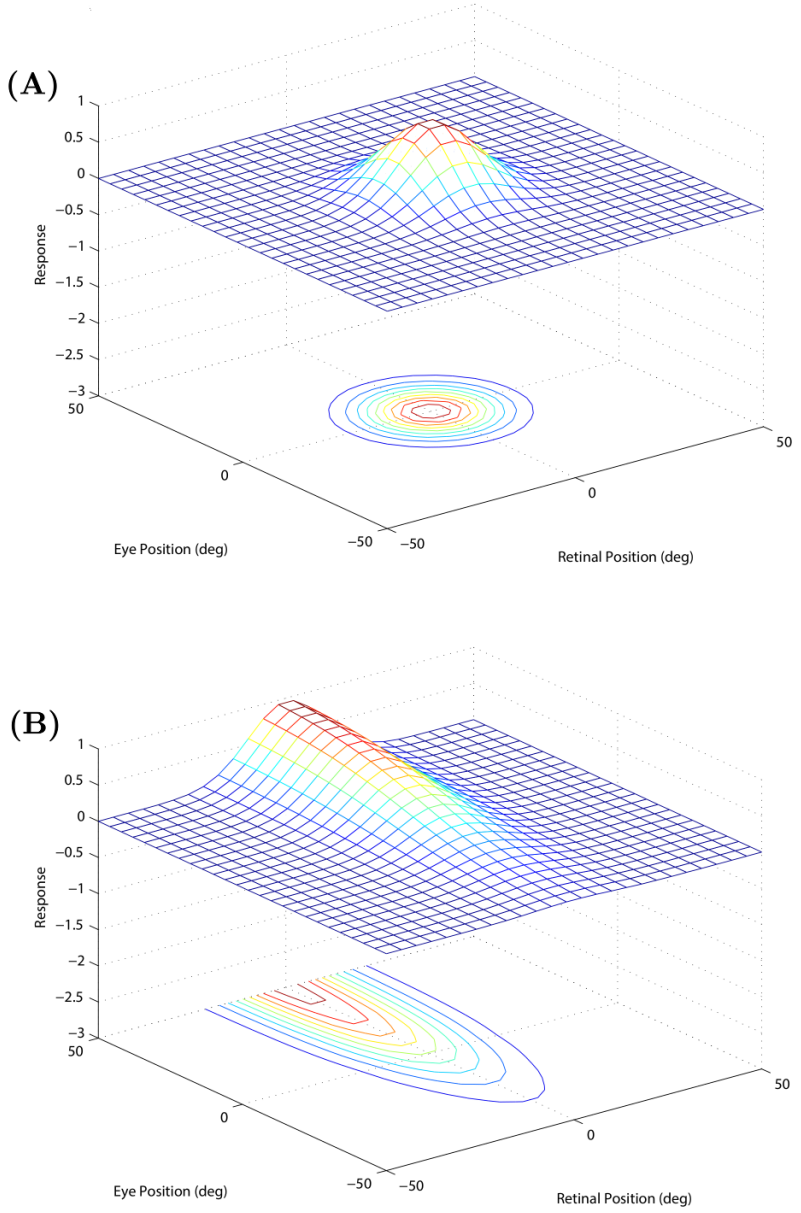


Figure 3.2: Examples of the two alternative response functions used to compute the firing rates of input neurons. For each type of response function, we plot the responses of an individual input neuron as a function of eye-position and retinal location of a visual target. Subplot 3.2**A** shows an example of a peaked response function, in which the firing rate is modulated by a peaked eye-position gain field as described by equation 3.1. Subplot 3.2**B** shows an example of a monotonic response function that is modulated by a sigmoidal (monotonic) eye-position gain field described by equation 3.2.

corresponding Gaussian retinal tuning curve is determined by the standard deviation σ. Each input neuron j is set to respond maximally to a unique combination of retinal target location (α_j) and eye-position (β_j). The entire two dimensional space consisting of possible combinations of eye-position and retinal target location is covered by the population of input neurons in integer steps of one degree in each dimension. This results in a total of $201 \times 61 = 12,261$ neurons in the input layer.

The second form of firing rate response function used to model the input neurons incorporates a sigmoid eye-position gain field as shown in figure 3.2**B**. This form of eye-position modulation is monotonic in the eye-position dimension, as has been observed in multiple visual areas (Andersen et al., 1990, 1985). Although the form of the modulation is sigmoidal whilst most empirical work has described it as planar, Pouget and Sejnowski (1997) showed the data is also compatible with a saturating sigmoidal gain formulation. The full response function is described by

$$v_j^I = \frac{1}{1 + \exp\left(\kappa_j(e - \beta_j)\right)} \times \exp\left(-\frac{\|r - \alpha_j\|^2}{2\sigma^2}\right) \tag{3.2}$$

In equation 3.2, the visual receptive field is modulated by a sigmoidal function of eye-position. For each input neuron j, the parameter κ_j determines the gain direction and saturation rate of the modulation, where κ_j is $-2\times$ the slope of the sigmoid. The inflection point β_j determines the eye-position where a firing rate response greater than 0.5 begins. The input neurons all have the same absolute saturation rate, that is $|\kappa_j| = |\kappa_m|$ for all j and m, but one half has a positive gain direction ($\kappa_j > 0$) whilst the other half has a negative gain direction ($\kappa_j < 0$). Each input neuron j is set to respond maximally to a unique combination of retinal target location α_j, eye-position β_j, and gain direction

and saturation rate κ_j. The population of input neurons evenly covers the entire three dimensional space resulting from such combinations. This results in an input population of $201 \times 61 \times 2 = 24,522$ neurons.

3.8.2 Output Layer

Three dynamical quantities were defined for each neuron i in the competitive output layer: an internal activation h_i, a memory trace value q_i, and an instantaneous firing rate v_i (Dayan and Abbott, 2001).

The internal activation is governed by the equation

$$\tau_h \frac{dh_i}{dt} = -h_i + \sum_j w_{ij} v_j^I \tag{3.3}$$

where w_{ij} denotes the strength of the synapse from the j^{th} input neuron to the i^{th} output neuron and τ_h is a time constant common for all output neurons.

The instantaneous firing rate is given by equation 2.8. The way of implementing competition within the competitive output layer in equation 2.8 has been previously used in competitive neural network models of the primate visual system with trace learning (Rolls, 2012). In the cortex, lateral inhibition is implemented via inhibitory interneurons (Dayan and Abbott, 2001).

3.8.3 Modification of Synaptic Weights by Trace Learning

Trace learning rules for synaptic modification encourage postsynaptic (output) neurons to bind together subsets of input patterns that tend to occur close together in time by incorporating a temporal trace of recent neuronal activity. The trace value q_i for the i^{th}

neuron in the output layer is given by

$$\tau_q \frac{dq_i}{dt} = -q_i + v_i \tag{3.4}$$

where v_i is the instantaneous firing rate of the neuron, and τ_q is a time constant common for all output neurons.

In the first part of this chapter, during training the strength of the synapse from the j^{th} input neuron to the i^{th} output neuron is governed by the standard trace learning rule defined by

$$\frac{dw_{ij}}{dt} = \varrho q_i v_j^I \tag{3.5}$$

where ϱ is the learning rate, v_j^I is the instantaneous firing rate of the j^{th} input neuron and q_i is the trace value of the i^{th} output neuron. However, later in this paper we will introduce a number of new, more powerful forms of trace learning, which are in fact needed to produce head-centred output neurons when the input neurons are modulated by a sigmoidal (monotonic) function of eye-position.

Finally, after each weight update during training, the length of the weight vector for each output neuron i, that is $\mathbf{w}_i = (w_{i1}, \ldots, w_{iN_I})$ where there are N_I input neurons, is renormalised according to equation 2.10.

3.8.4 Simulation of the Differential Model

The Forward-Euler scheme is used to numerically integrate the coupled differential equations given by equation 3.3 and equation 3.4-3.5. The numerical time step Δt is set to one tenth of the neuronal time constant τ_h. We checked that the simulation results remained similar if the time step was reduced or the number of training epochs increased.

During training and testing, the input patterns encoding the changing eye-position and retinotopic target location are simulated dynamically and sampled at 1kHz. Linear interpolation is used to compute the numerical inputs to the discretised Forward Euler model equations, which require input values at every numerical time step $\Delta t = \tau_h/10$.

3.9 Analysis

In order to analyse whether individual output neurons are predominantly responding in an eye-centred or head-centred frame of reference, we used a method of analysis originally developed by Mender and Stringer (2013).

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

To determine which reference frame an output neuron is responding in during testing, two separate metrics are applied that reflected to what degree the neuronal response is compatible with either an eye-centred or head-centred reference frame, and then the values of these two metrics are compared.

The head-centredness metric computed the degree to which the head-centred response vectors of a neuron remained stable across different eye-positions. The head-centredness

metric measured the degree of such stability for a given output neuron by averaging correlations between response vectors for different eye-positions, that is

$$\Pi = \frac{1}{\binom{E}{2}} \sum_{1 \leq i_1 < i_2 \leq E} \frac{\sum_{j=1}^{T} \left(\mathbf{R}[i_1, j] - \overline{\mathbf{R}}[i_1]\right) \left(\mathbf{R}[i_2, j] - \overline{\mathbf{R}}[i_2]\right)}{\sqrt{\sum_{j=1}^{T} \left(\mathbf{R}[i_1, j] - \overline{\mathbf{R}}[i_1]\right)^2 \sum_{j=1}^{T} \left(\mathbf{R}[i_2, j] - \overline{\mathbf{R}}[i_2]\right)^2}} \tag{3.6}$$

where

$$\overline{\mathbf{R}}[i] = \frac{1}{T} \sum_{j=1}^{T} \mathbf{R}[i, j] \tag{3.7}$$

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

A very similar analysis is done to quantify the compatibility of the responses of the output neuron with an eye-centred frame of reference. That is, a visual neuron is judged to respond in an eye-centred frame of reference to the extent that its eye-centred response vectors remain stable across different eye-positions. The eye-centred analysis proceeded as follows. To reiterate, each response vector $(\mathbf{R}[i, 1], \ldots, \mathbf{R}[i, T])$ is the result of testing over the same set of head-centred 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 is these portions that are 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 Appendix A. This gave the metric

$$\Omega = \frac{1}{\binom{E}{2}} \sum_{1 \le i_1 < i_2 \le E} \frac{\sum_{j=0}^{V-1} \left(\mathbf{R}[i_1, f_{i_1}+j] - \overline{\mathbf{R}}[i_1]\right)\left(\mathbf{R}[i_2, f_{i_2}+j] - \overline{\mathbf{R}}[i_2]\right)}{\sqrt{\sum_{j=0}^{V-1} \left(\mathbf{R}[i_1, f_{i_1}+j] - \overline{\mathbf{R}}[i_1]\right)^2 \sum_{j=0}^{V-1} \left(\mathbf{R}[i_2, f_{i_2}+j] - \overline{\mathbf{R}}[i_2]\right)^2}} \tag{3.8}$$

where

$$\overline{\mathbf{R}}[i] = \frac{1}{V} \sum_{j=0}^{V-1} \mathbf{R}[i, f_i + j] \tag{3.9}$$

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

A neuron is finally classified as head-centred if $\Pi > 0$ and $\Pi > \Omega$, and classified as eye-centred if $\Omega > 0$ and $\Omega > \Pi$. If neither of these conditions is met then the neuron remains unclassified.

3.10 Results

3.10.1 Self-organisation with peaked and monotonic gain fields

This experiment explores the feasibility of the self-organisation of head-centred receptive fields under the two different forms of eye-position gain modulation. Two models are trained and tested on the same stimuli, where one model has peaked eye-position modulation in the input population as shown in figure 3.2**A**, and the other model has sigmoidal modulation as shown in figure 3.2**B**. The training lasts for 20 epochs. During each training epoch, a visual target is presented for approximately $5s$ in each of the eight head-centred

training locations: $-63°, -45°, -27°, -9°, 9°, 27°, 45°$ and $63°$. For each period where the visual target is in a fixed head-centred target location, the eye-position is varied continuously through time as the model makes a series of saccades and fixations. During each such period, the model performs 14 saccades interleaved with 15 fixations, where each fixation lasts $300ms$. Each saccade is at a constant velocity of $400°/s$, and it is directed to a random eye-position within the range $[-24°, 24°]$. Each training epoch thus lasts for approximately $40s$, and the entire training of the network is completed after about $800s$ of simulated time. The model is tested as previously described. The parameters for the two model simulations are given in table 3.1.

Parameter	**Symbol**	**Value (peaked)**	**Value (monotonic)**
Number of target locations	M	8	8
Fixation sequence length	P	15	15
Number of training epochs	-	20	20
Width of peaked eye-position tuning curve	ρ	6°	-
Width of retinal tuning curve	σ	6°	6°
Gain magnitude	$\|\kappa\|$	-	0.0625
Output neuron population size	N	900	900
Input neuron population size	N_I	12261	24522
Trace time constant	τ_q	400ms	400ms
Activation time constant	τ_h	100ms	100ms
Activation function slope	φ	4.5	4.5
Activation function threshold	θ	0.4	0
Sparseness percentile	π	80%	90%
Learning rate	ϱ	0.05	0.05
Synaptic connectivity	ϕ	5%	5%

Table 3.1: Simulation parameters of self-organising models with either peaked or monotonic eye-position gain modulation.

Figure 3.3 compares the firing rate responses of the output neurons before and after training in the two models with either peaked or sigmoidal gain modulation of the visual input neurons. The responses of an output neuron from the model with peaked eye-

position gain modulation before training (figure 3.3**A**) exhibits no consistent structure in head-centred space across the different eye-positions. However, after training (figure 3.3**B**) there is a clear maximal response to the same head-centred location across all eye-positions. Therefore, the self-organisation process has made the response reference frame of this output neuron strongly head-centred. The responses of an output neuron from the model with sigmoidal eye-position gain modulation before training (figure 3.3**C**) also has an erratic and more eye-centred response prior to training due to the randomly assigned synaptic weights. However, unlike the peaked gain modulation model, training has the effect of making the neuron almost perfectly eye-centred. This is clearly seen by the receptive fields shifting in head-centred space in register with the eye-position shifts. Therefore, the self-organisation process has made this output neuron even more compatible with an eye-centred reference frame. The miniature scatter plots show the reference frame values of all neurons in the output layer, where each neuron is plotted as a point corresponding to that neuron's particular combination of head-centredness and eye-centredness. The miniature scatter plots confirm the same general effects across the entire populations of output neurons. That is, the subplot in 3.3**B** shows that a large proportion of the output neurons in the model with peaked gain modulation have a high head-centredness and low eye-centredness, and thus respond in a head-centred reference frame. While the subplot in 3.3**D** shows that a large proportion of the output neurons in the model with sigmoidal gain modulation have a low head-centredness and high eye-centredness, and thus respond in an eye-centred reference frame.

Figure 3.4 shows the synaptic weight vectors of the same output neurons as those shown in figure 3.3. Before training, there is no structure to the potentiated synapses of

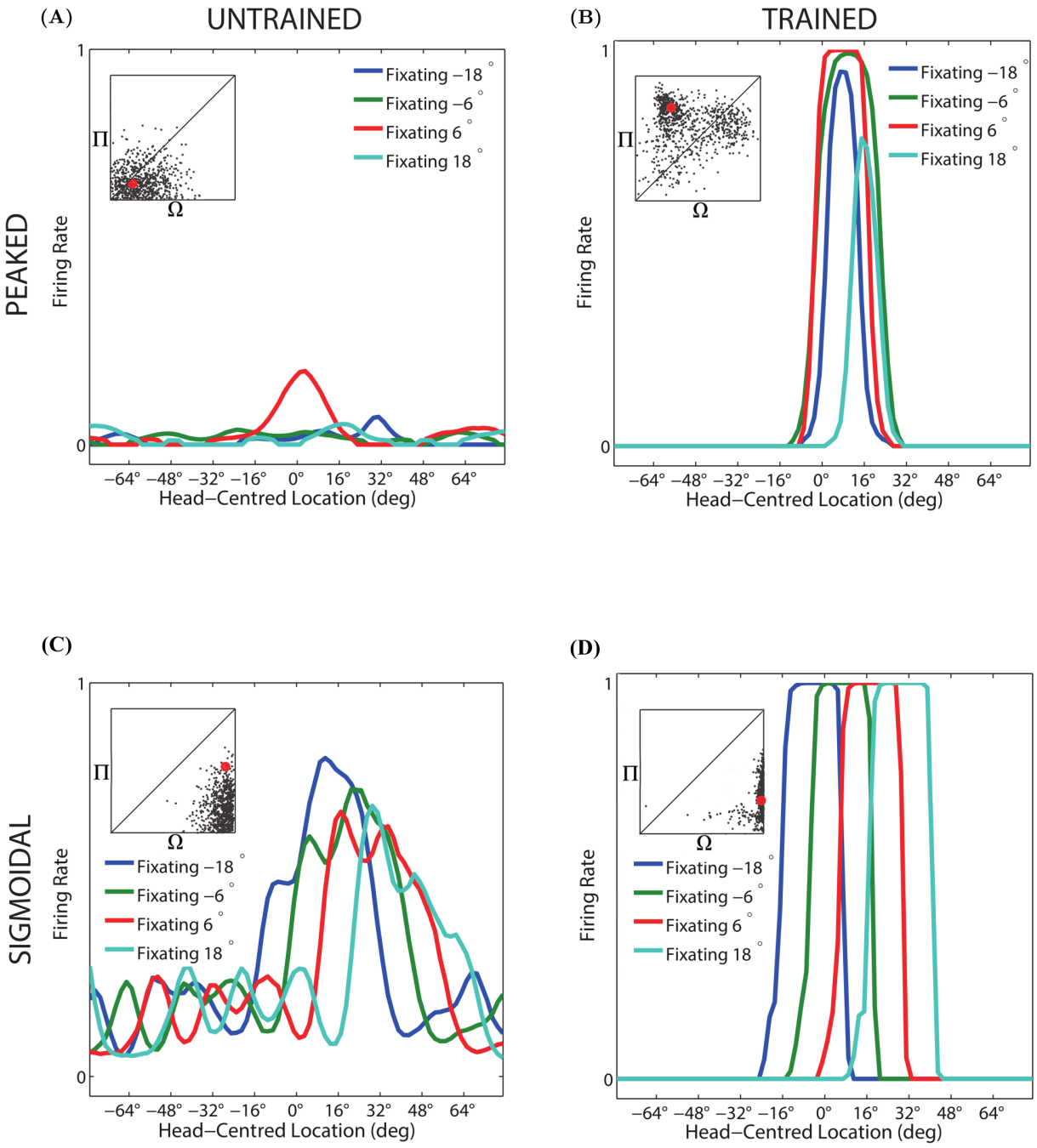

(A)
UNTRAINED
(B)
TRAINED
PEAKED
SIGMOIDAL
Firing Rate
Head–Centred Location (deg)
Fixating −18°
Fixating −6°
Fixating 6°
Fixating 18°
Π
Ω
1
0
−64° −48° −32° −16° 0° 16° 32° 48° 64°

Figure 3.3: Comparison of model performance with peaked and monotonic gain modulation of visual input neurons. The figure shows the firing rate responses of output neurons before and after training with the standard trace rule (equation 3.5). Specifically, each subplot shows the firing rate responses of a typical output neuron during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. The top row shows output neuron #95 from the model with peaked eye-position gain modulation before training (subplot 3.3**A**) and after training (subplot 3.3**B**). The bottom row shows output neuron #838 from the model with sigmoidal eye-position modulation before training (subplot 3.3**C**) and after training (subplot 3.3**D**). In each subplot, each curve corresponds to a fixed eye-position while a visual target is presented across the same range of head-centred locations. It is evident in subplot 3.3**B** that after training the model with peaked gain modulation of the input neurons, the output neuron responds reasonably consistently when the visual target is presented within the localised interval of head-centred space $[0°, 16°]$ regardless of the eye-position. The neuron is thus responding in a head-centred reference frame. However, in contrast, subplot 3.3**D** shows that after training the model with sigmoidal (monotonic) gain modulation, the responses of the output neuron in the head-centred visual space are much more dependent on the eye-position. Thus, this neuron is not representing the target position in a head-centred reference frame. The miniature scatter plots show the reference frame values of all neurons in the output layer, where each neuron is plotted as a point corresponding to that neuron's particular combination of head-centredness (ordinate) and eye-centredness (abscissa). The neuron whose firing rate responses have been plotted is shown in the scatter plot by a red mark. The miniature scatter plots confirm the same general effects across the entire populations of output neurons. That is, subplot 3.3**B** shows that a large proportion of the output neurons are clustered in the top left quadrant of the scatter plot, indicating a high head-centredness (ordinate) and low eye-centredness (abscissa). These output neurons are thus responding in a head-centred frame of reference. While subplot subplot 3.3**D** shows that a large proportion of the output neurons are clustered in the bottom right quadrant of the scatter plot, indicating a low head-centredness (ordinate) and high eye-centredness (abscissa). Thus, with monotonic gain fields acting on the input neurons, the population of output neurons have overwhelmingly learned to respond in an eye-centred reference frame.

the output neurons in terms of the preferences of the presynaptic input neurons (figure 3.4**A** and figure 3.4**C**), reflecting the random weighting assigned to an untrained network. After training, the synaptic weight vector of the output neuron from the model with peaked eye-position gain modulation shows a clear diagonal structure (figure 3.4**B**). This synaptic weight profile is consistent with a learned response to a particular location within the

Experiments with either peaked or monotonic input modulation				
	Peaked		**Monotonic**	
	Untrained	**Trained**	**Untrained**	**Trained**
Head-centredness	0.17 (0.12)	0.63 (0.15)	0.04 (0.16)	0.09 (0.11)
Eye-centredness	0.03 (0.10)	0.18 (0.13)	0.88 (0.10)	0.96 (0.02)
Proportion head-centred	26%	69%	0%	0%
Proportion eye-centred	73%	30%	100%	54%
Proportion unclassified	0%	1%	0%	46%

Table 3.2: Population summary statistics of response properties of output neurons in the models with either peaked or monotonic (sigmoidal) modulation of input neurons training with the standard trace rule (equation 3.5). Statistics are given before training and after training. The first two rows of the table show the mean and standard deviation (in parentheses) of the head-centredness and eye-centredness of the output population. In the case of monotonic modulation the value of head-centredness could not be determined due to the lack of enough head-centred neurons. The remaining rows of the table shows the proportion of output neurons that were classified as head-centred and the proportion of output neurons that could not be classified as either head-centred or eye-centred.

head-centred frame of reference, and is thus consistent with the observed head-centred responses of this neuron during testing (figure 3.3**B**). The synaptic weight vector of the neuron from the model with sigmoidal (monotonic) gain modulation exhibites an entirely different pattern of potentiation after training. In this case, the strenghtened synapses have an approximately horizontal structure that is concentrated on input neurons corresponding to a small portion of retinal preference space (figure 3.4**D**). As a result, this output neuron has learned an eye-centred response (figure 3.3**D**).

Figure 3.5 shows the reference frame values for all output neurons from both models with either peaked or sigmoidal gain modulation tested before and after training. It is clear that, for the model with peaked eye-position modulation, training has the effect of making the majority of output neurons head-centred, and also with a much larger head-

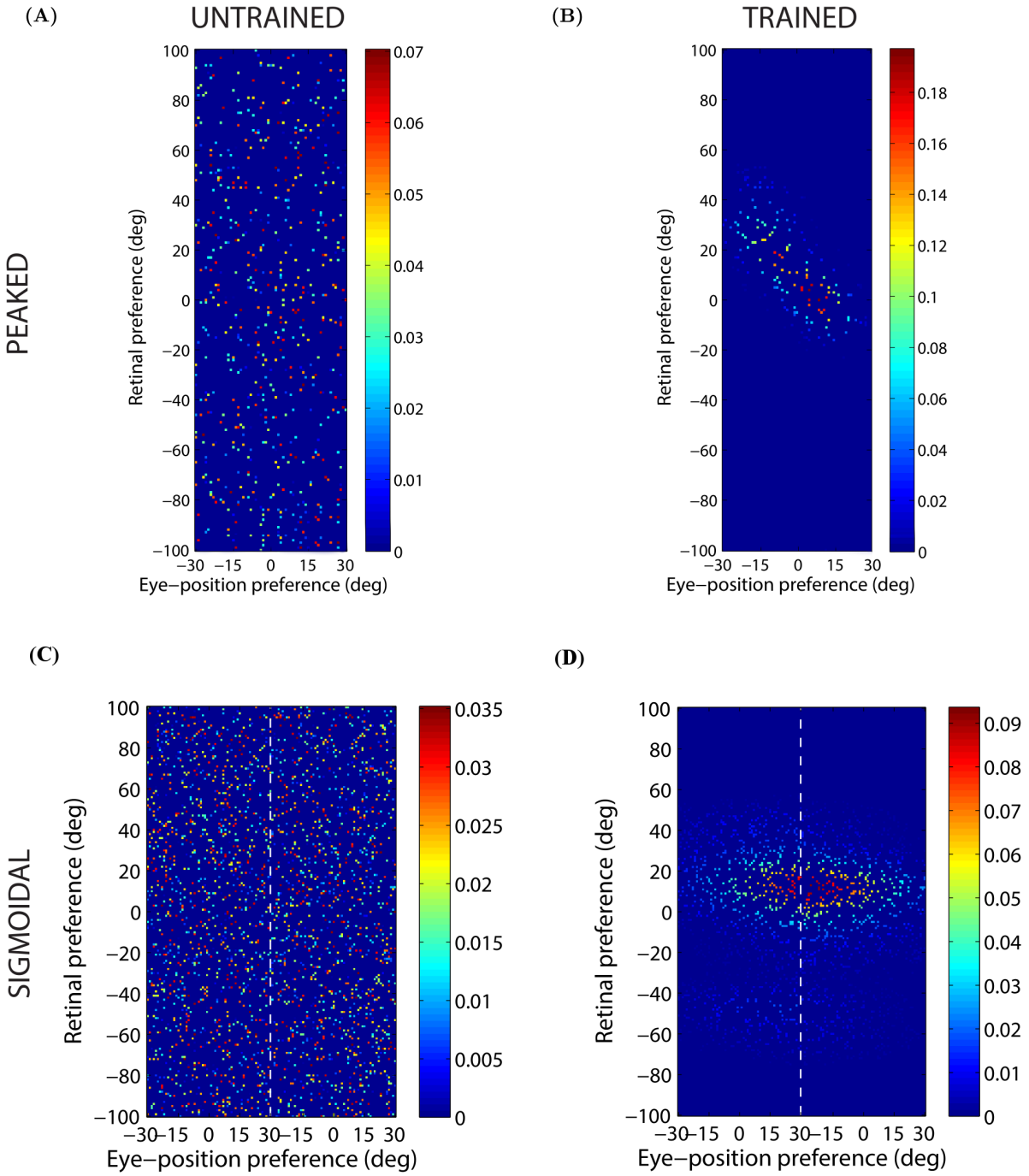

(A)
UNTRAINED
(B)
TRAINED
PEAKED
SIGMOIDAL
(C)
(D)
Retinal preference (deg)
Eye–position preference (deg)
100
80
60
40
20
0
−20
−40
−60
−80
−100
−30 −15 0 15 30
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01

Figure 3.4: Strengths of the afferent synapses from the input population to a typical output neuron during testing. Results are shown for the model with peaked eye-position gain modulation before training (subplot 3.4**A**)and after training (subplot 3.4**B**), and for the model with sigmoidal eye-position modulation before training (subplot 3.4**C**) and after training (subplot 3.4**D**). The output neurons correspond to those plotted in figure 3.4. In each plot, the afferent synapses have been arranged topographically by the preference of the input neuron for retinal location α_i and eye-position inflection point β_j. For the model with sigmoidal gain modulation, there are two input neurons for every combination of retinal preference and eye-position preference, but with opposite eye-position gain. Consequently, the input population has been separated by gain direction. The portion of each plot to the left of the white dashed line corresponds to input neurons with positive gain $\kappa_j > 0$, while the portion of each plot to the right of the white dashed line corresponds to those input neurons with negative gain $\kappa_j < 0$. It can be seen from subplot 3.4**B** that the output neuron in the trained network with peaked gain modulation has developed a diagonal weight structure, which is consistent with a learned response to a particular location within the head-centred frame of reference. In contrast, subplot 3.4**D** shows that the output neuron in the trained network with sigmoidal (monotonic) gain modulation has developed a more horizontal weight structure, which is consistent with a learned response to a specific location within the eye-centred reference frame.

centredness value. Specifically, before training 26% of output neurons are head-centred, and after training 69% are head-centred (table 3.2). Moreover, among the head-centred neurons, the average head-centredness rose from 0.17 before training to 0.63 after training. For the model with sigmoidal eye-position modulation, training has the effect of keeping the majority of output neurons eye-centred, and indeed increasing their average eye-centredness from 0.88 to 0.96 (table 3.2).

In summary, when the visual input neurons have peaked eye positon gain modulation, training the network has the effect of developing head-centred output neurons. However, when the input neurons have sigmoidal (monotonic) gain modulation, the training process makes most output neurons almost perfectly eye-centred.

Since a large proportion of visual neurons in the dorsal visual pathway have responses that are modulated by a monotonic function of eye-position, it is important to understand

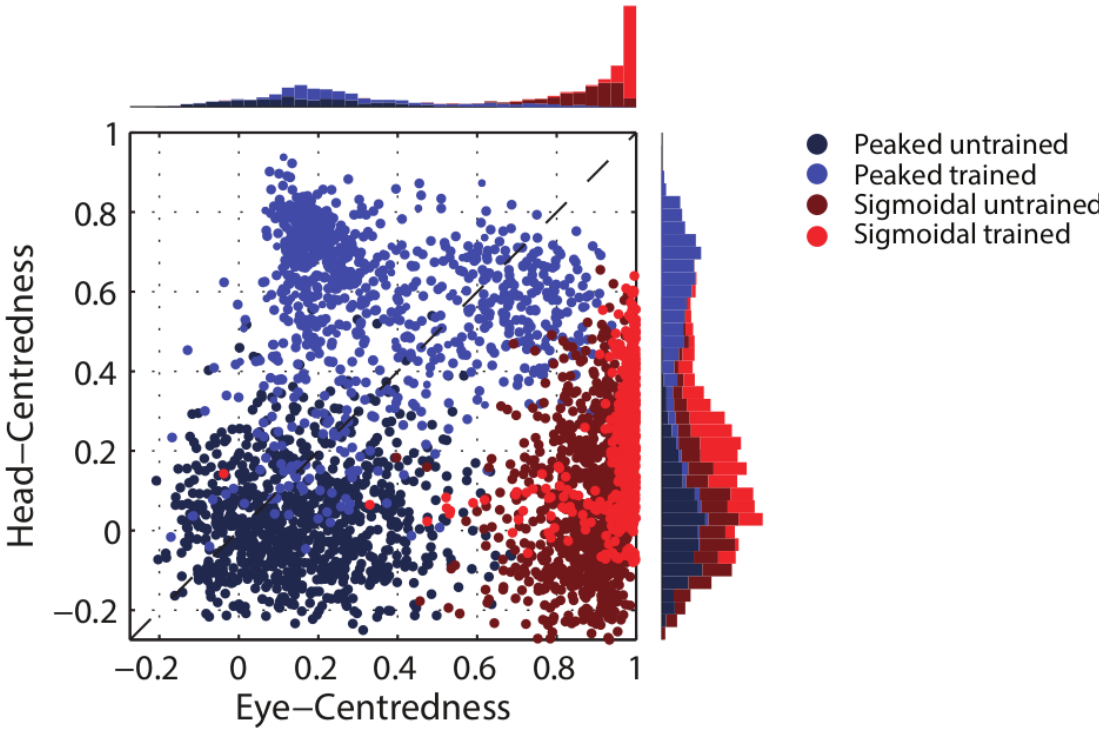


Figure 3.5: Scatter plot of eye-centredness and head-centredness values of output neurons from simulations with peaked and monotonic gain modulation. The scatter plot shows the eye-centredness and head-centredness values of all output neurons from four separate simulations corresponding to the models with peaked and sigmoidal (monotonic) gain modulation tested before and after training with the standard trace rule (equation 3.5). Each point in the scatter plot corresponds to an output neuron from the given simulation, plotted in terms of its eye-centredness (abscissa) and head-centredness (ordinate). The dashed diagonal line with positive unity slope separates those neurons which are classified as head-centred (above line) from those that are classified as eye-centred (below the line). Neurons that could not have been classified as eye-centred or head-centred are not plotted in the scatter plot. The marginal plots are histograms showing the frequency distribution of the numbers of output neurons with different values of eye-centredness (abscissa) and head-centredness (ordinate). It is evident that after training most of the output neurons from the network with peaked gain modulation have become head-centred, while nearly all of the output neurons from the network with monotonic gain modulation have remained eye-centred.

why monotonic gain fields make it more difficult for trace learning to produce head-centred output neurons. In the next section, we investigate this problem by carrying out a covariance analysis on the input patterns themselves.

3.10.2 Covariance Analysis of the Effects of Gain Modulation

The preceeding model simulations failed to develop head-centred output representations during self-organisation when the input population had sigmoidal eye-position modulation, despite succeeding with peaked eye-position modulation. This raises the question of what the difference is between the two different forms of input encoding from the perspective of competitive learning (Rolls and Treves, 1998). 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 may reveal what structure the weight vectors should converge towards under standard competitive learning conditions with a Hebbian learning rule (equation 2.9).

The covariance between input neurons may be computed in the same way for both input neurons with peaked gain described by equation 3.1 and input neurons with sigmoidal gain described by equation 3.2. 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

$$\mathrm{cov}_{\alpha,\beta}(\alpha^*, \beta^*) = \iint_{R \times E} \left(R_{\alpha,\beta}(r,e) - \overline{R}_{\alpha,\beta}\right)\left(R_{\alpha^*,\beta^*}(r,e) - \overline{R}_{\alpha^*,\beta^*}\right) dr de \tag{3.10}$$

$R_{\alpha,\beta}(r,e)$ 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 either equation 3.1 or equation 3.2. The term $\overline{R}_{\alpha,\beta}$ is the average response of the same neuron across all possible inputs in the $R \times E$

space, that is

$$\overline{R}_{\alpha,\beta} = \frac{1}{|R \times E|} \iint_{R \times E} R_{\alpha,\beta}(r, e) dr de \tag{3.11}$$

Figure 3.6**A** shows this covariance map for an input neuron with a peaked gain and preferences $\alpha = \beta = 0°$, and figure 3.6**B** shows the covariance map for an input neuron with sigmoidal gain and preferences $\alpha = \beta = 0°$ and $\kappa > 0$.

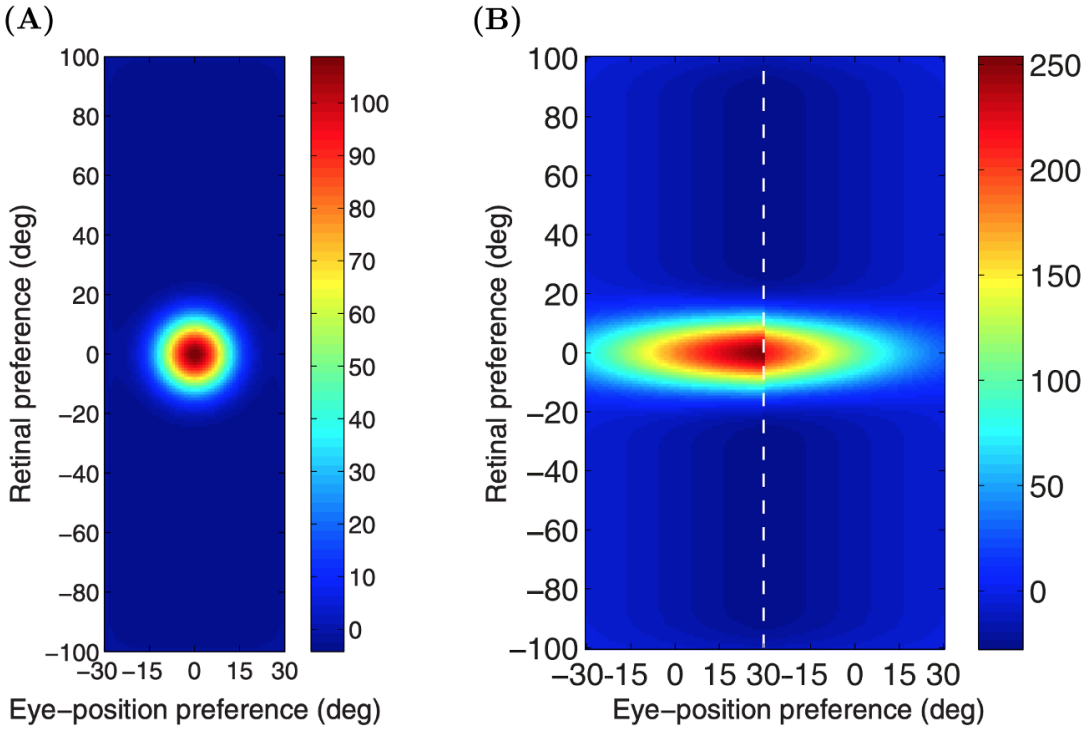


Figure 3.6: Covariance between a given input neuron and the rest of the input neuron population. Each plot shows the covariance between a given input neuron and the rest of the input neuron population in the form of a topographic map analogous to the weight vector maps shown above. Subplot 3.6**A** and subplot 3.6**B** show results for input neurons with peaked and monotonic gain, respectively. In both cases the input neuron has preferences $\alpha = \beta = 0°$, and in the monotonic case the input has positive gain ($\kappa > 0$).

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.

In the sigmoidal (monotonic) modulation case, the situation is more complicated. The strong covariance 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 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 eye-position to the right of the inflection point (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 inflection point 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 inflection point 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 eye-position to the right of the inflection point (i.e. smaller than) $0°$ cofire more frequently with the reference neuron.

The covariance maps shown in figure 3.6**A** and figure 3.6**B** 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 or

sigmoidal gain, respectively. These predictions are tested by running simulations with a Hebbian learning rule with both peaked and sigmoidal gain modulated input neurons. The Hebbian learning rule is implemented in the model by replacing the trace value q_i in the standard trace learning rule (equation 3.5) by the current firing rate v_i of the postsynaptic neuron. For each simulation, there are 200 training patterns corresponding to random locations in the $E \times R$ space. The activation time constant is reduced to $\tau_h = 30ms$ to avoid any trace effect during learning. Figure 3.7**A** and figure 3.7**C** show the synaptic weight vectors of two typical output neurons that developed after training with the Hebbian learning rule when the input neurons are modulated by either peaked or monotonic gain, respectively. It is clear that these synaptic weight vectors have a very similar structure to the corresponding covariance maps shown in figure 3.6. Thus, with the Hebbian learning rule, the underlying correlations between the activities of the input neurons with either peaked or sigmoidal gain shape the synaptic weight structure that develops during training. Most importantly, with monotonic gain, the synaptic weights are localised within the retinal preference dimension, but elongated within the inflection point dimension of eye-positions. This kind of synaptic weight structure leads to eye-centred output responses.

Next, comparison simulations are run with the trace learning rule. Figure 3.7**B** and Figure 3.7**D** show the synaptic weight vectors of two typical output neurons that developed after training with the trace learning rule when the input neurons are modulated by either peaked or sigmoidal gain, respectively. Figure 3.7**B** shows a diagonal band of potentiated synaptic weights, which correspond to input neurons representing the same head-centred location but with different combinations of retinal location and eye-position. Thus, with peaked gain, the trace learning rule is able to simply bind together clusters of input neurons

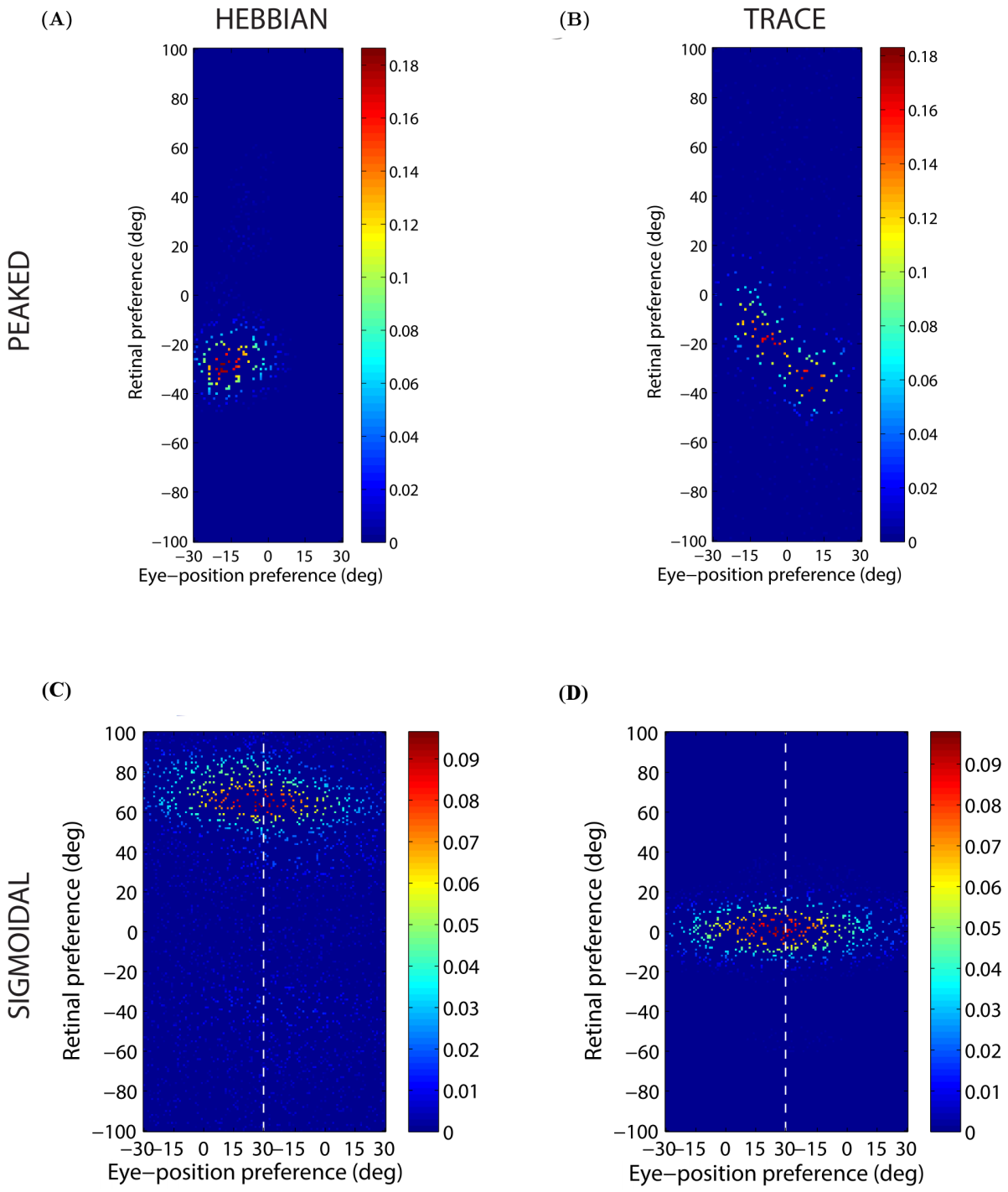

(A)
HEBBIAN
(B)
TRACE
PEAKED
SIGMOIDAL
(C)
(D)
Retinal preference (deg)
Eye–position preference (deg)
100
80
60
40
20
0
−20
−40
−60
−80
−100
−30 −15 0 15 30
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01

Figure 3.7: Weight vectors of two typical output neurons. 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 (subplot 3.7**A**) or the trace learning rule (subplot 3.7**B**). The bottom row shows the weight vectors of two typical output neurons when the input neurons have monotonic eye-position gain and the network is trained with either the Hebbian learning rule (subplot 3.7**C**) or the standard trace learning rule (subplot 3.7**D**).

along a diagonal line in the (retinotopic preference × eye-position preference) input space corresponding to a particular head-centred location. Output neurons will then respond to particular head-centred 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 3.7**D** shows a similar horizontal weight structure to that obtained with the Hebbian learning rule (figure 3.7**C**). In particular, in both of these last two cases, the weight vector is very similar to the covariance structure found among the input neurons with sigmoidal eye-position modulation (figure 3.6**B**). Thus, with sigmoidal gain, even if a trace learning rule is implemented, the output neurons still learn to represent eye-centred rather than head-centred locations. This is because developing head-centred output responses would require the trace learning rule to do more than simply bind input patterns together. With sigmoidal gain, 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 dimension of the eye-position inflection point. However, in practice the standard trace learning rule given by equation 3.5 is not strong enough to achieve this. Consequently, with the standard trace learning rule, these elongated clusters of input neurons with correlated activities continue to drive the development of eye-centred output neurons, as observed in the simulations

reported earlier in this chapter.

3.10.3 Introducing Plasticity into the Prewired Model

The failure of self-organisation to produce head-centred output neurons in a model with sigmoidal (monotonic) eye-position gain modulation suggests the following two important questions. First, does there actually exist a synaptic weight connectivity matrix that would support a mapping to head-centred output representations, even if in practice the self-organisation process using the standard trace learning rule (equation 3.5) originally implemented by Mender and Stringer (2013) fails to converge on this solution? Secondly, if it is possible to prewire a network with such a synaptic weight structure, then would the head-centred output representations be abolished by subsequently introducing synaptic plasticity either in the form of the standard trace learning rule (equation 3.5) or normal Hebbian learning (where the trace value q_i is replaced by the current firing rate v_i)? If so, then this would demonstrate an even deeper problem: the presence of such forms of synaptic plasticity in a network with sigmoidal gain modulation not only fails to drive the development of head-centred output representations, but would also abolish any existing head-centred representations.

To address the above two questions, we construct a manually prewired model that is designed to produce head-centred output neurons with input neurons that are modulated by sigmoidal functions of eye-position. The prewired model is constructed as follows. There are 24522 neurons in the input population, each corresponding to a unique combination of retinal-position preference (α_i), eye-position inflection point (β_j) and slope (κ_j). There are 900 neurons in the output population, each given a head-centred receptive field at one among nine head-centred locations, which are $-68°$, $-51°$, $-34°$, $-17°$, $0°$, $17°$, $34°$, $51°$

and 68°. Each neuron in the output population is postsynaptically connected to a randomly assigned subpopulation of the input population. There are only two synaptic weight values across all synapses, simply referred to as elevated and depressed. The strength of a synapse is elevated if the presynaptic input neuron responded maximally to a combination of eye-position and retinal location corresponding to a head-centred location that is closest to the head-centred location assigned to the output neuron. Otherwise the synapse is depressed. Therefore, the output neuron receives strong driving input to the extent that a visual target is near its assigned head-centred location. Specifically, the weight of a synapse with a postsynaptic neuron assigned to head-centred receptive field location h and a presynaptic neuron having retinal-preference α, eye-position inflection point β and $\kappa > 0$ is given by

$$w_{h,\alpha,\beta} = \begin{cases} 10 & h - \beta \leq \alpha \leq h + r/2 \quad \text{(Elevated)} \\ 1 & \text{else} \qquad\qquad\qquad\quad \text{(Depressed)} \end{cases} \tag{3.12}$$

where $r = 60°$ is the size of the eye-position dimension. Likewise when $\kappa < 0$ the weight is given by

$$w_{h,\alpha,\beta} = \begin{cases} 10 & h - r/2 \leq \alpha \leq h - \beta \quad \text{(Elevated)} \\ 1 & \text{else} \qquad\qquad\qquad\quad \text{(Depressed)} \end{cases} \tag{3.13}$$

Figure 3.8 shows the structure of the canonical weight vector produced by equations 3.12-3.13. Before testing the network, the synaptic weight vectors of all output neurons underwent the normalization step described by equation 2.10. To provide a baseline for comparison, a network with randomly wired synaptic connections is also tested in the same way. The parameters for both models are given in Table 3.3.

Figure 3.9 shows the eye-centredness and head-centredness values of output neurons in the manually prewired model as well as a randomly wired model for comparison. It is clear

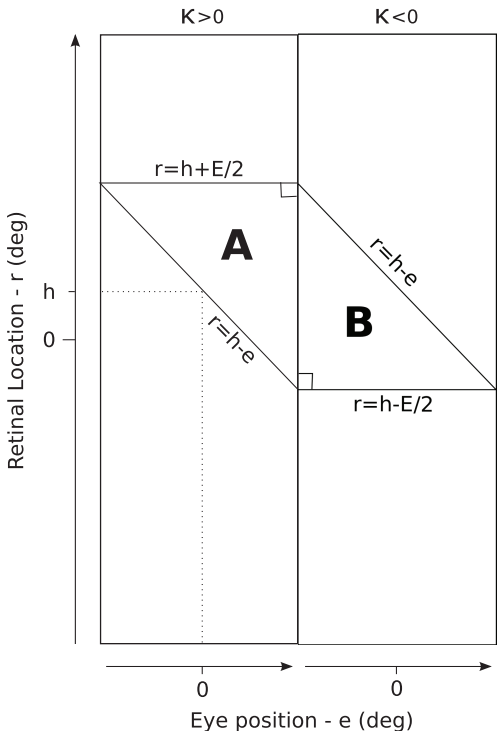


Figure 3.8: Synaptic weight structure of a network model that has been manually prewired in order to produce head-centred output neurons with input neurons that are modulated by a sigmoidal function of eye-position. The figure shows the structure of the canonical weight vector resulting from the prewiring equation 3.12 and equation 3.13. Each of the two rectangles represents the topographic organisation of one half of the input population in terms of retinal-preference (α_i) and eye-position inflection point (β_j), 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-centred receptive field at location h will have elevated connections from input neurons with preferences located in the right-angled triangles of the input space, labeled **A** and **B**.

that in the manually prewired model, the majority of output neurons are head-centred. In contrast, there are no head-centred neurons in the randomly wired model. In summary, the results from the manually prewired model demonstrate the existence of a synaptic weight matrix which allows the output neurons to perform the desired coordinate transformation to a head-centred reference frame even when the input neurons are modulated by a sigmoidal function of eye-position.

Parameter	Symbol	Value
Width of retinal tuning curve	σ	6°
Gain magnitude	$\lvert\kappa\rvert$	0.0625
Output neuron population size	N	900
Input neuron population size	N_I	24522
Activation time constant	τ_h	100ms
Activation function slope	φ	4
Activation function threshold	θ	0
Sparseness percentile	π	90%
Synaptic connectivity	ϕ	8.16%

Table 3.3: Simulation parameters of the network model that has been manually prewired in order to produce head-centred output neurons with monotonic modulated input neurons.

The next question to be explored is, what would be the effect of introducing synaptic plasticity, either in the form of the standard trace learning rule (equation 3.5) or normal Hebbian learning, into the manually prewired model while it is exposed to the kinds of visual input described above? This experiment would inform whether these forms of plasticity not only failed to drive the development of head-centred output representations during self-organisation, but would even abolish existing head-centred representations. The manually prewired model is therefore subjected to the same visual training stimuli as described above over 10 training epochs. The results presented here are from a simulation using the standard trace learning rule (equation 3.5). However, although not shown, further simulations with a Hebbian learning rule without an explicit memory trace gave qualitatively similar results.

The impact of introducing synaptic plasticity into the manually prewired models is inspected by plotting key summary statistics as a function of the number of training epochs in figure 3.10. There is a catastrophic drop in model performance after only the first epoch

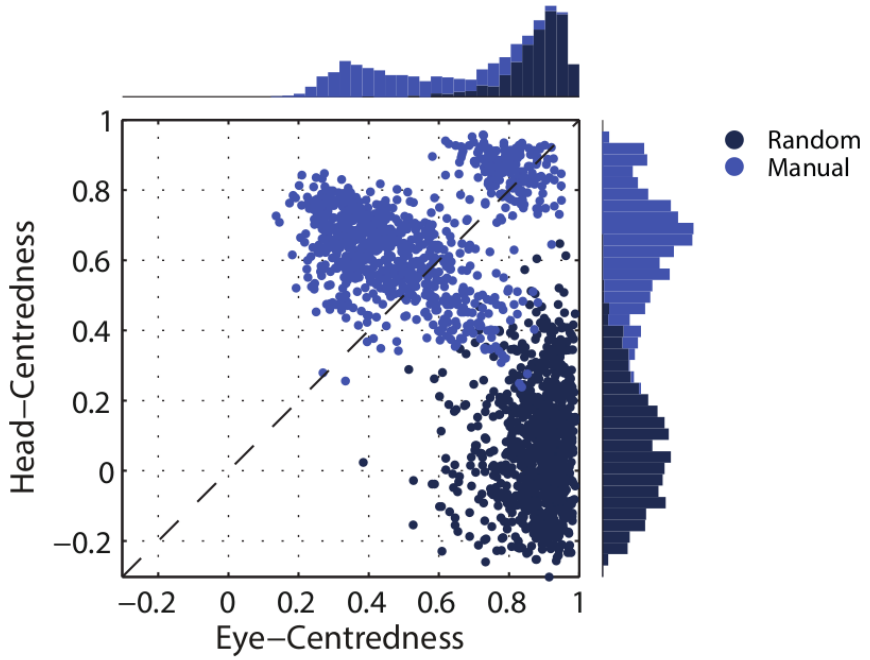


Figure 3.9: Performance of the prewired network model with monotonic modulated input neurons. The Figure shows the performance of the network model that has been manually prewired to produce head-centred output neurons with input neurons that are modulated by a sigmoidal function of eye-position. The scatter plot shows the eye-centredness and head-centredness values of all output neurons from the manually prewired model and a randomly wired model. Same conventions as in figure 3.5. It can be seen that the majority of the output neurons in the manually prewired model display head-centred responses.

of training, where the fraction of head-centred neurons decreased from $\sim 77\%$ to $\sim 0.3\%$, and the average head-centredness among head-centred neurons decreased from ~ 0.7 to ~ 0.26. Subsequent training epochs remained around these levels. After epoch 5 there is not a single head-centred neuron.

In summary, it is found that just a single training epoch switched most of the output neurons from being head-centred to eye-centred. We hypothesised that this is due to the same visually-guided learning dynamics described above in section 3.10.1 and in section 3.10.2, which come into operation when the retinotopic input neurons have monotonic

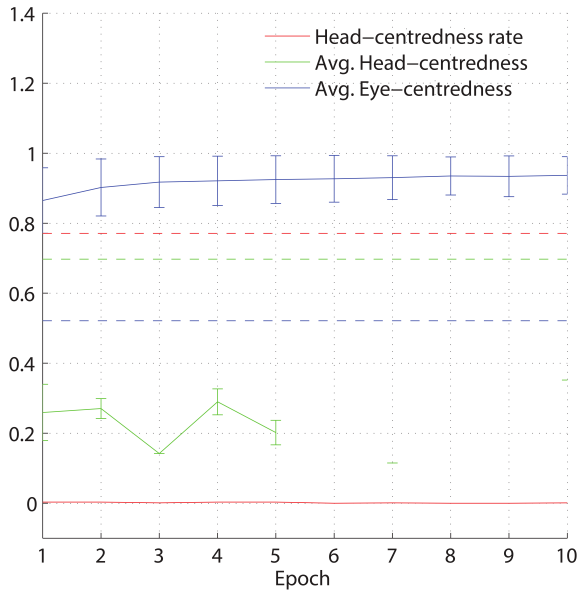


Figure 3.10: The effects of introducing synaptic plasticity into the network that has been manually prewired to produce head-centred output neurons when the input neurons that are modulated by a monotonic function of eye-position. The figure shows population analyses of the response properties of output neurons in the manually prewired model as the synaptic weights are further modified during ten training epochs with the standard trace learning rule (equation 3.5). Three key summary statistics are given. The head-centredness rate (red) is the fraction of head-centred neurons in the output population. The average head-centredness (green) is the average head-centredness among head-centred neurons, and becomes undefined if no head-centred neurons are found to exist. The average eye-centredness (blue) is the average eye-centredness among all output neurons. The dashed lines show these values for the manually prewired network before training, while the unbroken lines show the values through successive training epochs after synaptic plasticity has been introduced. The error bars are the standard deviations. It can be seen that by the end of the first training epoch the majority of the output neurons switched from being head-centred to eye-centred.

eye-position gain modulation. Thus, even if the synaptic weights are initially manually prewired to effect head-centred output responses, which might be suggested to happen in the brain through genetic specification, the introduction of just a limited amount of synaptic plasticity, either in the form of the standard trace learning rule (equation 3.5) or normal Hebbian learning, and visually-guided learning led to the output neurons rapidly switching

to eye-centred responses. The presence of even modest levels of such synaptic plasticity will quickly overwrite head-centred representations that have been set up through structured (e.g. genetic) prewiring. Thus, since plasticity is ubiquitous in primate cortex, this suggests that any explanation for the development of head-centred visual responses must utilise a more sophisticated visually-guided learning process than demonstrated by Mender and Stringer (2013), who considered only models with input neurons that were modulated by peaked functions of eye-position. Moreover, the loss of head-centred representations in the manually prewired model by introducing synaptic plasticity also represents a major challenge to the plausibility of previously published models, such as that of Salinas and Abbott (1995), which rely on an initial period of supervised learning to establish the required synaptic connectivity. The problem here is that when the supervisory training signal is eventually removed, the continued presence of associative plasticity may degrade and eventually abolish the head-centred output representations.

In the remainder of the paper we explore a variety of biologically plausible model variations that are aimed at discovering potential mechanisms by which head-centred output neurons may still develop through visually guided learning even when the network contains input neurons with monotonic modulation by eye-position. We begin by exploring the performance of the model when it incorporates a mixture of input neurons that are modulated by peaked and sigmoidal (monotonic) eye-position gain fields. After this, we explore the operation of the model with a number of more sophisticated, modified synaptic learning rules originally developed by Rolls and Stringer (2001) in the context of transform invariant visual object recognition, which maintain biological plausibility by continuing to rely on the locally available activities of the pre- and post-synaptic neurons. Moreover,

trace learning rules operate in an unsupervised manner with no requirement for an explicit training signal as needed for error correction rules such as backpropagation of error, which further adds to their biological plausibility. The choice of the modified versions of the trace learning rule used in the following sections of this paper is motivated by the superior performance of these learning rules reported by Rolls and Stringer (2001).

Physiological studies providing evidence for a form of memory trace in different brain areas, including the cortex, were reviewed by Rogerson et al. (2014). These authors discussed how changes in the neuronal excitability after a neuron has fired may be triggered by the transcription factor cyclic AMP-responsive element-binding protein (CREB). It was observed that higher levels of CREB modulate the probability that a given neuron will be involved in processing current sensory input by making the neuron more excitable. Such a mechanism may introduce a trace learning effect as follows. A first sensory input activates CREB in a subpopulation of neurons, which leads to an increase in the excitability of these neurons. This increase in excitability then biases the neuronal activation of many of the same neurons for the subsequent sensory inputs. The overall effect in the visual system, for example, could be to help to maintain the activity of neurons in higher neuronal layers while the retinal image transforms. This in turn could underpin a form of trace learning, which encourages individual neurons to learn to respond to sensory inputs that tend to occur close together in time. Additional experimental evidence of trace learning in the cortex was reported by Li and DiCarlo (2008). These authors investigated increases in transformation invariance and reductions in stimulus specificity among neurons in the primate visual system. Specifically, the study provided evidence that the responses of inferior temporal cortex neurons learned to respond to retinal images that occured close together

time. Computational evidence of how trace learning may naturally occur in a biologically realistic 'spiking' neural network model of the primate visual system was reported by Evans and Stringer (2012). These authors investigated the self-organisation of transform invariant visual representations using Spike Time Dependent Plasticity (STDP). In particular, they demonstrated that trace learning arose naturally in the network model during training if the synaptic time constants were sufficiently large. Lastly, Perry et al. (2006) investigated the contribution of trace learning in the learning of view invariant visual representations of objects in both computer simulations and psychophysical experiments using the same set of training stimuli. Both computational and psychophysical results provided evidence of trace learning mechanisms.

We finally conclude this chapter with the investigation of the performance of the model incorporating a mixed population of peaked and monotonic modulated visual input neurons with the synaptic weights also adjusted using a new modified version of the trace learning rule that we proposed.

3.10.4 Standard Trace Learning Rule with Mixed Peaked and Sigmoidal eye-position Modulation of Input Neurons

In this experiment it is investigated how mixing peaked and sigmoidal gain modulation in the input population in varying proportions would influence the development of head-centred output neurons in the self-organising model with the standard trace learning rule (equation 3.5). This is an important issue since all cortical areas with eye-position gain modulation exhibit a mixture of different forms of modulation (Andersen et al., 1990, 1985, Galletti et al., 1995). A series of simulations are conducted where each neuron in the input population is independently and randomly set to have either a peaked or

sigmoidal gain modulation. Specifically, each input neuron is changed from having peaked to sigmoidal modulation with a probability p, called the sigmoid modulation rate, and values of $p = 0, 0.1, \ldots, 1.0$ are explored.

The impact of varying the sigmoid modulation rate on the characteristics of the model is inspected by plotting key summary statistics as a function of p in figure 3.11. As expected, the head-centredness rate decreased as the sigmoid modulation rate increased, both in the trained and untrained models. However, as long as the sigmoid modulation rate is less than 30%, the trained model had a higher proportion of head-centred output neurons than the untrained model. In particular, for sigmoid modulation rates less than 20%, the fraction of head-centred neurons in the trained model did not drop below $\sim 15\%$, and the average head-centredness among head-centred neurons remained no less than ~ 0.58.

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

Given that the model incorporating a mixed population of input neurons failed to develop a significant proportion of head-centred output cells whenever $p > 0.2$, we next investigated whether the introduction of more powerful, modified synaptic learning rules could produce head-centred output neurons when the entire population of input neurons were again modulated by a sigmoidal function of eye-position. Our choice of the modified versions of the trace learning rule adopted in this paper was motivated by their superior

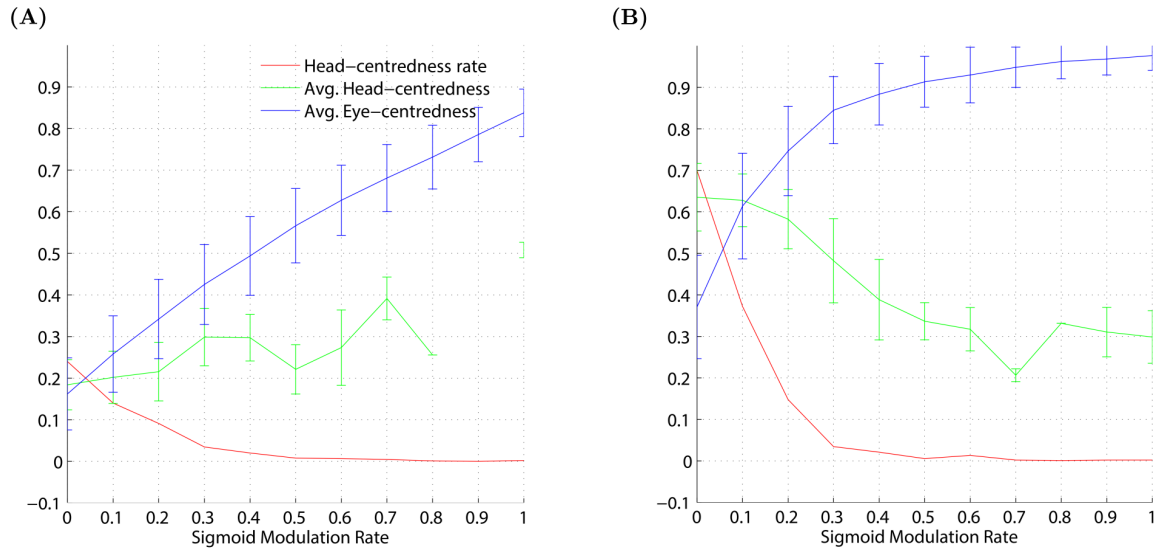


Figure 3.11: The effects of incorporating a mixed population of input neurons with both peaked and monotonic eye-position gain modulation. The plots show how the performance metrics vary with the monotonic modulation rate, p, which is the probability of each input neuron having a monotonic eye-position gain modulation. Results are presented showing the response characteristics of the output neurons before training (subplot 3.11**A**) and after training (subplot 3.11**B**). Conventions are similar to figure 3.10. It is evident that the head-centredness rate decreased as the sigmoid modulation rate increased, both in the trained and untrained models. However, as long as the sigmoid modulation rate is less than 30%, the trained model had a higher proportion of head-centred output neurons than the untrained model.

performance reported in previous work (Rolls and Stringer, 2001). These modified versions of the standard trace learning rule are still biologically plausible in terms of only using locally available signals to update the synaptic weights of connections. Moreover, we expected that the introduction of a delayed trace of synaptic activity and as well as an anti-Hebbian component incorporated in these modified versions of the trace learning rule would provide the model with a way of weakening the observed Hebbian-like training behaviour evidenced, for example, by the comparison of figure 3.11**D** and figure 3.11**C**, and therefore to facilitate the self-organisation of head-centred responses with visual input neurons with monotonic eye-position gain modulation. The absence of these components

makes other classic Hebbian-based learning rules (e.g. Oja's rule (Oja, 1982)) ineffective in this case.

3.10.5 Modified Learning Rule: Delayed Postsynaptic Trace with Anti-Hebbian Learning

Rolls and Stringer (2001) investigated how a set of modified more powerful versions of the trace learning rule can produce improved temporal binding and invariance learning. In particular, the authors showed that the performance of the trace learning rule is substantially improved by incorporating a trace of previous neuronal activity with an explicit time delay. This had the effect of removing the purely Hebbian term of the learning rule (Rolls and Stringer, 2001). In the next simulations, the learning rules proposed by Rolls and Stringer (2001) were adapted to differential formulations for the time-continuous scenario in which the simulations are performed. A Forward Euler scheme was used to numerically integrate the differential equations. In all simulations the numerical time step was kept as one tenth of the neuronal time constant τ_h. Unless explicitly mentioned, the learning rule was the only change from previous simulations.

This section presents simulation results showing the performance of a modified learning rule incorporating a delayed postsynaptic trace with an anti-Hebbian learning component in which the strength of each synaptic connection is decreased rather than increased in proportion to the product of the firing rates of the pre- and post-synaptic neurons (Kohonen, 2012). We investigate the impact of this learning rule on the self-organisation of the synaptic weights and firing rate responses of the output neurons when all input neurons were modulated by sigmoidal eye-position gain fields. The learning rule was defined by

$$\frac{dw_{ij}(t)}{dt} = \varrho\left(\beta q_i(t-\Delta T) - y_i(t)\right) v_j(t) \tag{3.14}$$

where $\varrho\beta q_i(t-\Delta T)v_j(t)$ is the *delayed-trace term* of the learning rule. This is the term which contains the tuning parameter β. Such a temporal delay ΔT in the plasticity at a synapse could arise due to a number of physiological factors. For example, a long axonal transmission delay ΔT between a pair of pre- and post-synaptic neurons would mean that plasticity at the synapse depends on the activity of the pre-synaptic cell when it reaches the synapse ΔT after the firing of the neuron. Axonal delays in the rabbit brain have been reported in the range $[0.1, 44]$ms (Swadlow, 1985, 1988, 1992) and in the range $[0.7, 5.2]$ms and $[0.7, 8.0]$ for neurons from the Frontal Eye Field (FEF) and from area LIP, respectively, of the primate brain (Ferraina et al., 2002). Such a delay mechanism has recently been implemented in neural network simulations to solve accurate path integration in the head direction system (Page et al., 2015, Walters et al., 2013) and provide an approach to solving feature binding in the visual brain (Eguchi et al., 2018, Isbister et al., 2018). The remaining term $-\varrho y_i(t)v_j(t)$ is minus the learning rate ϱ times the product of the post and presynaptic firing rates y_i and v_j, respectively. This term is referred to as the *anti-Hebbian term* of the learning rule. Various forms of anti-Hebbian learning have been reported in experimental neurophysiology studies in brain areas such as the hippocampus (Lamsa et al., 2007), primary motor cortex (Koch et al., 2013) and prefrontal cortex (Ruan et al., 2014). The value of ϱ used in the simulations below is given in table 3.1.

The behaviour of the learning rule shown in equation 3.14 is governed by scaling the parameter β. Scaling β up could result in the delayed-trace component dominating the learning rule and, therefore, resulting in the same trace-like training behaviour described in

previous sections. Similarly, scaling β down could make the anti-Hebbian term dominate the behaviour of the learning rule and consequently lead to a qualitative change in the final outcome of the training.

The parameters for the model are given in table 3.1. The model was trained according to the description given in section 3.6. Likewise, in all cases the model was tested on the same visual stimuli. The time delay ΔT used to compute the trace value of each output neuron and the parameter β were both tuned to optimise the performance of the model at developing head-centred output neurons. The time delay ΔT used to compute the trace value of each output neuron i was $50ms$. The parameter β was set to 2.2.

Figure 3.12 shows the firing rate responses of output neuron #168 before and after training, with results shown for four different eye-positions. The miniature scatter plots presented within each of the two subplots 3.12**A** and 3.12**B** show the reference frame values of output neuron #168 before and after training, respectively. Neuron #168 is indicated in the scatter plots by a red mark. Figure 3.12**A** shows that prior to training the response of output neuron #168 had no consistent structure in head-centred space across different eye-positions. However, Figure 3.12**B** shows that after training the output neuron responded maximally to the same head-centred location across all four eye-positions. This neuron has thus learned to respond in a head-centred reference frame.

The change in the synaptic weight structure of the same output neuron #168 due to training is shown in figure 3.13. Before training started the afferent synaptic weights were randomly assigned (figure 3.13**A**). Figure 3.13**B** shows that after training, in contrast to the horizontal structure previously obtained with the standard trace learning rule shown in figure 3.7**D**, the structure obtained with the modified learning rule 3.14 is similar to the

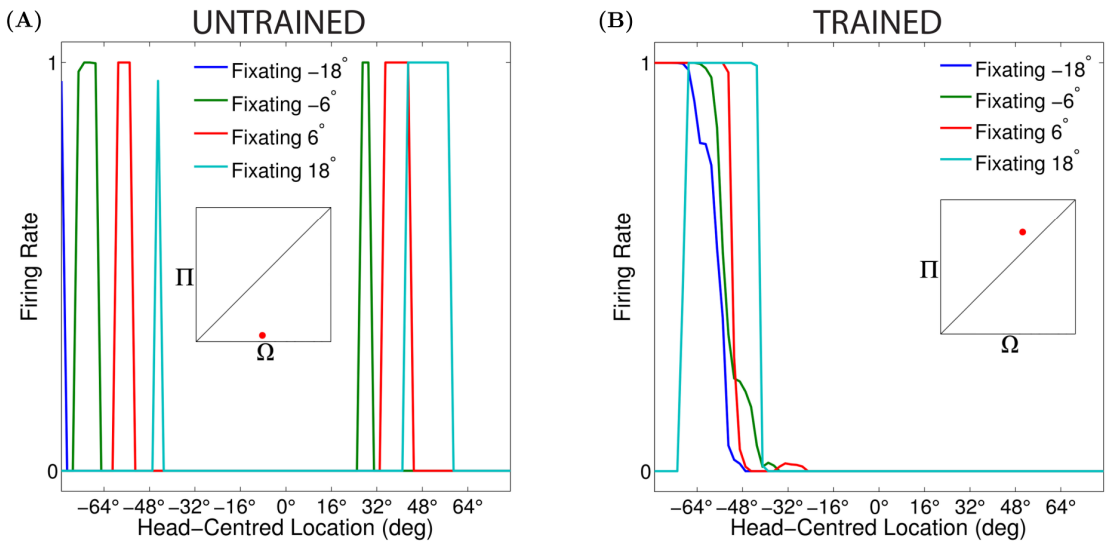


Figure 3.12: Simulation results showing the firing rate responses of a model incorporating a population of monotonic modulated input neurons trained with the modified learning rule 3.14: Delayed Postsynaptic Trace with anti-Hebbian Learning. The figure shows the firing rate responses of output neuron #168 before training (subplot 3.12**A**) and after training (subplot 3.12**B**) during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. In each subplot, each curve corresponds to a fixed eye-position while a visual target is presented across the same range of head-centred locations. The miniature scatter plot shows the reference frame values of all neurons in the output layer, where each neuron is plotted as a point corresponding to that neuron's particular combination of head-centredness (ordinate) and eye-centredness (abscissa). The neuron whose firing rate responses have been plotted is shown in the scatter plot by a red mark. After training it is evident that this neuron responds reasonably invariantly to a visual target presented at the same head-centred location regardless of the eye-position.

predicted structure shown in figure 3.8 for a head-centred neuron.

Figure 3.14 presents a population analysis of the reference frame response characteristics of the output neurons before and after training. In particular, the scatter plot in figure 3.14 shows that before training nearly all of the output neurons had head-centredness values close to 0 and were classified as eye-centred. However, after training the head-centredness values of many output neurons had dramatically increased, with quite a number of these neurons now classed as head-centred (table 3.4). Comparing the output population analysis

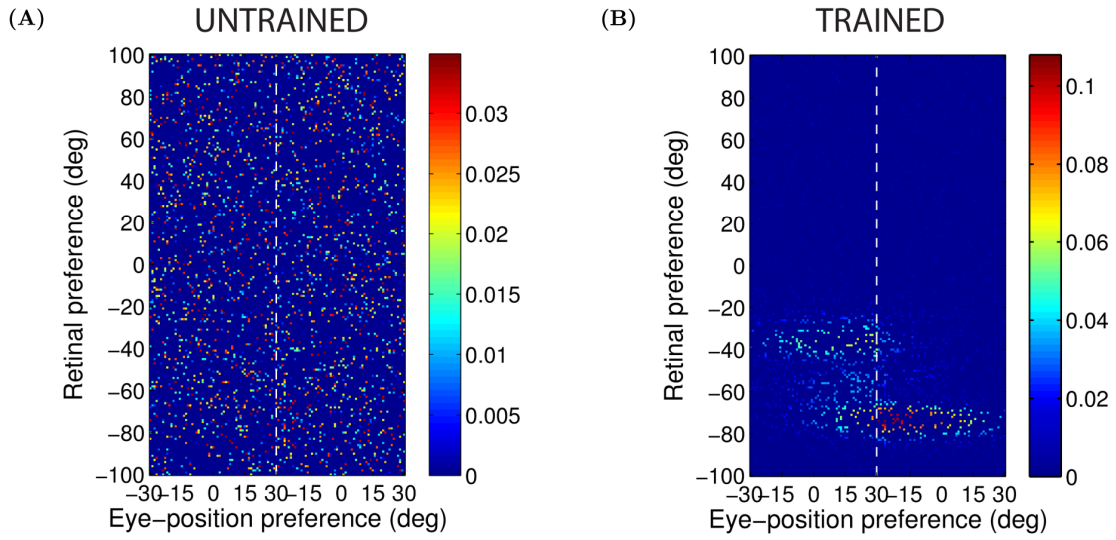


Figure 3.13: Simulation results showing the strengths of the afferent synapses of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.14: Delayed Postsynaptic Trace with anti-Hebbian Learning. The figure shows the strengths of the afferent synapses from the input population to output neuron #168 for the untrained (subplot 3.13**A**) and trained (subplot 3.13**B**) model. The output neuron corresponds to the one plotted in figure 3.12. In each plot, the afferent synapses have been arranged topographically by the preference of the input neuron for retinal location α_i and eye-position inflection point β_j. The portion of each plot to the left of the white dashed line corresponds to input neurons with positive gain $\kappa_j > 0$, whilst the portion of each plot to the right of the white dashed line corresponds to those input neurons with negative gain $\kappa_j < 0$. The synaptic weights for this output neuron after training (subplot 3.13**B**) have approximately the correct structure for a head-centred neuron as shown in figure 3.8.

of the model trained with the modified learning rule 3.14 shown in figure 3.14 with the performance of the model trained with the standard trace learning rule (equation 3.5) shown in figure 3.5 confirms that the new modified learning rule 3.14 is far more efficacious at driving the development of head-centred output neurons when the input neurons are modulated by a sigmoidal (monotonic) function of eye-position than the standard trace learning rule (equation 3.5).

In summary, these results showed that training the model with the modified learning

Experiments with modified learning rule 3.14		
	Untrained	**Trained**
Head-centredness	0.09 (0.11)	0.34 (0.21)
Eye-centredness	0.72 (0.23)	0.51 (0.27)
Proportion head-centred	0%	2%
Proportion eye-centred	70%	40%
Proportion unclassified	30%	58%

Table 3.4: Population summary statistics of response properties of output neurons in the models trained with the modified learning rule 3.14. Conventions as for table 3.2.

rule (equation 3.14) refined the response characteristics of many output neurons to be more compatible with a head-centred frame of reference, even when all of the input neurons had monotonic eye-position gain modulation.

3.10.6 Modified Learning Rule: Delayed Postsynaptic Firing Rate with Anti-Hebbian Learning

This section presents simulation results showing the performance of a learning rule which incorporates a delayed postsynaptic firing rate with anti-Hebbian learning (Rolls and Stringer, 2001). The same training procedure of previous simulations was used in the simulations presented in this section. The learning rule is defined by

$$\frac{dw_{ij}(t)}{dt} = \varrho \left(\beta y_i(t - \Delta T) - y_i(t)\right) v_j(t) \tag{3.15}$$

where y_i was the firing rate of output neuron i. In this case no postsynaptic trace value q_i is used to update the synaptic weights. The tuning parameter β works as described in section 3.10.5 for the anti-Hebbian learning rule with delayed trace (equation 3.14). The time delay ΔT in the firing rate of each output neuron and the parameter β were both

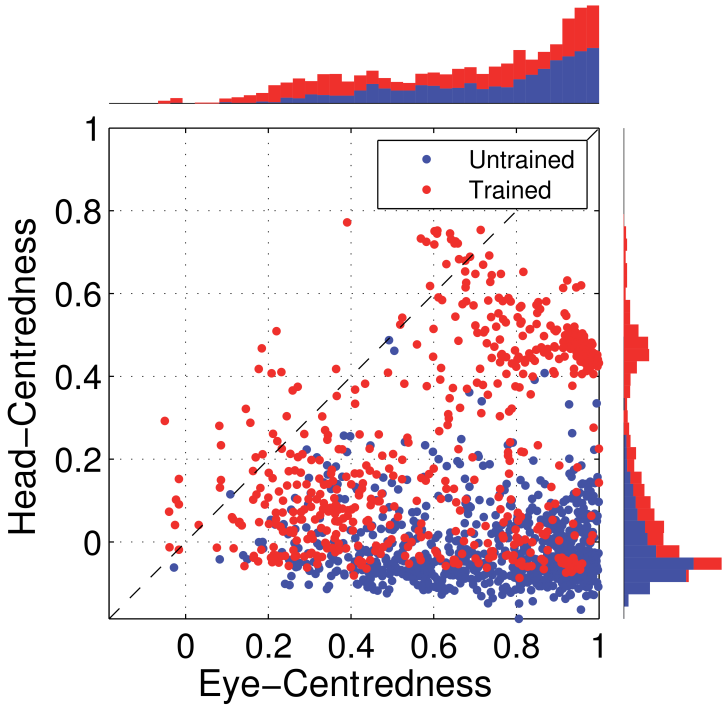


Figure 3.14: Simulation results showing the output reference frame response characteristics of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.14: Delayed Postsynaptic Trace with anti-Hebbian Learning. The scatter plot shows the reference frame response characteristics of all output neurons before and after training. Neurons that could not have been classified as eye-centred or head-centred are not plotted in the scatter plot. Each neuron is represented as a point corresponding to its combination of eye-centredness (abscissa) and head-centredness (ordinate) values. Data points for the untrained model are plotted in blue and data points for the trained model are shown in red. The marginal plots are histograms showing the frequency distribution of the numbers of output neurons with different values of eye-centredness (abscissa) and head-centredness (ordinate). The dashed diagonal line with positive unity slope separates those neurons which are classified as head-centred (above the line) from those that are classified as eye-centred (below the line). It can be seen that many of the output neurons have developed head-centred output responses after training.

tuned to optimise the performance of the model in driving the development of head-centred output neurons. The value of the time delay ΔT was set to $500ms$ and β was set to 2.4. Simulation parameters for the model, including the learning rate ϱ, are shown in table 3.1.

Figure 3.15 shows the firing rate responses of output neuron #876 before training (figure

3.15**A**) and after training (figure 3.15**B**). Figure 3.15**B** shows that after training output neuron #876 responded to the same head-centred location across different eye-positions. This was not the case for the same output neuron before training (figure 3.15**A**). Thus, it is evident that during training the neuron has learned to respond in a head-centred reference frame.

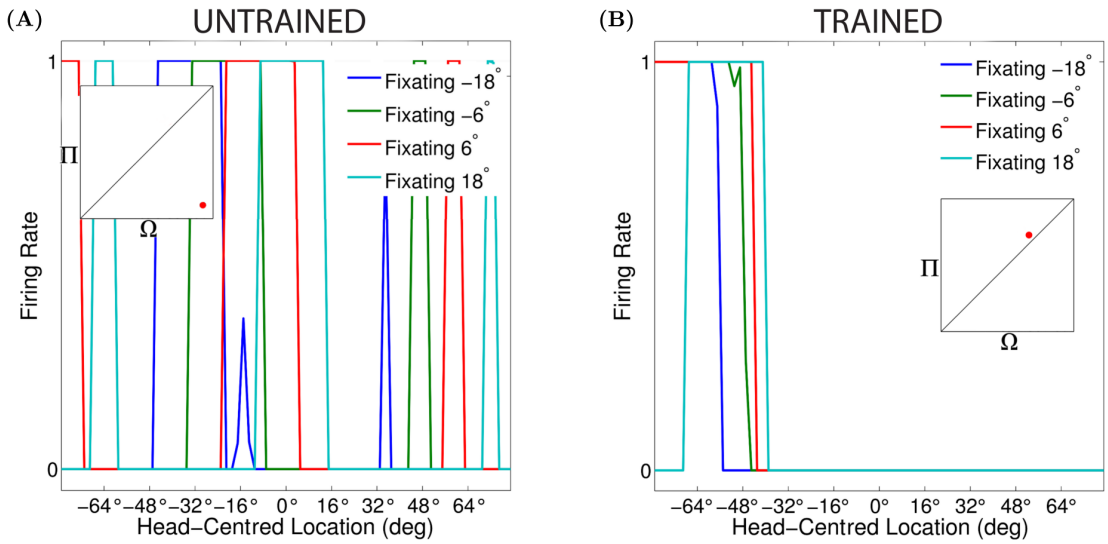


Figure 3.15: Simulation results showing the firing rate responses of a model incorporating a population of monotonic modulated input neurons trained with the modified learning rule 3.15: Delayed Postsynaptic Firing Rate with anti-Hebbian Learning. The figure shows the firing rate responses of output neuron #876 before training (subplot 3.15**A**) and after training (subplot 3.15**B**) during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. Conventions as for figure 3.12. The comparison of (subplot 3.15**A**) and subplot (subplot 3.15**B**) shows that the output neuron learned to respond to a specific head-centred location regardless of the eye-position after training.

Figure 3.16 shows how the synaptic weight structure of the same output neuron #876 changed due to training the model with learning rule 3.15. The final weight structure after training shown in figure 3.16**B** resembles the predicted weight structure shown in figure 3.8 for head-centred neurons.

Figure 3.17 shows the eye-centredness and head-centredness values of output neurons in

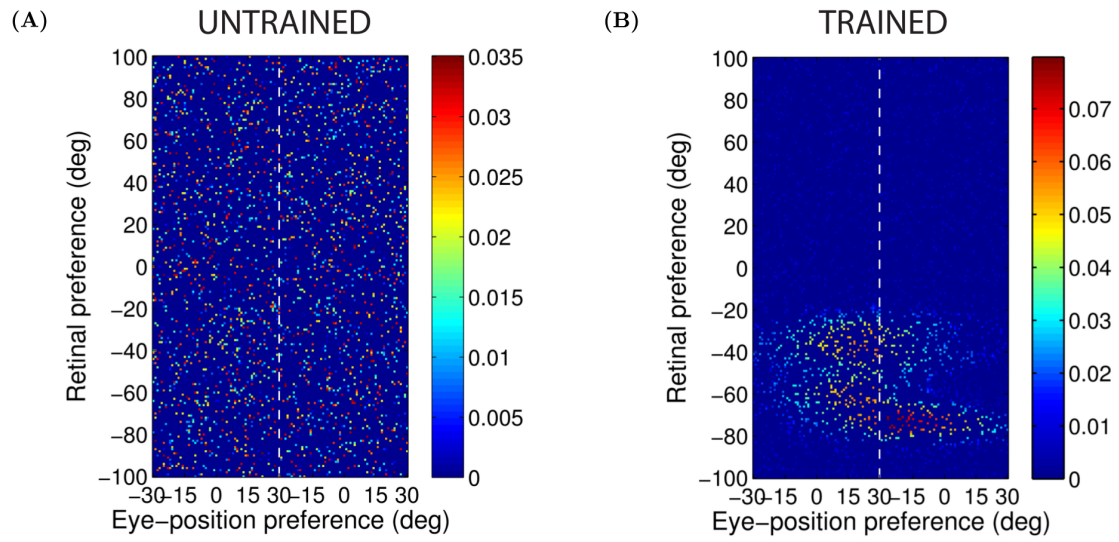


Figure 3.16: Simulation results showing the strengths of the afferent synapses of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.15: Delayed Postsynaptic Firing Rate with anti-Hebbian Learning. The figure shows the strengths of the afferent synapses from the input population to output neuron #876 for the untrained (subplot 3.16**A**) and trained (subplot 3.16**B**) model. The output neuron corresponds to the one plotted in figure 3.15. Conventions as for figure 3.13. The synaptic weight structure for this output neuron after training shown in subplot 3.16**B** has approximately the correct profile for a head-centred neuron (figure 3.8).

the untrained and in the trained model. It is clear that almost none of the output neurons in the untrained model had values of head-centredness higher than eye-centredness (table 3.5). However, after training the head-centredness values of many output neurons increased substantially, with a number of such neurons now having greater head-centredness than eye-centredness values (table 3.5). Such neurons are, therefore, classified as head-centred neurons (section 3.9). Comparing the output population analysis of the model trained with the modified learning rule 3.15 shown in figure 3.17 with the analysis of the model trained with the standard trace learning rule (equation 3.5) shown in figure 3.5 demonstrates that the new modified learning rule 3.15 is also more efficacious at producing head-centred

Experiments with modified learning rule 3.15		
	Untrained	**Trained**
Head-centredness	0.04 (0.07)	0.32 (0.21)
Eye-centredness	0.72 (0.22)	0.63 (0.26)
Proportion head-centred	0%	2%
Proportion eye-centred	70%	47%
Proportion unclassified	30%	51%

Table 3.5: Population summary statistics of response properties of output neurons in the models trained with the modified learning rule 3.15. Conventions as for table 3.2.

output neurons when the input neurons are modulated by a sigmoidal function of eye-position than the standard trace learning rule (equation 3.5) previously implemented by Mender and Stringer (2013).

In summary, these results showed that training the model with the modified learning rule 3.15 resulted in the development of head-centred output neurons, even when the whole input population had sigmoidal eye-position gain modulation. However, the comparison of figure 3.17 and figure 3.14 shows that learning rule 3.14, which incorporated a delayed postsynaptic trace $q_i(t - \Delta T)$, is in fact more efficacious at driving the development of head-centred output neurons than learning rule 3.15, which incorporated the delayed firing rate $y_i(t - \Delta T)$. Thus, the incorporation of the trace value $q_i(t - \Delta T)$ enhances the ability of the learning rule to perform temporal binding of input patterns corresponding to the same head-centred location.

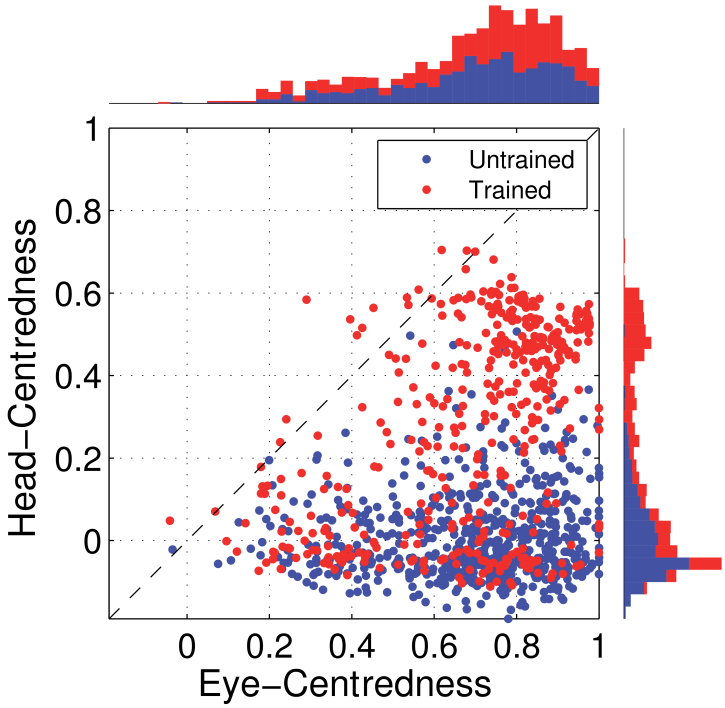


Figure 3.17: Simulation results showing the output reference frame response characteristics of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.15: Delayed Postsynaptic Firing Rate with anti-Hebbian Learning. The scatter plot shows the reference frame response characteristics of all output neurons before and after training. Conventions as for figure 3.14. It is evident that training had the effect of increasing the head-centredness values of most output neurons. Indeed many more head-centred output neurons are present in the trained model than in the untrained model.

3.10.7 Modified Learning Rule: Current Postsynaptic Trace with Anti-Hebbian Learning

This section presents simulation results showing the performance of a modified learning rule which incorporated the current postsynaptic trace with anti-Hebbian learning. Thus, this learning rule does not use an explicit time delay ΔT. The learning rule was given by

$$\frac{dw_{ij}(t)}{dt} = \varrho\left(\beta q_i(t) - y_i(t)\right) v_j(t) \tag{3.16}$$

where ϱ was the learning rate, β was the tuning parameter, q_i was the trace value of the output neuron i, and y_i and v_j were the post- and pre-synaptic firing rate values, respectively. What distinguishes the learning rule in equation 3.16 from the learning rule in equation 3.14 is the use of the trace value calculated at the same time t when the synaptic weight is updated.

Simulation parameters for the model, including the learning rate ϱ, are shown in table 3.1. The value of the tuning parameter β was set to 2.2 for optimal performance.

Figure 3.18 shows the firing rate responses of output neuron #328 for the untrained and trained models. Specifically, figure 3.18**A** shows that prior to training the response of output neuron #328 had no consistent structure in head-centred space across different eye-positions. However, figure 3.18**B** shows that after training the same output neuron reponded maximally to the same head-centred location across all four eye-positions. Thus, after training, neuron #328 responded in a head-centred frame of reference.

The change in the synaptic weight structure of the same output neuron #328 due to training is shown in figure 3.19. Before training the afferent synaptic weights were randomly assigned (figure 3.19**A**). After training, however, the synaptic weight structure of the same output neuron (figure 3.19**B**) is similar to the predicted weight structure for head-centred neurons shown in figure 3.8.

Figure 3.20 shows the eye-centredness and head-centredness values of output neurons in the untrained and in the trained model. In particular, figure 3.20 shows that training had the effect of increasing the head-centredness value for a large proportion of output neurons. Furrthermore, while almost no output neurons were classified as head-centred before training, a significant number of output neurons were classified as head-centred

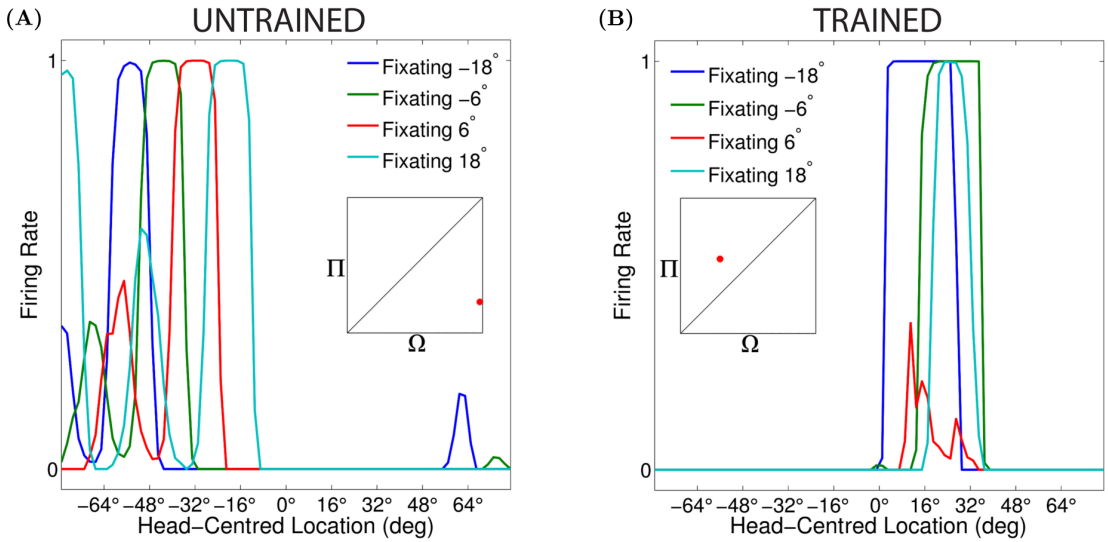


Figure 3.18: Simulation results showing the firing rate responses of a model incorporating a population of monotonic modulated input neurons trained with the modified learning rule 3.16: Current Postsynaptic Trace with anti-Hebbian Learning. The figure shows the firing rate responses of output neuron #328 before training (subplot 3.18**A**) and after training (subplot 3.18**B**)during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. Conventions as for figure 3.12. The comparison of subplot 3.18**A** and subplot 3.18**B** shows that the output neuron learned to respond to a specific head-centred location regardless of the eye-position after training.

after training (table 3.6). A comparison of the output population analysis of the model trained with the modified learning rule 3.16 shown in figure 3.20 with the performance of the model trained with the standard trace learning rule (equation 3.5) presented in figure 3.5 shows that the modified learning rule 3.16 is also significantly more capable of producing head-centred output neurons than the standard trace learning rule (equation 3.5) when all of the input neurons have sigmoidal gain modulation by eye-position.

In summary, these results showed that training the model with learning rule 3.16 drives the development of head-centred output neurons even when the whole input population had sigmoidal eye-position gain modulation. The comparison of figure 3.20 with figure 3.14 shows that learning rule 3.14, which incorporated a delayed postsynaptic trace $q_i(t - \Delta T)$,

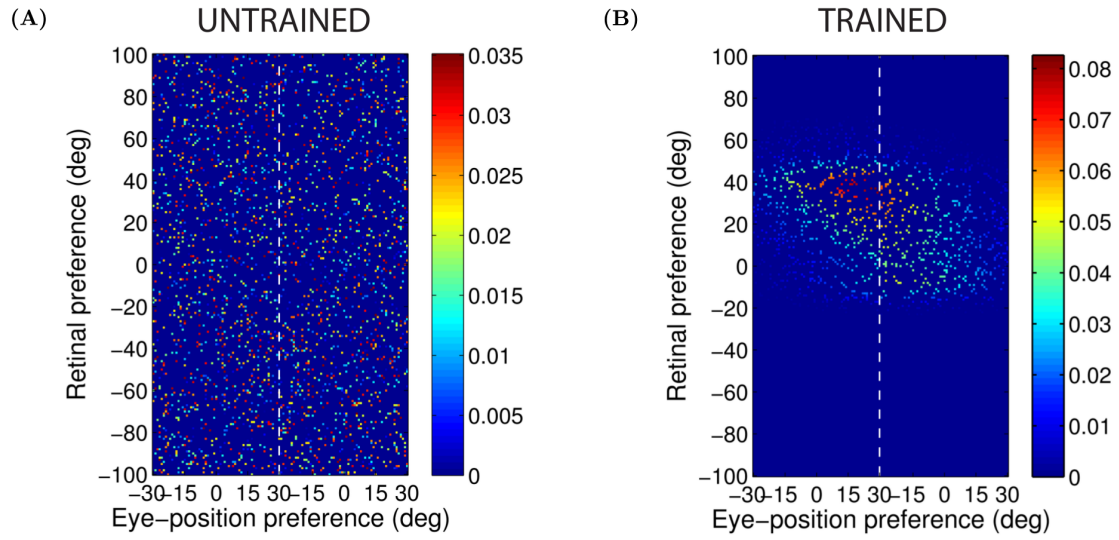


Figure 3.19: Simulation results showing the strengths of the afferent synapses of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.16: Current Postsynaptic Trace with anti-Hebbian Learning. The figure shows the strengths of the afferent synapses from the input population to output neuron #328 for the untrained (subplot 3.19**A**) and trained (subplot 3.19**B**) model. The output neuron corresponds to the one plotted in figure 3.18. Conventions as for figure 3.13. The synaptic weight structure for this output neuron after training shown in subplot 3.19**B** has the correct kind of profile for a head-centred neuron (figure 3.8).

is more effective at producing head-centred output neurons than learning rule 3.16, which incorporated the current trace $q_i(t)$ calculated at the same time t when the weights are updated.

3.10.8 Modified Learning Rule: Delayed Postsynaptic Trace Learning Rule

The simulation results presented in section 3.10.5, section 3.10.6 and section 3.10.7 showed, respectively, that training the model with learning rule 3.14, learning rule 3.15 and learning rule 3.16 successfully self-organised head-centred output responses and increased the head-centredness value of a large proportion of output neurons when all input neurons had

Experiments with modified learning rule 3.16		
	Untrained	**Trained**
Head-centredness	0.14 (0.10)	0.41 (0.14)
Eye-centredness	0.78 (0.21)	0.71 (0.15)
Proportion head-centred	0%	2%
Proportion eye-centred	93%	37%
Proportion unclassified	7%	61%

Table 3.6: Population summary statistics of response properties of output neurons in the models trained with the modified learning rule 3.16. Conventions as for table 3.2.

sigmoidal eye-position gain modulation. Importantly, section 3.10.1 showed this was not the case for the standard trace learning rule (3.5). Training the model with the standard trace learning rule successfully self-organised head-centred output neurons when all of the input neurons had peaked eye-position gain modulation, but failed when the modulation of the input population was altered from peaked to sigmoidal. Indeed, the simple introduction of a small proportion (e.g. with $p > 0.2$) of input neurons with sigmoidal eye-position gain modulation was enough to undermine the self-organisation of head-centred output responses (section 3.10.4).

The modified learning rules introduced in section 3.10.5, section 3.10.6 and in section 3.10.7 all had an anti-Hebbian term as their common component. Out of these learning rules, the best performance was observed with learning rule 3.14, which incorporated a postsynaptic delayed-trace term. In this case, an interesting question is whether the superior efficacy of this learning rule in driving the development of head-centred output responses was primarily due to the anti-Hebbian term or the postsynaptic delayed-trace term. In particular, is an anti-Hebbian term actually needed in this learning rule for

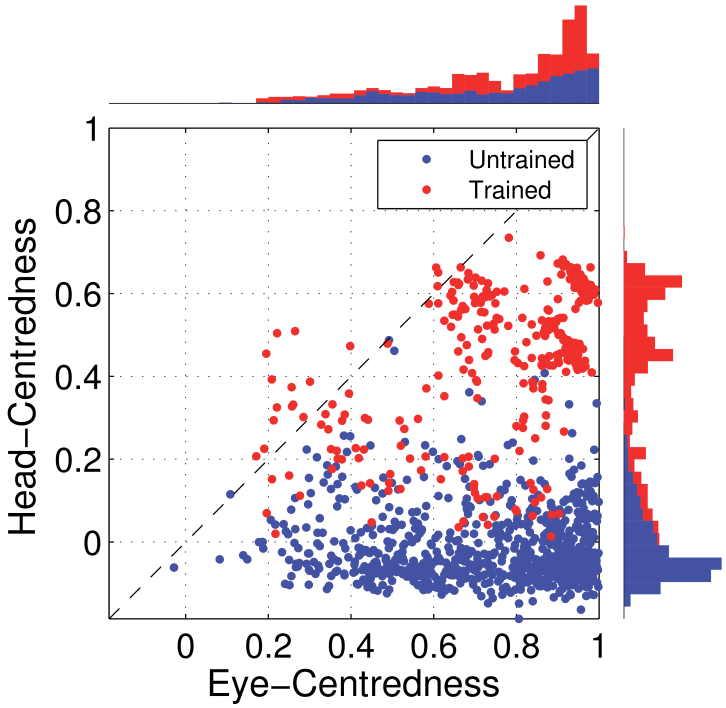


Figure 3.20: Simulation results showing the output reference frame response characteristics of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.16: Current Postsynaptic Trace with anti-Hebbian Learning. The scatter plot shows the reference frame response characteristics of all output neurons before and after training. Conventions as for figure 3.14. It can be seen that training increased the head-centredness values of most output neurons, with quite a number of head-centred output neurons present in the trained model.

the self-organisation of head-centred responses in the presence of input neurons with sigmoidal eye-position gain modulation, or could head-centred output responses develop using a learning rule that only incorporated a postsynaptic delayed-trace term?

This section addresses the above questions by investigating the performance of the model with the following learning rule

$$\frac{dw_{ij}(t)}{dt} = \varrho q_i(t - \Delta T)v_j(t) \tag{3.17}$$

where ϱ was the learning rate, q_i was the trace value of the output neuron i and v_j was the firing rate of the input neuron j. In equation 3.17 the trace value q_i is calculated at time $(t - \Delta T)ms$. This learning rule has no anti-Hebbian term, and relies purely on the postsynaptic delayed-trace term $q_i(t - \Delta T)$ to drive the development of head-centred output neurons.

Simulation parameters for the model, including the learning rate ϱ, are shown in table 3.1. The time delay ΔT used to compute the trace value of each output neuron i was set to $30ms$ for optimal learning performance.

Figure 3.21 shows how training changed the firing rate responses of output neuron #223. Prior to training the response of output neuron #223 had no consistent structure in head-centred space across different eye-positions (figure 3.21**A**), whilst after training the same output neuron responded maximally to the same head-centred location across all four eye-positions (figure 3.21**B**). Thus, neuron #223 learned to respond in a head-centred reference frame after training.

Figure 3.22 shows how the synaptic weight structure of the same output neuron #223 plotted in figure 3.21 changed due to training the model with learning rule 3.17. It is clear that the final synaptic weight structure after training (figure 3.22**B**) resembles the predicted weight structure for head-centred neurons (figure 3.8), even though this was not the case before training (figure 3.22**A**).

Figure 3.23 shows the eye-centredness and head-centredness values of all of the output neurons in the untrained model and in the trained model. After training, the head-centredness values of many output neurons had increased substantially (table 3.7). However, there was only a single output neuron, which was neuron #223, with a greater

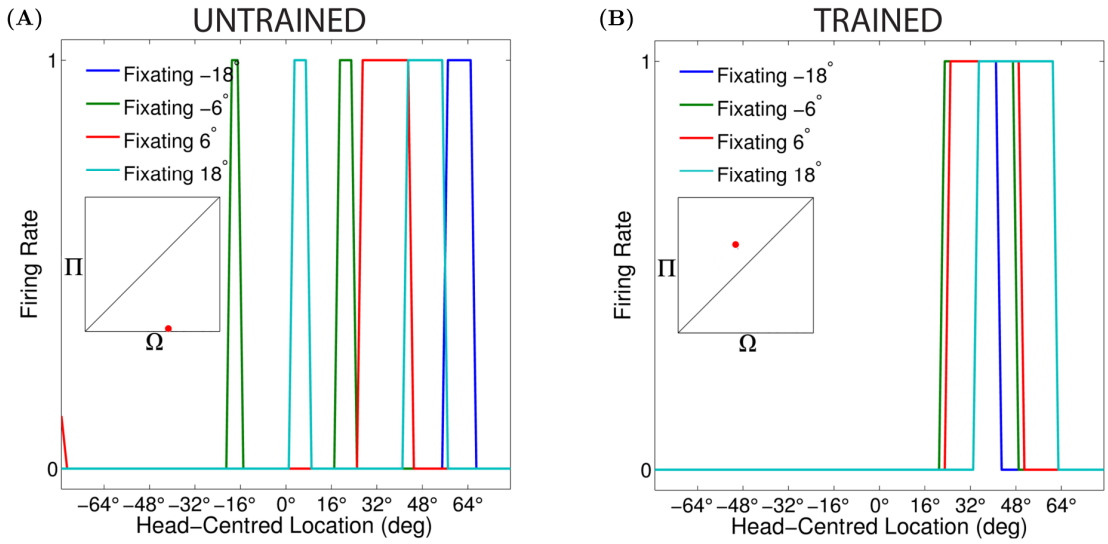


Figure 3.21: Simulation results showing the firing rate responses of a model incorporating a population of monotonic modulated input neurons trained with the modified learning rule 3.17: Delayed Postsynaptic Trace Learning Rule. The figure shows the firing rate responses of output neuron #223 before training (subplot 3.21**A**) and after training subplot 3.21**A** during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. Conventions as for figure 3.12. The comparison of subplot 3.21**A** and subplot 3.21**B** shows that the output neuron learned to respond to a specific head-centred location regardless of the eye-position after training. Subplot 3.21**B** shows that this output neuron learned to respond to a specific head-centred location regardless of the eye-position after training.

head-centredness value than eye-centreredness, and which was consequently classified as responding in a head-centred reference frame. Therefore, in the simulations presented in this paper, the anti-Hebbian term in learning rules (3.14), (3.15) and (3.16) appears to play an important and essential role in producing relatively large numbers of head-centred output neurons when all of the input neurons have sigmoidal eye-position gain fields.

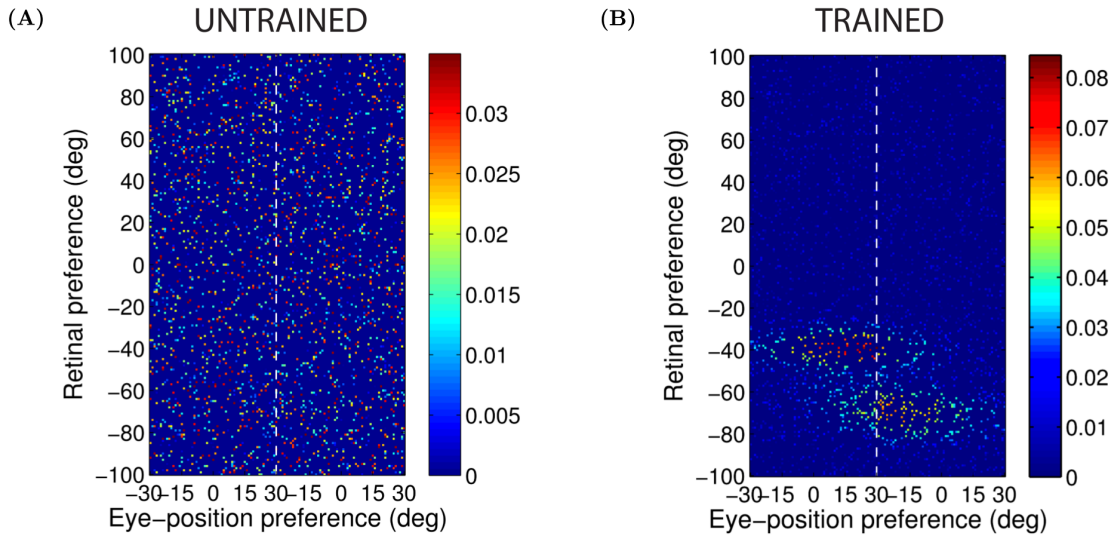


Figure 3.22: Simulation results showing the strengths of the afferent synapses of a model incorporating a population of sigmoidal modulated input neurons trained with the modified learning rule 3.17: Delayed Postsynaptic Trace Learning Rule. The figure shows the strengths of the afferent synapses from the input population to output neuron #223 for the untrained (subplot 3.22**A**) and trained (subplot 3.22**A**) model. The output neuron corresponds to the one plotted in figure 3.21. Conventions as for figure 3.13. The synaptic weight structure for this output neuron after training shown in subplot 3.22**A** has approximately the correct profile for a head-centred neuron (figure 3.8).

3.10.9 Delayed Postsynaptic Trace Learning Rule with Mixed Peaked and Monotonic eye-position Modulation of Input Neurons

In the previous section, it was found that an anti-Hebbian term was needed in the learning rule in order to produce relatively large numbers of head-centred output neurons if the entire population of input neurons had sigmoidal gain fields. However, experimental studies have demonstrated that the primate cortex contains a mixed population of visual neurons with either peaked or monotonic eye-position gain fields (Andersen et al., 1990, 1985, Galletti et al., 1995). This raises the question of whether learning rule (3.17), which relies solely on a postsynaptic delayed-trace term without any anti-Hebbian term, could produce

Experiments with modified learning rule 3.17		
	Untrained	**Trained**
Head-centredness	0.12 (0.18)	0.38 (0.19)
Eye-centredness	0.85 (0.14)	0.75 (0.21)
Proportion head-centred	0%	0%
Proportion eye-centred	70%	58%
Proportion unclassified	30%	42%

Table 3.7: Population summary statistics of response properties of output neurons in the models trained with the modified learning rule 3.17. Conventions as for table 3.2.

a much larger number of head-centred output neurons if the input population had a 50:50 mix of peaked and sigmoidal eye-position gain fields.

This section presents simulation results with the modified learning rule (3.17) when there is a 50:50 mixture of peaked and monotonic gain fields in the visual input population. Can the model produce a large number of head-centred output neurons under such conditions without an anti-Hebbian term in the learning rule? This is still a potentially challenging task for the model because in section 3.10.4 it was shown that the standard trace learning rule (3.5), which lacks an anti-Hebbian term, failed to produce significant numbers of head-centred output neurons when the proportion of sigmoidal gain modulated input neurons rose above just 20% of the overall input population. The model parameters for the simulations presented in this section are given in table 3.1.

Figure 3.24 shows how training changed the firing rate responses of output neuron #281. In particular, figure 3.24**B** shows that after training output neuron #281 responded maximally to the same head-centred location across all four eye-positions, although this was not the case prior to training (figure 3.24**A**). Hence neuron #281 responded in a

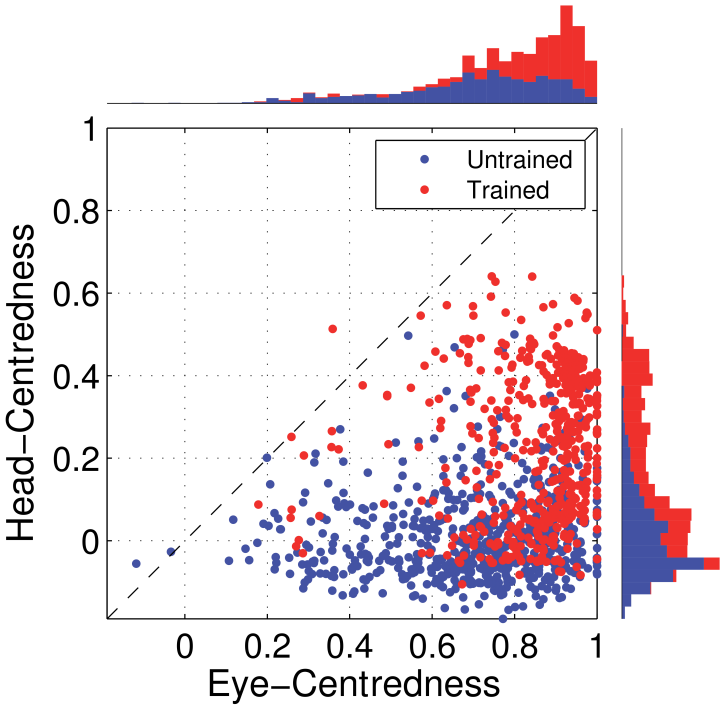


Figure 3.23: Simulation results showing the output reference frame response characteristics of a model incorporating a population of monotonic modulated input neurons trained with the modified learning rule 3.17: Delayed Postsynaptic Trace Learning Rule. The scatter plot shows the reference frame response characteristics of all output neurons before and after training. Conventions as for figure 3.14. It is evident that training had the effect of increasing the head-centredness values of most output neurons. Although, there is only a single head-centred output neuron present in the trained model.

head-centred frame of reference after training.

Figure 3.25 shows how the synaptic weight structure of the same output neuron #281 plotted in figure 3.24 changed due to training. Figure 3.25**B** shows that after training, the synaptic weight structure of the output neuron was clearly similar to the predicted weight structure for head-centred neurons shown in figure 3.8. The synaptic weight structure before training reflected the randomly assigned connection weights (figure 3.25**A**).

Figure 3.26 shows the eye-centredness and head-centredness values of output neurons

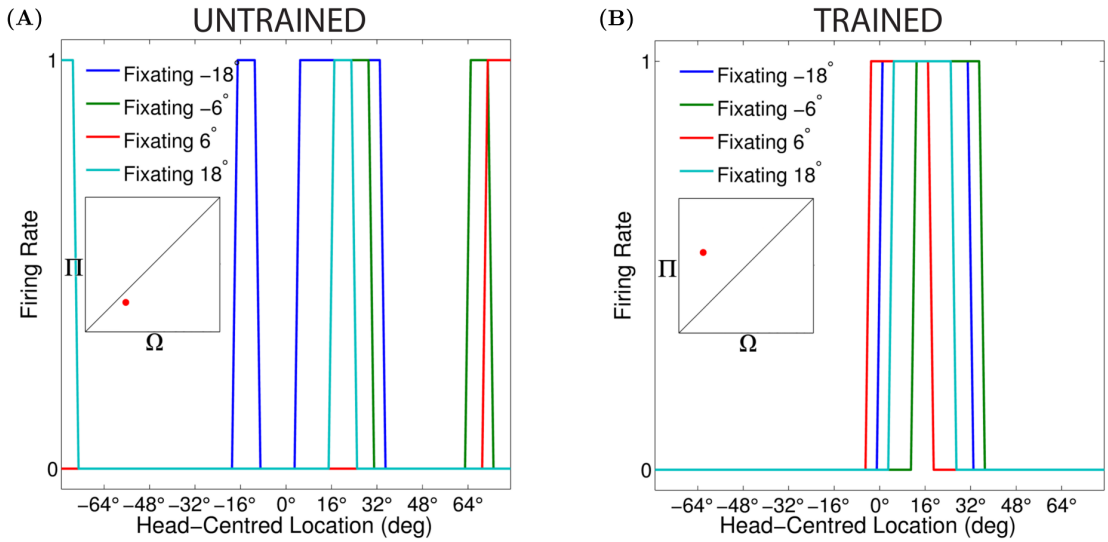


Figure 3.24: Simulation results of a model incorporating a mixed population of peaked and monotonic modulated input neurons, with sigmoidal (monotonic) modulation rate p set to 0.5, trained with the modified learning rule 3.17: Delayed Postsynaptic Trace Learning Rule. The figure shows the firing rate responses of output neuron #281 before training (subplot 3.24**A**) and after training (subplot 3.24**B**) during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. Conventions as for figure 3.12. Subplot 3.24**A** shows that this same output neuron learned to respond to a specific head-centred location regardless of the eye-position after training.

in the untrained and in the trained model. Figure 3.26 shows that training the model with learning rule 3.17, where the input population contained a 50:50 mix of neurons with peaked or sigmoidal gain modulation, had the effect of increasing the head-centredness value of most output neurons (table 3.8). Moreover, a large proportion of output neurons in the trained model are head-centred (i.e. with a head-centredness value greater than eye-centredness). Indeed, more head-centred output neurons were observed in the trained model in this section than in any of the trained models presented in previous section 3.10.5, section 3.10.6 and section 3.10.7.

In summary, these results demonstrated that it was possible for the model to produce large numbers of head-centred output neurons with a learning rule that relied purely on

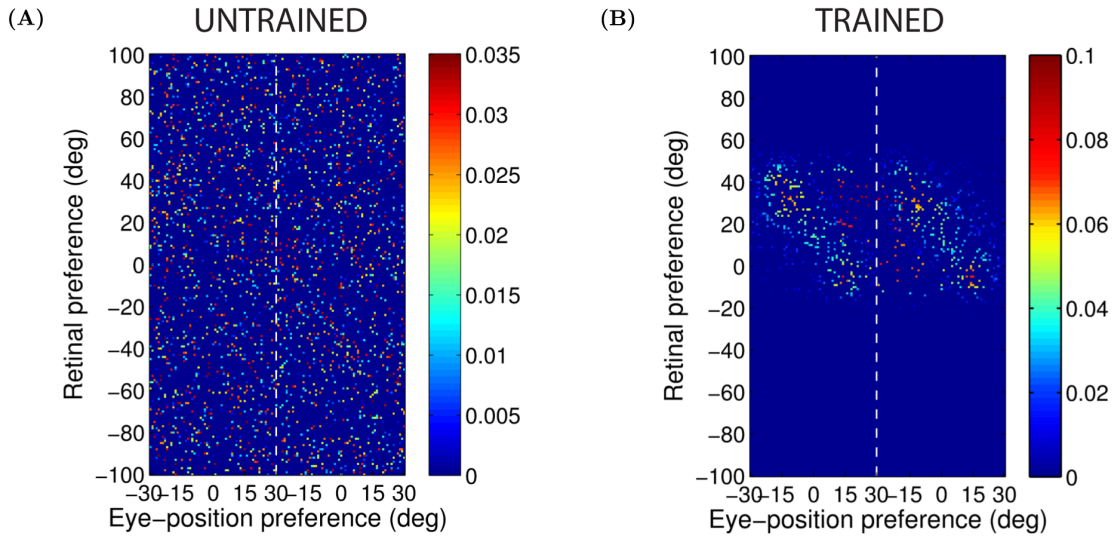


Figure 3.25: Simulation results of a model incorporating a mixed population of peaked and monotonic modulated input neurons, with sigmoidal (monotonic) modulation rate p set to 0.5, trained with the modified learning rule 3.17: Delayed Postsynaptic Trace Learning Rule. The figure shows the strengths of the afferent synapses from the input population to output neuron #281 for the untrained (subplot 3.25**A**) and trained (subplot 3.25**B**) model. The output neuron corresponds to the one plotted in figure 3.24. Conventions as for figure 3.13. The synaptic weight structure for this output neuron after training shown in subplot 3.25**B** has the correct kind of profile for a head-centred neuron (figure 3.8).

a postsynaptic delayed-trace term without any anti-Hebbian term if the input population contained a 50:50 mix of neurons that were modulated by either peaked or sigmoidal gain fields. Such a mixture of different forms of eye-position gain modulation is biologically compatible with experimental findings (Andersen et al., 1990, 1985).

3.11 Conclusion

In conclusion, the work presented in this chapter has shown that the existence of a synaptic weight structure that can perform a coordinate transformation does not guarantee that it will emerge through a process of self-organisation using any one particular form of

Experiments with learning rule 3.17 and mixed input modulation		
	Untrained	**Trained**
Head-centredness	0.07 (0.10)	0.52 (0.15)
Eye-centredness	0.22 (0.14)	0.48 (0.17)
Proportion head-centred	2%	6%
Proportion eye-centred	54%	38%
Proportion unclassified	44%	50%

Table 3.8: Population summary statistics of response properties of output neurons in the models trained with the modified learning rule 3.17 when the model had a 50:50 mixed of neurons that were modulated by either peaked or sigmoidal gain fields.. Conventions as for table 3.2.

trace learning rule. In particular, the standard trace learning rule originally proposed by Mender and Stringer (2013) failed to produce head-centred output neurons when the input neurons were modulated by a sigmoidal function of eye-position, and even failed with a biologically realistic 50:50 mix of input neurons with peaked and sigmoidal gain fields. In order to remedy this problem, we had to investigate the use of more sophisticated, yet still biologically plausible, learning rules that incorporated temporally delayed memory traces, as well as a mixture of both trace learning and anti-Hebbian learning terms. These new, modified learning rules were found to produce head-centred output neurons when the input population had sigmoidal gain fields. Moreover, it was also found that the delayed postsynaptic trace in learning rule (equation 3.17) was sufficient to drive the development of head-centred output neurons in the absence of anti-Hebbian learning, especially if the input population had a 50:50 mix of peaked and monotonic gain fields. This work thus makes an important contribution to understanding how head-centred visual neurons may develop in the brain through an unsupervised process of visually-guided learning given visual precursor

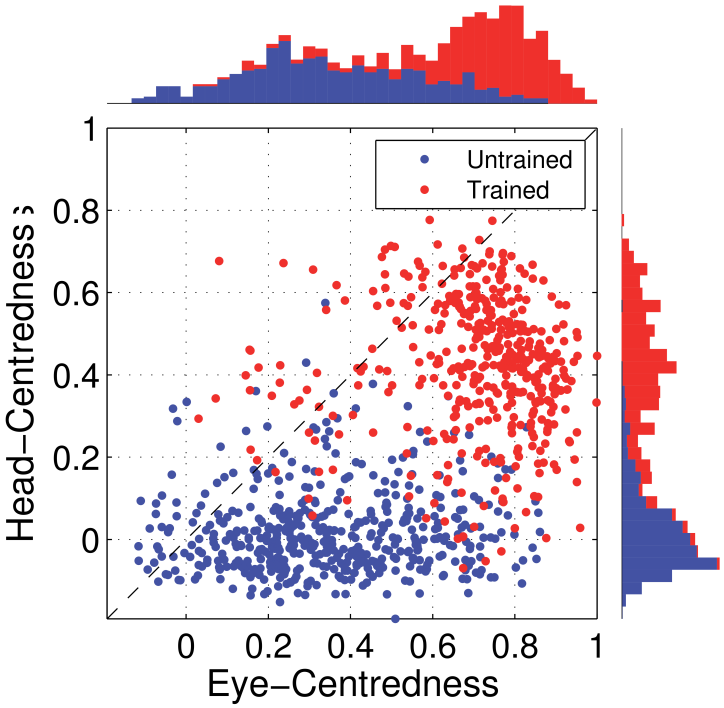


Figure 3.26: Simulation results of a model incorporating a mixed population of peaked and sigmoidal modulated input neurons, with sigmoidal modulation rate p set to 0.5, trained with the modified learning rule 3.17: Delayed Post-synaptic Trace Learning Rule. The scatter plot shows the reference frame response characteristics of all output neurons before and after training. Conventions as for figure 3.14. It is evident that training had the effect of increasing the head-centredness values of most output neurons, with many more head-centred output neurons present in the trained model.

neurons with sigmoidal (monotonic) eye-position modulation. Furthermore, although we have studied one particular kind of coordinate transformation, i.e. from eye-centred to head-centred visual representations, these findings may also apply to a range of other coordinate transformations with sensorimotor integration of monotonically encoded motor variables (Bremner and Andersen, 2012, Duhamel et al., 1997, Pesaran et al., 2006).

Having developed the neural network computer model with simulated data, we are now in a position to test specific experimental predictions. In particular, we predict that the

statistics of eye and head movements recorded from human test subjects during a natural visual search task would be appropriate to drive the development of head-centred visual neurons in the output layer of the model. This study is described in chapter 5.

Chapter 4

DN Mobile Rotation Tracking System: An Android application for recording head-rotation movements

4.1 Introduction

The DN Rotation Tracking System is a smartphone application developed to record human movement data for scientific studies. In particular, this version of the application has the capability to collect and combine data from a number of smartphone sensors to detect and record (e.g. head) rotation movements along the x-axis and along the y-axis. Previous methods of collecting similar data for scientific purposes relied, for example, on more ad hoc solutions Bartz (1966), personalised devices Minati et al. (2016) or expensive commercial devices David et al. (2018). By comparison, the use of wearable or mobile devices, such as smartphones, may overcome the intrinsic problems of other device-dependent solutions adopted by earlier studies Bartz (1966), David et al. (2018), Minati et al. (2016). For instance, recent technological developments have turned smartphones and wearable gadgets into powerful computing devices. In particular, an increasing number of enhanced sensors are included with new flagship devices produced by different manufacturers. However,

the significant potential of using such devices to collect data for scientific purposes, such as behavioural data from humans, has not been fully explored in previous work. Whilst the creation of mobile applications to collect behavioural data is clearly promising, the development of such applications is sufficiently difficult to be beyond the ability of a typical researcher. The DN Rotation Tracking System for Android is a scientifc tool created to address the challenge of using Android enabled mobile devices (e.g. smartphones) to collect behavioural data.

The DN Rotation Tracking System was built with a modular and flexible design to facilitate the addition of new functionaties in future versions of the application. In its current version, the DN Rotation Tracking System provides the following features:

- Clear, intuitive and consistent user interface.
- Visual indicators on the Home screen of the application with the status of the required hardware and software permissions.
- The Recording screen is only accessible from the Home screen when the hardware and software requirements are fulfilled.
- The file manager application can be accessed from the Home screen of the application to view and manage the recorded files.
- Validation of unique file names in real time with both textual and visual information on the Recording screen of the application.
- Textual status messages to inform the user of warnings, tips and errors occurred during the tasks performed by the application.

Moreover, the DN Rotation Tracking System has been successfully used to record head-rotation movements of human participants during a visual search task in the study discussed in the next chapter. The recorded data was then used to train our biological neural network models of coordinate transformation in the visual brain discussed in this thesis.

4.2 Motivation and significance

Mobile devices, such as smartphones, are promising tools to collect data for scientific purposes. This is particularly true, for example, in the case of human movement data, such as head-rotation movements during the natural exploration of visual scenes. This form of human movement data has been explored in a number of different studies. For instance, in Bartz (1966) the authors collected simultaneous eye and head movements to investigate the nature of compensatory eye movements in peripheral vision. To record head movements the authors constructed an epoxy skullcap for each human participant that took part in their experiment. A flexible cable was attached to each skullcap at the center of rotation of the head. The other end of the cable attached to the epoxy skullcap was connected to a potentiometer used to infer the rotation of the head. A different approach was used by Minati et al. (2016) when investigating how a range of different forms of behavioural data, including head-rotation movements, can be used to control a robotic arm. In particular, in Minati et al. (2016) the authors used electrooculogram (EOG), electroencephalogram (EEG), electromyogram (EMG) and head movements to control a robotic arm. In this case, head movement data was collected with a wearable device primarily sold as a meditation aid (muse 2014; InteraXon Inc., Toronto, Canada) that had to be modified by the authors. Finally, in David et al. (2018) the authors used a commercial VR headset (HTC

VIVE, HTC, Valve corporation) equipped with an SMI eye-tracker (SensoMotoric Instrument) to simultaneously record eye and head movements from human participants watching 360° videos. The experimental setup in David et al. (2018), however, required powerful computing resources, which included a computer with a NVIDIA GTX1080 GPU. It is clear that these data collection methods used in previous studies are intrinsically costly, rather impractical and, as a consequence, less likely to be used in future studies by other researchers. In contrast, the use of wearable and mobile devices provides an affordable, device-independent alternative for behavioural data collection. Moreover, an increasing number of enhanced sensors are included with new flagship devices produced by different manufacturers. Nonetheless, the significant potential to use such devices to collect data for scientific purposes, such as behavioural data from humans, has not been fully explored in previous work. Whilst the creation of mobile applications to collect behavioural data is clearly promising, the development of such applications is sufficiently difficult to be beyond the ability of a typical researcher.

The development of the DN Rotation Tracking System for Android enabled devices was motivated by the need for a reliable, cost-effective and non-device-dependent method to collect behavioural data. The primary research question that this application aims to address is whether or not sensors available on flagship devices (e.g. smartphone sensors) are powerful enough to be used for scientific data collection. In particular, we investigate whether flagship smartphone sensors can record head rotation movements from human participants during natural exploration of visual scenes, similar to the head movements investigated in previous experimental studies (Bartz, 1966, David et al., 2018, Minati et al., 2016). In its current version, the DN Rotation Tracking System collects data from different

sensors within the smartphone to record head rotation movements. Furthermore, due to the its modular design, new functionalities can be easily integrated to future versions of the application.

The DN Rotation Tracking System is released as an open source scientific tool for Android smartphones to support other researcher who may benefit from the type of behavioural data used in previous studies Bartz (1966), David et al. (2018), Minati et al. (2016). In fact, the DN Rotation Tracking System has been successfully used in the study described in chapter 5 to train our biological neural network models of coordinate transformation in the visual brain. Our main focus in this chapter, however, is the description of the design and implementaion of the DN Rotation Tracking System.

4.3 Software description

In its current version the DN Rotation Tracking System implements a number of functionalities to (i) allow continous collection, processing and recording of sensor data and to (ii) facilitate the handling of the recorded data files using the user's preferred file manager application. The collection, processing and recording of sensor data is performed by a specific component of the DN Rotation Tracking System. The specific component of the application accessing the sensor data is only activated when all hardware and software requirements are satisfied. This design ensures that the application is battery efficient by calling on and activating highly energy-consuming functionalities only when needed. Furthermore, facilitating the handling of the recorded data files within the DN Rotation Tracking System application allows the user to easily and safely move the data files with the recorded data from the default directory used by the DN Rotation Tracking System to

other directories locally available to the device, as well as to upload these data files to the cloud using the device's Internet access, or even share these data files using any sharing capabilities of the smartphone (e.g. Bluetooth etc).

The DN Rotation Tracking System was written in Java and developed using the Android Studio SDK platform. The Android Studio is the official integrated development environment for Google's Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development[1]. In its current version, the user can interact with the application and make use of the functionalities implemented in the following two core *Android activities*: the *Home activity* and the *Recording activity*. An Android activity is the technical term used amongst the developer community to refer to a screen of the user interface of any given Android application.

4.3.1 Operational states of the application

The DN Rotation Tracking System can be started by the user after the application is successfully installed on the device. Figure 4.1 shows the operational state diagram of the applicatoin and how the different state signals sent by the Android operating system, whether or not resulting from user interaction, may change the application state to either *App Suspended*, *App Activated* or *Device Switched off*.

Figure 4.1 shows that upon starting the application, an activation signal is sent by the Android operating system, which sets the application state as activated (i.e. App Activated state). The normal operation of the application occurs during the App Activated state. For example, it is during the App Activated state that the Home activity provides the user with information about the required hardware and software requirements to perform the data

[1]Available at: https://developer.android.com/studio

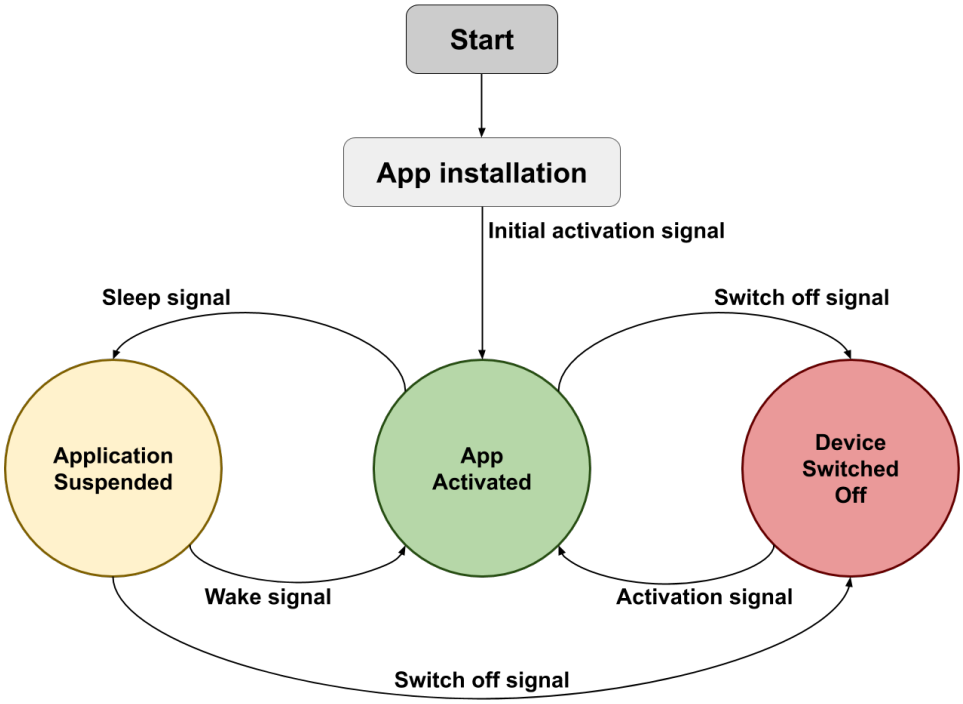


Figure 4.1: DN Rotation Tracking System Operational State Diagram. The diagram shows the different operational states of the DN Rotation Tracking System application and how the different signals sent by the Android Operating System, whether or not resulted by user interaction, may change the application state to either *App Suspended*, *App Activated* and *Device Switched off*.

collection. It is also during the App Activated state that the Recording activity handles the sensor data. Moreover, whenever the App Activated state is enabled, e.g. when the application is started by the user, or when the application returns from either of the other two operating states, the Android operating system state signals of each respective case are used to trigger the revalidation of the required hardware and software requirements. The App Activated state may be changed to either the App Suspended state of operation, if the application is suspended by the user, or to the Device Switched off state, if the device is switched off. In both cases, each respective state signal sent by the Android operating system is used to force the application to complete any eventual pending operations (e.g.

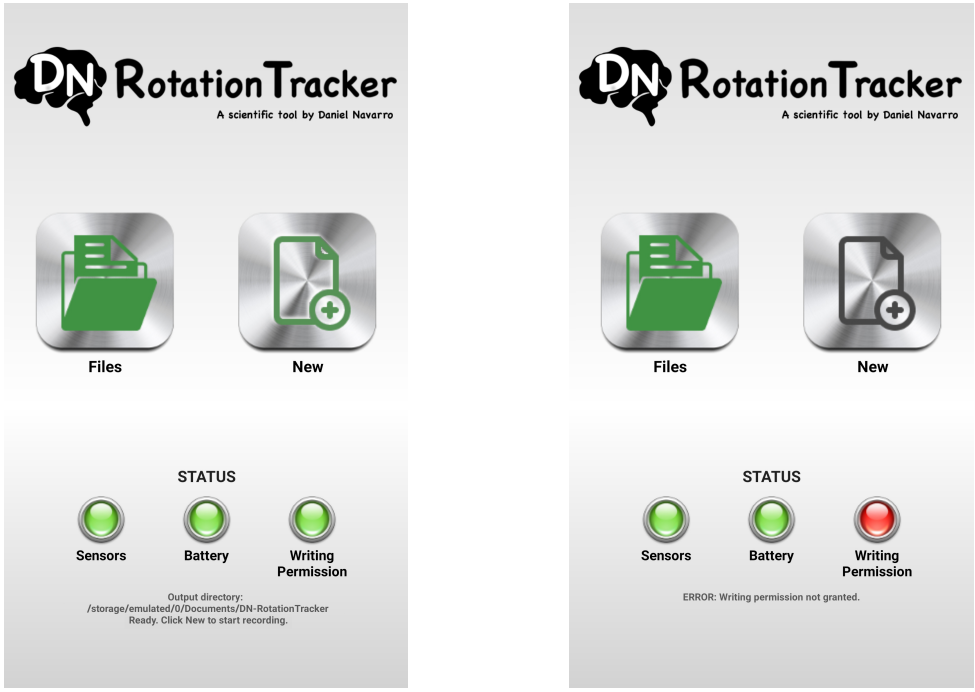
finish writing the data to the device, if necessary), and to release all sensors that might be in use by the application.

4.3.2 User interactive components and software functionalities

4.3.2.1 Home activity

Once the DN Rotation Tracking System is started and successfully loaded, the user is presented with the Home activity screen. Figure 4.2 shows an example of the Home activity screen.

Once the DN Rotation Tracking System is started and successfully loaded, the user is presented with the Home activity screen. The Home activity screen can be described in terms of the following two main segments (figure 4.2). The first segment provides the user with one action button that opens the default file manager application of the device (Files) and one action button that starts the Recording activity to start a new recording (New). The second segment (Status) displays and updates in real time both graphical and textual information about the hardware and software requirements for the successful use of the application. In its current version, the DN Rotation Tracking System has two hardware requirements and one software requirement that need to be satisfied before the button to start the Recording activity is enabled. Namely, the first hardware requirement of the application is the availability of both a magnetic field sensor and a accelerometer sensor. Such sensors are perhaps the most common smartphone sensors. If the DN Rotation Tracking System cannot access both of these sensors, then (i) the relevant textual information is displayed on the Status segment of the Home activity and (ii) the colour of the respective graphical information icon (Sensors) is turned to red. The second hardware requirement is the battery level, which needs to be above 20%. This is



(A) Home screen with no error. **(B)** Home screen with error.

Figure 4.2: Examples of the Home activity screen. The figure shows the two main segments of the Home activity screen. The first segment (middle of the screen) provides the user with one action button that opens the default file manager application of the device (Files) and one action button that starts the Recording activity to begin a new recording (New). The second segment (Status) displays and updates in real time both graphical and textual information about the status of the fulfilment of the hardware and software requirements for the successful use of the application. Subfigure 4.2**A** shows an example of the Home activity screen when all hardware and software requirements are fulfilled. In this case the action button that opens the Recording activity to start a new recording (New) is enabled and the relevant graphical and textual information are displayed on the Status segment of the Home activity screen. Subfigure 4.2**B** shows that if at least one of the hardware or software requirements are not fulfilled, then the action button that starts the Recording activity to start a new recording (New) is disabled whilst relevant graphical and textual description of the error are displayed on the Status segment of the Home activity screen.

not an unusual battery level requirement. In fact, the most recent versions of the Android operating system will display a low battery warning when the battery level reaches 15 or 20%. The requirement of a minimum battery level was included in this version of the application after consideration of feedback of users of an earlier version of the DN

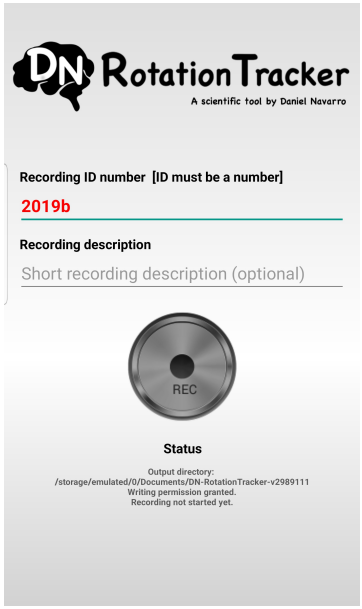
Rotation Tracking System. This particular hardware requirement is relevant because the continous real-time collection and processing of sensor data may be very energy-consuming, and therefore a minimum of battery is required before the button to start the Recording activity is enabled. If the battery level is lower than 20%, then the Status segment of the Home activity is updated with (i) the relevant textual information and (ii) the colour of the respective graphical information icon (Battery) is turned to red. The application is always informed of the battery level by the Android operating system. As a helpful warning for the user, the colour of the Battery icon changes to yellow when the battery level drops below 40%, accompanied by the relevant textual information suggesting that the device may need to be recharged soon.

Finally, the current version of the DN Rotation Tracking System has only one software requirement: the application needs reading and writing storage permission before the action button that starts the Recording activity (New) is enabled. The reason for this software requirement is that without storage permission the application would not be able to save any of the data that would be collected and processed if the Recording activity was started. Subfigure 4.2**A** shows an example of the Home activity screen when all hardware and software requirements are fulfilled. In this case the action button that starts the Recording activity (New) is enabled. Subfigure 4.2**B** shows that if at least one of the hardware or software requirements are not fulfilled, then the action button that starts the Recording activity (New) is disabled whilst relevant graphical and textual description of the error are displayed on the Status segment of the Home activity screen. The hardware and software requirements are re-verified whenever the state of the application becomes active (figure 4.1).

4.3.2.2 Recording activity

With all hardware and software requirements satisfied, the Recording activity can be started by the user from the Home screen activity (subsection 4.3.2.1). Two input fields are located on the top of the Recording activity screen. The main recording button is located between these two input fields and the Status segment on the bottom of the Recording activity screen. Figure 4.3 shows examples of the Recording activity screen.

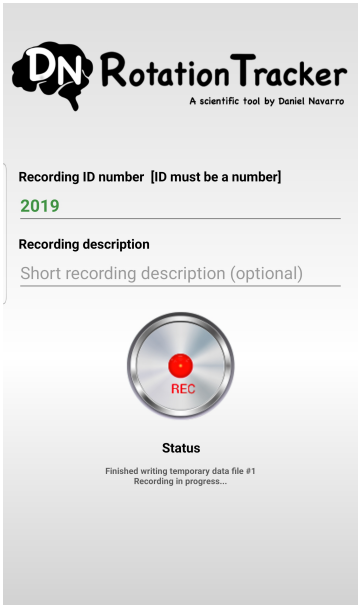
With all hardware and software requirements satisfied, the Recording activity can be started by the user from the Home screen activity (subsection 4.3.2.1). Two input fields are located at the top of the Recording activity screen. The main recording button is located between these two input fields and the Status segment on the bottom of the Recording activity screen. Figure 4.3 shows examples of the Recording activity screen. The first input field asks the user to insert the *recording ID*. The recording ID needs to be a numeric sequence that uniquely identifies the current recording session. The second input field allows the user to optionally insert a textual description of the recording session. The recording ID is validated in real time, and both graphical and textual information is displayed to the user if the recording ID is invalid. A recording ID is invalid if a non-numeric character is inserted or if the ID has already been used. Subfigure 4.3**A** shows an example of the Recording activity screen when an invalid recording ID is inserted (in this case, 2019*b*). The recording ID 2019*b* is invalid because it includes the non-numeric character *b* after the numeric sequence 2019. In this case, the font colour of the recording ID turns red and the textual message "*ID must be a number*" is displayed. Subfigure 4.3**B** shows that after deleting the character *b* from 2019*b*, the font colour of the new recording ID 2019 becomes green. This means that the numeric recording ID 2019 has not yet been used and



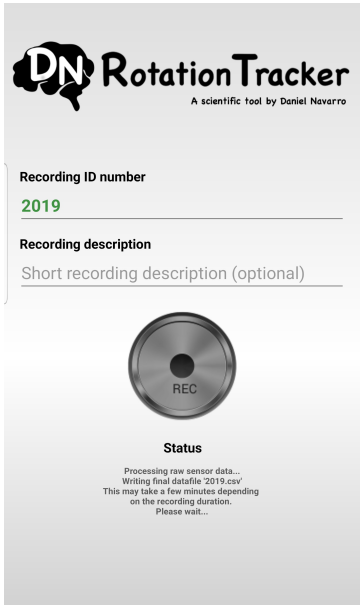
(A) Invalid recording ID.



(B) Valid recording ID.



(C) Recording in progress.



(D) Recording finished.

Figure 4.3: Examples of the Recording activity screen. Subfigure 4.3**A** and subfigure 4.3**B** show, respectively, the Recording activity screen with an invalid and with a valid recording ID. Subfigure 4.3**C** shows the Recording activity screen when recording is in progress. Finally, subfigure 4.3**D** shows the Recording activity screen when the user stops the recording.

is, therefore, successfully validated.

The comparison of subfigure 4.3**A** and subfigure 4.3**B** shows that the recording button is initially disabled when then Recording activity screen is started or when the recording ID is not successfully validated. In other words, the recording button is enabled only when the inserted recording ID is valid. The textual description of the recording session is optional and, therefore, is not validated by the application and does not have any effect on whether or not the recording button is enabled or disabled. Once the recording ID is validated, a new recording starts when the user activates the enabled recording button (subfigure 4.3**B**). Subfigure 4.3**C** shows that when recording is in progress the recording button lights up and the Status segment on the bottom of the Recording activity screen actively displays information about the tasks being performed by the application during the recording session. The recording session is finished when the user deactivates the recording button. Subfigure 4.3**D** shows that at the end of the recording session the recording button is disabled and a textual message on the Status segment of the Recording activity screen requests the user to wait patiently until the Home activity screen is automatically loaded after the application performs the final processing of the sensor data that has been collected.

4.3.3 Recording rotation movements

4.3.3.1 Defining rotation movements

In its current version, the DN Rotation Tracking System records rotation movements based on the orientation of the mobile device. Figure 4.4 shows the orientation of a mobile defined as a combination of the following three angular quantities: azimuth, pitch, and roll. Azimuth is the angle between the positive x-axis and magnetic north defined within

$[0°, 360°]$. Pitch is the tilted angle between the z-axis and the y-axis. Finally, roll is the tilted angle between the z-axis and the x-axis.

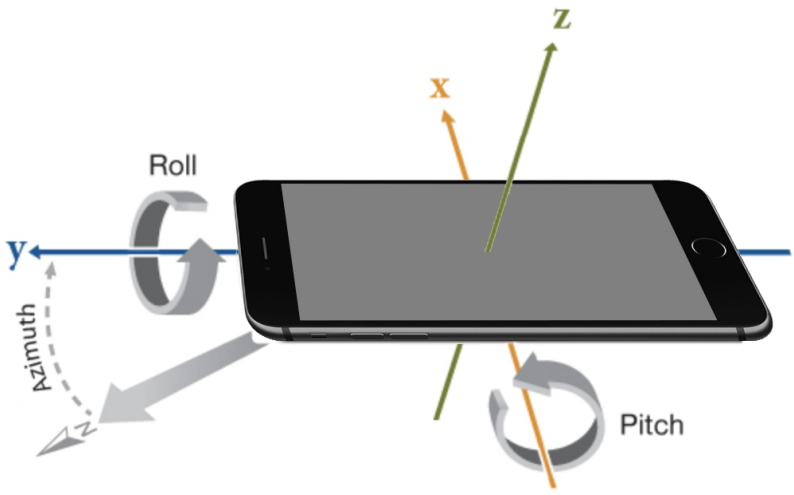


Figure 4.4: Orientation of a mobile device. The figure shows the orientation of a mobile device defined as combination of angular quantities azimuth, pitch, and roll. In its current version, the DN Rotation Tracking System records rotation movements across the x-axis (azimuth) and across the y-axis (pitch).

In its current version, the application collects and combines data from the magnectic field sensor and the accelerometer sensor to record rotation movements *in degrees* across the x-axis (azimuth) and across the y-axis (pitch) (figure 4.4). The initial position of the device when the recording button is activated by the user (subsection 4.3.2.2) is defined as the initial position (i.e. 0°) for both azimuth and pitch. Rotation movements across the x-axis are positive for right movements and negative for left movements. Rotation movements across the y-axis are positive for movements heading up and negative for movements heading down.

4.3.3.2 The challenge of continuously recording real-time sensor data

Continuously collecting and recording real time data from smartphone sensors requires particular attention to the use of the available mobile computing resources, such as CPU time and RAM memory. For this reason it is strongly recommended that precautionary actions are taken to prevent disruption to the application during a recording session. Examples of such precautionary actions would be (i) disabling notifications, (ii) disabling alarms and (iii) activating *airplane mode* when the smartphone is used as a recording device. The DN Rotation Tracking System efficiently collects real-time sensor data by implementing a combination of a *dual-buffer system* with extended *Android AsyncTask*[2] functionalities. The dual-buffer system efficiently uses both CPU time and RAM memory by immediately saving the raw sensor data (i.e. movement and timestamp data read from the sensors without any processing) into a temporary buffer loaded in memory. The buffer size is appropriately chosen to assure that only a small portion of the available RAM memory is used at all times by the application when recording is in progress. As soon as the buffer is full, the following two actions happen simultaneously: (i) a new independent AsyncTask process is started to write the buffered data to the disk in a temporary text file and a (ii) new empty buffer is used to store the most recent data read from the sensors. The use of independent AsyncTask processes allows the application to continously and efficiently collect and record sensor data. This means that the maximum duration of any given recording session depends only on the amount of physical storage space available on the device and on the amount of battery time offered by the device.

[2]More information about the Android AsyncTask is available at: https://developer.android.com/reference/android/os/AsyncTask.

2019.csv

```
** CREATION DATE: Thu 10 Sep 2019 02:58:45 BST
** FILE TYPE: CSV
** APPLICATION VERSION: DN ROTATION TRACKER v2.0
** DEVICE NAME:  SAMSUNG
** DEVICE MODEL: SM-G965F
** RECORDING ID: 2019
** RECORDING DESCRIPTION: NO RECORDING DESCRIPTION WAS PROVIDED.


00,0.000000,0.000000
02,0.009301,0.001001
04,0.973266,0.977954
06,0.970279,0.994483
08,0.967239,0.998050
09,0.972278,0.919161
10,0.970146,0.950489
12,0.975730,0.953042
14,0.967109,0.850533
16,0.964618,0.714889
```

Figure 4.5: Example of a CSV file created by the DN Rotation Tracking System. The final output file created after a recording session with the DN Rotation Tracking System starts with a header and is followed by the recorded data. The header includes the date of the recording, the file type, the name and version of the application, the name and the model of the device, the recording ID and the optional recording description provided by the user. The recorded data saved on the CSV file consists of timestamps in milliseconds and rotation movements in degrees.

Although the recorded data is saved as one single text file when the recording is complete (figure 4.5), several sequencial temporary text files with the buffered data may be created during a recording session. These temporary files are processed by the application and automatically deleted at the end of the recording session. The recording session is finished when the user deactivates the recording button. When the recording button is deactivated, the application starts one final independent AsyncTask process to read the raw sensor data saved on the temporary text files and to create the final output file (figure 4.5). The buffered timestamps are converted to milliseconds and the recorded rotation movements to are converted to degrees. One single comma-separated values (CSV) file is created with the processed data (i.e. timestamps and rotation movement data). All temporary files are

deleted after processing is complete. Finally, a header is included to the CSV file with the following information about the recording session: (i) the date of the recording, (ii) the file type (i.e. CSV), (iii) the name and version of the application, (iv) the name of the device, (v) the model of the device, (vi) the recording ID and finally (vii) the recording description.

4.4 Conclusion

The DN Rotation Tracking System is the first open source Android application designed as a scientific tool to record human movement data. In particular, the current version of the DN Rotation Tracking System collects and combines data from smartphone sensors to record rotation movements. The form of rotation movement recorded by the DN Rotation Tracking System is investigated by a number of earlier and more recent experimental studies Bartz (1966), David et al. (2018), Minati et al. (2016). Clear practical and financial challenges of recording this form of behavioural data have been observed in these previous studies. However, recent technological developments and the increasing number of enhanced sensors in modern smartphones have turned these devices into powerful computing machines. Although smartphone sensors have been used in previous studies Coskun et al. (2015), Novakova and Pavlis (2017), the potential to use such sensors to collect data for scientific purposes has not yet been fully explored. In most cases, the development of such applications is sufficiently difficult to be beyond the ability of a typical researcher. The DN Rotation Tracking System for Android is a scientifc tool created to overcome the challenge of using Android enabled mobile devices (e.g. smartphones) to record rotation movements. The DN Rotation Tracking System allows researchers to easily incorporate this form of

behavioural data to newer studies by using smartphone devices. The application uses a combination of a dual-buffer system with Android AsyncTask functionalities to record real-time rotation movements in a cost-effective manner. At the end of each recording session the recorded data are saved as a CSV file to facilitate the immediate analysis of the data by the researcher. Processing of text files, particularly CSV files, is common amongst the scientific community. Additional information is added as a header to each CSV file created at the end of a recording session (figure 4.5).

Moreover the data collected by the DN Rotation Tracking System has been successfully used in a recent computational study to train a biological neural network model of coordinate transformation performed by the visual brain discussed in the next chapter. This study showed that the natural statistics of the head movements recorded by the application, combined with eye movement data simultaneously recorded with the Eyelink II System (SR ResearchLtd., Ontario, Canada), lead the neural network model to successfully learn to perform a coordinate transformation of visual responses from an eye-centred to a head-centred frame of reference.

The DN Rotation Tracking System was developed with an user-friendly interface to record behavioural data, in particular rotation movements, and can be used in its current version in different research areas. The positive feedback from testing users during the development of the application and the successful use of the application in the computational study discussed in the next chapter encourages the adoption of the DN Rotation Tracking System in other projects. Furthermore, due to its modular design, new functionalities can be easily added to future versions of the application. In particular, the dual-buffer system combined with Android AsyncTask functionalities implemented in the current version of

the DN Rotation Tracking System can be used to extend the data collection capabilities of the application. Moreover, the application provides the user with the contact details of the developer to facilitate the request of new features and to instigate the establishment of future collaboration with other researchers within the scientific community.

Chapter 5

Self-organisation of head-centred responses using eye and head movements recorded from human participants

5.1 Introduction

In this chapter we show how head-centred representations of visual targets may develop through visually guided learning when the model is trained with eye and head-movements recorded from human participants during a visual search task. We have hypothesised that if a primate performs eye movements more frequently than head movements to adjust its gaze, then retinal images that occur close together in time will tend to correspond to different positions of the eyes, but the same position of the head. The simulation results in chapter 3 showed that artificially created eye and head-movements with these properties allow the trace learning to encourage output neurons in our model to learn to respond to the position of a visual target in the same head-centred location across different retinal positions. The main motivation of the present study is to investigate the experimental prediction in chapter 3 of whether the statistics of eye and head movements recorded from

human participants during a natural visual search task would be appropriate to drive the development of head-centred output visual neurons in our model. Our experimental task required human participants to explore a series of visual scenes with freedom of eye and head movements. The eye and head movements of each participant during the task were recorded using eye-tracking and head-tracking devices. The combination of recorded eye and head movement was used to train our neural network model using the *delayed-trace learning rule* investigated in chapter 3 in a more biologically realistic training condition consisting of a mixture of peaked and monotonic modulated visual input neurons.

In chapter 3 we showed how our neural network mode used the standard trace learning rule (equation 2.9) to robustly self-organise head-centred output responses when all visual input neurons had peaked modulation by eye-position, but would completely fail when the modulation of the input population was changed from peaked to monotonic. A mathematical analysis of how the change in the form of modulation of the input neurons would affect the self-organisation of the model led us to introduce in chapter 3 a number of new, more powerful forms of the standard trace learning. These new biologically plausible learning rules enabled the model to develop head-centred output neurons when the input neurons were modulated by a monotonic function of eye-position. Moreover, we showed how the neural network model self-organised head-centred responses using the particular modified version of the standard trace learning rule in equation 3.17, the delayed-trace learning rule, under a more biologically realistic scenario where the input population had a 50:50 mix of peaked and monotonic eye-position gain fields.

The simulation results discussed in this chapter show that the recorded eye and head-movements from human participants during a visual search task successfully drives our

artificial neural network model to self-organise head-centred visual representations in the presence of both peaked and monotonic modulated input neurons and thus provide an important contribution to understanding how head-centred visual neurons may develop in the brain.

5.2 Materials and Methods

5.2.1 Experimental task

5.2.1.1 Participants

A total of forty human participants aged 18 or over years (21 males and 19 females) took part in this study. All participants had normal or corrected to normal vision.

All participants gave written informed consent and the study procedures were approved by the Medical Sciences Interdivisional Research Ethics Committee of the University of Oxford (Ref: R53648/RE001). This study was conducted in accordance with the Declaration of Helsinki (Code of Ethics of the World Medical Association).

5.2.1.2 Eye and head-movement recording

Eye movements were recorded using the HREF (head-referenced) data recording option of the Eyelink II System (SR Research Ltd., Ontario, Canada) at a sampling rate of $500Hz$. The EyeLink II proprietary algorithm was used to calibrate the eye-tracking system for each participant. Data were only recorded when the level of calibration accuracy was flagged by the Eye Link II calibration software as *good*. The HREF data recording option of the Eyelink II System creates a pair of HREF coordinates that defines a point in a plane at a distance d_0 of 15000 arbitrary units from the eye (figure 5.1) (SR Research Ltd., 2009).

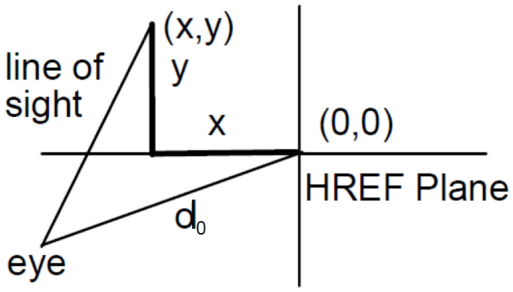


Figure 5.1: Example of how HREF coordinates may be interpreted. HREF coordinates define points in a plane at a distance d_0 of 15000 arbitrary units from the eye (SR Research Ltd., 2009).

In particular, head-movements do not affect HREF coordinates and HREF coordinates do not directly relate to the gaze position with respect to the visual stimuli. Instead, the origin point of the HREF coordinate system is arbitrary and the difference between any two subsequent coordinate pairs measures their relative distance. Therefore, the angle ψ *in degrees* between two HREF coordinate pairs (x_1, y_1) and (x_2, y_2) is given by

$$\psi = c \times \arccos\left(\frac{d_0^2 + x_1x_2 + y_1y_2}{\sqrt{(d_0^2 + x_1x_1 + y_1y_1)(d_0^2 + x_2x_2 + y_2y_2)}}\right) \tag{5.1}$$

where c is a constant set to 57.296, used to convert the value of the arccos from *radians* to *degrees* (SR Research Ltd., 2009).

With ψ defined by equation 5.1, the rotation of the eye (α_x) in degrees along the x dimension is given by

$$\alpha_x = \psi \times \sin\left(\arctan\frac{(x_2 - x_1)}{(y_2 - y_1)}\right) \tag{5.2}$$

where (x_1, y_1) and (x_2, y_2) are two subsequent HREF coordinate pairs.

Similarly, the rotation of the eye (α_y) in degrees along the y dimension is given by

$$\alpha_y = \psi \times \cos\left(\arctan\frac{(x_2 - x_1)}{(y_2 - y_1)}\right) \tag{5.3}$$

where (x_1, y_1) and (x_2, y_2) are two subsequent HREF coordinate pairs and ψ is defined by equation 5.1 (SR Research Ltd., 2009).

Head movements with respect to the body were recorded using the DN Rotation Tracking System described in chapter 4 installed on a smartphone (Samsung S7 Edge SM-G935F). The smartphone was positioned and fixed on top of the head of each participant, together with the head-mounted eye-tracker Eye-Link II. Additionally, an application using the Microsoft Kinect sensor (*DN Head-Tracker*) was also developed by the author to validate the head-tracking data recorded by the DN Rotation Tracking System.

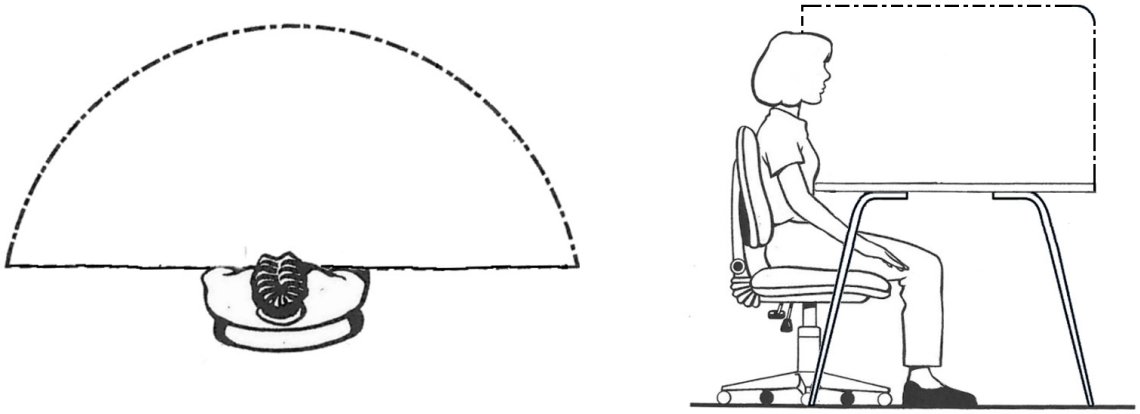
5.2.1.3 Stimuli and task

During the visual search task in our study we used the *Where's Wally?©* *The Cake Factory* scene with permission from Walker Books Ltd.©. The *Where's Wally?©* *The Cake Factory* scene was originally published in 1997 on *Where's Wally: The Wonder Book by Martin Handford*. This *Where's Wally?©* scene allowed for complex visual search instructions and that the small size of individual features of the image meant that foveating eye-movements would be required. Participants were not previously familiar with the *Where's Wally?©* *The Cake Factory* scene.

5.2.1.4 Procedure

The experimental setup is shown in figure 5.2. The *Where's Wally?©* scene was fixed on a semi circular table frame and positioned around each participant. The dimensions of the *Where's Wally?©* poster was approximately $220cm$ wide by $85cm$ height. Figure 5.2**A** shows the poster fixed (broken line) on the semi circular support. Participants sat on an adjustable office chair during our experimental task. Figure 5.2**B** shows the office chair

positioned with the semi circular poster in our experimental setup.



(A) Semi circular presentation of the poster. **(B)** Adjustable chair and visual scene.

Figure 5.2: Experimental setup. Subfigure 5.2**A** shows how the *Where's Wally?*© poster (broken line) is fixed to a semi circular support and positioned around the participant. Subfigure 5.2**B** shows the experimental setup with the poster (broken line) and the adjustable office chair used during the visual search task.

The height of the *Where's Wally?*© poster was chosen in such a way that by setting the height of the office chair according to the height of each participant the need for head movements along the y dimension would be minimised, or completely eliminated. Visual inspection during each recording section showed that head movements along the y dimension were minimum. The reasons for the adjustment of the office chair to each participant are twofold. First, the participants should feel comfortable during the data recording session to ensure that the recorded data correspond to natural movements during the visual search task. Second, the previous model in chapter 3 was trained on head rotation along the x dimension, and therefore, this was the movement that we used in this study. The use of movements along the x dimension allow the changes in the simulation results discussed in this chapter to reflect the use of natural recorded data when compared to the simulation results previously discussed in chapter 3. The trunk of the participant

was always in contact with the support holding the *Where's Wally?*© poster to prevent trunk rotation during the recording session. All participants were instructed not to rotate their body or move the office chair during the data recording session, but to use only eye and head-movements to visually explore the *Where's Wally?*© poster during the visual search task. Participants, however, were not instructed to move their eyes or head in any particular manner.

Participants performed a series of small visual search tasks during a short period of time (approximately $10-15$min). Such series of visual search tasks consisted of (i) finding *Wally*©, (ii) counting the number of characters wearing hats, (iii) counting the number of characters wearing yellow shirts and (iv) counting the number of characters wearing blue shirts. The main purpose of these simple visual search tasks was to encourage the participants to keep actively exploring the visual scene during the recording session. Participants were informed that they were free to keep exploring the poster until notified that the data recording session has ended. At the end of the data recording session text files with the eye-movement and head-movement data were created and saved, respectively, by the eye-tracking and head-tracking systems described in previous session.

5.2.1.5 Physical properties of the recorded eye and head movement data

This subsection describes some of the physical properties of the recorded eye and head movement data[1]. The underlying properties of the recorded eye and head movement data used to train the neural network model described in the next section will have had an impact on how the model learned to self-organise head-centred visual output representations.

[1]The work described in this subsection was carried out in collaboration with Anna-Katharina Hauperich, who contributed figure 5.3, figure 5.4 and figure 5.5.

Therefore it is of interest to examine the properties of the eye and head movements contained that were recorded from human participants during the visual search task previously described in chapter 3. The properties of the recorded data presented in this section are particularly relevant, for example, for comparing the performance of the neural network model when trained with new recorded data in the future, or to provide information for modelling data similar to the set used here.

The eye and head movement data were analysed for the occurrence of saccades and fast head movements and the spread of visual exploration of the scene. Both eye and head movements were first smoothed with a Savitzky-Golay filter (order: 3, frame length: $22ms$) to improve saccade detection. Saccades in the eye movement data were extracted using a detection method by Engbert and Mergenthaler (2006), while any head movement with a velocity greater than one standard deviation above the mean was considered a *fast head movement*. To evaluate if saccade detection was adequate, the characteristically linear main sequence relationship between peak velocity and amplitude, as described by Zuber and Stark (1965), is plotted in figure 5.3.

Saccade distance is given as the diagonal of the smallest rectangle covering all of the movement whilst the peak velocity was calculated using the first derivative of the smoothed position as given by the Savitzky-Golay-filter. The mean size of saccades across participants was 7.5 degrees.

We also provide three examples of different exploration patterns in the form of heat maps (see figure 5.4). The heat maps chosen in figure 5.4 is a perceptual colourmap distributed by Kovesi (2015) under the name CET-C2[2]. These were created by adding the

[2]Available at: https://peterkovesi.com/projects/colourmaps/

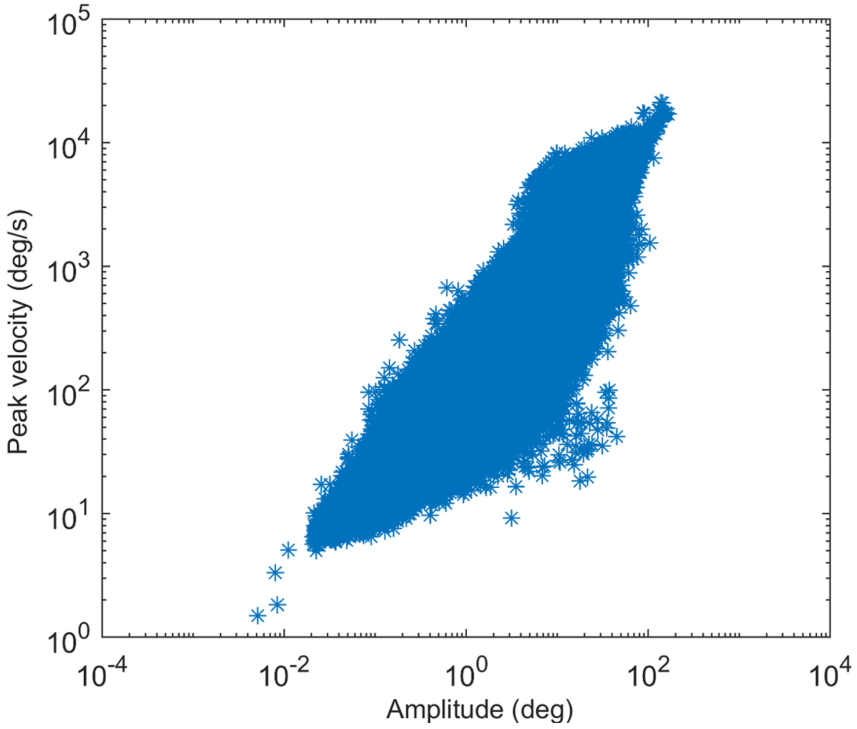
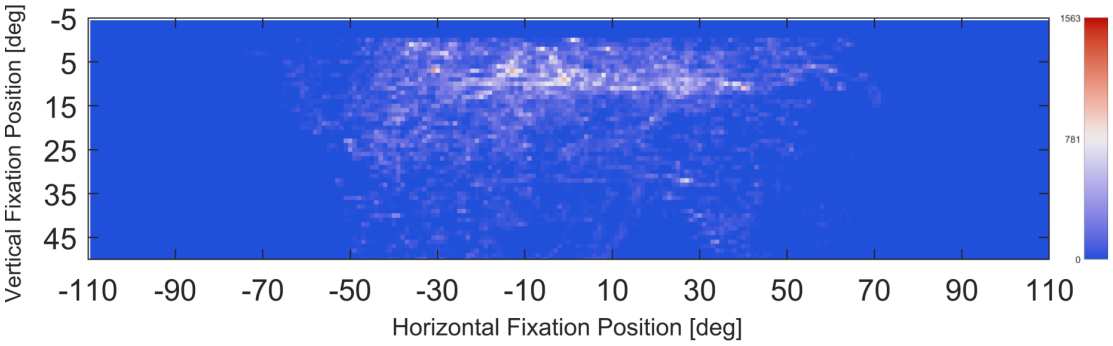
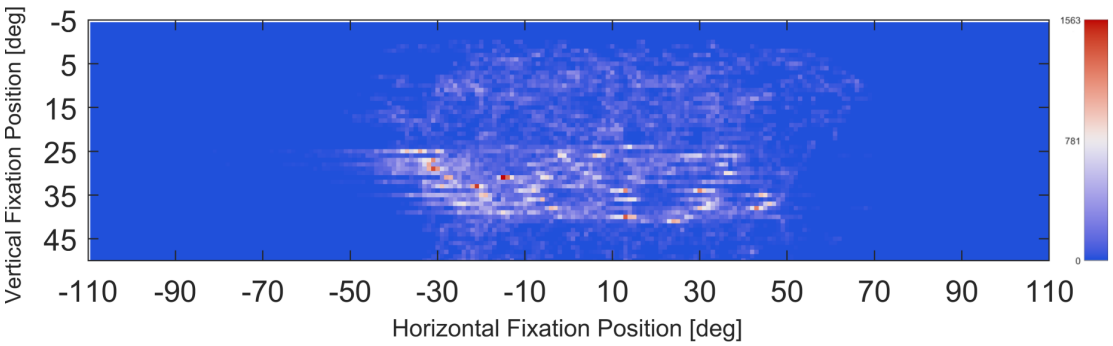


Figure 5.3: Main sequence for the recorded saccades across all participants. The plot shows the relationship between the peak saccade velocity and the saccade amplitude. The saccade amplitude is defined as the diagonal of the smallest rectangle covering all of the movement. The peak velocity is calculated using the first derivative of the smoothed position as given by the Savitzky-Golay-filter. The mean size of saccades across participants was 7.5 degrees.

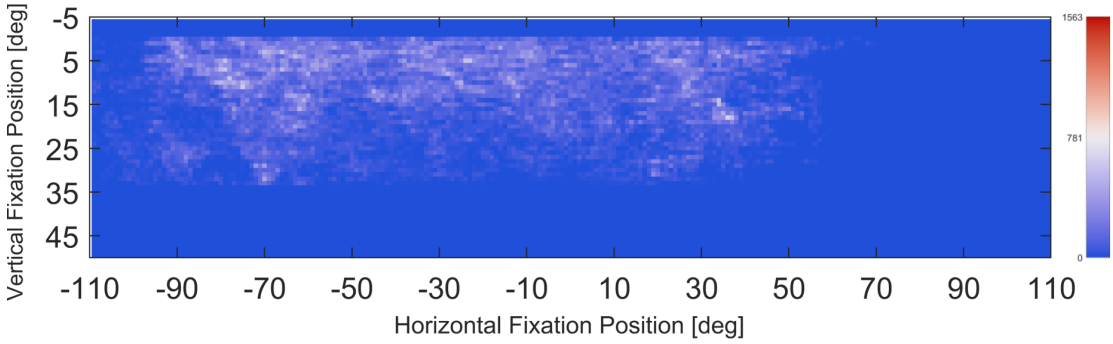
horizontal head movements to the eye movement traces. Figure 5.4 shows the position of fixation points within the visual space being viewed by each participant during the visual search task. The scale of the heat map is adjusted according to the participant with the highest number of fixations within one square degree amongst the three participants SA, MA and AN (i.e 1563 for participant MA). This strategy of scaling the heat map allows, for example, a comparison of the differences in terms of how much these three participants returned their gaze to the same location of the visual scene. All participants followed individual exploration patterns during the visual search task, but extensive exploration of the scene is evident in all of them.



(A) Participant SA.



(B) Participant MA.



(C) Participant AN.

Figure 5.4: Heat maps showing three examples of different exploration patterns during the visual search task. Subfigure 5.4**A**, subfigure 5.4**B** and subfigure 5.4**C** show, respectively, the heat maps of the pattern of exploration performed by participant SA, participant MA and participant AN. These subfigures show that the strategy of exploration of the visual scene varied across different human participants during the visual search task. Extensive exploration of the visual scene was observed for all human participants.

To compare the relative contribution of eye vs head movements to fast shifts in the visual image, the frequencies of saccades and fast head movements are shown in figure 5.5. As is to be expected for a visual scene of the size and detail used for the visual search task, the exploratory saccades are more frequent than fast exploratory head movements. This is the key behaviour required by our hypothesis for how head-centred visual neurons may develop in the brain and in the neural network models presented next.

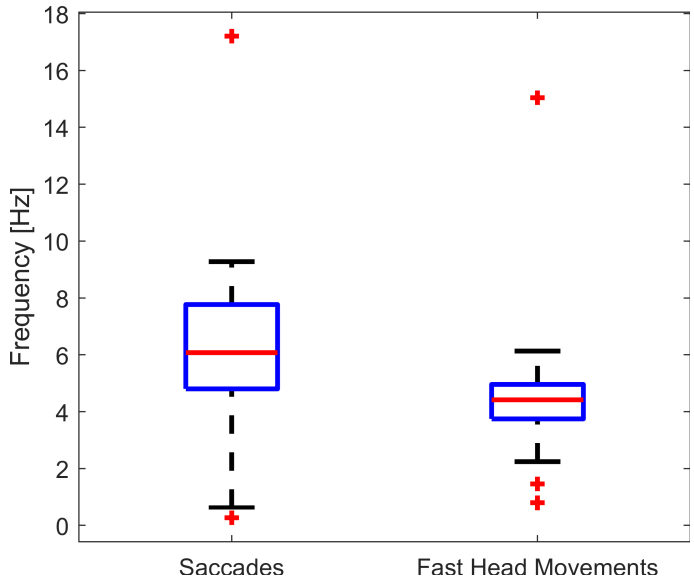


Figure 5.5: Box plots of the frequencies of saccades and fast head movements. These box plots show the frequencies of saccades and fast head movements across all human participants during the visual search task.

5.3 Network Architecture

The architecture of the neural network model consists of one layer of input neurons and one layer of output neurons, as presented in chapter 3 (figure 3.1). Neurons in the input layer simultaneously encode the position of the eye of the agent and the retinal location of the visual target. These visual input neurons are modelled as retinotopic neurons with eye-

position gain modulation. Each neuron in the input population is set to respond maximally to a unique combination of eye-position and retinal location of the visual target. The spaces of preferred eye and retinal position represent, respectively, the range $[-30°, 30°]$ of eye-positions in orbit and the range $[-90°, 90°]$ of retinal target locations. Feedforward synaptic connections project from neurons in the input layer to neurons in the output layer. A total of $N = 900$ neurons in the output layer competed to represent patterns in the input layer (Rolls and Treves, 1998). Output neurons received the same number of afferent connections from a random subset of input neurons. Specifically, each output neuron received synaptic connections from a random sampling of $\phi = 5\%$ of input cells.

5.3.0.1 Input Layer

There are $N_I = 24552$ neurons in the input layer. The input population had a $50 : 50$ mixture of neurons with response functions modulated by either *peaked* eye-position gain fields, which has been reported in cortical area PO by Galletti et al. (1995), or *monotonic* eye-position gain fields, which has been observed in multiple visual areas (Andersen et al., 1990, 1985). Figure 3.2 shows an example of both peaked and monotonic modulated response functions. The response function used to model the instantaneous firing rate v_j^I of each input neuron j encodes the current retinotopic location of the visual target. Such response function is modulated by the position of the eye. The range of firing rate values is defined within $[0, 1]$. The response function modelling the instantaneous firing rate v_j^I of each input neuron j modulated by peaked eye-position gain field is described by equation 3.2.

5.3.0.2 Output Layer

There are 900 neurons in the output population. The dynamics of each neuron i in the competitive output layer was modelled by the following three quantities: an internal activation h_i (equation 3.3), an instantaneous firing rate v_i^I (equation 2.8), and a memory trace value q_i (equation 3.4).

5.3.0.3 Modification of the Synaptic Weights

Before each simulation the strengths of the feedforward synaptic connections are initialised to random weights in the interval $[0, 1]$. Then the synaptic weight vector of each output neuron is renormalised as is typical in competitive networks (Rolls and Treves, 1998).

During learning the synaptic weights are modified by either the standard Hebbian learning rule (equation 2.9) or the associative trace learning rule defined by equation 3.17. The trace learning rule for synaptic modification encourages output neurons to bind together subsets of input patterns that tend to occur close together in time by incorporating a temporal trace of recent neuronal activity. After each weight update during training using the standard Hebbian learning rule, the length of the weight vector for each output neuron is renormalised according to equation 2.10. This renormalisation prevents unbounded growth of the synaptic weights during training (Dayan and Abbott, 2001). The work of Royer and Paré (2003) provides experimental evidence for renormalisation of synaptic weights in the brain.

5.4 Training the Network

The coupled differential equations 3.3, 3.4 and 3.17 are numerically integrated using the Forward-Euler scheme. The time step Δt is set to one tenth of the neuronal time constant τ_h. Linear interpolation was used to compute the numerical inputs to the discretised Forward Euler model equations, which required input values at every numerical time step $\Delta t = \tau_h/10$. We have verified that the simulation results remained similar if the number of training epochs was increased or if the time step was reduced.

Our original self-organising neural network model described in chapter 3 was trained with artificially generated eye and head-movements. In this chapter we investigate whether training our previously published neural network model with natural statistics of eye and head-movements recorded from human participants during a visual search task would drive the development of head-centred output neurons. Training the network consisted of periods of time when a visual target is fixed to a position in the simulated visual environment whilst the network is simulated making the series of eye and head movements recorded from human participants during the experimental visual search task. A total of M evenly spaced training locations in the environment for the visual target were defined within $[-63°, 63°]$. Each training epoch had M of these periods where a visual target was fixed to a different location in the environment and the experimentally recorded data were used to train the model. Each training period individually corresponded to one of the M chosen visual target locations in the environment. Therefore, during training the network was presented with input sequences representing the fixed position of the visual target in the environment whilst the network performed the experimentally recorded sequences of eye and head movements during the visual search task. In the previous simulations discussed

in chapter 3 both the eye and head movements were artificially controlled and biologically unrealistic. For instance, the model in chapter 3 used saccades to 15 different and uniformly sampled eye-positions within $[-24°, 24°]$. The saccades between successive eye-positions were performed at a constant velocity of $400°/s$ and the duration of each fixation was set to $300ms$. In contrast, the training conditions in this chapter are naturally much more challenging due to the biological realism of the experimentally recorded eye and head movements.

5.5 Testing the Network

After training is complete the responses of each output neuron were recorded for all combinations of the following E eye fixation positions and T visual target locations in the environment. The following E eye-positions are used during testing: $-18°, -6°, 6°$ and $18°$. For each of these eye-positions the visual target is shifted in increments of $2°$ every $330ms$ to the next one of the $T = 80$ target locations in the environment within $[-79°, 79°]$. At the end of each fixation period the firing rates of all output neurons are saved and used to analyse some of the receptive field properties of the output neurons, such as reference frame of response. The neural network model is tested with the same combinations of eye-position and visual target location used to test the model in chapter 3 to assess the robustness of the self-organisation process to learn representation of the location of the visual target in novel head-centred locations (i.e. head-centred locations that may not have been presented to the network during training).

5.5.1 Analysis of Network Performance

In order to analyse whether individual output neurons are predominantly responding in an eye-centred or head-centred frame of reference we use the eye-centredness (Ω) and head-centredness (Π) metrics described in chapter 3. The mathematical details of the analysis are presented in section 3.9. Output neurons with $\Omega > 0$ and $\Omega > \Pi$ are classified as eye-centred, and output neurons with $\Pi > 0$ and $\Pi > \Omega$ can be classified as head-centred. If neither of these conditions is met then the neuron remains unclassified. The analysis of the weight vector structure of head-centred neurons presented in chapter 3 with idealised training data revealed the structure of the weight vector for head-centred output neurons when all input neurons were modulated by either monotonic (figure 3.2**B**) or peaked (figure 3.2**A**.) gain fields. The plots in figure 3.7**B** and in figure 3.8**B** show the type weight vector structures that should be expected for a head-centred output neuron after training.

5.6 Results

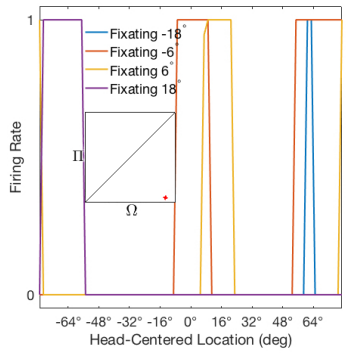
This experiment investigates whether the natural statistics of recorded eye and head-movements from human participants during a visual search task drives the development of head-centred output responses when the input population had a 50:50 mixture of peaked and monotonic gain fields (figure 3.2). The training lasts for 20 epochs. During each training epoch, a visual target is presented for approximately $5s$ in each of the eight training locations in the simulated visual environment: $-63°, -45°, -27°, -9°, 9°, 27°, 45°$ and $63°$. For each period where the visual target is in a fixed location in the environment, the network performed the series of eye and head movements that were experimentally recorded from human participants during the visual search task described in the Materials and

Methods section. The retinal location of the visual target was derived from the recorded position of the eye in the orbit and the position of the head. Each training epoch thus lasts for approximately $40s$, and the entire training of the network is completed after about $800s$ of simulated time. The model is tested as previously described. The parameters for the model simulations are given in table 5.1. This section presents simulation results of models trained with recorded eye and head-movements from human participants SA, MA and AN.

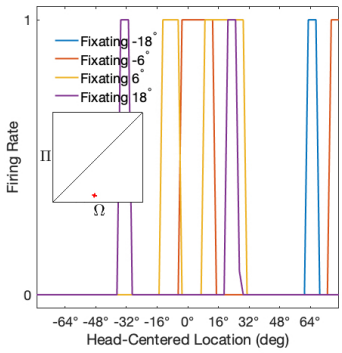
Parameter	**Symbol**	**Value**
Number of target locations	M	8
Number of training epochs	-	20
Width of peaked eye-position retinal tuning curve	ρ	6°
Width of retinal tuning curve	σ	6°
Gain magnitude	$\|\kappa\|$	0.0625
Output neuron population size	N	900
Input neuron population size	N_I	24522
Trace time constant	τ_q	400ms
Activation time constant	τ_h	100ms
Trace delay	ΔT	50ms
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 5.1: Simulation parameters of the self-organising neural network model.

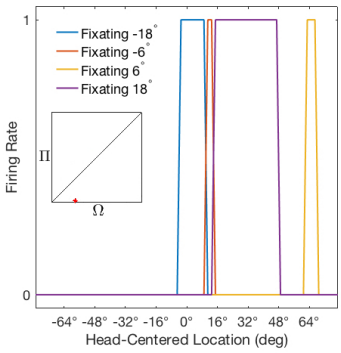
Figure 5.6 compares the firing rate responses before and after training of a typical output neuron in each of these three models. Particularly, figures 5.6**A**, 5.6**B** and 5.6**C** show that the firing rate responses of output neurons in the untrained models SA, MA and AN, respectively, exhibit no consistent structure in head-centred space across the different eye-positions. This reflects the randomly assigned synaptic weights before training. However,



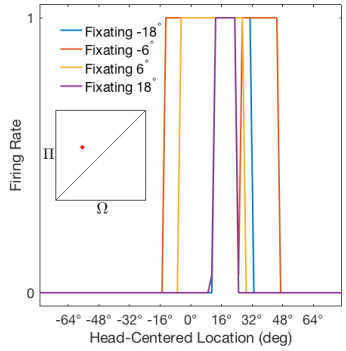
(A) Untrained neuron #618 (model SA).



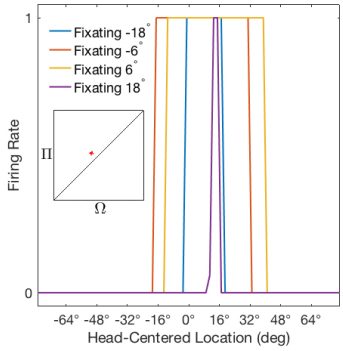
(B) Untrained neuron #771 (model MA).



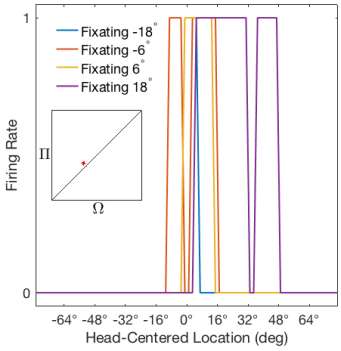
(C) Untrained neuron #279 (model AN).



(D) Trained neuron #618 (model SA).

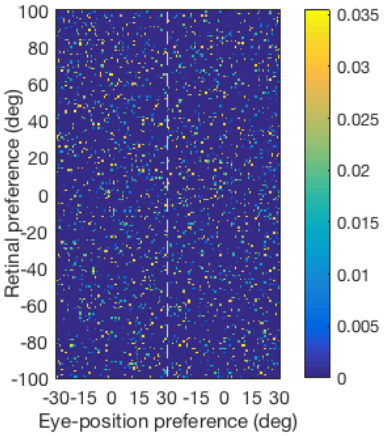


(E) Trained neuron #771 (model MA).

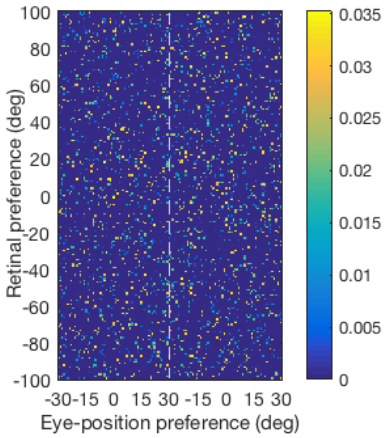


(F) Trained neuron #279 (model AN).

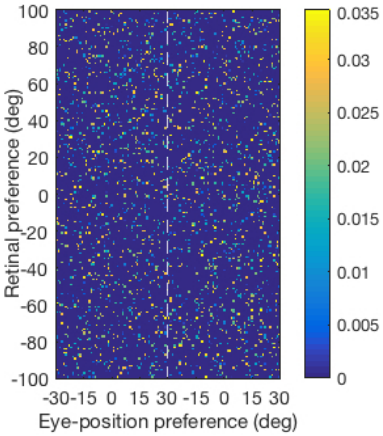
Figure 5.6: Comparison of model performances with a 50:50 mix of peaked and monotonic gain modulated visual input neurons using recorded data from human participants SA, MA and AN. The figure shows the firing rate responses of output neurons of untrained and trained models using the delayed-trace learning rule (equation 3.17), where training was driven using experimentally recorded eye and head movement data from human subjects SA, MA and AN. Specifically, each subplot shows the firing rate responses of typical output neurons during testing for four different eye-positions: $-18°, -6°, 6°$ and $18°$. The top row shows output neuron #618, #771 and #279 from the models before training. The bottom row shows these same three output neurons #618, #771 and #279, but from the model trained using recorded data from participants SA, MA and AN, respectively. It is evident in subplots 5.6**A**, 5.6**B** and 5.6**C** that before training these output neurons responded to different head-centred locations when the eyes shifted to different positions. However, subplots 5.6**D**, 5.6**E** and 5.6**F** show that after training these same output neurons learned to respond more consistently to a localised interval of head-centred locations regardless of the eye-position.



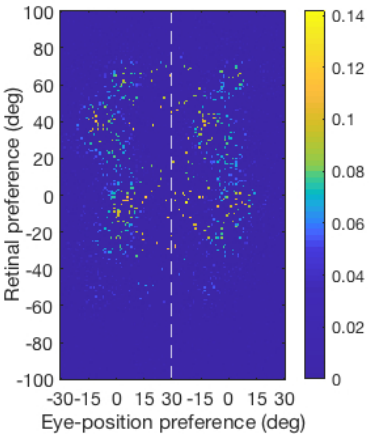
(A) Untrained neuron #618 (model SA).



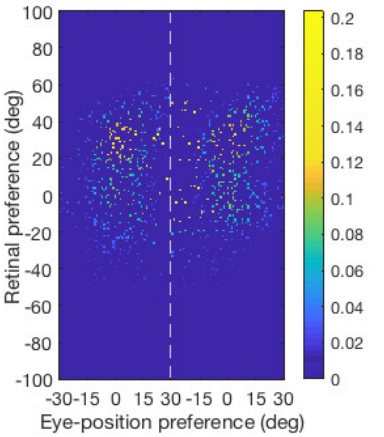
(B) Untrained neuron #771 (model MA).



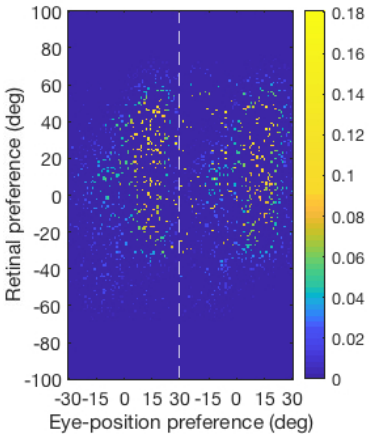
(C) Untrained neuron #279 (model AN).



(D) Trained neuron #618 (model SA).

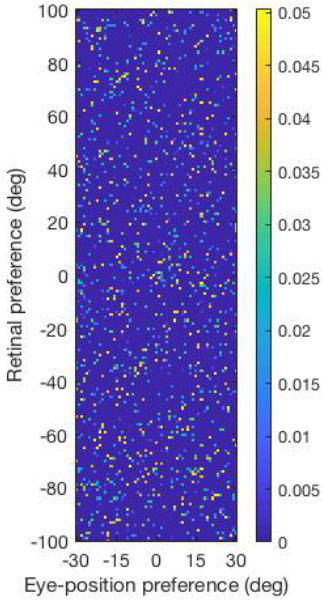


(E) Trained neuron #771 (model MA).

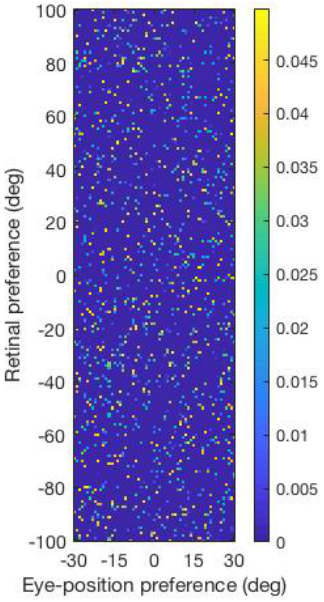


(F) Trained neuron #279 (model AN).

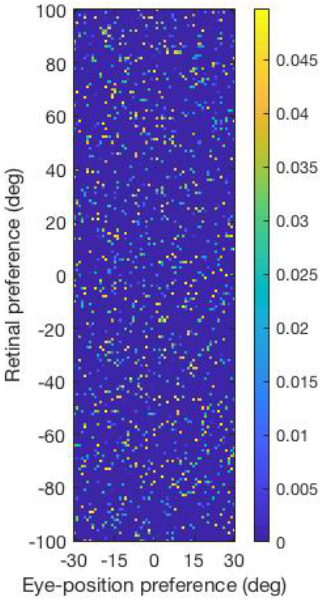
Figure 5.7: Strengths of the afferent synapses from monotonic gain modulated visual input neurons to three typical output neurons in models trained with recorded data from human participants SA, MA and AN. Results are shown for models with a 50:50 mix of peaked and monotonic gain modulated visual input neurons trained with the delayed-trace learning rule (equation 3.17). The output neurons correspond to those plotted in figure 5.6. Same conventions as in figure 3.5. The comparison of subplots 5.7**D**-5.7**F** with subplots 5.7**A**-5.7**C** show how training changed the synaptic weight structures from input neurons with monotonic modulation to output neurons #618, #771 and #279 to a structure that resemble the expected weight structure for a head-centred output neuron.



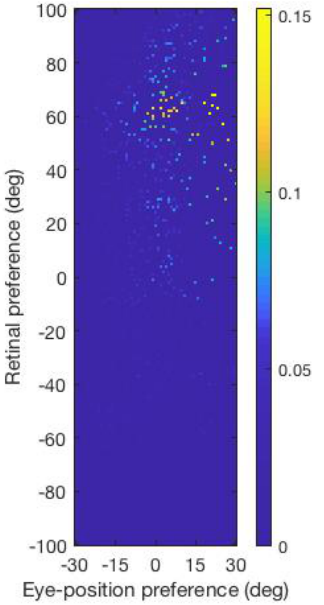
(A) Untrained neuron #618 (model SA).



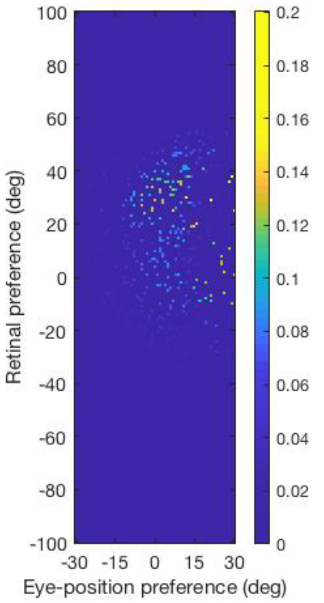
(B) Untrained neuron #771 (model MA).



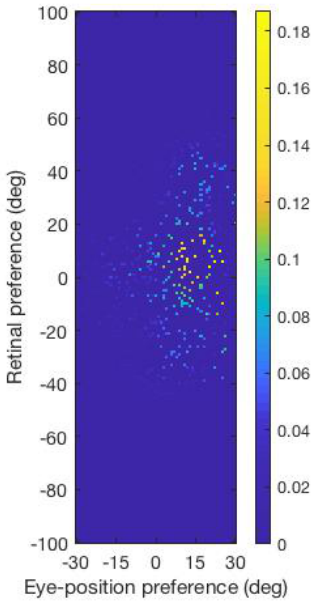
(C) Untrained neuron #279 (model AN).



(D) Trained neuron #618 (model SA).



(E) Trained neuron #771 (model MA).



(F) Trained neuron #279 (model AN).

Figure 5.9: Strengths of the afferent synapses from peaked gain modulated visual input neurons to three typical output neurons in models trained with recorded data from human participants SA, MA and AN. Results are shown for the model with a 50:50 mix of peaked and monotonic gain modulated visual input neurons trained with the delayed-trace learning rule (equation 3.17). The output neurons correspond to those plotted in figure 5.6. In each plot, the afferent synapses have been arranged topographically by the preference of the input neuron for retinal location α_i and eye-position β_j. Each combination of both retinal and eye-position preferences are associated with one input neuron modulated by peaked gain field. Subplots 5.8**A**-5.8**C** show that there was no clear weight vector structure before training. This reflected the randomly assigned afferent synaptic weights from input neurons to output neurons. However, it can be seen from subplots 5.8**D**-5.8**F** that after training the output neurons #618, #771 and #279 have developed synaptic weight structures approximately consistent with a learned response to a particular location within the head-centred frame of reference.

figures 5.6**D**, 5.6**E** and 5.6**F** show that after training there is a clear maximal response to the same head-centred location across all eye-positions for each of these output neurons. Particularly, the head-centredness of output neuron #618 in the model MA, output neuron #771 in the model SA and output neuron #279 in the model AN, respectively, increased with training from -0.04 to 0.55, -0.01 to 0.53 and -0.08 to 0.51. Therefore, training the neural network model with recorded data and the delayed-trace learning rule defined by equation 3.17 has made the response reference frame of these output neurons strongly head-centred. The miniature scatter plots show the reference frame values of each selected neuron in the output layer for each respective model, where each neuron is plotted as a point corresponding to that neuron's particular combination of eye-centredness (abscissa) head-centredness (ordinate). The miniature scatter plots confirm that there was an increase in the head-centredness of these output neurons. This is clearly seen by the receptive fields shifting in head-centred space in respect with the eye-position after training.

Experiments with 50:50 mixed input modulation using experimentally recorded data

Delayed-trace learning rule (equation 3.17)

	Model SA		Model MA		Model AN	
	Untrained	Trained	Untrained	Trained	Untrained	Trained
Head-centredness	0.01 (0.10)	0.31 (0.20)	0.01 (0.10)	0.18 (0.12)	0.01 (0.10)	0.16 (0.11)
Eye-centredness	0.03 (0.09)	0.74 (0.17)	0.03 (0.09)	0.43 (0.21)	0.03 (0.09)	0.37 (0.20)
Proportion head-centred	7%†	3%	7%†	1%	7%†	3%
Proportion eye-centred	46%	26%	46%	29%	46%	42%
Proportion unclassified	47%	71%	47%	70%	47%	55%

Standard Hebbian learning rule (equation 2.9)

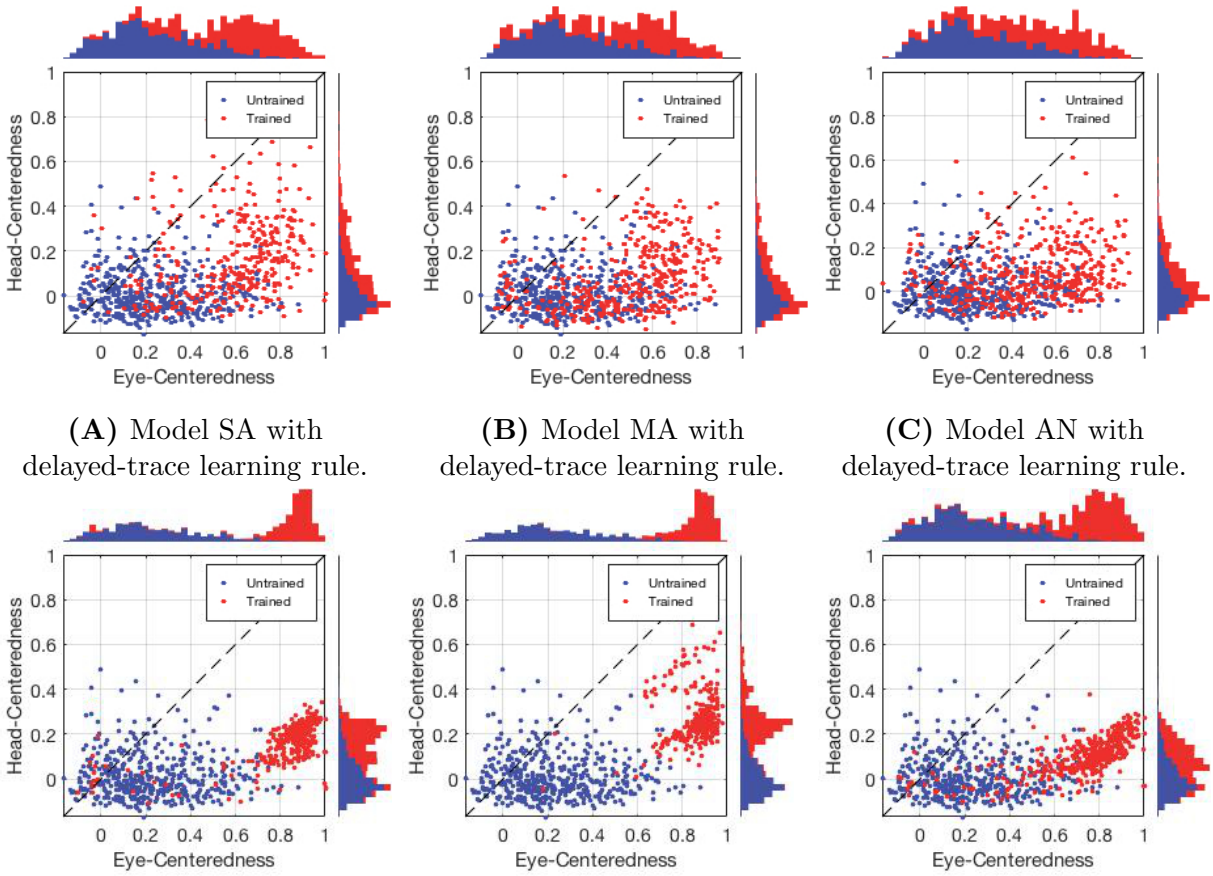
	Model SA		Model MA		Model AN	
	Untrained	Trained	Untrained	Trained	Untrained	Trained
Head-centredness	0.01 (0.10)	0.14 (0.11)	0.01 (0.10)	0.27 (0.11)	0.01 (0.10)	0.10 (0.08)
Eye-centredness	0.03 (0.09)	0.82 (0.21)	0.03 (0.09)	0.86 (0.09)	0.03 (0.09)	0.73 (0.21)
Proportion head-centred	7%†	1%	7%†	0%	7%†	1%
Proportion eye-centred	46%	39%	46%	35%	46%	46%
Proportion unclassified	47%	60%	47%	65%	47%	53%

Table 5.2: Population summary statistics of the response properties of output neurons in three models: Model SA, Model MA and Model AN, where these models refer to simulations run with experimental data recorded from three human participants SA, MA and AN, respectively. Each model was trained with the modified learning rule 3.17 (top) or the standard Hebbian learning rule (equation 2.9) (bottom). The input population of each neural network model had a 50:50 mix of neurons that were modulated by either peaked or monotonic (sigmoidal) gain fields. Conventions as for table 3.2. †Some output neurons before training had slightly higher head-centredness than eye-centredness values, and so were classified as responding in a head-centred frame of reference. However, on visual inspection of the response profiles there was no output neuron consistently responding in a head-centred reference frame before training. After training, the proportion of neurons classified as head-centred were found on visual inspection to consistently respond in a head-centred frame of reference.

Figure 5.7 and figure 5.9 show the synaptic weight vectors of the same output neurons as those shown in figure 5.6 from input neurons modulated by monotonic and peaked gain fields, respectively. Figures 5.7**A**, 5.7**B**, 5.7**C**, 5.8**A**, 5.8**B** and 5.8**C** show that before training there is no clear structure to the potentiated synapses of the output neurons in terms of the preferences of the presynaptic input neurons modulated by peaked or by monotonic gain fields. This reflects the random weighting assigned to the untrained neural network model. Figure 5.7**D**, 5.7**E** and 5.7**F** show the synaptic weight vector of each output neuron after training when the input neurons were modulated by monotonic gain fields. Figure 5.8**D**, 5.8**E** and 5.8**F** show the synaptic weight vector of each output neuron after training when the input neurons were modulated by peaked gain fields. These synaptic weight profiles after training are consistent with a learned response to a particular location within the head-centred space, and are thus consistent with the respective head-centred firing rate responses of these neurons during testing shown in figures 5.6**D**, 5.6**E** and 5.6**F**.

A population analysis of the reference frame response characteristics of the output neurons before and after training for each model is presented in figure 5.10. The output reference frame response characteristics before and after training each model with the delayed-trace learning rule (equation 3.17) are shown in subplots 5.10**A**-5.10**C**. The output reference frame response characteristics before and after training each model with the standard Hebbian learning rule are shown in subplots 5.10**D**-5.10**F**. The comparison of each respective plot shows that training with the delayed-trace learning rule (equation 3.17) resulted in many more head-centred output neurons than training with the standard Hebbian learning rule (table 5.2).

In particular, subplots 5.10**D**-5.10**F** show that training the model with the standard



(A) Model SA with delayed-trace learning rule.

(B) Model MA with delayed-trace learning rule.

(C) Model AN with delayed-trace learning rule.

(D) Model SA with Hebbian learning rule.

(E) Model MA with Hebbian learning rule.

(F) Model AN with Hebbian learning rule.

Figure 5.10: Simulation results showing the output reference frame response characteristics of models with a 50:50 mix of peaked and monotonic gain modulated visual input neurons trained with recorded data from human participants SA, MA and AN. Subplots in the top row show the population analysis of models before and after training with the delayed-trace learning rule (equation 3.17) and subplots in the bottom row show the population analysis of models before and after training with the standard Hebbian learning rule (equation 2.9). Same conventions as in figure 3.5. It can be seen that training the model with the delayed-trace learning rule resulted in many more head-centred output responses after training than with the standard Hebbian rule. In fact, training the model with the standard Hebbian rule resulted in all output neurons responding in an eye-centred frame of reference.

Hebbian rule resulted in the majority of output neurons responding in an eye-centred frame of reference. Although before training some output neurons had slightly higher head-centredness than eye-centredness values, resulting from randomly assigning a combination

of afferent synaptic weights from input neurons to output neurons, on visual inspection of the response profiles there was no output neuron consistently responding in a head-centred reference frame in the untrained model. However, after training the proportion of neurons classified as head-centred were found on visual inspection to consistently respond in a head-centred frame of reference. The output population analysis of the model trained with the delayed-trace learning rule (equation 3.17) using recorded data from human participants SA, MA and AN shown in figures 5.10**A**, 5.10**B** and 5.10**C** confirms that the experimentally recorded eye and head-movements from human subjects during a visual search task has driven the model to self-organise head-centred visual responses.

5.7 Conclusion

In chapter 3 we presented a biologically plausible neural network model of coordinate transformation from an eye-centred to a head-centred reference frame using trace learning rule. Trace learning incorporates a memory trace of recent neuronal activity into the synaptic learning rule (see Rogerson et al. (2014) for a review of physiological studies providing evidences of memory trace in different brain areas, including the cortex). Trace learning in our model encourages output neurons to learn to respond to the position of a visual targets in the same head-centred location across different retinal positions. Particularly we have previously shown that our model robustly self-organised head-centred responses when all visual input neurons had peaked modulation by eye-position, but completely fail when only the modulation of the input units was changed from peaked to monotonic.

In fact, in chapter 3 we showed that the introduction of only a small proportion (approximately 20%) of monotonic modulated input neurons prevented the development of

head-centred output neurons. If all neurons in the input population were modulated by a monotonic function of eye-position, then the majority of output neurons learned to respond in an eye-centred rather than a head-centred reference frame. A theoretical and computational analysis of the influence of monotonic modulated input neurons on the self-organisation of the model led us to propose a set of modified versions of the standard trace learning rule (equation 2.9) that enabled the model to self-organise head-centred responses in the presence of monotonic modulated input neurons. These modified versions of the standard trace learning rule are still biologically plausible in that they rely on locally available signals to update the synaptic weights of connections. In particular, we showed in chapter 3 how the *delayed-trace learning rule* given by equation 3.17 was able to drive the development of head-centred visual responses in a model with a 50:50 mix of peaked and monotonic gain modulated visual input neurons. This is a more biologically realistic scenario as a mixture of eye-position gain modulation has been observed in the brain.

However, the eye and head-movements used to train our original neural network model were artificially created and controlled (chapter 3). In contrast, the training conditions in this work are naturally much more challenging due to the use of experimentally recorded eye and head-movements from human participants during a visual search task. The simulations in this work investigated whether the natural statistics of eye and head-movements recorded from human participants would be appropriate to drive the self-organisation of head-centred output visual neurons using the delayed-trace learning rule defined by equation 2.9 with a 50:50 mix of peaked and monotonic gain modulated visual input neurons. We have found that the natural statistics from recorded eye and head-movements from human subjects enabled the model to develop head-centred responses amongst neurons

in the output layer. Thus, this work confirms that our neural network model can self-organise head-centred visual representations when trained with eye and head-movements of human participants under ecological training conditions. The simulation results discussed in this chapter showed that the model was driven by experimentally recorded eye and head-movements of three human participants during a visual search task. The simulation results discussed in this paper are a subset of a total of forty human participants who took part in the visual search task. The experimentally recorded data enabled our model to self-organise head-centred visual output responses under the same simulation conditions described in this paper. A natural consequence of using experimentally recorded data is that there was an uneven coverage of both retinal and head space. This means that it is more difficult for the model to respond to every possible hand-centred location consistently across all head-centred testing locations when compared to the use of artificially generated, and more carefully controlled, eye and head-movements in chapter 3. The visual coverage of the head-centred space should be complete with more extensive visual training. Obviously, under more ecological conditions humans are exposed to much longer visual training timescales.

This is the first time that our neural network model of coordinate transformation was trained with recorded eye and head-movements from human participants, instead of artificially created eye and head-movements. By showing how the neural network model self-organised head-centred visual responses using realistic eye and head-movements, this work makes an important contribution to understanding how head-centred visual responses may develop in the brain through an unsupervised process of visually-guided learning given visual precursor neurons with peaked and monotonic eye-position gain modulation. More-

over, although we investigated one particular type of coordinate transformation (i.e., from eye-centred to head-centred visual representations), our findings may also apply to a range of other coordinate transformations with sensorimotor integration of monotonically encoded motor variables (Bremner and Andersen, 2012, Duhamel et al., 1997, Pesaran et al., 2006).

Chapter 6

Conclusions

This chapter provides a summary of the findings in this thesis and discusses how our findings instigate future work.

The simulation results presented in this thesis investigated the development of how neural response properties in the visual brain may develop using biologically plausible learning mechanisms. Furthermore, the performance of the neural network models investigated in this thesis was assessed under simulated conditions similar to the experimental tasks that revealed the neural properties that we studied. When plasticity was introduced to the neural network models, then the performance of the models were assessed before and after training. The performance of the models was explained not only in terms of the changes in the neural responses before and after training, but more importantly in terms of the agreement of the trained responses with experimental studies in primates.

The two main neural properties investigated in this thesis were the emergence of gain modulation by the position of the eye (chapter 2) and the development of head-centred visual responses (chapter 3 and chapter 5). Gain modulation by eye-position refers to the modulatory effect that the position of the eye in the orbit has on the firing rate of some visual neurons responding in a given reference frame. Head-centred responses refers to the

visual receptive field of neurons that respond maximally to a visual stimulus in the same head-centered location across all eye positions. These two properties have been the subject of study in a large number of experimental and theoretical studies. In particular, both eye position gain fields and head-centred responses are relevant for sensorimotor integration (e.g. oculomotor systems) that builds neural representations to guide motor behaviour.

Chapter 2 shows simulation results of both a hardwired and a self-organising neural network model of how multiplicative gain modulation of visual responses by the position of the eye may develop in the brain. In particular, our simulation results address the problem with the use of a linear function to model the dependence of the visual responses on eye position by previous modelling studies (Salinas and Abbott, 1996). This issue is important because the response profiles of eye position neurons in parietal areas have been reported to be a better fit to Gaussian rather than to linear functions (Andersen, 1987, Dayan and Abbott, 2001, Marĉelja, 1980, Wang et al., 2007). Thefore the simulations in chapter 2 investigate how multiplicative gain modulation by eye position may develop using inputs from eye position neurons with Gaussian response profiles. We started by studying a hardwired model of how cortical visual neuron responses depend on both the position of the eyes and the retinotopic location of the visual target. The simulation results discussed in chapter 2 showed a variety of different functional forms of gain modulation by eye position of visual responses in the hardwired model that were similar to experimentally recorded cortical responses. In particular, some of the responses were approximately linear for all eye positions when the visual stimulus was at the neuron's preferred retinal location. We then investigated how the functional form of the eye position gain modulation of visual responses, such as its linearity, would be affected by varying the width ρ of the eye position

Gaussian tuning curve and revealed that increasing ρ resulted in more linear responses. Such simulation results showed that different functional forms of gain modulation by eye position similar to those seen in parietal cortex, including responses that were linear across different eye positions, can be observed with more biologically realistic Gaussian tuning curves modelling the dependence of visual responses on the position of the eye and on the retinal location of the visual stimulus.

A further relevant contribution of the simulations discussed in chapter 2 is the investigation of how such a variety of different functional forms of gain modulation by eye position, as well as multiplicatively gain modulated visual responses, would self-organise in a neural network model of cortical neurons. The self-organising model in chapter 2 consisted of an output population of visual neurons that receives synaptic connections from a population of input neurons. The Gaussian responses of the input neurons depended on either the retinal location of the visual stimulus or the position of the eye. During training the synaptic connections were updated using a Hebbian learning rule (equation 2.9) whilst the network was presented with randomised combinations of eye position and retinal stimulus location. The simulation results in chapter 2 showed that the model successfully self-organised the forms of visual responses observed in the parietal cortex, including peaked and monotonic gain modulation by eye position. In particular, a form of eye-position gain modulation that was multiplicative, similar to the effects shown for the hardwired model, was observed in the self-organising model after training. Finally, we showed that increasing the width ρ of the eye position Gaussian tuning curve resulted in the self-organisation of responses that varied more monotonically with eye position. Further simulation results in this thesis shows how the different forms of eye position gain modulation discussed in chapter

2, particularly monotonic modulation, deeply influences the self-organisation process of head-centred output responses.

In this thesis the development of head-centred visual responses in computer simulations via a biologically plausible process of visually-guided learning was based on a hypothesis that requires the following four model features. The first feature is the coupled encoding of the position of the eye and the retinal location of the visual target by a population of input neurons. The second feature is the competition of output neurons via inhibitory interactions. The third feature is the use of an associative trace learning rule to temporally bind input patterns that occur close together in time. Finally, the fourth feature is the existence of periods during the exploration of a visual scene where the eyes move whilst both the head and the visual target are stationary. This hypothesis was originally tested in Mender and Stringer (2013). Chapter 3 shows that the self-organising neural network model studied by Mender and Stringer (2013) is able to develop head-centred output representations during training using the standard trace learning rule (equation 3.5) with peaked eye-position gain modulated input neurons. However, the self-organising model with purely monotonic eye-position gain modulated input neurons develops eye-centred output neurons when using the standard trace learning rule. The cause of this contrasting behaviour between peaked and monotonic gain fields is the correlations between the activity 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 and Treves, 1998). In the self-organising model with peaked gain modulation, 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 input space corresponding to a particular head-centred location. Output neurons will then respond to particular head-centred locations regardless of eye-position or the retinal location of a visual target. However, the situation is quite different with monotonic gain modulated input neurons. Due to monotonic saturation of the input neuron response function in the eye-position dimension of the input space, the subsets of input neurons that are most frequently co-active are localised in the retinotopic preference dimension but elongated in the eye-position inflection point 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-centred output responses. However, if a trace learning rule is implemented, the output neurons still learn to represent eye-centred rather than head-centred locations. This is because developing head-centred 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-centred output neurons. This finding holds for any input neuron response function with a monotonic gain in the eye-position dimension, be it purely linear (Zipser and Andersen, 1988) or linear rectified (Salinas and Abbott,

1995).

We next showed in chapter 3 that a manually prewired neural network model with monotonic eye-position gain modulated input neurons could perform the coordinate transformation. This is achieved by elevating the weight of synaptic connections from all input neurons which responded strongly, for some eye-position, to a visual target in the head-centred location to which the postsynaptic output neuron is assigned. This result established the feasability of the coordinate transformation within the given model architecture with monotonic gain fields acting on the input population. However, when synaptic plasticity based on the standard trace learning rule (equation 3.5) is suddenly introduced into the manually prewired model, it is found that just a single epoch of visually-guided training is sufficient to dramatically reduce the prevalence of head-centred neurons in the output population, and by the 5^{th} epoch all neurons are eye-centred for all subsequent epochs of training. This showed that monotonic gain modulation would, in so far as the synapses are plastic, even undermine the functioning of a cortical circuit which is somehow prewired, perhaps by genetic hardwiring or by an initial period of supervised learning as suggested by Salinas and Abbott (1995), to produce the head-centred visual representations. Given the ubiquitous presence of associative synaptic plasticity in cortex, it is therefore a substantial challenge to explain how head-centred visual neurons might develop and persist in the brain given the relatively large proportion of precursor eye-centred visual neurons that have monotonic eye-position gain fields.

It has been shown that neurons in many parietal areas have a mixture of different forms of gain modulation, not all of which are planar. In fact, 41% and 56% are not planar in areas LIP and 7a, respectively (Andersen et al., 1990). Consequently, we explored

how various degrees of prevalence of monotonic gain modulation would influence the self-organisation of the model with the standard trace learning rule (equation 3.5). It is found that so long as there is no more than approximately 20% monotonic gain modulation in the input population then the fraction of head-centred neurons in the trained model did not drop below $\sim$ 15%. However, the decrease in performance, and eventual collapse of the model, is very severe when the presence of monotonic gain modulation in the input population is increased. In particular, with a more biologically realistic 50:50 mix in peaked and monotonic gain modulated input neurons, the model failed to develop a significant population of head-centred output neurons with the standard trace learning rule (equation 3.5).

In order to find a biologically plausible way in which the self-organising model could develop relatively large numbers of head-centred visual output neurons when at least half or even all of the input neurons had monotonic gain fields, we next explored the performance of the model architecture using a variety of more sophisticated learning rules that may incorporate temporally delayed memory traces, as well as a mixture of both trace learning and anti-Hebbian learning terms. These new, modified learning rules were previously developed by Rolls and Stringer (2001) in the context of modelling transform invariant visual object recognition. The new learning rules continue to be biologically plausible because they rely on locally available biological quantities, namely the activities of the pre- and post-synaptic neurons, to guide the modification of synaptic weights. The different modified learning rules were found to successfully drive the self-organisation of head-centred output responses when all of the input neurons had monotonic eye-position modulation. The modified learning rules were able to substantially improve the temporal binding of the

standard trace learning rule by incorporating a trace of previous neuronal activity with an explicit time delay ΔT. This was evidenced by better performance being achieved when the learning rule incorporated a trace value of the postsynaptic neuron from an earlier time ΔT in the past instead of the trace value at the current time. Furthermore, the new learning rules were able to overcome the high correlations between overlapping input patterns with monotonic gain fields through the additional incorporation of anti-Hebbian learning, which had been previously found to offer a significant performance enhancement by Rolls and Stringer (2001). However, the delayed trace learning rule (equation 3.17) incorporated a delayed postsynaptic trace without an anti-Hebbian term (section Modified Learning Rule: Delayed Postsynaptic Trace Learning Rule). Simulation results with this learning rule showed that head-centred output responses developed in the trained model. Thus, anti-Hebbian learning is not necessarily a requirement for the development of head-centred output responses. The performance of the delayed trace learning rule (equation 3.17) was then investigated under a more biologically realistic scenario in which the input population consisted of a 50:50 mixture of neurons with either peaked or monotonic eye-position gain modulation. These results demonstrated that the model could produce large numbers of head-centred output neurons with a learning rule that relied purely on a postsynaptic delayed-trace term without any anti-Hebbian term if the input population contained a 50:50 mix of neurons modulated by either peaked or monotonic gain fields. These last findings should be contrasted with the performance of the model with the standard trace learning rule (equation 3.5), which failed to develop a significant number of head-centred output neurons with such a mixed input population.

However, the eye and head-movements used to train the neural network model in chap-

ter 3 were artificially created and controlled. Having developed the neural network computer model with simulated data, we were in a position to test some specific experimental predictions. In particular, we predicted that the statistics of eye and head movements recorded from human test participants during a natural visual search task would be appropriate to drive the development of head-centred visual neurons in the output layer of the model. Chapter 4 presented a smartphone application, the DN Rotation Tracking System, developed by the author to record human movement data for scientific studies. In particular, the current version of the application has the capability to collect and combine data from a number of sensors to detect and record 2-dimensional rotation movements, such as head rotation movements during the exploration of a natural scene. Clear practical and financial challenges of recording the form of rotation movement recorded by the DN Rotation Tracking System have been found in previous studies, e.g. Bartz (1966), Minati et al. (2016) and David et al. (2018). However, modern smartphones have become powerful computing machines due to recent technological developments and increased number of enhanced sensors, for example. Although smartphone sensors have been used in previous studies (Coskun et al., 2015, Novakova and Pavlis, 2017), the potential to use such sensors to collect data for scientific purposes has not yet been fully explored. In most cases, the development of such applications is sufficiently difficult to be beyond the ability of a typical researcher. The DN Rotation Tracking System for Android is a scientifc tool created to address the challenge of using Android enabled mobile devices (e.g. smartphones) to record rotation movements. The application uses a combination of a dual-buffer system with Android AsyncTask functionalities to record real-time rotation movements in a cost-effective manner. At the end of each recording session the recorded data are saved as a

CSV file to facilitate the immediate analysis of the data by the researcher. CSV is a particularly friendly file format for researchers, who are used to processing such text files. The DN Rotation Tracking System allows researchers to easily incorporate this form of human movement data to newer studies by using smartphone devices. Moreover, the application provides the user with the contact details of the developer to facilitate the request of new features and the establishment of future collaboration with other researchers within the scientific community. This is an additional important contribution of the work presented in this thesis. Head-rotation movements using the DN Rotation Tracking System were simultaneously recorded with eye-movements using the Eyelink II System (SR ResearchLtd., Ontario, Canada) during the visual search task discussed in chapter 5.

The simulations in chapter 5 investigated whether the natural statistics of eye and head-movements recorded from human participants would be appropriate to drive the self-organisation of head-centred output visual neurons using the delayed-trace learning rule defined by equation 3.17 with a 50:50 mix of peaked and monotonic gain modulated visual input neurons. We have found that the natural statistics from recorded eye and head-movements from human participants enabled the model to develop head-centred responses amongst neurons in the output layer. When compared to the results of the simulations in chapter 3, the training conditions of the simulations discussed in chapter 5 are naturally much more challenging due to the use of experimentally recorded eye and head-movements from human participants during a visual search task. The simulation results presented in chapter 5 confirmed that our neural network model can self-organise head-centred visual representations when trained with eye and head-movements of human participants under ecological training conditions. In particular, chapter 5 showed simulation results when the

model was driven by experimentally recorded eye and head-movements of three human participants during a visual search task. The simulation results discussed in chapter 5 are a subset of a total of forty human participants who took part in the visual search task. The experimentally recorded data enabled our model to self-organise head-centred visual output responses under the same simulation conditions described in chapter 3. A natural consequence of using experimentally recorded data is that there was an uneven coverage of both retinal and head space. This means that it is more difficult for the model to respond to every possible head-centred location consistently across all head-centred testing locations when compared to the use of artificially-generated, and more carefully controlled, eye and head-movements (chapter 3). The visual coverage of the head-centred space should be complete with more extensive visual training. Obviously, under more ecological conditions humans are exposed to much longer visual training timescales.

This was the first time that our neural network model of coordinate transformation was trained with recorded eye and head-movements from human participants, instead of artificially created eye and head-movements. By showing how the neural network model self-organised head-centred visual responses using realistic eye and head-movements, the simulation results in this thesis make an important contribution to understanding how head-centred visual responses may develop in the brain through an unsupervised process of visually-guided learning given visual precursor neurons with peaked and monotonic eye-position gain modulation. Moreover, although we investigated one particular type of coordinate transformation (i.e., from eye-centred to head-centred visual representations), our findings may also apply to a range of other coordinate transformations with sensorimotor integration of monotonically encoded motor variables (Bremner and Andersen, 2012,

Duhamel et al., 1997, Pesaran et al., 2006).

The simulation results discussed in this thesis showed for the first time how head-centred output responses may develop through a biologically plausible self-organisation process of visually-guided learning with monotonically modulated input neurons. In particular, we showed that our model is robust enough to self-organise head-centred output responses with real eye and head movements recorded from human participants during a visual search task. The models investigated in this thesis were rate-coded. Rated coded neural network models, such as the models used in this thesis, use the form of sigmoid functions in equation 2.8 or equation 3.2, for example, to model the rate by which a neuron fires.

In other words, rate-coded models do not take into account the actual timings of individual spikes of each neuron, but only the frequency of spikes of a neuron over time. The use of rate-coded models allowed us to perform the theoretical analysis of the influence of monotonic modulation on the self-organisation process discussed in chapter 3. Although the use of monotonic gain modulated input units makes the rate-coded models studied in this thesis more biologically plausible than previous work, the theoretical analysis in chapter 3 showed that, with rate-coded models, the development of head-centred output responses requires that some form of trace-like learning rule disrupt and break apart highly overlapping output representations (chapter 3).

This finding with rate-coded models motivates future studies with 'spiking' neural networks, in which the timings of individual spikes emitted by neurons are explicitly represented. We hypothesise that in a spiking neural network model the spatiotemporal spike patterns encoding the activity of retinotopic input neurons with monotonic modulation by eye position could help to reduce the effective overlap between these input patterns that

is present within a rate-coded model: in a spiking neural network model clusters of input neurons with correlated activities from a purely rate-coded perspective would be differentiated by their own unique spike time activities (Izhikevich, 2006). That is, encoding overlapping input patterns, which contain large common subsets of active input neurons, using a spike timing code could effectively reduce the similarity in these input representations from the point of view of receiving neurons in higher layers, the responses of which would be sensitive to the relative timings of incoming spikes. In this scenario, even if two input patterns had almost all of their active input neurons in common with each other, receiving neurons in the next layer would see these input patterns as completely different if the active input neurons emitted their spikes in a different temporal order for the two input patterns. This could improve the performance of the self-organisation process, resulting in a greater number of head-centred output neurons after training.

Furthermore, Eguchi et al. (2018) and Isbister et al. (2018) studied an approach to solving feature binding in the visual brain with spiking neural network models of invariance learning. The simulation results by these authors showed that a type of neuron, called a 'binding neuron', naturally emerged that explicitly encoded the hierarchical binding relationships between lower-level and higher-level visual features represented in successive layers of the network. In the context of this thesis, similar 'binding' neurons could emerge within a spiking neural network model of coordinate transformation to bind the visual features encoded in an earlier body-centred reference frame (e.g. eye-centred) to visual features encoded within a later reference frame (e.g. head-centred). Consequently, we hypothesise that the form of 'binding' neurons reported by (Eguchi et al., 2018, Isbister et al., 2018) may be relevant for representing the correspondences between visual features

across different body-centred reference frames in the dorsal visual system.

Appendix A

Appendix: Eye-Centredness Reference Frame Analysis

During testing, the visual target was located in head-centred 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 centred 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-centred and eye-centred space across multiple eye positions.

The set of head-centred 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

$$\min \bigcap_i R_i = t_1 - e_1$$
$$\max \bigcap_i R_i = t_T - e_E \tag{A.3}$$

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 centred 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

$$\begin{aligned} V &= T - \frac{\Delta e}{\Delta h}(E + i) - \frac{\Delta e}{\Delta h}(i-1) \\ &= T - \frac{\Delta e}{\Delta h}(E + i - (i-1)) \\ &= T - \frac{\Delta e}{\Delta h}(E + 1) \end{aligned} \tag{A.8}$$

Bibliography

Andersen, R. A. (1987), 'Inferior parietal lobule function in spatial perception and visuomotor integration', *Comprehensive Physiology* .

URL: *https://onlinelibrary.wiley.com/doi/full/10.1002/cphy.cp010512*

Andersen, R. A. and 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.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/6827308*

Andersen, R., Bracewell, R., Barash, S., Gnadt, J. and 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. and 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*

Bartz, A. E. (1966), 'Eye and head movements in peripheral vision: nature of compensatory

eye movements', *Science* **152**(3729), 1644–1645.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/5936888*

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P. and Friston, K. J. (2012), 'Canonical microcircuits for predictive coding', *Neuron* **76**(4), 695–711.

URL: *https://doi.org/10.1016/j.neuron.2012.10.038*

Bremner, L. R. and Andersen, R. A. (2012), 'Coding of the reach vector in parietal area 5d', *Neuron* **75**(2), 342–351.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/22841318*

Breveglieri, R., Bosco, A., Canessa, A., Fattori, P. and 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.

URL: *https://link.springer.com/chapter/10.1007/978-3-642-02267-8_35*

Brotchie, P. R., Andersen, R. A., Snyder, L. H. and 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*

Buneo, C. A. and Andersen, R. A. (2006), 'The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements', *Neuropsychologia* **44**(13), 2594–2606.

URL: *http://www.sciencedirect.com/science/article/pii/S0028393205003337*

Buneo, C. A., Jarvis, M. R., Batista, A. P. and Andersen, R. A. (2002), 'Direct visuomotor transformations for reaching', *Nature* **416**(6881), 632–636.

URL: *http://dx.doi.org/10.1038/416632a*

Chang, S. W., Papadimitriou, C. and Snyder, L. H. (2009), 'Using a compound gain field to compute a reach plan', *Neuron* **64**(5), 744–755.

URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2811884*

Coskun, D., Incel, O. D. and Ozgovde, A. (2015), Phone position/placement detection using accelerometer: Impact on activity recognition, *in* '2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)', IEEE, pp. 1–6.

URL: *https://ieeexplore.ieee.org/document/7106915*

Daddaoua, N., Dicke, P. and Thier, P. (2014), 'Eye position information is used to compensate the consequences of ocular torsion on v1 receptive fields', *Nature communications* **5**, 3047.

URL: *https://www.nature.com/articles/ncomms4047*

David, E. J., Gutiérrez, J., Coutrot, A., Da Silva, M. P. and Callet, P. L. (2018), A dataset of head and eye movements for 360 videos, *in* 'Proceedings of the 9th ACM Multimedia Systems Conference', ACM, pp. 432–437.

URL: *https://hal.archives-ouvertes.fr/hal-01994923/document*

Dayan, P. and Abbott, L. F. (2001), *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, ISBN 0-262-04199-5.

URL: *http://www.gatsby.ucl.ac.uk/ dayan/book/*

De Meyer, K. and Spratling, M. W. (2011), 'Multiplicative gain modulation arises through unsupervised learning in a predictive coding model of cortical function', *Neural computation* **23**(6), 1536–1567.
URL: *https://dx.doi.org/10.1162/NECO_a_00130*

Duhamel, J.-R., Bremmer, F., BenHamed, S. and Graf, W. (1997), 'Spatial invariance of visual receptive fields in parietal cortex neurons', *Nature* **389**(6653), 845–848.
URL: *https://www.nature.com/articles/39865*

Eguchi, A., Isbister, J. B., Ahmad, N. and Stringer, S. (2018), 'The emergence of polychronization and feature binding in a spiking neural network model of the primate ventral visual system', *Psychological review* **125**(4), 545.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/29863378*

Engbert, R. and Mergenthaler, K. (2006), 'Microsaccades are triggered by low retinal image slip', *Proceedings of the National Academy of Sciences* **103**(18), 7192–7197.
URL: *https://www.pnas.org/content/103/18/7192*

Evans, B. and Stringer, S. (2012), 'Transform-invariant visual representations in self-organizing spiking neural networks', *Frontiers in Computational Neuroscience* **6**(46).
URL: *https://www.frontiersin.org/articles/10.3389/fncom.2012.00046/full*

Ferraina, S., Paré, M. and Wurtz, R. H. (2002), 'Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements', *Journal of Neurophysiology* **87**(2), 845–858.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/11826051*

Földiák, P. (1991), 'Learning invariance from transformation sequences', *Neural Computation* **3**(2), 194–200.

URL: *https://ieeexplore.ieee.org/document/6796614*

Freedman, E. G. and Sparks, D. L. (1997), 'Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys', *Journal of Neurophysiology* **77**(5), 2328–2348.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/9163361*

Galletti, C., Battaglini, P. and 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.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/12106184*

Galletti, C., Battaglini, P. and Fattori, P. (1995), 'Eye position influence on the parieto occipital area po of the macaque monkey', *European Journal of Neuroscience* **7**(12), 2486–2501.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/8845954*

Galletti, C. and Battaglini, P. P. (1989), 'Gaze-dependent visual neurons in area v3a of monkey prestriate cortex', *Journal of Neuroscience* **9**(4), 1112–1125.

URL: *https://www.ncbi.nlm.nih.gov/pubmed/2703870*

Galletti, C., Battaglini, P. P. and Fattori, P. (1993), 'Parietal neurons encoding spatial locations in craniotopic coordinates', *Experimental Brain Research* **96**(2), 221–229.

URL: *https://link.springer.com/article/10.1007/BF00227102*

Guo, K. and Li, C.-Y. (1997), 'Eye position-dependent activation of neurones in striate cortex of macaque', *Neuroreport* **8**(6), 1405–1409.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/9172144*

Hoffmann, K.-P. (1998), 'Eye position encoding in the macaque posterior parietal cortex', *European Journal of Neuroscience* **10**(1), 153–160.

Isbister, J. B., Eguchi, A., Ahmad, N., Galeazzi, J. M., Buckley, M. J. and Stringer, S. (2018), 'A new approach to solving the feature-binding problem in primate vision', *Interface focus* **8**(4), 20180021.

Izhikevich, E. M. (2006), 'Polychronization: computation with spikes', *Neural computation* **18**(2), 245–282.

Koch, G., Ponzo, V., Di Lorenzo, F., Caltagirone, C. and Veniero, D. (2013), 'Hebbian and anti-hebbian spike-timing-dependent plasticity of human cortico-cortical connections', *Journal of Neuroscience* **33**(23), 9725–9733.

Kohonen, T. (2012), *Self-organization and associative memory*, Vol. 8, Springer Science & Business Media.

Kovesi, P. (2015), 'Good colour maps: How to design them'.

Kwisthout, J. and Van Rooij, I. (2013), 'Bridging the gap between theory and practice of approximate bayesian inference', *Cognitive Systems Research* **24**, 2–8.
URL: *https://dx.doi.org/10.1016/j.cogsys.2012.12.008*

Lal, R. and Friedlander, M. J. (1990), 'Effect of passive eye position changes on retinogeniculate transmission in the cat', *Journal of Neurophysiology* **63**(3), 502–522.

Lamsa, K. P., Heeroma, J. H., Somogyi, P., Rusakov, D. A. and Kullmann, D. M. (2007), 'Anti-hebbian long-term potentiation in the hippocampal feedback inhibitory circuit', *Science* **315**(5816), 1262–1266.

Li, N. and DiCarlo, J. J. (2008), 'Unsupervised natural experience rapidly alters invariant object representation in visual cortex', *science* **321**(5895), 1502–1507.

M, R. and S, F. (2015), 'Three problems for the predictive coding theory of attention', *Midas Online* .

URL: *http://mindsonline.philosophyofbrains.com/wp-content/uploads/2015/09/2015-Ransom-and-Fazelpour-Three-Problems-for-the-Predictive-Coding-Theory-of-Attention-extended-abstract.pdf*

Marĉelja, S. (1980), 'Mathematical description of the responses of simple cortical cells', *The Journal of the Optical Society of America* **70**(11), 1297–1300.

Mazzoni, P., Andersen, R. A. and Jordan, M. I. (1991), 'A more biologically plausible learning rule for neural networks', *Proceedings of the National Academy of Sciences* **88**(10), 4433–4437.

Mender, B. M. and 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. (2014), Self-organizing Models of Primate Supraretinal Visual Representations, PhD thesis, University of Oxford.

Mender, B. M. W. and Stringer, S. M. (2013), 'A model of self-organizing head-centered visual responses in primate parietal areas', *PloS one* **8**(12), e81406.
URL: *http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081406*

Merriam, E. P., Gardner, J. L., Movshon, J. A. and Heeger, D. J. (2013), 'Modulation of visual responses by gaze direction in human visual cortex', *Journal of Neuroscience* **33**(24), 9879–9889.
URL: *http://www.jneurosci.org/content/33/24/9879*

Minati, L., Yoshimura, N. and Koike, Y. (2016), 'Hybrid control of a vision-guided robot arm by eog, emg, eeg biosignals and head movement acquired via a consumer-grade wearable device', *IEEE Access* **4**, 9528–9541.

Morris, A. P. and Krekelberg, B. (2019), 'A stable visual world in primate primary visual cortex', *Current Biology* **29**(9), 1471–1480.

Mullette-Gillman, O. A., Cohen, Y. E. and 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*

Navarro, D. M., Mender, B. M., Smithson, H. E. and Stringer, S. M. (2018), 'Self-organising coordinate transformation with peaked and monotonic gain modulation in the primate dorsal visual pathway', *PLoS one* **13**(11), e0207961.

Navarro, D. M., Smithson, H. E. and Stringer, S. M. (2019), 'The development of head-

centred visual representations in the brain: A computer modelling study using natural recorded data', *Journal of Vision (submitted)* .

Novakova, L. and Pavlis, T. L. (2017), 'Assessment of the precision of smart phones and tablets for measurement of planar orientations: A case study', *Journal of Structural Geology* **97**, 93–103.

O'dhaniel, A., Cohen, Y. E. and Groh, J. M. (2005), 'Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus', *Journal of neurophysiology* .

O'Donohue, T. L., Millington, W. R., Handelmann, G. E., Contreras, P. C. and 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*

Oja, E. (1982), 'Simplified neuron model as a principal component analyzer', *Journal of mathematical biology* **15**(3), 267–273.
URL: *https://link.springer.com/content/pdf/10.1007/BF00275687.pdf*

Page, H. J., Walters, D. and Stringer, S. M. (2015), 'Architectural constraints are a major factor reducing path integration accuracy in the rat head direction cell system', *Frontiers in computational neuroscience* **9**, 10.

Perry, G., Rolls, E. T. and Stringer, S. M. (2006), 'Spatial vs temporal continuity in view invariant visual object recognition learning', *Vision Research* **46**(23), 3994–4006.

Pesaran, B., Nelson, M. J. and Andersen, R. A. (2006), 'Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning', *Neuron* **51**(1), 125–134.

Pouget, A. and 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*

Pouget, A. and Snyder, L. H. (2000), 'Computational approaches to sensorimotor transformations', *Nature Neuroscience* **3 Suppl**, 1192–1198. PMID: 11127837.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11127837*

Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R. and Rao, C. R. (1973), *Linear statistical inference and its applications*, Vol. 2, Wiley New York.

Rogerson, T., Cai, D. J., Frank, A., Sano, Y., Shobe, J., Lopez-Aranda, M. F. and Silva, A. J. (2014), 'Synaptic tagging during memory allocation', *Nature Reviews Neuroscience* **15**(3), 157.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3992944*

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. and Stringer, S. M. (2001), 'Invariant object recognition in the visual system with error correction and temporal difference learning', *Network (Bristol, England)* **12**(2), 111–129. PMID: 11405418.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11405418*

Rolls, E. T. and Treves, A. (1998), *Neural Networks and Brain Function*, Oxford University Press Oxford.

Royer, S. and Paré, D. (2003), 'Conservation of total synaptic weight through balanced synaptic depression and potentiation', *Nature* **422**(6931), 518–522.

Ruan, H., Saur, T. and Yao, W.-D. (2014), 'Dopamine-enabled anti-hebbian timing-dependent plasticity in prefrontal circuitry', *Frontiers in neural circuits* **8**, 38.

Salinas, E. and 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. and Abbott, L. F. (1996), 'A model of multiplicative neural responses in parietal cortex', *Proceedings of the National Academy of Sciences* **93**(21), 11956–11961.
URL: *http://www.pnas.org/content/93/21/11956.abstract*

Spratling, M. W. (2009), 'Learning posture invariant spatial representations through temporal correlations', *Autonomous Mental Development, IEEE Transactions on* **1**(4), 253–263.

SR Research Ltd. (2009), 'Eyelink®II user manual version 2.14'.
URL: *http://sr-research.jp/support/manual/EyeLink%20II%20Head%20Mounted%20User%20Manual%202.14.pdf*

Swadlow, H. A. (1985), 'Physiological properties of individual cerebral axons studied in vivo for as long as one year', *Journal of Neurophysiology* **54**(5), 1346–1362.

Swadlow, H. A. (1988), 'Efferent neurons and suspected interneurons in binocular visual cortex of the awake rabbit: receptive fields and binocular properties', *Journal of neurophysiology* **59**(4), 1162–1187.

Swadlow, H. A. (1992), 'Monitoring the excitability of neocortical efferent neurons to direct activation by extracellular current pulses', *Journal of neurophysiology* **68**(2), 605–619.

Trotter, Y. and Celebrini, S. (1999), 'Gaze direction controls response gain in primary visual-cortex neurons', *Nature* **398**(6724), 239.

Wallis, G. and Rolls, E. T. (1997), 'Invariant face and object recognition in the visual system', *Progress in neurobiology* **51**(2), 167–194.

Walters, D., Stringer, S. and Rolls, E. (2013), 'Path integration of head direction: updating a packet of neural activity at the correct speed using axonal conduction delays', *PLoS One* **8**(3), e58330.

Wang, X., Zhang, M., Cohen, I. S. and Goldberg, M. E. (2007), 'The proprioceptive representation of eye position in monkey primary somatosensory cortex', *Nature neuroscience* **10**(5), 640–646.

Xing, J. and Andersen, R. A. (2000), '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. and 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*

Zuber, B. L. and Stark, L. (1965), 'Microsaccades and the Velocity-Amplitude Relationship for Saccadic Eye Movement', *Science* **150**(3702), 1459–1460.