NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Allocation of Computational Resources in the Nervous System

A dissertation

submitted for the degree of

Doctor of Philosophy

by

Santiago Jaramillo

Supervisor: Barak A. Pearlmutter

Department of Electronic Engineering

National University of Ireland, Maynooth

October 2006

Contents

List of Figures

Summary

The nervous system integrates past information together with predictions about the future in order to produce rewarding actions for the organism. This dissertation focuses on the resources underlying these computations, and the task-dependent allocation of these resources. We present evidence that principles from optimal coding and optimal estimation account for overt and covert orienting phenomena, as observed from both behavioral experiments and neuronal recordings.

First, we review behavioral measurements related to selective attention and discuss models that account for these data. We show that reallocation of resources emerges as a natural property of systems that encode their inputs efficiently under non-uniform constraints. We continue by discussing the attentional modulation of neuronal activity, and show that: (1) Modulation of coding strategies does not require special mechanisms: it is possible to obtain dramatic modulation even when signals informing the system about fidelity requirements enter the system in a fashion indistinguishable from sensory signals. (2) Optimal coding under non-uniform fidelity requirements is sufficient to account for the firing rate modulation observed during selective attention experiments. (3) The response of a single neuron cannot be well characterized by measurements of attentional modulation of only a single sensory stimulus. (4) The magnitude of the activity modulation depends on the capacity of the neural circuit. A later chapter discusses the neural mechanisms for resource allocation, and the relation between attentional mechanisms and receptive field formation.

The remainder of the dissertation focuses on overt orienting phenomena and active perception. We present a theoretical analysis of the allocation of resources during state estimation of multiple targets with different uncertainties, together with eye-tracking experiments that confirm our predictions.

We finish by discussing the implications of these results to our current understanding of orienting phenomena and the neural code.

Acknowledgements

First and foremost, I would like to thank my supervisor Barak Pearlmutter for his guidance and support throughout these years. I also wish to thank the faculty, administrators, and fellow students at the Hamilton Institute, NUI Maynooth, for significant help during the execution of this research.

Special thanks are extended to Rodolfo Llinás and the people at the Center for Neuromagnetism, New York University School of Medicine. Being part of the neuroscience community at NYU, and learning from many outstanding researchers at this institution, were both extraordinary experiences.

The courses at the Santa Fe Institute for Complex Systems, the Marine Biological Laboratory in Woodshole, and the Frankfurt Institute for Advance Studies constituted a significant element of my education. I am grateful to organizers, faculty and students who made possible these amazing courses.

I also owe a great deal to all my friends who shared part of their lives with me during my years as a graduate student. Albuquerque, New York City and Maynooth would not have been the same without you all.

I would like to express my gratitude to all those who take part of the Open Source community. My work was considerably dependent on your contributions, and it would not have been as enjoyable without the tools the community continues to provide.

Finally, I wish to thank my family for their unconditional support and understanding.

Declaration

I hereby certify that this material, which I now submit for assessment on the program study leading to the award of Doctor of Philosophy in Electronic Engineering is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: ID No.: 62134141

Date:

Chapter 1

Introduction

Summary

This chapter presents an outline of the contents of this dissertation, explains why orienting phenomena and resource allocation were selected as research topics, and describes the approach followed to investigate them. A brief history of relevant work is also included, highlighting landmarks that will be discussed in detail in subsequent chapters.

Understanding how the signals in the nervous system subserve behavior is one of the main quests of 21st century science. The work presented in this dissertation investigates various hypotheses concerning how the nervous system encodes information about our perceptions and actions. This exploration uses of a combination of theoretical, computational and experimental approaches. Our interest is, for the most part, on questions related to the human nervous system, but in our attempt to achieve a better understanding of neural processing animal models have proven to be of extreme importance and will be used frequently in our discussion. We begin by describing the main concepts and terminology that will be used throughout this dissertation.

1.1 Resources and allocation

The term *computational resources* refers here to the ability of the nervous system to transform signals. The concept *resource* is used to emphasize that there is a certain amount of computation that can be done per unit of time, and a certain amount of information that can flow through the system, limited for example by the number of receptors and noise level. Energy, time and the number of sensory receptors, connections and muscles are the physical quantities underlying these resources. Computational resources are to be found at all stages of processing, from sensory to motor systems, including "intermediate" steps involved in performing mental calculations or recalling memories. *Allocation* refers to the process of distributing these resources in order to represent, transform and transmit different signals involved in perception and action.

As an example of allocation of resources, let us focus on the optic nerve of humans in the context of information transmission. Composed of approximately 1.2 million fibers per eye, its information capacity can be estimated, assuming independence and given statistics of the firing rates and noise characteristics, to be almost 10 Mbps (Koch et al., 2006). The signals going through these axons depend on the position of the eyes, implying that the nervous system is selecting the content that will be transmitted to the next stage of processing, or equivalently, allocating transmission resources to certain features of a visual stimulus. This process, when movement of the sensors takes place, is called *overt* allocation. In contrast, *covert* allocation does not require physical movement (except perhaps for ions moving across membranes making the signal transmission possible), and occurs for example when the retina adapts to different intensity levels and contrasts so as to employ better the dynamic range of the retinal cells (Baccus and Meister, 2002). In this dissertation, we will focus mainly in allocation phenomena related to orienting attention, both covertly and overtly.

Orienting phenomena can be *voluntary* or *automatic*. It is voluntary, for example, when we try to find a friend in a crowd, or when we focus on one conversation in a noisy environment filtering out other sounds that reach our ears (usually known as the *cocktail party* effect). Orienting is *automatic*, for instance, when our focus changes after hearing someone say our name. These two types of phenomena are usually referred to as *endogenous* (task

driven) and *exogenous* (stimulus driven), respectively. In our discussion, we will use the term orienting not only to indicate phenomena in which we focus on a particular location, but also for those situations in which we focus on particular features (*e.g.*, when we use the fact that our friend in the crowd is wearing a red hat), or when focusing on one out of many overlapping objects (like the effect we get when looking through a reflective window). These are commonly known as *spatial* attention, *feature-based* attention and *object-based* attention, respectively.

The phenomena introduced above are usually presented under the name of "attention". It is therefore necessary to clarify that orienting is just one type of attentional phenomenon, and to interchange the two terms would be misleading. To explain the differences it is best to follow Michael Posner's proposal of the three attentional systems (Raz and Buhle, 2006): The first type is referred to as the *alerting* system and it is related to our readiness in response. The second one, called *executive*, deals with conflict resolution, like in the Stroop task (MacLeod, 1991). The third is called *orienting* and involves the *selective* attention phenomena described above. This dissertation focuses on topics related to the third attentional system and in cases were the word "attention" is used, it will refer to either covert or overt orienting.

The thesis supported by this dissertation states that principles of optimality account for orienting phenomena. Our investigation includes a discussion of behavioral data as well as neuronal activity measurements, an analysis of the underlying neural mechanisms, and principled attempts to explain the *raison d'être* of orienting phenomena. In this discussion, the nervous system is seen not as a straight channel from sensors to muscles, but instead as a system that combines beliefs and current observations in order to make decisions that eventually bring reward to the organism.

1.2   Motivation: the neural code

Researchers working in the multidisciplinary field of neuroscience are driven by different motivations including philosophical questions in regard to what we "are", using this knowledge to create machines that behave as we do, trying to cure the myriad of diseases related to the nervous system, or improving human performance, among others. My personal motivation de-

rives from the believe that a deeper understanding of the nervous system will allow the development of techniques that will revolutionize human-to-human and human-machine communication.

Understanding the functioning of the nervous system requires the development of a set of general principles that give rise to the behavior observed in humans and other animals. These principles should be powerful so as to predict, accurately enough, the outcome of any experiment or real world situation, given the state of the system (which includes influences from the past). These principles can be stated at different levels, from molecular to cellular to whole brain. A level of particular interest in our context is that one from which we can predict percepts, mental images, and decisions, based on neuronal activity alone. In other words, we are looking for the coding principles used by the nervous system to represent our sensations, memories and actions, or as it is commonly called: the *neural code*.

In our study of the neural code, we must keep in mind some ideas that have challenged traditional views of neural processing. First, neuronal activity may not represent only the value of physical quantities (or its estimated mean), but also the uncertainty associated with these values, or even, the whole probability distribution of particular features (Knill and Pouget, 2004). Secondly, the code may not be fixed with respect to a stimulus or action, and adaptation at different scales may have different influences on it. Striking evidence for this idea comes from electrophysiological recordings during selective attention tasks showing that even when stimuli remained fixed, neuronal activity, and therefore the code, change as a function of attentional state (Reynolds and Chelazzi, 2004). If we intend to understand the code, we must understand the adaptation phenomena that influence it. Third, the stimulus best encoded by a neuron is not necessarily the one corresponding to the peak of the neuron's tuning curve, since two nearby stimuli are most easily discriminated when they fall in the region of highest slope of the tuning curve. The relative importance between these two alternatives seems to depend on the experimental context and level of variability of the neuronal response (Butts and Goldman, 2006). Fourth, the precise time of action potentials has been shown to encode additional information not represented by the firing rate alone (Rieke et al., 1996).

How will we know if we have found what we are after? The following

three conditions seem to be necessary: **(1)** Given a stimulus (or action or mental "image"), we must be able to predict the neuronal activity at specific stages of processing. **(2)** Given neuronal activity, we must be able to estimate the stimulus (or action or mental image) associated with it. **(3)** We must be able to modify perception (or action or imagery) by directly stimulating the neural circuit. Some examples of these conditions have already been demonstrated. The book by Rieke et al. (1996) contains various examples of predicting spike trains and firing rates of the H1 neuron of the blowfly for different visual stimuli with changing velocities. Stanley et al. (1999) were able to reconstruct some features of natural images from the activity of LGN neurons in the cat. Romo et al. (1998) showed that monkeys could perform equally well tasks with vibrating stimuli on their fingertips, and tasks in which the stimuli were replaced by microstimulation of the primary somatosensory cortex. These examples do not imply that our quest is over, but show that these conditions are achievable to a certain extent, and that they provide a better understanding of neural coding principles.

1.3   Historical context

Orienting phenomena have been discussed in the literature for millennia, from Aristotle's comments on simultaneous stimuli (Aristotle, 2004, Ch. 7) to William James' infamous chapter from "The principles of psychology" (James, 1890, Ch. 11) to researchers publishing at least seven books on *attention* in the last ten years (Styles, 1997; Pashler, 1998, 1999; Parasuraman, 2000; Braun et al., 2001; Posner, 2004; Itti et al., 2005). Experimental paradigms involving covert orienting date back to the late 1800s and the studies of stereoscopy with instantaneous illumination. In the words of Hermann von Helmholtz: *"In this experiment the attention is entirely independent of the position and accommodation of the eyes or, indeed, of any known variations in or on the organ of vision. Thus it is possible, simply by a conscious and voluntary effort, to focus the attention on some definite spot in an absolutely dark featureless field. In the development of a theory of the attention, this is one of the most striking experiments that can be made."* (Helmholtz, 1924, Vol 3, p. 455).

Attention literature often mentions the arrival of behaviorism as partly responsible for the absence of research about internal processes, including

covert selective attention, during the period from the 1920s to the 1950s. We should however note that it was during this period that techniques such as electroencephalography (in the 1920s) and nuclear magnetic resonance (in the mid 1940s) were developed, together with the introduction of the Stroop task in 1935.

The psychology of selective attention entered a new stage in the 1950s. Significant work during this decade was conducted by Donald Broadbent who, based on his studies on the auditory system, suggested an early-selection theory in which filtering of unattended stimuli occurs before reaching awareness (Broadbent, 1954). Later experiments showed that some "ignored" information could actually reach consciousness or at least influence behavior. Anne Treisman was one of the most influential researchers suggesting alternative accounts (Treisman, 1969). These theories marked the beginning of the early *vs.* late selection debate, which is discussed in depth in Driver (2001). Current resolutions of the debate suggest that selection occurs at many levels of processing depending on the capacity for transmitting or transforming information at each stage (Lavie, 2001, 2005).

In parallel to psychological studies, two electrophysiological methods were being used in the 1960's and 1970's to study the neural correlates of selective attention. The first, single cell recordings in non-human mammals, showed activity modulation in the auditory cortex (Hubel et al., 1959; Hocherman et al., 1976), the superior colliculus (Goldberg and Wurtz, 1972), the frontal eye fields (Wurtz and Mohler, 1976) and the parietal cortex (Lynch et al., 1977). The second method, electroencephalography (EEG) in humans, was performed mostly during auditory tasks, and helped characterize changes in the components of event-related potentials as a function of attentional state (Picton and Hillyard, 1974).

During the 1980s and until today, researchers continued to study covert and overt orienting phenomena with techniques ranging from single cell recordings to whole brain imaging and psychophysics, describing neural correlates of selective attention in visual, auditory and other modalities in humans, other mammals and even insects (van Swinderen, 2005).

1.4   Remaining questions

In an interview by Michael Gazzaniga (Gazzaniga, 1996, p.45), Michael Posner stated the main questions an attention researcher should focus on: What is the anatomy of the networks that subserve the operations of attention? What is their real-time activation? How do they develop? And what goes wrong with them in brain injury and psychopathology? In addition, Posner and other researchers have shown the importance of studies of attention in the scientific understanding of consciousness (Posner, 1994). A decade later, these questions remain as the main goals of attention research.

In the context of the neural code, more detailed questions can be posed based on those above. In particular, we would like to know how the transformations that sensory pathways produce on their inputs are modulated by changes in the attentional state of the subject. We would like to characterize the response of neurons and populations not just with respect to stimuli but also with respect to context, given for example by the attentional state. In addition, we would like to be able to infer the attentional state of a subject from the neural activity. Preliminary results related to this goal have already been reported (Crowe et al., 2005).

Researchers are also interested on finding those stimulus attributes that capture attention, and characterizing how we explore changing environments. The exact features that draw attention for each modality are still under debate (see for instance Franconeri et al., 2005), but principled models, applied to visual saliency, are starting to provide interesting predictions (Itti and Baldi, 2006).

The ultimate question is probably *why have attention at all?* Or in other words, why did orienting mechanisms evolve? What are the advantages of having them? And, is it actually possible to separate sensory and action processing from attentional mechanisms, or are they so dependent on each other that it does not make sense to think of one without the other? The next chapters will evaluate some of the possible answers to these questions.

1.5   The modeling approach

Most of the contributions presented in this dissertation are derived from modeling work. Models can be of different types as defined by Dayan and

Abbott (2001, Preface): *Descriptive models* summarize large amounts of experimental data, but their primary purpose is to describe phenomena, not to explain them. *Mechanistic models*, on the other hand, address the question of how nervous systems operate on the basis of known anatomy, physiology and circuitry. Such models often form a bridge between descriptive models at different levels. *Interpretive models* use general principles to explore the behavioral and cognitive significance of various aspects of the nervous system functioning, addressing the question of why nervous systems operate as they do. We can think of the different modeling approaches as follows: descriptive models tell *what* happens, mechanistic models describe *how* it happens, and interpretive models explain *why* it happens.

This dissertation is concerned mostly with interpretive and high-level mechanistic models, but models of all types are presented as background. Ad-hoc methods in which quick "hacks" are added to previous models to account for extra features are avoided. We favor instead principled approaches where some cost function is optimized to account for receptive field formation, coding strategies and orienting phenomena in a unified fashion.

Models are by no means proof of the validity of a hypotheses, but help reassure the relevance of the general principles suggested. One advantage of the modeling approach, as opposed to just theorizing with words, is that it forces you to state the assumptions explicitly and to be clear about how the relevant processes actually work (O'Reilly and Munakata, 2000). Theoretical approaches complement experimentation by suggesting new measurements and predicting their outcomes. By developing new theories consistent with our observations, we refine at each step our understanding of nature.

1.6   Dissertation outline

The dissertation is organized as follows: First, evidence for limitations in perceptual and computational resources in the nervous system is presented in Chapter 2. This chapter describes performance measurements during attention tasks and presents models that account for these observations. The third chapter describes the model used throughout the first part of the dissertation, focusing on its input-output behavior. The model demonstrates

that allocation of resources emerges as a natural property of systems that implement principles of optimal coding under non-uniform relevance of the input features. The next two chapters discuss the effects of attention on neuronal activity. Chapter 4 reviews the literature concerning measurements of attentional modulation of neuronal activity, from whole brain to single cell. This chapter also includes a description of models that account for these phenomena. In the following chapter we use the model introduced earlier to show that, in addition to resource allocation effects, many phenomena related to the attentional modulation of neuronal activity can be accounted for by principles of optimal coding. Predictions with regard to this modulation are derived from simulations. Chapter 6 investigates the neural mechanisms underlying attentional phenomena, describes experimental observations that shed light into the origin of attentional control signals, and explores the relation between attentional mechanisms and the development of receptive fields.

The remainder of the manuscript explores overt mechanisms in relation to the resource allocation hypothesis. In Chapter 7 we introduce the concept of *active perception*, and derive optimal strategies for the case of simultaneous tracking of two targets. Predictions from this model are tested in Chapter 8, where we discuss results from experiments in which eye-positions are measured while subjects simultaneously track multiple dynamic targets.

In the concluding chapter we summarize the contributions presented in this dissertation, evaluate their implications, and suggest future work in relation to orienting systems and the neural code.

Chapter 2

Behavioral measurements of orienting systems

Summary

This chapter introduces traditional experimental paradigms employed in the study of the psychology of attention. In addition, it discusses the relation between covert and overt orienting, and presents computational models that account for phenomena observed during selective attention tasks.

2.1 Covert attention improves performance

It is not difficult to realize that directing our senses towards an object of interest increases our ability to detect it and identify it. A more intriguing phenomenon is the change in performance associated with covert orienting. In this case, it is not our body but our "mind" which turns its focus to a particular location or stimulus attribute. Researchers have shown that this orienting-without-movement also improves our perception of the attended features.

Attending to something does not always imply physically orienting our senses towards it. For instance, we are able to focus on different instru-

ments of an orchestral piece without having to move our head or body. Playing team sports provide another example. In this case we visually track the movement of multiple objects simultaneously, and without having to follow a single player we can keep track of their locations as they move around and cross their paths.

Improvements in performance when attending to a stimulus can be either in the speed or the accuracy of its detection or discrimination. In experiments by Downing (1988), subjects were cued to expect a target at a particular location away from the fixation point, and then asked to perform tasks at this and other locations. The tasks consisted of luminance detection, brightness discrimination, orientation discrimination and form discrimination. Results from these experiments showed that perceptual sensitivity (measured by the area under the R.O.C. or the d' statistic) was higher for regions closer to the cue location, indicating that visual performance can indeed improve without having to move the eyes (Fig. 2.1A). Other experiments showed that spatial attention manipulated either by cueing or by removing concurrent tasks improves visual thresholds differently for different tasks: from improvements in contrast detection thresholds of about 20%, to improvements of up to 70% in orientation and spatial frequency discrimination thresholds (Lee et al., 1999). In addition to changes in sensitivity, researcher have shown that detection latencies are reduced when subjects receive a cue that indicates where in the visual field the signal will occur (Fig. 2.1B) (Posner et al., 1980; Corbetta et al., 1993).

Analogous phenomena have been observed in non-visual modalities. In the auditory system, for instance, knowing what frequency to listen for improves detection and discrimination, but does not seem to reduce reaction time. In contrast when information is available about the spatial locus of a sound, reaction times are shorter although detection and discrimination may not be any better (Scharf, 1998). Johansen-Berg and Lloyd (2000) provide a review of the psychology of attention in the somatosensory system.

2.2 Divided attention

Once it was shown that we could allocate processing resources to particular attributes and spatial locations, researchers started examining the ability of
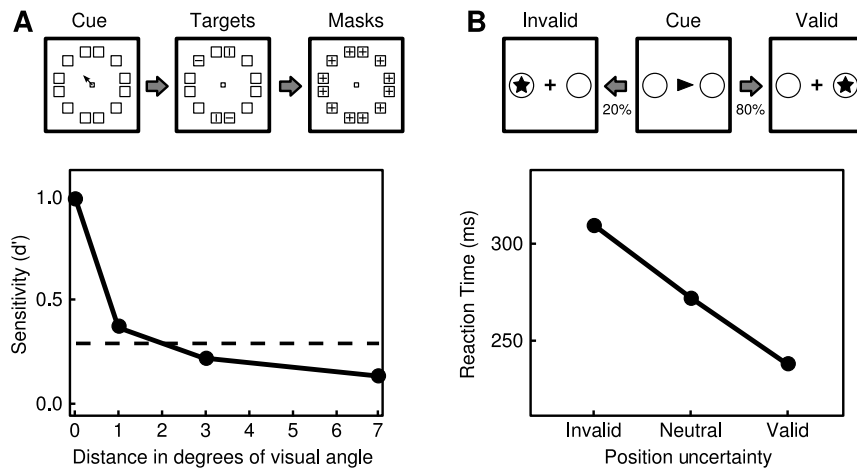
Figure 2.1: Covert attention improves performance. (**A**) Peripheral cueing improves detection of targets close to cued position. An arrow was displayed indicating that a target was likely to occur in that location. Subjects had to maintain fixation in the center of the display and report the content of the cued box and 3 other boxes after they had been masked. The plot shows the sensitivity at different distances from the cued location. The dashed line indicates sensitivity in neutral (no cue) trials. Reproduced from Downing (1988). (**B**) The validity of the cue has an effect on the detection time of visual targets. First, a cue was presented indicating that a target was likely to occur at that location. Reaction times are plotted comparing valid, invalid and neutral trials. Reproduced from Posner et al. (1980).

attending to multiple modalities at the same time, or similarly, the possibility of splitting resources between non-contiguous regions of the feature space.

Pashler (1994) reviews the literature on concurrent tasks, focusing on simple experiments in which performing one task interferes with the performance of another. In particular, he evaluates the slowing effect that occurs when a subject tries to perform two speeded tasks at the same time, the so-called "psychological refractory period".

When the tasks are not completely separate but instead subjects monitor two different channels for the presence of a signal requiring a speeded response, an interesting effect occurs: when signals are presented simultaneously in both channels detection is faster, on average, than when presented on either channel alone. This effect is known as the "redundant signals effect" (Miller, 1982).

In spatial attention studies there is a common analogy relating the focus of attention with a spotlight. This analogy suggests asking if this beam can

be split or has to be continuous in space. Recent experiments suggest that the spotlight may indeed be divided between spatially separated locations excluding regions in between (Awh and Pashler, 2000; Müller et al., 2003).

2.3   Why have covert orienting if we can move our bodies?

One of the most common paradigms in attention research is *visual search* (Wolfe, 1998). In these experiments, subjects have to look for the presence of a target immersed in distractors. Usually, reaction time and accuracy are measured and compared against those for displays where no target is present. In the laboratory, visual search tasks usually require the subject to fixate on a center point and perform the search by covertly attending to the stimuli. Interestingly, when comparing the same task with and without eye movements, reaction times did not differ significantly (Klein and Farrell, 1989). This observation motivates questions concerning the relation between saccadic eye-movements and covert attention.

The premotor theory of attention is the traditional explanation relating eye-movements and covert allocation, and it states that subjects use spatial attention in the programming and execution of saccadic eye movements. It is supported by experiments showing a deviation on saccade trajectories when covertly attending to a different location than the saccade target (Sheliga et al., 1994), and experiments in which detection accuracy was shown to be highest when the location of the target coincided with the location of the saccade (Hoffman and Subramaniam, 1995). Visuospatial attention is then regarded as an important mechanism in generating voluntary saccadic eye movements. Some of these ideas may need to be reevaluated to take into account experiments in which dissociation between covert and overt attention in both pop-out and voluntary conditions has been shown to be possible (Hunt and Kingstone, 2003a,b).

Studies of the dynamics of attention have shown the high control subjects have over the covert allocation of resources, with the ability of reliably improving performance by attending to particular moments in time (Nobre, 2001). Other studies have shown that the dwell time, *i.e.*, the time spent at each item in a display, can be as short as 50 ms in visual search tasks as well as tasks in which targets are presented sequentially without

13

masks (Egeth and Yantis, 1997), even if sequential presentation with masking gives times as long as several hundred milliseconds per item (Duncan et al., 1994). Furthermore, measurements of reaction times for two-target comparisons at different distances while fixating, have shown that it does not take longer for attention to move greater distances (Egeth and Yantis, 1997).

In other modalities, for example during auditory tasks, attentional effects are less dependent on the position of the body and sensors and still provide spatial and feature selectivity (Scharf, 1998). All the evidence suggests that covert visual attention is a fast mechanism for allocating resources and it allows for selective phenomena where movements of the sensors are irrelevant, like pre-attending to a particular color, or looking through a window while ignoring the reflection that it produces. It seems intuitive that these "action without movement" mechanisms, which are advantageous for perception, evolved together with the sensory systems.

2.4   Saliency

A phenomenon that has played an important role in attention research is that of saliency. It refers to those properties of the incoming sensory signals that draw attention more strongly than others. This phenomenon is not necessarily dependent on overt orienting. Salient features can draw resources without producing movements of the body, and in some cases strongly salient features or stimuli can be automatically detected without the need of a serial search, and effect called "pop-out".

The traditional characterization of pop-out effects has been derived from the slopes of the curves that result from plotting reaction time *vs.* the number of items in the display in visual search tasks. The modern view in visual search research uses the terms *efficient* and *inefficient* search, as opposed to serial and parallel, to indicate that there is a continuum of these slopes depending on the properties of targets and distractors (Wolfe, 1998).

2.5   Measurements of overt orienting

As shown in previous sections, measurements of covert attention require indirect indexes from reaction times and accuracy of detection and discrim-

ination. In contrast, in the study of overt orienting we can directly measure resource allocation by recording eye and body positions. Researchers looking at overt orienting would like to characterize those stimulus features that guide eye-movements during given tasks in addition to achieving good predictions on these movements (Liversedge and Findlay, 2000).

In this dissertation, we focus mainly on saccadic eye-movements, without exploring in detail other movements and compensatory mechanisms like the vestibular-ocular reflex (to compensate for head movements), smooth pursuit (to smoothly track moving targets) or vergence (to maintain an image of an object at a given depth in the same position in both retinae).

Advances in virtual environments and eye-tracking technologies allow for measurements of eye movements during natural behavior while subjects freely explore complex environments (Hayhoe and Ballard, 2005). Studies making use of these technologies have found further evidence about the importance of the task in eye-movements, and the relation between the task and learning where and when to fixate. In addition, these studies suggest that internal rewards guide eye and body movements and that these phenomena can be appropriately modelled by principles from reinforcement learning theory (Sprague and Ballard, 2004).

Chapter 8 presents experiments in which eye-movements are recorded during simultaneous tracking of multiple targets, in order to test hypotheses related to active perception.

2.6 Modeling orienting phenomena

In an attempt to understand orienting phenomena, models at many levels of detail have been developed. These models are expected to reproduced observed phenomena, and to provide a starting point for the design of new experiments, generating predictions on the outcome of new measurements. This section reviews models of selective attention and orienting, focusing on their input-output behavior. Some of these models generate further predictions regarding the neuronal activity associated with orienting phenomena; these aspects will be presented in later chapters of this dissertation.

Visual search has been one of the phenomena that has received most interest when deriving models of attention. In this respect, two influential proposals are worth mentioning. First, the Feature-Integration Theory
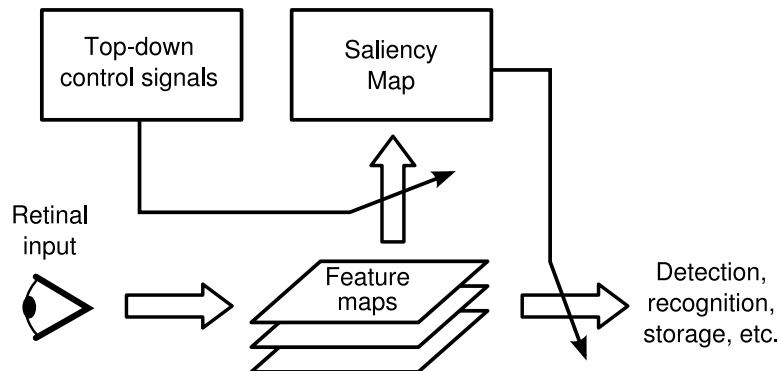
Figure 2.2: Schematic representation of the different modules of traditional models of attention. Feature maps represent orientations, simple motion, colors, intensities and other basic features of the input image. These features compete for representation in the saliency map. This competition is biased by top-down signals depending on the requirements of the task. The saliency map gates the flow of information between basic features and higher level modules for recognition and storage.

(Treisman and Gelade, 1980) was introduced to account for many phenomena observed when comparing detection and discrimination of features against conjunction of these features. According to this theory, features like color, orientation and brightness are registered early, automatically, and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention. This theory predicted the differences in slope observed in feature *vs.* conjunction search paradigms. It also predicted phenomena related to texture segregation, illusory conjunction, identity and location, and interference from unattended stimuli. Later studies revealed that some conjunction searches were too efficient to be described as "serial" searches, requiring a reevaluation of the theory (Wolfe, 1998).

The second theory, the Guided-Search model, argued that attention could be guided to likely conjunctions by combining information from more than one pre-attentive processes. For instance, two sources of information (coding for "red" and "vertical") could combined into a saliency map, so that objects having both features will be doubly activated. These conjunction will be then efficiently found if attention is directed to the locus of greatest activation (Wolfe, 1998).

Connectionist models of attention, consisting of networks of simple neuron-like processing elements, have implemented these ideas, provid-

ing a concrete description of the assumptions involved and accounting for an increasing number of phenomena. Mozer and Sitton (1998) review the main features of these models. Most of these models have detailed architectures inspired in the visual system, with topographic organization and defined receptive and projective field sizes. The main model of attentional selection presented in this review can recognize letters in parallel, or sequentially select locations to control the flow of information preventing the recognition system from being overloaded. The model is composed of an object recognition network with primitive feature maps together with an attentional (saliency) map that gate the flow of information in the recognition network. This modeling approach contains many of the features that have been traditionally used in mechanistic models of attention (Fig. 2.2): feature maps, saliency maps, mechanisms for inhibition of return and explicit gating of information flow (Koch and Ullman, 1985).

This type of model reproduces cue-validity effects observed in psychophysical experiments in humans (Posner et al., 1980). First, the stimulus is presented to the network by activating the primitive feature maps. If the network is started from an inactive attentional state, locations containing the stimuli will activate until one of them becomes preferred through competition, and the primitive feature units representing this location are allowed to pass their activity through the recognition network. To model validity effects, the network is started with a cued location represented by activity in the attentional map prior to the stimulus presentation. This will generate faster responses to stimuli in cued locations. For the invalid cue condition, the dynamics of the attentional network and in particular the winner-takes-all mechanism for selecting a region, will take longer in getting activated, generating a slower response in the detection/recognition of the stimulus. The model also reproduced other phenomena like changes in reaction time when the presentation interval between the cue and the target are reduced, and the slowing effects of distractors on reaction times. The results from this and other models suggest that spatial attention should be regarded as the result of competition among locations.

With small modifications, the model presented above can also reproduce phenomena observed during visual search tasks. In particular, it produces a flat curve of reaction time *vs.* the number of items in the display for single feature search, while giving a higher slope curve when the search is

performed for conjunction of features such as a red vertical target among red horizontal and blue and red vertical distractors.

Another proposal that includes some of the ideas described above has been called *biased competition* (Desimone and Duncan, 1995). This type of model proposes that first, the relevance of particular input features is defined according to the task; then, a short-term description of relevant features, called the attentional template, is used to control competitive bias between inputs, favoring those that match the description.

Abstract mathematical approaches like the Theory of Visual Attention (TVA) (Bundesen, 1990) use these principles to account for attentional phenomena, and have been shown to have a neural interpretation (Bundesen et al., 2005). The equations in TVA jointly describe two mechanisms of attentional selection: *filtering*, for the selection of objects, affecting the number of cells in which an object is represented, and *pigeonholing* for the selection of features, which implements a multiplicative scaling of the level of activation in cells coding for particular features.

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_z w_z} \tag{2.1}$$

$$w_x = \sum_j \eta(x,j)\pi_j \tag{2.2}$$

In this model, a module representing visual short term memory encodes categorizations of the form "object x belongs to category i". The rate equation or filtering (2.1) defines a value $v(x,i)$ representing this categorization, which depends on the attentional weight w_x of object x and a perceptual decision bias β_i associated with a particular category i. Each category has also a corresponding measure of importance of attending to objects that belong to that category, named *pertinence* and denoted by π_j. In the weight equation or pigeonholing (2.2), weights are calculated from pertinence values together with the evidence $\eta(x,i)$ that the particular object belongs to each category. The operation of these equations is better described with the following example (Bundesen et al., 2005): Consider the task of reporting red digits from a mixture of red and black digits. To select the red objects, the pertinence value π_{red} of the visual category red is set high, keeping other values low. This speeds up the processing of all types of categorizations of red objects. To perceive the identity of red digits, one perceptual

decision-bias parameter β_i per digit is set high, keeping other parameters low. This speeds up the processing of categorizations with respect to digit types. This combined effect of pertinence and decision-bias parameters is what makes the processing of red digits faster with respect to any other characterization. This model and its extensions has been shown to reproduce a wide variety of psychological effects on visual discriminability and bias, selection from multi-object displays, effects of divided attention, joint effects of numbers of targets and distractors in partial report, and consistent practice in search, among others.

A review of computational models of visual attention by Itti and Koch (2001) focuses on bottom-up phenomena and complements the list of modeling principles by highlighting the importance of five features: the dependence of perceptual saliency on context, the existence of an explicit and unique topological saliency map, inhibition of return (and allowing attention to be released from the current location) , the tight interplay between covert attention and eye movement, and how scene analysis and object recognition constrain the selection of locations.

Other approaches to attention modeling have been derived from statistical theories of estimation and inference, which conceptualize top-down modulations as changes in priors. For example, Yu and Dayan (2005) present a network that performs inference on a visual orientation variable while an attentional signal sets a prior for a location variable. This model display changes in reaction times depending attentional state, consistent with the cue-validity effects presented above.

Computational modeling of eye-movements have used ideas from statistical inference and reinforcement learning, and suggest that eye movements serve to reduce uncertainty about the environmental variables that are task relevant (Sprague and Ballard, 2004; Jaramillo and Pearlmutter, 2006d). A good example is presented by Najemnik and Geisler (2005) who derived the ideal Bayesian observer that defines the optimal way of making saccades in order to find a target in a scene. They observed that humans performed close to optimal leading them to hypothesize that humans compute the analogous of a posterior probability map and use it to determine the next fixation. This way, subjects can reach near-optimal performance, despite poor memory for visual details and poor integration across fixations. Along similar lines of thought, a mathematical formulation of *sur-*

prise has shown to be useful at predicting eye-movements (Baldi, 2005; Itti and Baldi, 2006).

Chapter 3

Resource allocation in multi-layer perceptron networks

Summary

This chapter extends the discussion on computational models started in the previous section. Here, we describe a family of models that are based on principles of optimal coding under non-uniform feature relevance.

Main contribution

We show that resource allocation emerges naturally in models that implement principles of optimal coding under non-uniform relevance. These model are shown to account not only for spatial-based phenomena of selective attention, but also for phenomena in which features are attended independent of their location.

3.1 Optimal coding under non-uniform relevance

Because of limitations in the "hardware" that encode and transmit sensory stimuli in the nervous system, the neural representation of physical quantities is in general *lossy*. In other words, when a stimulus contains more

information than the neural hardware and code can handle, some of this information has to be discarded. There is therefore a tradeoff in the fidelity of the representation of different features of a stimulus, with some of these features being more relevant than others depending on the task at hand. This uneven relevance implies that the strategy for allocating representational resources that maximizes performance may differ for each task.

Following these ideas, we developed a model in which the encoding of a stimulus could have different quality requirements for each feature of the input, and these requirements may change depending on context. In order to test the performance of the encoding procedure, a decoder that reconstructed the input after it has been encoded in a lossy fashion, was also modelled (Fig. 3.1). An error measure between the input and output defined the quality requirements of the reconstruction, with a different error measure representing each one of the possible states of the system. The particular error used at each instant depended on an additional modulatory signal, denoted as **A** in Fig. 3.1, which informed the system about the current state. The encoder-decoder system was trained to minimize the error function for different inputs and states. We show that once the model has found the optimal codes given its resources, it reproduces orienting phenomena by allocating resources according to the task, as well as producing, when implemented in a neural fashion, emergent modulation of the activity of its units that is consistent with electrophysiological recordings (Chapters 4 and 5).

The following sections present a auto-encoder multi-layer perceptron network that serves as an implementation of the principles described above. The cost function is chosen so that the error for each input unit is weighted non-uniformly depending on the task. Each error function was associated with a value of an additional signal that informed all layers about the current attentional state. The stimulus is presented as a monochromatic image for illustration purposes only, since the ideas presented here apply to different modalities and do not attempt to simulate a particular stage in the processing pathway. Once the error minimization procedure optimized and fixed all the connection strengths between units in the network, allocation of resources resulted only from the non-linearities of the system according to the modulatory signal, given that the error function was not explicitly represented anywhere in the network.

3.2 Feedforward network with modulatory input

Network structure. An auto-associative network was constructed consisting of five layers connected in a feedforward fashion (Fig. 3.2B). The number of units in each layer was 256–20–10–20–256, respectively. Each unit received inputs from all units in the previous layer, in addition to two attentional signals (displayed as a single arrow per layer in Fig. 3.2B). This attentional input was the same for all layers. There were no lateral connections within units in a layer. The activity in the input and output layers was represented as an image of 16×16 pixels.

Unit model. A firing rate model was used in which the output of each unit was calculated as the weighted sum of the inputs passed through a saturating nonlinearity, as follows:

$$
\begin{aligned}
r_i &= S\left(\sum_j w_{ij}\, r_j + b_i\right) \\
&= S\left(\sum_{j'} w_{ij'}\, r_{j'} + \sum_{j''} w_{ij''}\, r_{j''} + b_i\right)
\end{aligned}
\tag{3.1}
$$

where the sum is over all units from the previous layer (indexed by j') and the attentional inputs (indexed by j''). The saturating function was:

$$
S(x) = a \tanh(bx) \qquad \text{with} \quad a = 1.716, \quad b = 0.667
\tag{3.2}
$$

as suggested by Haykin (1999, p.179). The activity of unit i is denoted r_i and the parameters w_{ij} correspond to the strength of the connection from unit j to unit i. Note that each unit i also included a bias term b_i; this term can be replaced for a constant input for which its weight is also optimized. The connection strength w_{ij} from unit j to unit i was real valued and unbounded.

Optimization. The optimization process consisted on finding the set of weights and bias parameters that minimize the cost function $E = \langle E_{\mathbf{p}} \rangle$ where

$$
E_{\mathbf{p}} = \sum_k c_k(\mathbf{p})\,(y_k(\mathbf{p}) - d_k(\mathbf{p}))^2
\tag{3.3}
$$

and k indexes locations in the 16×16 grids holding the stimulus and its reconstruction, $c_k(\mathbf{p})$ defines the importance of that particular location (analogous to the intensity of an attentional spotlight), $y_k(\mathbf{p})$ is the output of the network, $d_k(\mathbf{p})$ is the desired output, which is in our case the same as the input, and $\mathbf{p}$ represents the complete pattern of information coming into the system at one point in time, *i.e.* the input image as well as the top-down attentional signal. The expectation (angle brackets) is taken over the input patterns $\mathbf{p}$.

The gradient was calculated using backpropagation (Rumelhart et al., 1986) of the weighted error defined in Equation 3.3, and optimization used online gradient descent with a weight decay term of 10^{-6}. All weights were plastic during learning. See Appendix A.2 for the derivation of the learning rule.

To limit the capacity of the system, which is theoretically unbounded for real-valued units (except for effects of numerical precision), zero-mean Gaussian noise with standard deviation of 0.1 was added to each bottleneck unit's total input during optimization.

3.3   Spatially-driven allocation

3.3.1   Simulation parameters

Stimulus statistics.   The set of patterns used during optimization consisted of 20,000 16×16 pixel monochromatic images. Pixel intensity values had zero mean and standard deviation $\sigma = 1/3$. The images were created by convolving (filtering) white Gaussian noise images with a rotationally symmetric 2D Gaussian with $\sigma_{filter} = 2$. Edge effects were avoided by extracting only the 16×16 center of the resulting image. These images were then scaled to have the desired variance. See Appendix A.1 for examples.

Attentional input.   The attentional signal consisted of a two-element vector with values in the range $[-1, 1]$. For each optimization step, this input was randomly drawn from a uniform distribution over the possible range.

Optimization parameters.   In our simulations the learning rate was set to $\eta = 0.005$ and the attention coefficients in the penalty function formed a
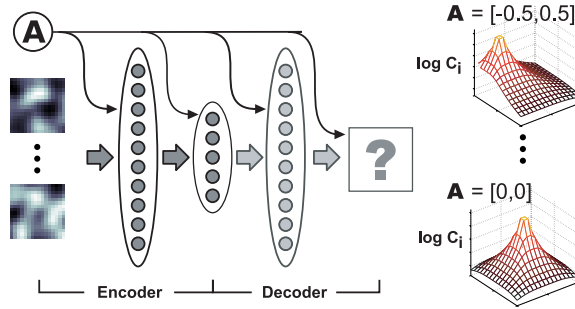
Figure 3.1: Inputs during optimization of the connection strengths. The auto-encoder network received, at each step, a random input image (left) together with a random attentional input (right) associated with a particular error weighting soft-mask.

simple soft mask:

$$c_k(\mathbf{p}) = \frac{1}{1 + m^2 \|\vec{k} - \vec{a}(\mathbf{p})\|^2} \tag{3.4}$$

with $\vec{a}(\mathbf{p})$ being the attentional input (in our case, a two-dimensional vector representing the center of attention) and $\vec{k}$ being a location in the plane. The width of the attentional spotlight was set by m, which was held constant at $m = 12$ in our simulations. Fig. 3.1 shows some examples of input values and corresponding attentional masks.

3.3.2 Results

Performance of the system at encoding and decoding the input was measured using random patterns independently generated from those in the training set. For performance comparison, a network that used a flat penalty function $c_k(\mathbf{p}) = 1$ was also trained (Fig. 3.2C). The results presented here correspond to measurements on the system after the optimization procedure has found the appropriate connection strengths, thus, any modulation is due only to changes in activity and not to changes in the structure or connection strengths of the network.

Fig. 3.2A shows an example of the reconstruction of one test pattern. Here, the input image remained fixed as the attentional input changed. The dashed circles indicate those regions for which preferential reconstruction was requested. The reader must be reminded that the attentional input consisted only of two additional values (which the network will interpret as the center of attention), and that these signals entered each layer the
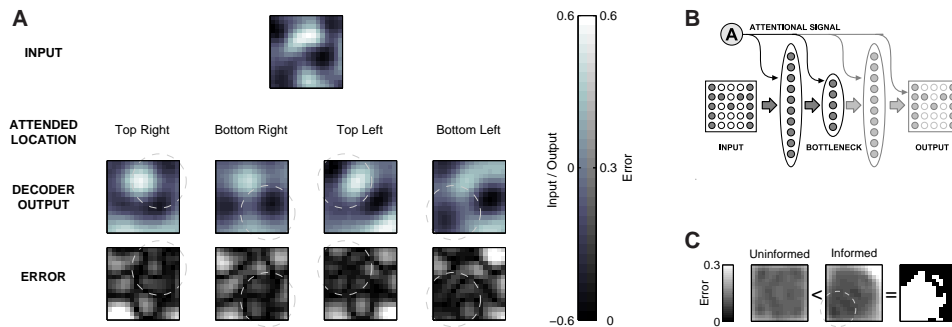
Figure 3.2: Reallocation of resources was observed when attentional signal changed. (**A**) Example of reconstruction of a single input pattern and four different attentional states as indicated by the dashed circles. The error was calculated as the absolute intensity difference between input and output. Error is lower for attended locations. (**B**) Structure of the feedforward network used in the simulations. For each unit, attentional inputs are indistinguishable from sensory inputs. (**C**) Average error for one attentional state compared against a system with no attentional signal. The white region in the image on the right indicates lower error when the system makes use of the attentional signal.

same way as feedforward (sensory) inputs. The regions represented by the dashed circles were not explicitly represented by any signal in the network during this simulation (only during optimization, see Section 3.2). This example shows that the quality of the output is not the same for all regions, and that it depends on the value of the attentional signal, achieving a lower error for those regions closer to the center of attention.

In addition, we tested how these results compared to those from a classical model in which there is no informing signal. Fig. 3.2C compares the average reconstruction performance of a system that ignores the informing signal with a system attending to the bottom-left corner of the input image. The difference shows that reconstruction was better for attended regions but worse for unattended ones, suggesting that coding resources were indeed reallocated.

3.3.3 Discussion and limitations

The key result from these simulations is that uneven allocation of resources emerges as a natural property of a system when designed to encode its inputs so as to achieve minimal error under non-uniform fidelity requirements. This approach contrasts modeling studies in which explicit gating of information flow is implemented in order to obtain selective perfor-

mance in higher stages of processing (Mozer and Sitton, 1998). We must note in addition that selective performance could be achieved by a system that uniformly encodes all features and later throws away those that are irrelevant, wasting the resources assigned to them. Instead in the model presented here, resources that were in other conditions assigned to some input features are now used to improve representation of relevant inputs. Other emergent properties of this model concerning the modulation of activity of its units will be presented in Chapter 5. Details on the mechanisms, including the fact that after optimization of the connection strengths top-down signals are indistinguishable from sensory inputs, will be discussed in depth in Chapter 6

Our simulations incorporated a number of simplifications, most of which were made for ease of exposition or computational efficiency:

- During optimization, the penalty function was set to a single attentional spotlight. This is not a requirement of the general model, and other functions could be used to define performance demands. For example, non-spatial goals can be incorporated requiring higher fidelity in the reconstruction of a feature that occurs in the input regardless of its location (Section 3.4).

- For display purposes, simulated stimuli represented visual patterns. Efficient representation of stimuli and attentional modulation phenomena are present in many (if not all) modalities and the model explored here applies to these non-visual scenarios.

- The data displayed were collected after the optimization procedure had been run and the connection strengths had been fixed at their optimal values. In nature, we would expect this type of adaptation to continue as the system performs the task.

- The goal of the network in the simulations was to find an efficient representation for the stimulus. Biological networks likely transform stimuli not to merely compressed representations, but to representations that serve to inform future actions.

- As noted by Arbib (2003, p.21), backpropagation of error *"is an example of 'neurally inspired' modeling, not modeling of actual brain structures;*

and there is no evidence that backpropagation represents actual brain mechanisms". We are not suggesting that the optimization procedure chosen for these simulations is the one used by the nervous system. The only requirement is that the system associates an additional signal with different cost functions. Evaluating the biological plausibility of optimization procedures is outside the scope of this chapter, but it is important to note that biologically plausible learning rules that could achieve the same results as backpropagation of errors have been suggested in the literature (Mazzoni et al., 1991).

The limitations presented above are specific to the current simulation and not to the general model proposed here, or to the predictions it makes. The model has focused on endogenous phenomena, in which covert orienting is defined by the task, but the model does not require the additional input to be a top-down, and could as well be a bottom-up signal (*e.g.*, according to saliency) in order to produce stimulus driven attentional phenomena. The undefined origin of the attentional signal will not be explored further in this chapter, but it plays an important role in the remaining of the dissertation.

The resource allocation hypothesis implies that improving performance on one task can occur only at the cost of reducing the amount of resources available for other concurrent tasks. An alternative idea is that attention reduces uncertainty about the location of targets improving performance by decreasing the number of noise sources. Luck et al. (1996) evaluated these hypotheses using spatial cueing experiments in which the uncertainty of the target location was eliminated. They concluded that resource allocation still occurs when enough time is allowed for this mechanism to take place.

The next section extends the results presented above by introducing a different cost function, which makes allocation to be driven by particular features of the stimulus instead of a region of interest.

3.4 Non-spatially driven allocation (feature-based)

3.4.1 Simulation parameters

Following the network architecture described in Section 3.2, we designed cost functions that could account for feature-based phenomena. In this case

ERROR	H	V
Att H	0.33	0.38
Att V	0.42	0.35
Flat	0.37	0.38
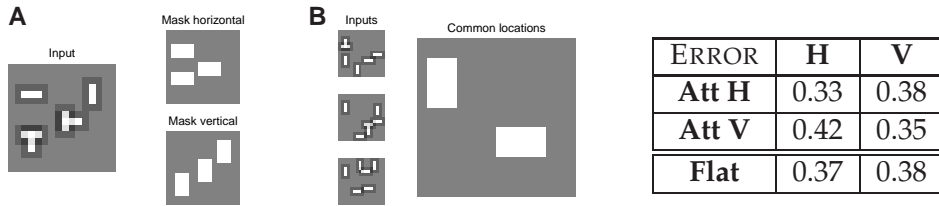
Figure 3.3: Feature-based allocation. (**A**) Example of input pattern and associated attentional masks. (**B**) Performance was tested by generating many patterns for which one vertical and one horizontal bars were fixed while other bars were randomly positioned for each pattern. The table on the right shows the average errors over these common locations (H and V) for different attentional states (AttH and AttV). Results for a system that ignores the attentional input are also included for comparison (bottom row).

the attentional signal does not inform the system about the location of interest but about particular features of interest. We simulated a system that learns to focus on either horizontal or vertical lines. This system differs from that one presented in previous sections in the following:

- The attentional input is now unidimensional and takes only two values, -1 or $+1$.
- The training set is now composed of patterns containing only horizontal and vertical bars (sometimes overlapping) with 3 pixels of length. In the monochromatic scale we use to visualize stimuli, bars are always white with a dark surrounding, and the background is always gray (corresponding to a value of zero).
- The mask used during optimization is a sharp mask around the bar stimuli and its surrounding (Fig. 3.3A).
- The attentional value -1 is associated to a mask favoring vertical bars. Similarly, a value of $+1$ favors horizontal bars.

The number of units per layer was kept the same as in previous simulations: 256–20–10–20–256. Twenty thousand random patterns and their corresponding masks were used for training, using the algorithm described in Appendix A.2. The learning rate was set at $\eta = 5 \times 10^{-4}$.

3.4.2 Results

Fig. 3.3A shows an example of input patterns and corresponding attentional masks. To test the performance of the system for different attentional states, we created a set of patterns in which a particular location always

had a vertical bar, and another location always had a horizontal bar. The rest of the image contained bars located randomly for each pattern in the training set, as shown in Fig. 3.3B. The table in Fig. 3.3 presents the average error of reconstruction at the locations containing the horizontal and vertical stimuli for the two attentional states (attend horizontal or attend vertical). The table also compares these error with those obtained when the system is optimized with a flat mask, *i.e.*, ignoring the attentional signal. As expected, reconstruction error is lower when the particular feature is attended. If we take the average error for a horizontal bar when attending to horizontal bars, we note that it is lower than the error to vertical bars in this condition, and to the error of horizontal bars when attending to vertical bars. The dual is also true. When comparing against a flat mask, we expect the error to fall between the error values for the two attention conditions. Results for the flat mask indicated lower or equal error compared to the unattended condition, and higher error when compared to the attended condition, suggesting that resources are being reallocated.

3.4.3   Discussion

Results from the feature-based selection simulations show that a simple principle of associating an additional signal to different error functions can reproduce not only spatial selection effects but also location-independent feature selection phenomena.

Traditional models of feature-based attention have required maps that explicitly code for each feature of interest. These maps subsequently influence the processing system by enhancing activity corresponding to particular features (Mozer and Sitton, 1998; Itti and Koch, 2001). Here we have shown it is not necessary to explicitly define such maps, yet the optimization procedure can still find the appropriate codes and produce differential fidelity between features.

The same limitations presented for the case of spatial-based selection apply here. These are again specific to the simple implementation presented here, and not to the general ideas that these principles encompass. Further simulations are required to evaluate the performance of this system in tasks in which either simple features or conjunction of features must be identified (Chapter 2).

In summary, these results are further evidence that allocation of re-

sources emerges naturally in neural systems optimized to perform tasks with non-uniform relevance on the input features.

3.5   Concluding remarks

This chapter described models in which principles of optimal coding under non-uniform fidelity requirements are implemented by a multi-layer perceptron network performing various tasks. We introduced the models, focusing only on their input-output behavior and flexibility. Further evaluation of the activity of the units in the model networks, together with predictions with regard to the modulation of this activity, will be presented in Chapter 5.

The most important result at this level of analysis is that resource allocation emerges naturally from the coding principles implemented here. A system could also obtain differential performance by simply encoding all features uniformly, and later selecting out irrelevant stimuli, discarding some resources. Instead, better performance can be obtained if those resources are used for representing, transmitting or transforming relevant signals. This new take on the debate between early *vs.* late selection suggests that whenever the hardware allows it, it is advantageous to allocate resources early to features of interest.

Chapter 4

Attentional modulation of neuronal activity

Summary

Neuronal activity in response to a fixed stimulus has been shown to change as a function of attentional state. This chapter summarizes results from these experiments and describes models that account for these observations.

4.1 BOLD/fMRI and PET measurements

Functional brain imaging techniques based on metabolic activity have been used since the early 1990s to describe neural correlates of selective attention phenomena in humans. These correlates correspond to changes in the activation of particular brain regions as inferred from positron emission tomography (PET) or from blood oxygenation level dependent (BOLD) signals measured by a magnetic resonance scanner. The tasks contrasted in these studies differed either on the attended modality or on the attended attribute of the stimulus within a single modality. For example, early PET experiments by Corbetta et al. (1990, 1991) showed modulation of activity

as a function of the attended visual stimulus feature: shape, color or speed. These changes in activity occurred in regions of the extrastriate cortex that appear to be specialized for processing the feature of interest. Their measurements also showed differences in activation between selective attention (attending to only one feature) and divided attention (attending to multiple features) tasks.

By the late 1990s, functional MRI was the technique of choice of cognitive neuroscientists and many fMRI selective attention studies have been published since. Early studies showed differential activation in the human homologous of area MT when subjects were paying attention to stationary versus moving dots (O'Craven et al., 1997) or to different features (color and speed) of moving dots (Beauchamp et al., 1997).

Imaging studies allowed researchers to test the topographical effects of selective attention (Tootell et al., 1998; Brefczynski and DeYoe, 1999) and to map the network of regions involved in attentional functions (Corbetta and Shulman, 2002). As an example, Fig. 4.1 presents results from the retinotopy study by Brefczynski and DeYoe (1999). The figure shows the modulation of coherence between the BOLD signals and modelled responses during visual selective attention tasks.

Using these imaging techniques, researchers have found attentional modulation of activity in a great variety of cortical and subcortical areas involved in sensory processing, including structures as early as the LGN of the thalamus (O'Connor et al., 2002). More details on the functional imaging of visual selective attention and the relation of these findings to those from other techniques can be found in review papers by Kanwisher and Wojciulik (2000); Kastner and Ungerleider (2000) and Pessoa et al. (2003).

4.2 Modulation of brain rhythms and synchrony

Using electro- and magnetoencephalography (EEG/MEG) measurements, which provide higher temporal resolution than the techniques introduced in the last section, researchers have reported changes in brain rhythms associated with attention and cognition. For instance, Gruber et al. (1999) used EEG measurements to show that the power of signals between 31 and 51 Hz (a subset of the gamma band) increased when subjects were covertly attending to a rotating pattern in one half of the screen as opposed
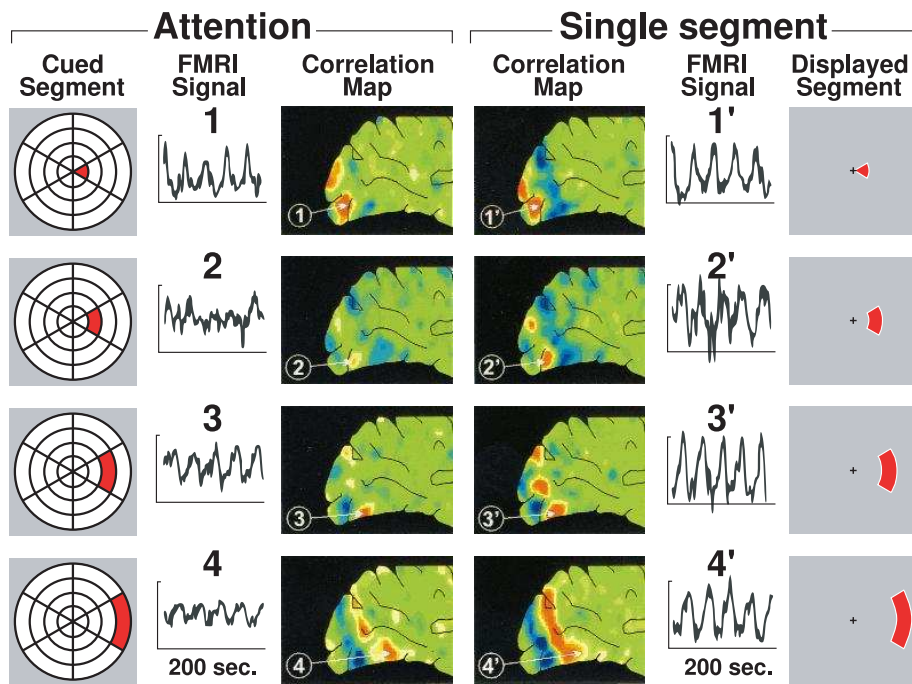
Figure 4.1: Retinotopy of visual attention. Response amplitude for each voxel was estimated as the covariance between the fMRI response and the idealized response waveform. Maps show correlation coefficient. Higher correlation (indicated by the arrows) on areas in V1 corresponded retinotopically to attended locations. The stimulus consisted of a circle divided in segments at different angles and eccentricities, each one containing colored oriented stripes that changed randomly. The subject's task was to monitor changes on a cued segment while fixating in the center of the display. High coherence areas during the cued experiment (left) corresponded to those areas with high coherence when the segment was presented in isolation (right). Reproduced from Brefczynski and DeYoe (1999).

to attending to a static pattern in the other half. These changes were more prominent in parieto-occipital regions contralateral to the attended location, and interestingly, they were higher when rotation on the left hemifield was attended. Another study from the same group showed that the steady-state visual evoked potential can also be modulated by selective attention (Müller et al., 1998).

Modulation has also been shown to occur in non-visual modalities. For instance, using MEG during somatosensory stimulation, Bauer et al. (2006) observed an increase in induced activity in the range 60-95 Hz, mostly in somatosensory cortex contralateral to the attended hand. Furthermore, the measurements showed an increased suppression of alpha (7.5-12.5 Hz) and

beta (14-26 Hz) rhythms in occipital sensors contralateral to the attended side.

Researchers have hypothesized that the synchronization of oscillatory responses of spatially distributed, feature selective cells might be a way to establish relations between features in different parts of the visual field (Gray et al., 1989; Engel et al., 2001). Extending this hypothesis, it has been suggested that localized changes in synchronization may also serve to amplify behaviorally relevant signals. Some studies have started to provide neural correlates of attention based on changes of the level of synchronization between neurons activated by the attended features. In experiments by Fries et al. (2001), multi-unit activity (MUA) and local field potentials (LFPs) were recorded from multiple V4 sites in monkeys. They measured the power spectra of the spike-triggered averages of the LFPs in two conditions: one in which the monkey was attending inside the receptive field (RF) of the recorded neuron, and one attending outside the RF. The spectra showed an increase of synchronization in the range 35-60 Hz (gamma band) and a reduced synchronization in lower frequencies (below 20 Hz) during the attended condition.

Changes in the synchrony of action potentials between neurons have also been observed in the somatosensory system. In this case, different levels of synchrony were observed for pair of neurons when comparing tasks that require monkeys to attend to either tactile or visual stimuli (Steinmetz et al., 2000).

4.3   Modulation of evoked potentials and evoked fields

Steven Hillyard from the University of California in San Diego has studied the effects of selective attention on the electromagnetic evoked responses in humans for over 30 years. An early study (Picton and Hillyard, 1974) measured the attentional modulation of the different components of the event related potential (ERP) during an experiment in which subjects either attend or ignore an auditory stimulus. This study concluded that the amplitude of the N1 (90 ms) and P2 (170 ms) components was modulated by attention but there was no clear modulation of earlier components, indicating that human auditory attention is not mediated by a peripheral gating mechanism.

With the advent of high spatial resolution imaging techniques like PET and fMRI, researchers were able to characterize the location together with the dynamics of the attentional effects. Heinze et al. (1994) combined PET with ERP recordings and showed that visual inputs from attended locations receive enhanced processing in the extrastriate cortex (fusiform gyrus) at 80-130 ms after stimulus onset. Hillyard and Anllo-Vento (1998) provide a review of this and later studies evaluating the modulation of ERP components during spatial and non-spatial attentional tasks.

During the last decade, researchers have continued to take advantage of the fast temporal resolution of EEG and MEG to test hypotheses regarding the dynamics of selective attention and orienting processes. A review paper by Luck et al. (2000) discusses some of these studies, and notes the discrepancy between estimates for the duration of attentional shifts that has been measured in the visual system (from 50 ms to 500 ms).

4.4 Firing rate modulation

Electrophysiological recordings of single cells was one of the first techniques that allowed researchers to measure neural correlates of attentional effects. These experiments are of most relevance for our discussion since they allow exploring the effects of attention at the level of small circuits and provide the main method for characterizing the response of individual neurons. Early studies using these techniques focused on the modulation of firing rates when the modality of interest was either attended or ignored. For instance, Hubel et al. (1959) recorded from the auditory cortex of cats and found cells that responded to noises only when the animal was looking at the sound source.

With the development of better stimulation and recording techniques, researchers started dissociating areas presumably responsible for directing attention from those receiving the influence of these attentional signals. Lynch et al. (1977) presented measurements from the parietal cortex of monkeys during visual tasks supporting the hypothesis that there exists within the parietal lobe a neuronal mechanism for directing visual attention, for the fixation of gaze on objects of interest, for maintaining visual grasp of the object if it moves slowly, and for loosening fixation and initiating rapid saccadic movements towards new objects of interest presented

within the visual field. Earlier studies had also shown the involvement
of the superior colliculus (Goldberg and Wurtz, 1972) and the frontal eye
fields (Wurtz and Mohler, 1976) in the generation and execution of sac-
cades.

These, together with later studies suggested that neurons in posterior
parietal cortex and the pulvinar in the thalamus were involved in covert
attention, whereas the frontal eye fields, prefrontal cortex, caudate nucleus,
and superior colliculus showed selective enhancement of a visual response
only before the execution of an eye movement to the stimulus location (Cor-
betta et al., 1993). This "limited involvement" ideas are continuously being
challenged by experiments showing that some of these areas, *e.g.*, the su-
perior colliculus, also contribute to the control of covert spatial attention
(Müller et al., 2005).

To further dissociate the effects of saccades from attention itself, Moran
and Desimone (1985) designed an experiment in which monkeys had to
maintain fixation while performing a task away from the fixation point.
This paradigm for measuring the effects of covert attention on neuronal
activity has been used extensively, as summarized below. In these exper-
iments the same stimulus is presented while attention is directed either
inside or outside the receptive field (RF) of the cell(s) being measured. In
addition, and more relevant for our discussion, some experiments test the
effects of directing spatial attention to one of many locations inside the re-
ceptive field. The following sections describe and compare measurements
from different regions of the visual system of monkeys when performing
covert attention tasks.

4.4.1 Primary visual cortex: V1

Moran and Desimone (1985) reported that when presenting one stimulus
inside and one outside the receptive field of cells in V1, attention did not
have a significant effect on the response of these cells. Because of the size
of the RFs in V1 they could not test the effects of having multiple stim-
uli inside. In contrast, Motter (1993) reported that activity in area V1 (as
well as V2) could be modulated by directing attention inside or outside the
receptive field of the cells, and argue that the presence of competing stim-
uli is necessary to observe these effects. McAdams and Maunsell (1999)
supported these findings showing that 31% of their recorded neurons in

primary visual cortex displayed attentional effects.

Further evidence of attentional modulation of firing rates in V1 comes from studies of curve tracking (Roelfsema et al., 1998; Khayat et al., 2006). In these experiments, a monkey was presented with two curves, one of which was connected to the fixation point. The animal had to maintain fixation for an initial period and then make a saccade to the end of the target curve. The response of V1 neurons before the saccade was stronger when their receptive fields covered the target curve compared to the ignored curve.

4.4.2   V2 and the ventral stream: V4 and IT

Moran and Desimone (1985) also measured changes in activity of cells in areas V4 and IT when attention was directed to a preferred stimulus versus a non-preferred stimulus, with both presented inside the receptive field of the cells. Responses were attenuated when attention was directed to the non-preferred stimulus compared to those when attending the preferred stimulus. They observed a stronger effect for cells in V4 than for those in IT. With their particular stimulation protocol, they did not observe significant effects when comparing attention directed inside *vs.* outside the RF of cells in these areas. Motter (1993), in contrast, showed that the activity of 45% of the recorded cells in V4 and 39% in V2 was modulated depending on whether attention was directed inside or outside the receptive field. These modulatory effects were dependent on having multiple competing stimuli in the display, and most cells did not display attentional modulation when a single stimulus was presented. This may explain the apparent contradiction with the observations from Moran and Desimone (1985). These experiments also showed that while the firing rate was higher when attention was directed inside *vs.* outside the receptive field, these changes occurred from either an increase or a decrease of activity with respect to baseline.

In a later study, Motter (1994) evaluated the effects of non-spatial attention on the modulation of activity in V4. In this case a cue of a particular color was presented before an orientation discrimination task. Only a bar that matched the cue color was relevant for the task. The activity of V4 cells was higher when the target matched the color of the cue compared to a control condition in which the monkey performed a discrimination task

at the fixation point. In contrast, activity was lower when the stimulus in the receptive field was the ignored object.

Luck et al. (1997) tested if the modulatory effects in areas V2 and V4 were different between sequential *vs.* simultaneous presentation of the competing stimuli. The measurements indicated a reduced effect (in magnitude and number of cells modulated) when stimuli were presented sequentially. These experiments confirmed that most cells had larger responses when attending to the preferred stimulus when multiple stimuli were presented inside the RF, but not when the other stimuli were presented outside the RF. In addition, they observed an increase in the baseline firing of cells in V4 (from 10.1 to 14.4 spikes/sec on average) when attention was directed inside the RF before any stimulus was presented.

With the interest of exploring if attention has an effect only in the processing of an attended target or also in the region of space it occupies, Connor et al. (1997) designed experiments in which attention was directed to different positions outside but close to the receptive field of cells in V4, while a bar stimulus was presented at different locations inside the RF. These experiments showed that changes in attention not only enhanced the responses to behaviorally relevant targets but also produce a complex modulation of the responses to other stimuli in the surrounding space. These results are discussed in more detail in the next chapter when we compare them to predictions from our models.

In order to test the biased competition hypothesis (Desimone and Duncan, 1995), which predicts smaller attentional effects in the absence of competing stimuli, Reynolds et al. (1999) presented further measurements from cells in V2 and V4 when two stimuli (of different preference) were presented inside the receptive field. They observed that with attention outside the RF, the level of activity of the cells for the pair of stimuli was between the activity generated by each stimulus alone. Once attention was directed to one stimulus, the activity changed as if this stimulus was presented in isolation.

Spitzer et al. (1988) compared the modulation of activity in V4 cells by changing the difficulty of the task. They observed that 81% of the cells responded stronger during the difficult task and that 77% of the cells showed a narrowing of their orientation tuning curve during this condition, concluding that increasing the amount of attention directed towards a stimulus

enhances the responsiveness and selectivity of those neurons that process it. Later experiments by McAdams and Maunsell (1999) showed that while most neurons with significant attention effects had higher rates in the attended conditions, the tuning curves did not change shape during different attentional states. They clarify that the discrepancy with previous studies may lie in how width is measured. From their measurements they inferred that attention was producing a multiplicative scaling of the tuning curves similar to the effect of changing contrast of the visual stimulus.

4.4.3 The dorsal stream: MT and MST

Attentional modulation of firing rates has also been observed in the dorsal stream of the visual system. Treue and Maunsell (1996, 1999) recorded from cells in the middle temporal (MT) and medial superior temporal (MST) areas of monkeys while presenting patterns of moving dots. Neurons in these areas, traditionally associated with the processing of motion, showed consistent modulation in their firing rates when attention was directed to preferred or non-preferred stimuli inside their receptive field. They also found modulation in responses when one stimulus was presented inside and another outside the receptive field, as attention was shifted between them. Treue and Martinez Trujillo (1999) further characterized this modulation by measuring changes in the tuning curve of cells in area MT, and showed that attention increases the gain of these neurons without changing the tuning width.

4.4.4 Other areas selective to visual stimuli

Higher areas where neurons have been shown to be selective to visual stimuli also display modulation of activity by attention. In the posterior parietal cortex (PPC) for instance, researchers have found neurons that code for the saliency of visual stimuli (Constantinidis and Steinmetz, 2005; Gottlieb et al., 1998), and neurons for which activity changed as a monkey followed an imaginary path without moving the eyes (Crowe et al., 2005). In addition, cells in the prefrontal cortex (PFC) show differential activation between targets and non-targets, even if they rarely show different response between non-targets (Everling et al., 2002, 2006).

4.4.5 Non-visual modalities

Modulation of firing rate has been observed likewise in other modalities. Hsiao et al. (1993) recorded from neurons in primary and secondary somatosensory cortex (SI and SII) of monkeys while they performed either a tactile or a visual task. Most neurons showed a significant modulation of their response when tactile stimuli presented to the fingers of the monkeys were relevant for the task. This modulation included both a increase or a decrease in firing rate for neurons in SII, but only an increase in rate for neurons in SI. Additional studies in the somatosensory system have shown that the modulation of firing rates in SII is multiplicative, without affecting the ratio between the variance and the mean rate (Sripati and Johnson, 2006).

In the auditory system, modulation of firing rates depending on the relevance of sound stimuli has been observed (Hocherman et al., 1976). In these experiments, the response of single units to identical stimuli were compared between conditions in which either a sound or a light was the relevant cue. The measurements indicated that response magnitude differed between the two conditions. Additionally, researchers have shown rapid changes in the spectrotemporal response field (STRF) of neurons in the auditory system of ferrets when comparing task *vs.* non-task conditions (Fritz et al., 2003).

4.5 Microstimulation studies

Researchers have used microstimulation techniques to study the causal relation between activity in different brain areas during selective attention and orienting tasks. Moore and Armstrong (2003) stimulated the frontal-eye fields (FEF) and recorded responses of neurons in area V4 when presenting preferred or non-preferred stimuli. They noticed that activity in V4 was enhanced after brief stimulation of corresponding sites in the FEF, even if the magnitude of the stimulation was not enough to evoke a saccade. Sub-threshold stimulation of the FEF also produced changes in performance during visual detection of peripheral targets (Moore and Fallah, 2001, 2004). In order to test the contributions of the superior colliculus (SC) to the control of covert spatial attention, Müller et al. (2005) evaluated vi-

sual performance while stimulating the SC. They found that microstimulation improved performance in a spatially selective fashion, even if monkeys maintained their gaze fixed. Using similar techniques, Cavanaugh and Wurtz (2004) had found that microstimulation of the cells in SC could produce improvements in change detection and reaction times in monkeys. In addition to these studies in the visual system, Winkowski and Knudsen (2006) showed microstimulation of gaze control circuits in the forebrain of the barn owl enhanced responses to auditory stimuli in neurons of the midbrain. This enhancement was space-specific and stimulation could also suppress responses at locations that did not correspond to those encoded by the stimulated forebrain sites.

4.6 Computational models

Early computational studies of attention focused on modeling behavioral data rather than modeling neuronal activity (Section 2.6). As the characterization of attentional modulation of neuronal activity was detailed further, and models increased their biological realism, predictions on the effects of attention on the activity of neurons at different stages were developed.

Olshausen et al. (1993) published a model that allowed for position- and scale-invariant representation of visual objects. A key element of the model was an attentional mechanism that dynamically routed information throughout the neural circuit. The model predicted changes in position and size of the receptive fields in higher areas of the visual ventral pathway as attention was shifted or rescaled. In addition, the model qualitatively matched modulation effects observed for cells in area V4 when attention is directed to effective or ineffective stimuli (Moran and Desimone, 1985).

Reynolds et al. (1999) suggested a simple model to account for changes in activity of cells in areas V2 and V4 when competing visual stimuli where presented inside or outside their receptive field. The model is defined by the following equations:

$$E = x_1 w_1^+ + x_2 w_2^+ \qquad \text{Total excitatory input} \qquad (4.1)$$

$$I = x_1 w_1^- + x_2 w_2^- \qquad \text{Total inhibitory input} \qquad (4.2)$$

$$\frac{dy}{dt} = (B - y)E - yI - Ay \qquad \underset{t \to \infty}{y} = \frac{BE}{E + I + A} \qquad (4.3)$$

where y represents the activity of the modulated cell, and the variables x_i represent activity from neurons that code for the features to discriminate, *e.g.*, orientation. The strength of the connections is define by w_i^+ and w_i^-, for excitatory and inhibitory synapses, respectively. Additional constant parameters that determine the passive decay (A) and the cell's maximum response (B) are included. In this model, attention is assumed to increase the efficacy of the synapses w_i from the cell(s) activated by the attended stimuli projecting to the cell of interest. The model was evaluated for different sets of randomly selected weights, by measuring in each case a *sensory interaction index* (activity of a stimulus pair minus activity of a reference stimulus) and a *selectivity index* (activity of a probe minus activity of a reference stimulus). Simulation results reproduced the effects observed in neurons from areas V2 and V4 under different attentional conditions.

The Theory of Visual Attention presented in Section 2.6 was extended by Bundesen et al. (2005), giving the equations a neural interpretation. This interpretation allowed for the reproduction of activity modulation effects when presenting single *vs.* multiple stimuli in the receptive field, effects of luminance contrast, multiplicative modulation of firing rates, and baseline shifts.

Other researchers have used information-theoretic approaches to predict changes in tuning curves that optimize some coding criterion like the Fisher information (Dayan and Abbott, 2001, Section 3.3). Nakahara et al. (2001) explored, from the perspective of population coding, the effects of height and base rate of the tuning curves on the encoding accuracy as measured by the Fisher information. They observed that when curves have experimentally observed parameters, only an increase in base and height produced an increase in information, even though relaxing this constraint allowed for cases in which decreasing tuning height and base rate in some of the neurons improved encoding accuracy. Similarly, Schwabe and Obermayer (2005) implemented a model of a cortical microcircuit and evaluated the changes in parameters that would maximize Fisher information. They concluded that changing the gain of the excitatory neurons was the mechanism that produce results most consistent with attentional modulation of activity.

The effects of attention have also been modelled by circuits that implement statistical inference. A model presented by Yu and Dayan (2005) re-

produced, in addition to the cue-validity effects discussed in Section 2.6, the multiplicative modulation of orientation tuning curves observed for cells in V4 (McAdams and Maunsell, 1999). Another network that performs statistical inference and accounts for various attentional phenomena is presented by Rao (2005). In his model, neurons combine prior knowledge with sensory information in a probabilistic fashion according to Bayes' rule. A first set of neurons represent the probabilities on location, while another set represents the probabilities on feature identity, and one last layer (intermediate representation) combines all these together with image information. Spatial attention is simulated by increasing the prior probability $P(L)$ for a desired location. Results from the model are consistent with three phenomena in visual cortex: (1) a multiplicative modulation of orientation tuning curves in V4 when the subject attends to a stimulus inside the receptive field of the cell (McAdams and Maunsell, 1999); (2) the restoration of response in V2 and V4 when attending to a target in the presence of distractors, closely approximating the response when the target is presented alone (Reynolds et al., 1999); and (3) the influence of attention on neighboring unattended locations as measured by the response of V4 neurons (Connor et al., 1997).

In the next chapter, we describe how attentional modulation of neuronal activity emerges naturally from models that implement principles of optimal coding under non-uniform relevance of the input features.

Chapter 5

Activity modulation emerges from optimal coding principles

Summary

In this chapter we evaluate the effects of attention over the activity of the encoding units in the model introduced in Chapter 3. We compare the results from these simulations to single cell measurements from the dorsal and ventral pathways of the visual system of monkeys.

Main contribution

We show that the type of activity modulation found in single cell measurements in monkeys during visual attention tasks can be accounted for by a simple feedforward model that implements principles of optimal coding.

In 1988, David Zipser and Richard Andersen showed that an artificial neural network could be trained to perform the coordinate transformation required to read a target independently of eye position (Zipser and Andersen, 1988). Interestingly, when they measured the properties of hidden units in the network, they discovered responses and effects similar to those from neurons in the posterior parietal cortex, including gain-modulated receptive fields (Salinas and Thier, 2000).
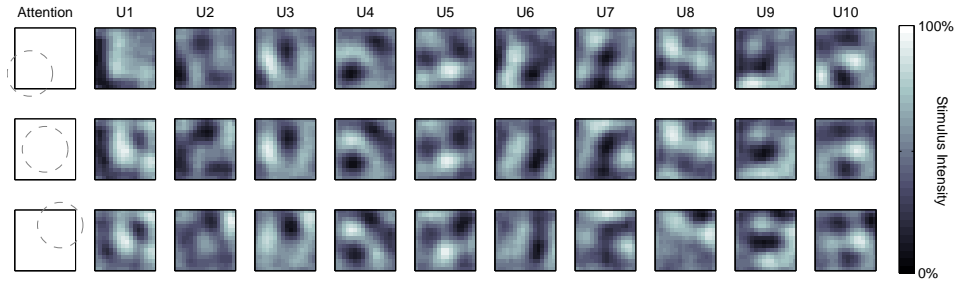
45

Figure 5.1: Preferred stimuli were dependent on attentional state. Color-coded images represent the preferred stimulus for each unit in the bottleneck layer (columns) for three attentional states (rows) as indicated by the dashed circles. Higher contrast was observed on attended regions.

The work presented in this chapter uses a similar approach in the context of covert spatial attention. Here, we investigate the attentional modulation of activity in the encoding units of the model described in Section 3.2, trained in the same fashion as it was done for the spatial attention simulations. All measurements evaluated here were taken after the connection strengths had been optimized, thus any modulation in the activity of the units emerges from the nonlinear interactions between the inputs to the system, and not from changes in the weight parameters.

5.1 Attentional dependence of preferred stimuli

We start by evaluating the preferred stimuli for each unit in the bottleneck layer. Here, "preferred" refers to the the stimulus that maximally drives each unit, and it is calculated using a method analogous to reverse correlation from spikes (de Boer and Kuyper, 1968). We first generated 10^6 random white Gaussian images and found the activation of the unit of interest for each of these patterns, keeping the attentional state fixed. The preferred stimulus was calculated as the average of these random patterns weighted by the activity produced in the unit of interest. The anti-preferred stimulus was defined as the negative of the preferred stimulus.

Subsequently, we tested if preferred stimuli depend on attentional state. Fig. 5.1 shows results from three conditions, as indicated by the dashed circles. These results suggest that preferred stimuli change slightly depending on the attentional state. In particular, preferred stimulus for each unit appear to contains sharper edges in regions closer to the center of attention.
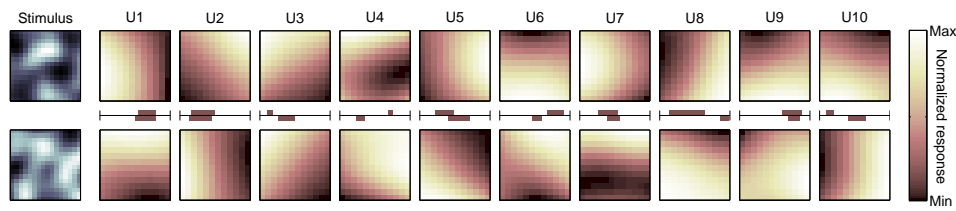
46

Figure 5.2: Activity of model units was modulated by the attentional signal. Color-coded images represent the activity of each bottleneck unit for a fixed stimulus as the center of attention is directed to different regions of the input space. Two different stimuli (top and bottom rows) are used for comparison. Maps are scaled according to the range of activities observed for that particular stimulus. The bars between the two rows display the range of activity for each of the two conditions with respect to the absolute limits of the activity of the units.

5.2 Attention-dependent modulation of activity

In this section we investigate how the activity in the encoding neurons is affected by the attentional signals. Fig. 5.2 shows maps of activity created by fixing the stimulus and moving the center of attention around the image. The intensity at each point in the map represents the activity of one unit in the bottleneck (normalized in the range of activities observed for that particular stimulus). The bars between the two rows of maps represent the range of activities achieved in each condition, with respect to the minimum and maximum possible value for model neurons. The pattern of activity modulation was clearly different for each unit. Furthermore, the modulation of a unit's activity was dependent on the stimulus. For example, for unit U1, higher changes in activity occurred when attention changed from left to right or from top to bottom depending on the stimulus. This effect occurred even when, in the case of U1, the activity ranges were very similar for the two conditions.

We related these findings to experimentally observed modulation of neuronal activity during selective attention tasks. First, we replicated the analysis presented in Treue and Maunsell (1999), in which combinations of preferred and anti-preferred stimuli were presented inside the receptive field (RF) of a cell and activity was measured as attention changed between these two regions. Fig. 5.3A shows the response from a neuron in the medial superior temporal (MST) area. The preferred stimuli for this neuron was a pattern of dots moving in one direction (indicated by the arrow pointing up). The attended stimulus is indicated by the dashed el-
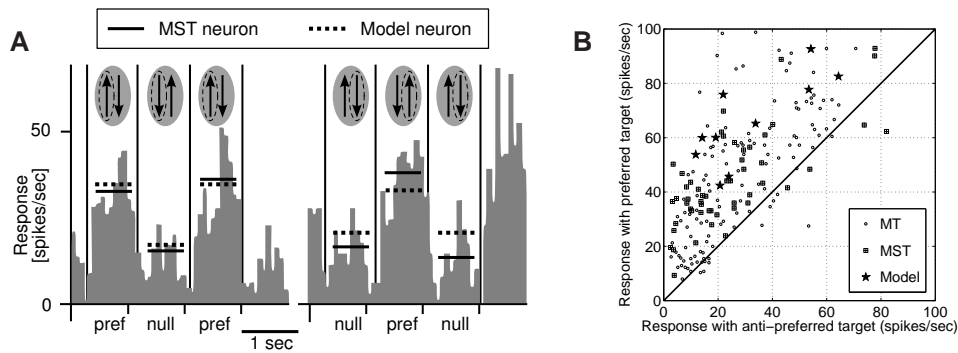
Figure 5.3: Activity modulation matched experimental measurements from area MST. (**A**) Neuronal response to two stimuli inside the receptive field of the cell, one preferred (arrow up) and one anti-preferred (arrow down). Attentional focus is indicated by a dashed ellipse. The histograms in gray show the firing rate of the cell as a function of time. Mean responses are indicated by solid horizontal lines (MST cell) and dotted lines (model unit) for each condition. (**B**) Scatter plot showing the response to anti-preferred stimuli *vs.* the response to preferred stimuli for each unit. Points above the diagonal indicate higher activity when attending to the preferred stimulus. Based on figures from Treue and Maunsell (1999).

lipse. The histograms in gray show the firing rate of the cell as a function of time, and average responses are indicated by solid horizontal lines for each condition. Overlaid, we show the response of one neuron from our model (dotted lines) under similar conditions. The stimuli consisted of combinations of the left and right halves from the preferred and anti-preferred stimuli, calculated with attention directed to the center of the input space. The maximal firing rate for the model neuron was set to 50 spikes/sec to obtain comparable magnitudes. The modulation of activity of the model neuron matched that of the visual neuron. In particular, when attention was shifted from preferred to non-preferred features of the same stimulus, activity decreased dramatically.

The scatter plot in Fig. 5.3B shows the firing rates for neurons in areas MT and MST when attention was directed towards the preferred stimulus (*y* axis) versus the firing rates obtained while attending to the anti-preferred stimulus (*x* axis). Points above the diagonal indicate higher activity when attending to the preferred stimulus. Values for all model units, indicated by stars, fall above the diagonal and within the range of experimentally observed values. For this plot, the maximal firing rate of all model units was set to 100 spikes/sec.

Further exploration of these effects is shown in Fig. 5.4. Responses to

48
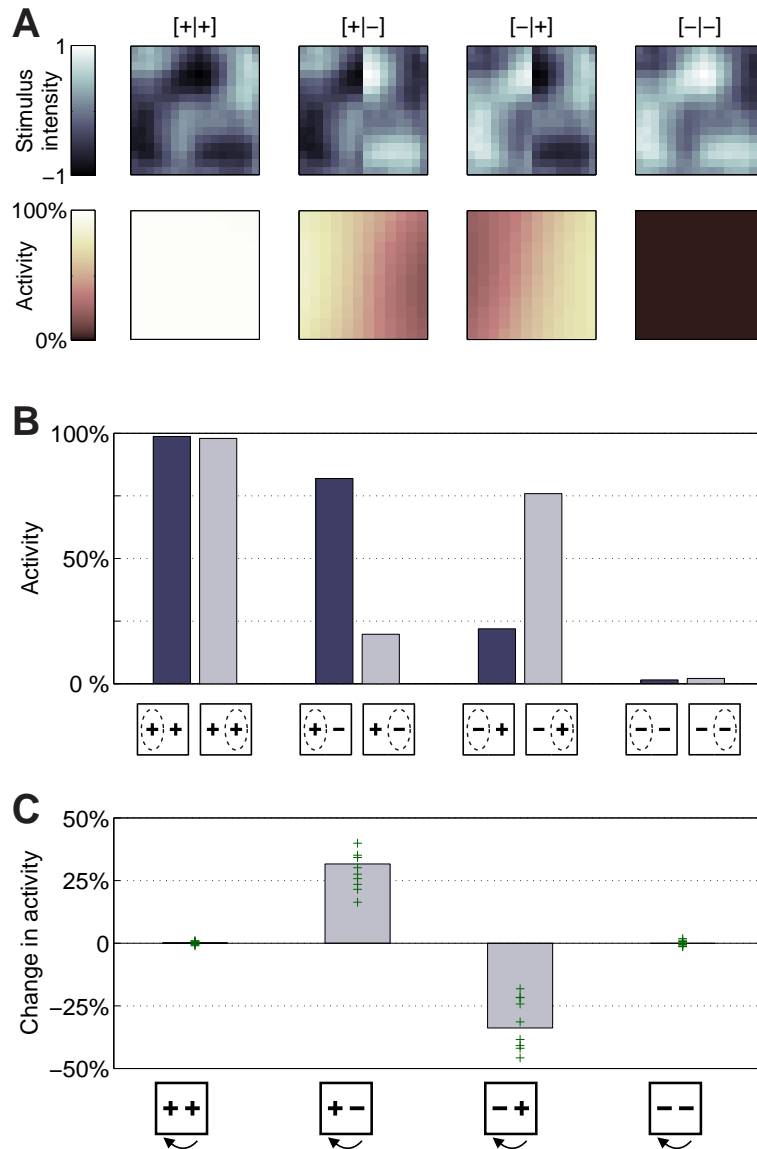
Figure 5.4: All model units showed similar activity modulation. (**A**) Combination of preferred (**+**) and anti-preferred (**–**) stimuli for unit U2 (top) and attentional modulation of activity in U2 for these inputs (bottom). (**B**) Comparison of activity in U2 for two attentional states as indicated by the dashed ellipses. (**C**) Changes in activity for each unit in the bottleneck as attention is shifted from right to left. The bars correspond to the average across units.

the four combinations of preferred and anti-preferred half-stimuli are presented, first for a single unit, and then for all units in the bottleneck layer. Fig. 5.4A shows the stimuli and corresponding attention maps for unit U2. The stimuli consisted of combinations of the left and right halves from the preferred and anti-preferred stimuli. Attention maps showed a clear change in the activity of the unit as attention moved from right to left. The simulation also showed that changes in features far from the attended region have smaller effect on activity than changes presented at the attended location (compare dark bars of Fig. 5.4B). As shown in Fig. 5.3, when attention was shifted from preferred to non-preferred features (of the same stimulus) activity decreased dramatically. In contrast, the effect of attention when both halves were preferred or anti-preferred was very small. These observations were consistent for all units as shown by Fig. 5.4C.

The model was also compared to experiments in which the response of cells in area V4 were measured for four attentional conditions, while a bar of fixed orientation was displaced inside the receptive field of the cell (Connor et al., 1996, 1997). Fig. 5.5A shows the response of one V4 cell. These plots contain various features that are common in attentional modulation:

(1) The response of the cell for a fixed stimulus depends on the attended location.

(2) The stimulus that elicits the strongest response depends on attention.

(3) The cell's response when attention is fixed depends on the stimulus; here the position of the bar.

(4) This dependence on stimulus position differs between attentional states; for instance, the left and right panels in Fig. 5.5A display different trends as the stimulus position is changed.

Fig. 5.5B shows the results from a model neuron under similar conditions. In this case, the input image is composed of a non-preferred stimulus with a small region belonging to the preferred stimulus at different positions, indicated by the numbers 1–5. Attention is directed to the borders of the image, as indicated by the circles. The model exhibited all features observed in the experimental data. Two shift indexes were also calculated for each neuron, after Connor et al. (1997). The *fractional shift* measures the proportion of total response that shifted from one side of the RF to the other when attention is shifted. This index is bounded between −1 and 1,
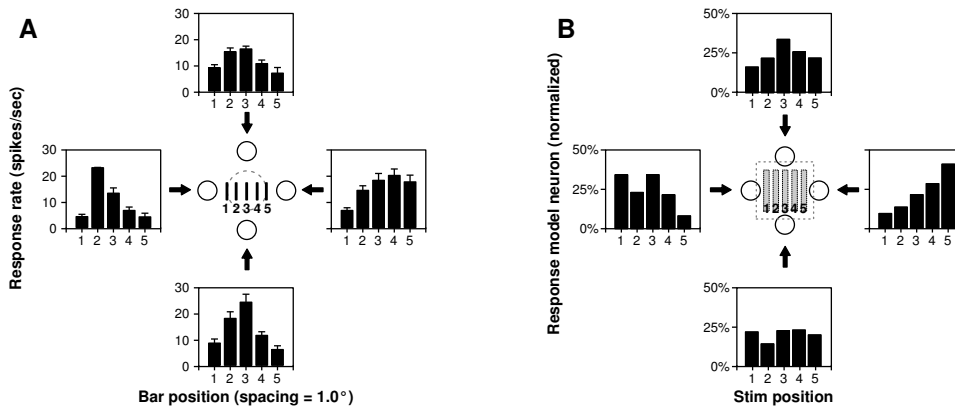
Figure 5.5: Activity modulation matched experimental measurements from area V4. (**A**) Response of one V4 neuron to a bar stimulus placed at five different positions inside the receptive field, as indicated by a dashed circle. Attention was directed to one of the four circles outside the receptive field. Reproduced from Connor et al. (1996). (**B**) Response of one model neuron as a region of a non-preferred stimulus is replaced by the preferred stimulus in five different locations. Attention is directed to the border of the input space as indicated by the circles.

with a positive value indicating shifts in the direction of attention. Connor et al. (1997) reported mean values of 0.16 or 0.26 depending on whether 5 or 7 bar positions were in use. All our model neurons had a positive fractional shift, with a mean value of 0.22. The second index, the *peak shift*, measures the distance between positions generating maximum responses. Mean experimental reported values were 10% or 25% of the RF size, depending on whether 5 or 7 bar positions were in use. Our model neurons had non-negative peak shifts with a mean value of 25% of the total position variation.

5.3 The magnitude of the attentional modulation depends on the system's capacity

The average modulation of activity of bottleneck units in the model as attention was shifted from left to right is shown in Fig. 5.6, with panel A showing how this varies with the number of bottleneck units, and panel C showing how this varies with the amount of injected noise. Gray open circles correspond to the absolute value of the modulation averaged over 1000 random test patterns, for each unit in the bottleneck. The solid circles

show the mean across units, with error bars indicating the standard error. Panels B and D show the corresponding reconstruction error for each set of parameters, plotting the mean errors for the attended versus unattended halves of the sensory input. Note that noise was added to the bottleneck units only during optimization, and for this reason changes in the modulation magnitude cannot be attributed to noise in the measurements.

The magnitude of the modulation decreased as the number of bottleneck units increased. As expected, errors in the ignored region were higher than those in the attended region, but decreased as the number of bottleneck units was increased, gradually closing the gap. As the noise level in the bottleneck units was increased, the magnitude of the modulation and of the reconstruction error both increased. The error for the attended region increased faster than for the unattended region. These two plots display an abrupt transition as the noise level becomes very high: the trend of the activity modulation, as well as the error values, changes, with the error values for the attended and unattended regions becoming equal.

5.4   Modulation of orientation tuning curves

Simulations represent only a small patch of the visual field, with each cell's receptive field filling the entire patch. Therefore, attention could not be directed outside the receptive field. Despite this restriction, it was still possible to analyze orientation tuning curves of the encoding cells (Ferster and Miller, 2000) and compare these effects to experimental results on the attentional modulation of these features (McAdams and Maunsell, 1999). For some units in the middle layer, the preferred stimulus could be approximated by a two-dimensional Gabor image (Dayan and Abbott, 2001, p.62). For these units, orientation curves were created by rotating a Gabor pattern and plotting the activity of the selected neuron as a function of rotation angle. Fig. 5.7 shows the resulting curve for units U6 and U3. These curves displayed attentional dependency. The plots in each panel of Fig. 5.7 correspond to the orientation tuning when attention is directed to the center of the rotating pattern (filled squares) or to a corner of the image away from the pattern (open circles). The observed changes for unit U6 were consistent with a multiplicative effect of the attentional signal on the unit's activity, *e.g.*, changes in activity were higher when activity was high. The model
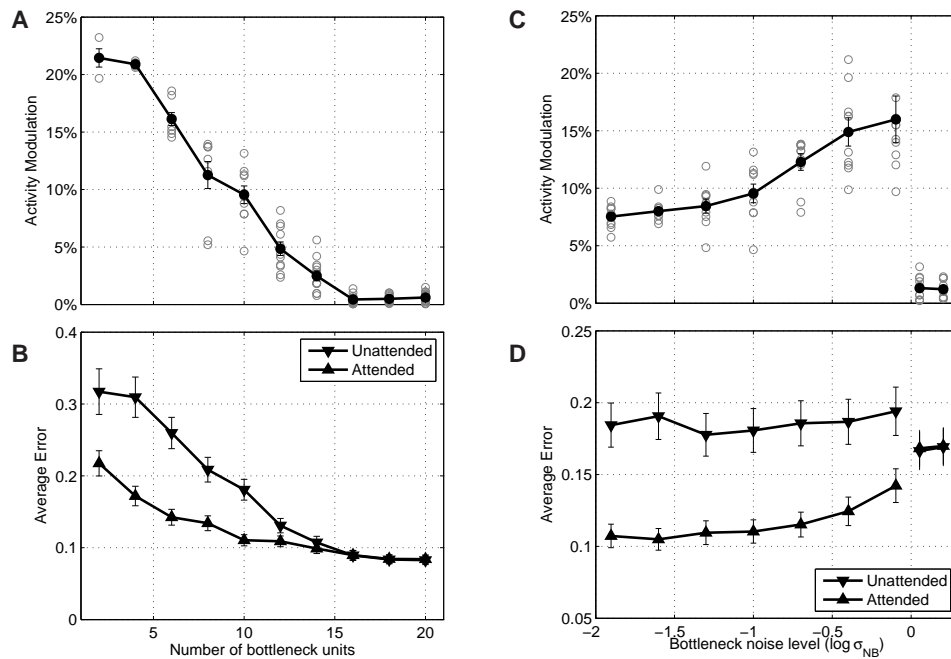
Figure 5.6: Magnitude of attentional modulation was dependent on the system's capacity. (**A**) Modulation of activity when attention was shifted from left to right of random stimuli, for networks with different number of units in the bottleneck. The noise level was fixed at $\sigma_{NB} = 0.1$. Open gray circles represent the average modulation for each unit. Solid circles show the average over all units with the corresponding standard error. (**B**) Average reconstruction error for the attended and unattended halves. Error was calculated as the average absolute difference between input and output pixel values. (**C**,**D**) Same as (A,B) but using a fixed number of bottleneck units, 10, and varying the noise level.
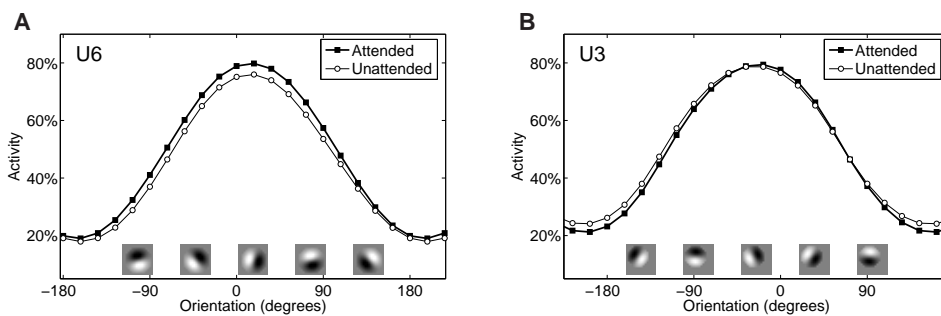
Figure 5.7: Attentional modulation of tuning curves. A Gabor pattern that resembled the preferred stimuli for units U6 (panel **A**) and U3 (panel **B**) was found and used to create orientation tuning curves. Each plot shows tuning curves for two attentional states, one centered at the rotating pattern (solid squares) and one away from the pattern (open circles). Activity in unit U6 was consistent with a multiplicative modulation, whereas modulation of activity in unit U3 was stronger for low firing rates.

did not include any multiplicative interaction, so any effect appeared as the result of activity from previous layers and the non-linear saturating function at each unit. Changes in activity for unit U3 resembled an inverted multiplicative effect in which modulation is stronger for lower firing rates (anti-preferred pattern). Because of the symmetry in the activity of the units in out model, a change in the sign of connection strengths in U3 would produce an effect analogous to that of U6.

5.5 Discussion and limitations of the model

The results obtained above support the idea that many phenomena related to the attentional modulation of neuronal activity can be accounted for by optimal coding principles under resource constraints and nonuniform requirements on representation fidelity. In addition, the simulations predict a complex interaction between attentional processes and neuronal responses, in that the characterization of the response of single neurons will greatly depend on attentional state. Lastly, the model predicts an increase in the magnitude of the modulation as the complexity of the stimulus increases with respect to the capacity of the system.

All the effects presented in this chapter were measured once the synaptic strengths had been fixed after optimization. Changes in the inputs were sufficient to obtain modulatory effects. The limitations presented in Section 3.3.3 for the particular implementation of the coding principles in a feedforward network, apply here likewise. Nevertheless, the model in its simplicity is still powerful enough to account for attentional modulation of activity in the ventral and dorsal pathways of the visual system, and to generate testable predictions. The simulation are limited to explorations inside the receptive field (RF) of a cell, and cannot therefore make detailed predictions about the form of RFs. Extended models would be necessary to predict attentional modulation of realistic RFs. For instance, incorporating sparsity constraints into the optimization procedure should give rise to localized RFs (Olshausen and Field, 1996), allowing prediction of attentional modulation and spatial shifts.

The simulation allowed synapses from a single neuron to be both excitatory and inhibitory. This is not a common feature in biological neural systems, where a combination of excitatory and inhibitory neurons could

achieve the same functionality. In addition, in order to allow standard effective learning algorithms to be applied, the simulation was at the level of firing rates rather than spikes and no short-term plasticity of any sort was included. It is sensible to assume that the effects obtained here with a simple model will be observed as well in more complex models that follow similar principles.

Some researchers have stated that changes in RFs reflect mechanisms for encoding position and that attention-centered coordinates are probably used at some stages of processing (Connor et al., 1997). Our results suggest that for each attentional state, the population code varies significantly, but further explorations are necessary to test if the changes correspond actually to a change into attention-centered coordinates.

The modulation of tuning curves observed from the model is not strong, possibly due to attention being shifted only between regions nearby inside the RF. Nevertheless, the results suggest that a multiplicative effect can be achieve without the need of explicit multiplicative interactions in the circuit. Fig. 5.7 compared only two attentional conditions, center *vs.* corner; when attention was directed to different regions of the image (not shown), effects of attending away were complex and sometimes in different directions depending on the attended location. Complex modulation of this type has been observed in experiments with multiple stimuli inside the receptive field (Connor et al., 1997).

An influential modeling study that made concrete predictions regarding attentional changes of neuronal activity was developed by Olshausen et al. (1993). In that study, control neurons dynamically modified the synaptic strengths of the connections in a model network of the visual ventral pathway. The network selectively routed information into higher cortical areas producing invariant representation of visual objects. This model predicted changes in position and size of receptive fields as attention was shifted or rescaled. These phenomena are partially supported by results from Connor et al. (1997). Their model also qualitatively matches modulation effects observed by Moran and Desimone (1985) with stimuli inside and outside the classical receptive field of V4 neurons. In comparison to their model, in which modulatory effects were obtained by explicitly modulating the synaptic strengths of the connections, attentional modulation in our model emerges from the nonlinearity of the units and general objective

of the network without requiring changes in the synaptic strengths.

More recent studies incorporate principles of statistical inference into models of attention. For instance, Yu and Dayan (2005) and Rao (2005) present networks that implement Bayesian integration of sensory inputs and priors, and which replicate behavioral as well as electrophysiological measurements. In these studies, spatial attention is equated to prior information on the location of the features of interest. The Bayesian inference approach to modeling attention should be regarded as complementary to that taken here. The transformations performed by the model units in the present work are defined by the solution to an optimal coding problem; and under certain conditions, these computations would be equivalent to those in inference-based networks. In fact coding, statistical modeling of distributions, and inference from partial data are, mathematically speaking, very closely related.

5.6   Concluding remarks

The model presented here accounts for attentional modulation of neuronal responses in a framework that includes both attention and receptive field formation, and as a consequence of an underlying normative principle (optimal coding) rather than by tuning a complex special-purpose architecture. The model shows that reallocation of resources can emerge even in a simple feedforward network, and challenges the traditional characterization of neuronal activity. These results are consistent with the notion that attentional modulation is not, at its root, due to specific local architectural features, but is rather a ubiquitous phenomenon to be expected in any system with shifting fidelity requirements.

The fact that our simulation shows modulation effects consistent with physiological recordings suggests that we should not necessarily expect explicit gating circuitry in neural systems responsible for attentional phenomena. Furthermore, the informing signals do not have to explicitly represent the "attentional space", *i.e.*, a spatial attention effect is not necessarily mediated by a topographic input.

Our model strongly predicts that the stimulus that maximally drives a neuron depends on attentional state. Moreover, the behavior of a single neuron in this model cannot be well characterized by measurements of

attentional modulation of only a single sensory stimulus. This prediction is consistent with the experimental observations discussed above, and it should be possible to test it more explicitly using currently available experimental techniques. The model also suggests that stronger modulations are expected when the complexity of the input increases, relative to the capacity of the system.

Chapter 6

Neural mechanisms for resource allocation

Summary

Attentional modulation of neuronal activity may arise from different mechanisms. Here we evaluate the different hypotheses and supporting evidence for each one.

Main contribution
A model in which no special architectural requirements are needed for resource redistribution is presented. In this model, attentional control signals enter the system the same way as sensory signals.

In previous chapters we presented evidence that neuronal activity can change depending on the attentional state of the subject even when stimulation stays constant. In addition, we suggested that this modulation subserves a resource allocation that allows for efficient use of the neural hardware available. We focused on cortical areas known to process sensory information; these regions presumably receive contextual information from higher-level areas about the relevance of different features of the stimuli. These phenomena raise several question concerning the interaction between mod-
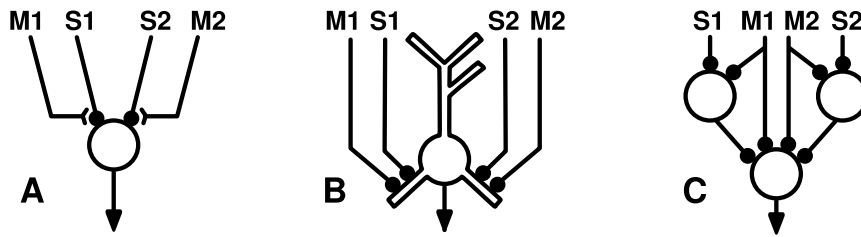
Figure 6.1: Neural mechanisms underlying the modulation of neuronal responses. (**A**) Modulatory signals M1 and M2 modify the synaptic efficacy for inputs S1 and S2. (**B**) The nonlinearities of the dendritic tree produces different effects according to the relative location between modulatory and sensory inputs. (**C**) All inputs enter neurons in a similar fashion, and the nonlinearity of the network produces a modulatory effect.

ulatory signals and sensory processing circuits, and about the origin of these signals. This chapter provides a discussion of these topics.

6.1 How do attentional signals enter sensory processing circuits?

There are different ways in which signals that code for feature relevance and attentional state could interact with sensory processing pathways. The following sections describe three possibilities that have been proposed in the literature (Fig. 6.1). These are supported by anatomical and physiological studies as well as models that evaluate their computational capabilities.

6.1.1 Changes in synaptic strengths

A common hypothesis that explains how attentional signals interact with sensory processing circuits argues that modulatory signals change the synaptic strengths of particular connections in the system. These ideas were employed by Olshausen et al. (1993) to create shifter circuits in which attentional signals define the routes that sensory signals can follow in a hierarchical model of the ventral visual pathway. Control signals that modify synaptic efficacy were also used in a model by Reynolds et al. (1999) to account for observations from areas V2 and V4. A schematic representation of this model is presented in Fig. 6.1A.

The biophysical mechanism that could allow for gating as suggested by these models is still under debate. Presynaptic inhibition and (post-

synaptic) shunting inhibition have been suggested as possible mechanisms (Olshausen et al., 1993). Neuromodulators like acetylcholine provide another alternative for changes in synaptic efficacy but it remains to be tested if their effects are fast enough to account for the rapid changes observed in attentional tasks. Nevertheless, neuromodulators seem to play an important role in attention systems (Voytko et al., 1994; Yu and Dayan, 2005).

6.1.2 Dendritic computation

It has been suggested that the computational capabilities of single neurons are enough to produce the modulation necessary to account for attentional phenomena. These proposals are based on two ideas: first, that dendritic trees of cortical pyramidal neurons introduce nonlinear interactions between the inputs depending on the site of the synapse (London and Häusser, 2005); and second, that top-down signals project to a different cortical layer, and therefore a different part of the dendritic arbor of cortical neurons, compared to bottom-up signals (Kandel et al., 2000, Chap. 17).

Experiments in which signals are injected at different locations of the dendritic tree of a cell have shown a great variety of nonlinear effects in the potential at the cell soma and the initiation of axonal action potentials (Larkum et al., 1999, 2001; Polsky et al., 2004). Based on these observations, researchers suggest that pyramidal cortical neurons are better modelled by a three-layer network of sigmoidal units with additive and multiplicative interactions between them (Hausser and Mel, 2003).

To further demonstrate the importance of the effects of multiple sites of integration, Körding and König (2001) showed that models including two integration sites allow for biologically realistic implementation of supervised and unsupervised learning rules.

6.1.3 Modulatory and sensory signals are indistinguishable

A third alternative for how top-down signals could influence the processing of bottom-up information is presented in Fig. 6.1C. Here, it is the global effect of the nonlinear units that allows for complex interactions between the different inputs that enter the system indistinguishable from each other. This is the type of interaction implemented by the model presented in Chapters 3 and 5.

Fig. 6.2 shows an example of preferential representation of variables in a feedforward network or sigmoidal units, where a attentional signal defines which channel should be represented with higher fidelity.

Other models of attention, based on oscillations (Niebur et al., 1993) or temporal correlations among neurons (Niebur and Koch, 1994; Tiesinga, 2005), are also consistent with this architecture. The model introduced by Yu and Dayan (2005) provides further supports for these ideas. In their network, a top-down signal represents prior information about location of the stimulus, modulating additively the activity coming from neurons that code for the log likelihood of spatial and orientation variables. Multiplicative effects are then achieved in further layers that represent marginal probabilities of the variable to estimate.

Relevant to our discussion is the work by Schwabe and Obermayer (2005) in which different mechanisms are evaluated in the context of attention and perceptual learning. In this study, a biologically realistic recurrent neural model was designed to encode a visual feature (stimulus orientation in this case) by the activity of its output units. An ideal estimator was used to measure the quality of this representation, by means of maximizing the Fisher information so that the variance of the estimate was minimal. Different parameters of the model were varied. Optimal values of afferent synapses, recurrent synapses, additive feedback inputs, and gain of excitatory neurons required uniform changes over all synapses. Only for the last mechanism the modulation of the population response to a particular orientation was strictly multiplicative, consistent with studies of selective attention. This suggests that changes in gain may be used by the nervous system in order to improve representation of stimuli, and that these changes produce a multiplicative effect in the response of neurons involved.

6.2   The origin of attentional control signals

Where do attentional control signals come from? Without having to recur to the idea of a homunculus sitting inside the brain and controlling every actions, one can still determine areas involved in the control of attention. These areas receive signals about the current stimulus as well as signals
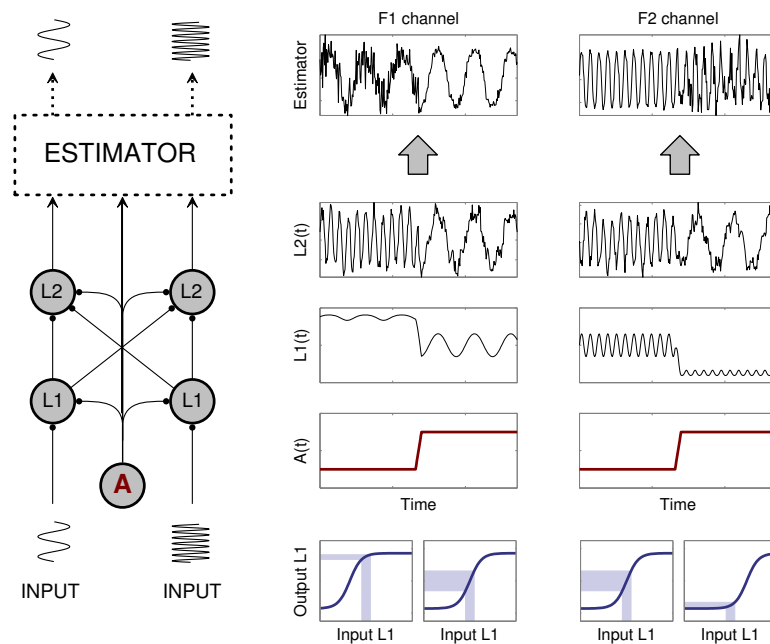
Figure 6.2: Allocation mechanisms in a noisy feedforward network. The output of the second layer (L2) provides an encoded representation of the input signals. The plots on the right panels represent the activity of each unit as a function of time. Changes in the attentional signal produce changes in the quality of the reconstruction given by the estimator. The lower panels show how the sigmoidal nonlinearity of the units in the first layer (L1) can produce a multiplicative effect.

representing previous stimuli, fed back from other brain regions. Some of these control areas seem to be primarily involved in overt orienting and saccadic control, while others represent relevant features as if implementing a saliency map. Here we review experimental evidence relating these areas to attentional processes.

The superior colliculus (SC) of the midbrain has been shown to play an important role in the control the saccadic movements, containing cells that code for both the destination of saccades and their timing (Liversedge and Findlay, 2000). Descending projections into the SC come from the frontal eye field (FEF) region and the posterior parietal cortex (PPC). Neurons in the lateral intra-parietal (LIP) region of the parietal cortex show higher responses to behaviorally significant visual stimuli in their receptive field compared to task-irrelevant stimuli. In comparison to neurons in FEF or SC in which this effect depends on the execution of a saccade, the increase in firing rate in LIP neurons does not change whether the behavioral re-

sponse is a saccade or a bar release (Colby and Goldberg, 1999). These observations suggest that neurons in LIP represent a saliency map, *i.e.*, a topographically organized representation of the relevance of the stimuli in the display (Gottlieb et al., 1998). Projections from area LIP to lower areas of the visual pathway like V4, could therefore be responsible for some of the attentional effects presented in Chapter 4.

The pulvinar, a sub-structure of the thalamus, has also been suggested to be involved in the control of attention. Studies showed that unilateral deactivation of the Pdm portion of the pulvinar, which projects to the PPC, reduced the saliency of contra-lesional stimuli. Analogous results were obtained from the lateral pulvinar (PL), which projects to areas V4 and IT. Desimone and Duncan (1995) suggested that deactivation of the pulvinar may just be depriving cortex from excitatory inputs reducing visual saliency. They argue that together with evidence that bilateral pulvinar lesions have no effect on the ability of monkeys to find a target embedded in distractors, these findings indicate that the pulvinar may not have a necessary role in the gating of attention.

6.3 Attention and receptive field formation

6.3.1 Receptive field formation

The *classical receptive field* of a neuron in the visual system is defined as the region in which a single stimulus can evoke a response, usually measured as a spike count. By definition, stimuli in the surrounding regions cannot generate a response, but they can greatly modulate the response of a stimulus inside the classical receptive field (Allman et al., 1985). Analogous definitions are also given for other sensory modalities, using the corresponding feature space; for example, the space of sound frequencies for auditory neurons. A more detailed characterization of neuronal responses requires specifying not just the region of influence but the magnitude of the influence at each point in this region and its temporal dynamics (Dayan and Abbott, 2001, Chap. 2).

In the primary visual cortex of primates, the receptive fields of simple cells (cells in which contributions from different locations within the visual

field sum linearly) can be characterized as being *spatially localized*, *oriented* and *bandpass* (selective to different spatial scales), similar to the basis functions of certain wavelets transforms (Field, 1994). How and when this selectivity appears is still debated among researchers. In addition to genetically determined factors, some of these properties depend on activity during development, as has been concluded from visual deprivation experiments, and from studies in which animals are raised with limited visual features, for example with visual stimuli of fixed orientation (Price and Willshaw, 2000, Chap 7).

Modeling studies showed that some of the properties of simple cells in primary visual cortex can emerge from networks trained with unsupervised learning algorithms(Hancock et al., 1992; Law and Cooper, 1994), but these early attempts failed to produce a set of receptive fields that spanned the whole image space and contains the three properties mentioned above.

6.3.2 Sparsity constraints and redundancy reduction

Olshausen and Field (1996) showed that two global objectives together are sufficient for receptive fields with the right properties (in this case for cells in primary visual cortex) to emerge. The first one states that information must be preserved, and the second one that the representation must be sparse. They formulated the problem of searching for a sparse code as an optimization problem where the cost function to be minimized is

$$E = \sum_{x,y} \left[I(x,y) - \sum_i a_i \, \phi_i(x,y) \right]^2 + \lambda \sum_i S(a_i) \qquad (6.1)$$

The first term represents the error in the reconstruction of an image $I(x,y)$, using the basis functions $\phi_i(x,y)$ and the coefficients a_i. The second term represents a measure of how non-sparse the representation is. These two terms are balanced through the parameter λ, and $S(a)$ is a nonlinear function like $|a|$ or $-e^{-a^2}$ that gives a higher weight to coefficients with larger magnitudes.

Similar results were obtained by Bell and Sejnowski (1997) applying the Infomax algorithm for independent components analysis to natural images. This approach is in principle equivalent to Barlow's redundancy reduction problem (Barlow, 2001).
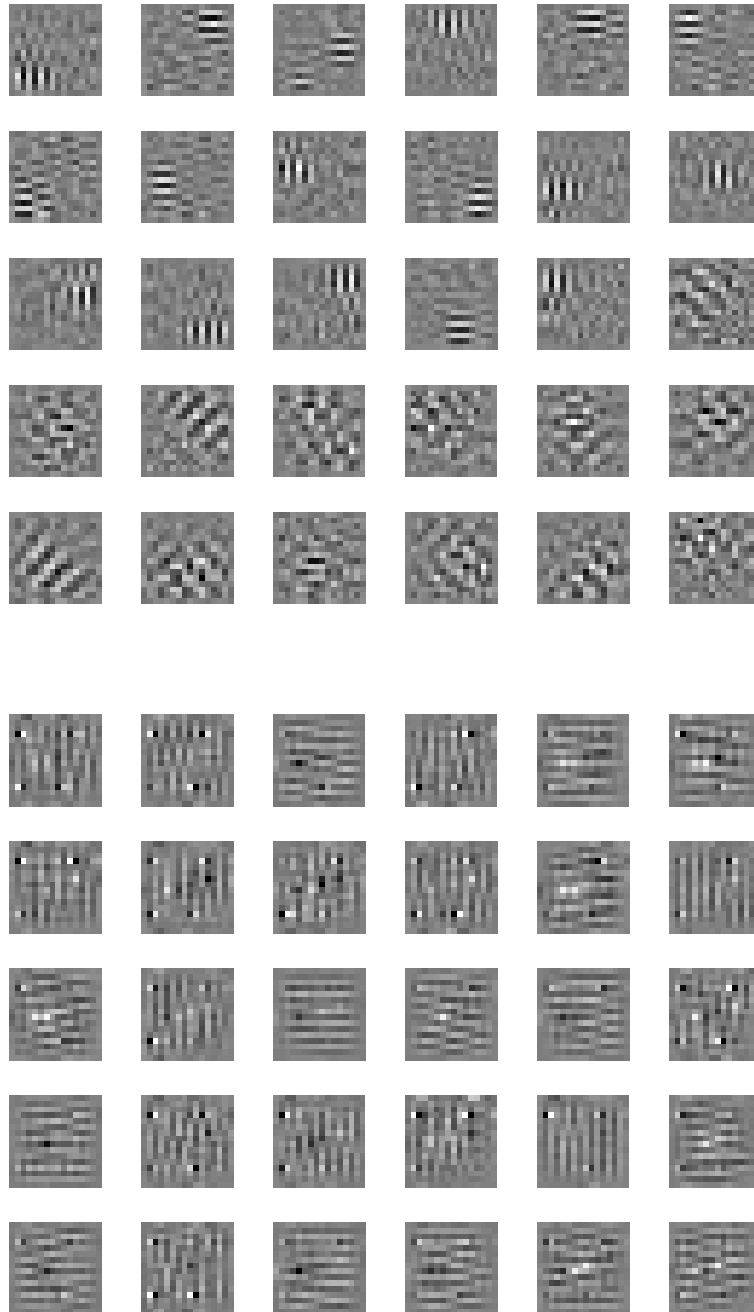
Figure 6.3: Non-uniform relevance produced localized preferred stimuli. The figure shows preferred stimuli for coding units in a network of 30 units in the bottleneck. (Top) Optimization using a sharp mask as described in Section 3.4. (Bottom) Optimization using a flat attentional mask, *i.e.*, ignoring the attentional input.

Evidence from electrophysiology confirm the predictions from these models (Vinje and Gallant, 2000). These studies suggest that the nervous system has the ability to adapt (either in an evolutionary long time scale, or a developmental shorter time scale) so as to find the transformations that allow for efficient coding of natural stimuli. These phenomena could occur even at shorter scales, as suggested by findings of perceptual learning and cortical plasticity in the adult brain (Gilbert et al., 2001).

6.3.3   Non-uniform relevance produces localized RFs

A feedforward multilayer network of sigmoidal units optimized to represent efficiently input patterns consisting of oriented bars, does not generate localized receptive fields. In contrast, when information about the relevance of features is added to the system, of the form described in Chapter 3, preferred stimuli for the encoding units display not only orientation selectivity but also spatial preference.

Fig. 6.3 shows the resulting preferred stimuli when using a sharp attentional mask compared to using a flat mask (equivalent to having no modulatory signal) for a network and stimuli analogous to those described in Section 3.4. Only when the network makes use of the attentional signal, localized preferred stimuli emerge in the encoding units.

These results suggest that mechanisms for selection of features may play a role in the development of receptive fields of sensory neurons.

Chapter 7

Allocation of resources during active perception

Summary

With this chapter, we shift our focus to orienting phenomena that require movement of the sensors. Here, we derive strategies that a perceptual system must follow in order to perform optimally, given particular constraints and goals. We analyze in detail the case in which the state of multiple dynamic targets must be estimated continuously.

Main contribution

The case of simultaneous tracking of two targets with Markovian dynamics is analyzed in detail. Optimal strategies are derived for different sets of parameters of the dynamics and noise of the targets.

7.1 Perception is an active process

We don't just see, we *look*. This process of actively exploring the world around us is not limited to our sense of vision. We set our body and mind

to listen, touch or sniff, not simply hear, feel or smell. The particulars of this orienting process depend in great degree on the task at hand. During these tasks, we allocate perceptual resources towards features of interest at the expense of neglecting other features.

We can describe this perceptual actions as serving an inference process in which we attempt to collect information about relevant variables from a changing world. Consequently, we can argue that knowledge about the dynamics of the stochastic process to be estimated, and the ability to assign more resources to represent features of interest, can be both used to improve statistical inference of the underlying process. The following example relates these concepts to a real world scenario: Imagine a basketball player who has to keep track of other players in the court. To do this, he/she must decide where to look, how often to change gaze, and how to track more than one player without having to look directly at any. Effective gazing strategies presumably depend on the predictability and relevance of each of the targets being tracked.

Studies of bottom-up attention, saliency and active perception (Bajcsy, 1988; Itti and Koch, 2001; Hayhoe and Ballard, 2005) have evaluated problems related to those described above, and have hypothesized that we orient our senses towards features of interest in order to maximize the gathering of information relevant for a particular task. In this chapter we approach the modeling of these phenomena by evaluating different resource allocation strategies for simultaneous estimation of two independent signals contaminated with noise.

In the simultaneous tracking task we have selected, the statistics of the dynamics of the signals to be estimated are assumed to be known, but the signals themselves are *hidden* and only noisy observations are available. Resources in our case refer to the quality of the observation, where a constraint on the available resources sets a limit on the average quality across channels. Results from this analysis allow for predictions of orienting strategies in simple dynamic environments.

The ideas presented in this chapter also relate to the concept of bit allocation for signal coding and compression, in which a limited number of bits is distributed between different features (*e.g.* subbands) according to their perceptual relevance and statistical characteristics (Gersho and Gray, 1992, Chapter 8).
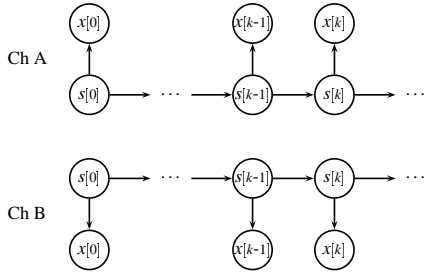
Figure 7.1: Two channel Markov model. States $s[k]$ are binary and independent between channels. The probability distribution of the observations given the states $p(x[k]|s[k])$ is Gaussian. The sum of the observation variances (noise) across channels is the same at each time k.

7.2 Optimal strategies for simultaneous tracking

Consider multiple independent noisy channels in parallel with a common power constraint. The goal is to distribute the total power among the channels so as to minimize some error function between the original signals and their estimates from the noisy observations. If the signals are known to have partially predictable dynamics, an optimal strategy should presumably combine prediction and power allocation.

In the following sections we describe the simple binary model for which we evaluate the performance of an allocator/estimator over different sets of parameters. Optimal strategies are found by exploring the parameter space through computer simulations. We compare two different conditions for allocation: one that requires fixing the signal-to-noise ratio (SNR) for each channel and maintaining it during estimation, and a second one in which SNR is dynamically reassigned at each time step according to the changing uncertainties of the estimation for each channel. For the remainder of this chapter we will keep the power of the signals fixed and vary instead the noise level for each channel.

7.3 Analysis of a two-channel system

The goal in our particular setup is to estimate, at each time step, two independent binary signals each one contaminated by Gaussian noise. The signals are generated by a Markov process with known transition probabilities, and the only link between the two channels is a constraint on the total noise of the system. The allocation problem consists of deciding how much of the total noise should be assigned to each channel so that the average number of errors in either channel is minimal.

The system is illustrated in Fig 7.1. States variables s_i are binary and change according to symmetric transition probability matrices T_i (a different one for each channel). Because of this symmetry condition the system dynamics can be described by one variable p_i per channel.

$$s_i[k] \in \{-1, 1\}, \quad i = A, B \tag{7.1}$$

$$T_i = \begin{bmatrix} p_i & 1 - p_i \\ 1 - p_i & p_i \end{bmatrix} \tag{7.2}$$

$$p_i = P(s_i[k] = s_i[k-1]) \tag{7.3}$$

The observations are drawn from a Gaussian distribution with mean equal to the current state for each channel:

$$x_i[k] \sim \mathcal{N}\left(s_i[k], \sigma_i^2\right) \tag{7.4}$$

The constraint on the noise level can be interpreted as having a common power source, or a constant total signal-to-noise ratio, and it is given by:

$$\sigma_{\text{total}}^2 = \sigma_A^2 + \sigma_B^2 \tag{7.5}$$

We will refer to the ratio between noise variances as θ:

$$\theta = \frac{\sigma_A^2}{\sigma_B^2} \tag{7.6}$$

Inference of the state (for each channel) at each time k is done by calculating the maximum a-posteriori (MAP) estimate for the individual state, given the observation sequence:

$$\hat{s}_i[k] = \underset{q \in \{-1,1\}}{\arg\max} \, P(s_i[k] = q \mid x_i[0], \dots, x_i[k]) \tag{7.7}$$

$$= \underset{q \in \{-1,1\}}{\arg\max} \, \frac{P(s_i[k] = q, \tilde{x}_i[0], \dots, \tilde{x}_i[k])}{P(\tilde{x}_i[0], \dots, \tilde{x}_i[k])} \tag{7.8}$$

Here we have used $\tilde{x}_i[k]$ instead of $x_i[k]$ so we can write down expressions for the observations as probability mass functions. The variables $\tilde{x}_i[k]$ can be interpreted as discretized versions of the observations, and are only

used for describing the algorithm, which is unaffected by this approxima-tion. Since the marginal of the observations does not affect the maximum, the problem can be solved by finding the joint probability between the state at time k and the observation sequence. This value corresponds to the for-ward variable $\alpha_{[k]}$ described by Rabiner (1989). In vector form we have:

$$\alpha_i[k] = \begin{bmatrix} P(s_i[k] = -1, \tilde{x}_i[0], \ldots, \tilde{x}_i[k]) \\ P(s_i[k] = 1, \tilde{x}_i[0], \ldots, \tilde{x}_i[k]) \end{bmatrix} \tag{7.9}$$

$$= \frac{\breve{\alpha}_i[k]}{\sum_q \breve{\alpha}_i[k](q)} \quad , \quad q = -1, 1 \tag{7.10}$$

$$\breve{\alpha}_i[k] = T_i\, \alpha_i[k-1] \odot f(x_i[k], s_i[k], \sigma_i^2) \tag{7.11}$$

$$\text{with} \quad \breve{\alpha}_i[0] = \pi_i \odot f(x_i[k], s_i[k], \sigma_i^2) \tag{7.12}$$

Here, $\odot$ denotes the element-by-element (or Hadamard) product, π_i is a 2-element vector that represents the initial state probabilities, and $f(x, s, \sigma^2)$ is a 2-element vector containing the likelihood (for each possible state) that the sample x came from the Gaussian distribution defined in eq. (7.4).

The estimation method described here maximizes the expected number of correct individual states by choosing the most likely state at each time step for each channel. Other optimality criteria imply different methods, *e.g.*, the Viterbi algorithm (Forney, 1973) finds the single best state sequence for a given observation sequence. In this study, we will focus only on the individual state MAP estimate described above.

Our goal is to evaluate the performance of this estimation method for different allocation strategies. Since we are interested on simultaneous tracking of the two channels, we measure performance by calculating the average number of time steps in which errors occur in either channel:

$$\bar{\mathcal{E}} = E\left\{ \hat{s}_A[k] \neq s_A[k] \lor \hat{s}_B[k] \neq s_B[k] \right\} \tag{7.13}$$

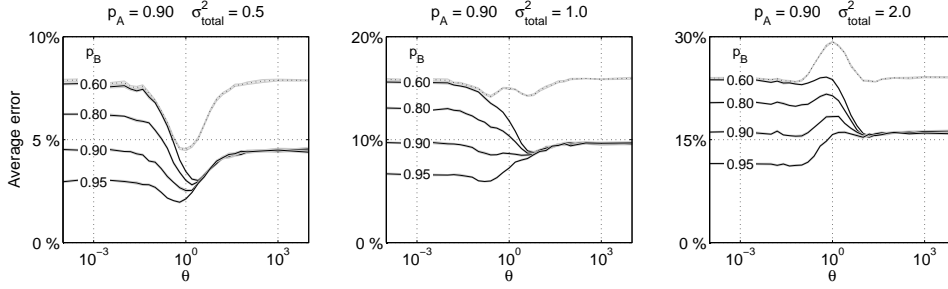where $E\{\cdot\}$ represents the expectation over time. A lower $\bar{\mathcal{E}}$ indicates better performance.

Figure 7.2: Fixed allocation. Average error as a function of noise ratio θ. Each panel presents results for a given SNR using one p_A and many p_B. The dotted line corresponds to the performance of an estimator that ignores the dynamics of the signals. Gray areas around the curves represent the standard error of the mean of the measurements.

7.3.1 Fixed allocation

In the fixed case, we want to find the constant ratio θ that minimizes the average error $\bar{\mathcal{E}}$, given the system parameters and total noise power:

$$\theta^* = \underset{\theta = \sigma_A^2 / \sigma_B^2}{\arg\min} \bar{\mathcal{E}} \tag{7.14}$$

One way to solve this problem is to derive an expression for the probability of making an error when estimating the states $P(\hat{s}_i[k] \neq s_i[k])$ at each time step k for each channel i, calculate the expectation over time of errors made in either channel, and then solve the minimization problem either analytically or by numerical methods. We took instead an empirical approach in which we simulated the system for fixed values of p_i and σ_{total}^2, and find the value of θ that gives minimal error. Results for different noise ratios θ were compared to the performance achieved by an estimator that assumed the process to be white and ignored the dynamics of the system.

The system was simulated using the fixed allocation method for different transition probabilities and three SNR scenarios ($\sigma_{total}^2 = 0.5, 1, 2$). The average error was measured following eq. (7.13) as we varied the noise ratio θ. Results for $p_A = 0.9$ are shown in Fig. 7.2. This figure includes the performance given by an estimator that assumes a white process, *i.e.*, an estimator that ignores the dynamics of the system and simply sets a decision boundary at $x_i[k] = 0$.

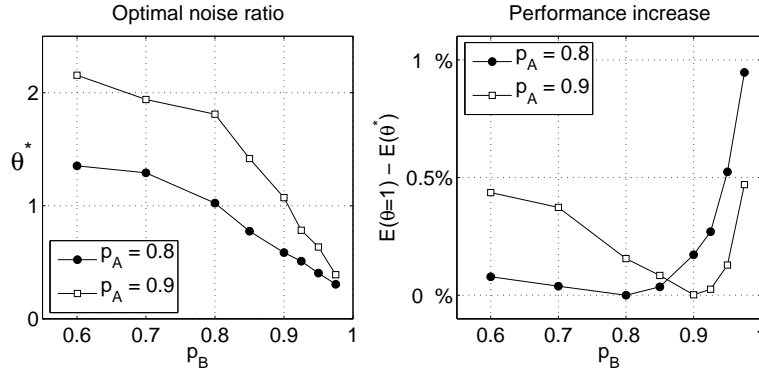For each total noise level we obtained a different behavior. In low SNR

Figure 7.3: Optimal fixed allocation in high SNR ($\sigma^2_{\text{total}} = 0.5$). The left panel shows the value of θ that produced the minimal average error. The right panel shows the difference between the optimal error and the error achieved when both channels were assigned the same amount of noise.

(right column of Fig. 7.2) the best strategy was to assign most resources to one of the channels and to *guess* (estimate with very high observation noise) the other one. The curves confirm the intuition that more resources (lower variance) should be assigned to the least predictable channel.

In contrast, when SNR is high, there is a non-extreme ratio at which the error is minimum. This ratio depends on the relative values of the transition probabilities as shown in the left panel of Fig. 7.3. Results imply that when both channels have identical dynamics it is optimal to split the resources equally across the channels, which is not the case when SNR is low. The right panel of Fig. 7.3 shows the difference between the error obtained with $\theta = 1$ and the minimal error. The improvement in performance when using the optimal ratio is lower than 1% for the transition probabilities shown here. This result implies that, when the SNR is high, it may not be worth trying to find the optimal allocation ratio, but simply distributing resources equally across channels.

7.3.2 Dynamic allocation

Many strategies for dynamic allocation could be considered. Here we explore one in which the ratio is changed at each step according to the relative uncertainties of the two channels. In this case, we define the certainty of our estimate as:

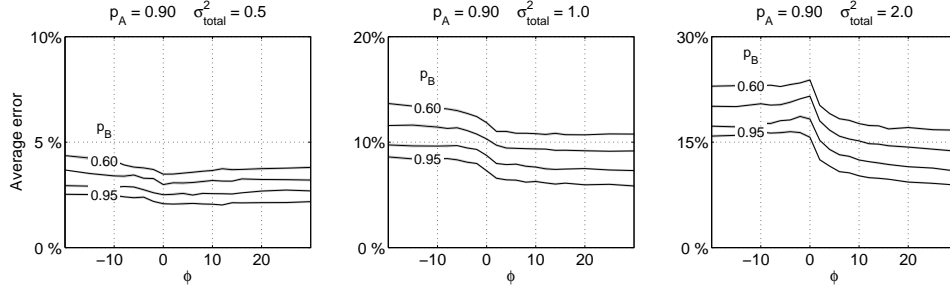$$c_i[k] = \left| 0.5 - \alpha_i[k]^{(1)} \right| \tag{7.15}$$

Figure 7.4: Dynamic allocation. Average error as a function of exponent ϕ. Each panel presents results for a given SNR using one p_A and many p_B.

where $\alpha_i[k]^{(1)}$ corresponds to the first element of the joint probability vector $\alpha_i[k]$ for channel i at time k. The value of $c_i[k]$ represents how close to 1 is the probability of being in one state, indicating a level of certainty about that estimate. An alternative would be to use the inverse of the entropy of each channel. Given these certainty values, we set the ratio at each time step as:

$$\theta[k] = \frac{\sigma_A^2[k]}{\sigma_B^2[k]} = \left(\frac{c_A[k-1]}{c_B[k-1]} \right)^\phi \tag{7.16}$$

And we want to find the optimal exponent:

$$\phi^* = \arg\min_\phi \bar{\mathcal{E}} \tag{7.17}$$

The exponent ϕ enables a non-linear relation between the certainty ratio and the noise ratio at the next time step . Note that this function includes the case in which the noise ratio is always 1 (ignoring the certainties) and the case in which resources are completely moved from one channel to the other at each time step.

The dynamic allocation method was evaluated using similar parameters to those for the fixed case but using instead the exponent ϕ as the abscissa, and calculating the ratio $\theta[k]$ at each time step. Results for $p_A = 0.9$ and three SNR scenarios are shown in Fig. 7.4.

For high SNR ($\sigma_{total}^2 = 0.5$), performance remained almost constant as we varied the exponent ϕ, tending to be lower for high values of ϕ. For low SNR, the curves indicate that a better performance was achieved as we increased the exponent. This implies that even for very small differences between the certainties on state estimates, we should assign all re-

74
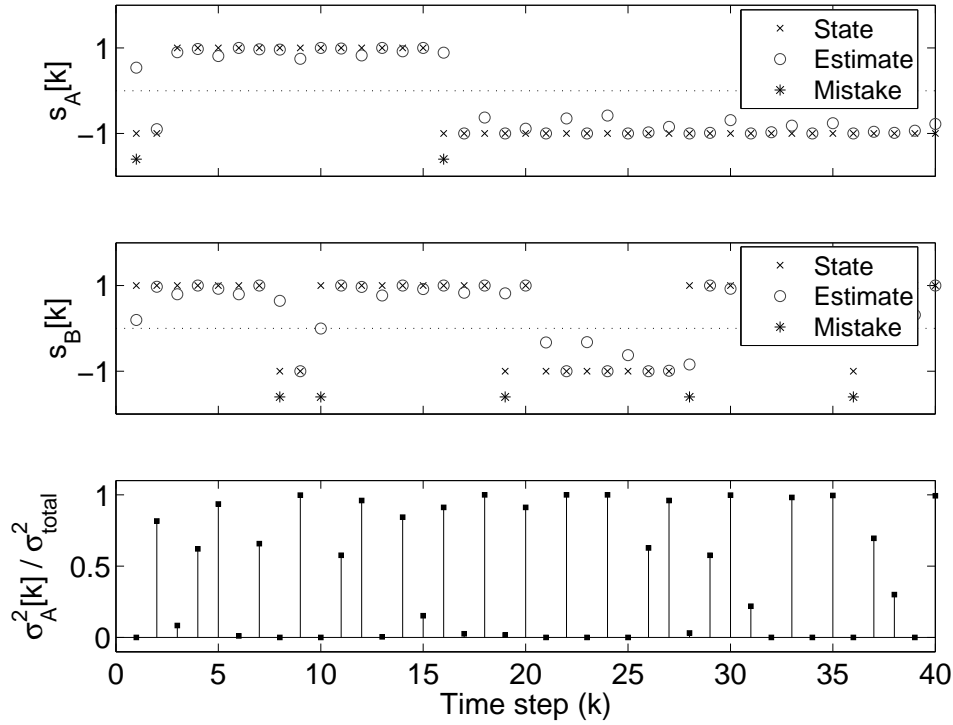
Figure 7.5: Example of dynamic allocation. Top two panels show the states and soft-estimates of channels A and B. Soft-estimates represent the probability of being in a particular state scaled to the range $[-1, 1]$. Bottom panel displays the (normalized) amount of noise assigned to channel A at each time step.

sources to only one channel, that one with lower certainty. Thus, according to eq. (7.16), a higher variance is assigned to the channel with higher certainty.

The extreme case as the exponent increases gives rise to a strategy in which resources are completely shifted from one channel to the other at each time step. This is illustrated in Fig. 7.5, which shows an example of the dynamic estimation procedure for $\phi = 20$, $p_A = 0.9$, $p_B = 0.8$ and $\sigma^2_{\text{total}} = 2$. Note the alternating behavior of the allocated variance σ^2_A.

7.3.3 Comparing fixed *vs.* dynamic strategies

In addition, we compared the performance of fixed and dynamic strategies, and find those condition in which one was more advantageous than the other. For $p_A = 0.9$, we calculated the minimal error achieved with each method and plotted them as a function of p_B in Fig. 7.6.

Figure 7.6: Performance comparison between fixed and dynamic allocation methods. All curves are calculated for $p_A = 0.9$. Error bars correspond to the standard error of the mean.

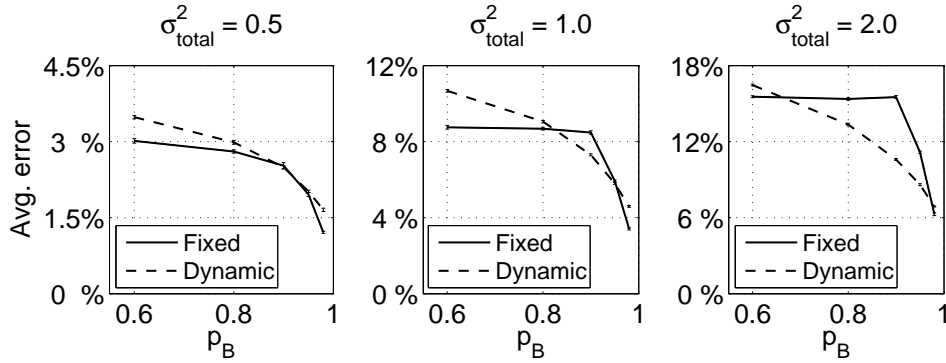For high SNR, performance for both methods is very high and relatively similar. In this case the error is already low enough, leaving no space for improvement. For low SNR, in contrast, there are significant differences in performance from both methods and a clear region in which dynamic allocation is better. This implies that in cases when the dynamics of both channels is very similar, it is better to switch from one channel to the other at each time step (best dynamic strategy) than giving all resources to only one channel and guess the other (best fixed strategy). For this particular case, the average error was lowered by 5%.

7.3.4 Discussion

The modeling results described above suggest that strategies for allocating resources in simultaneous tracking tasks depend on the uncertainties of the targets to be estimated. These uncertainties are defined by the intrinsic dynamics of the targets and by the observation noise. These results may depend on the particular cost function of the task, and further work is necessary to determine if strategies for constraints different to those evaluated here differ significantly. The assumptions made when defining the allocation and estimation algorithm, will be discussed in the remainder of this section.

For the simulations presented above, we used an estimation procedure that was not derived from the general optimization problem described initially, since a slightly different error measure was used. We limited our-

selves to find the optimal noise ratio for a given estimator, but a more general goal would be to devise a theory that encompasses both estimation and allocation. The estimation method used here was selected for its simplicity and low storage requirements and because the quantities it relies on relate directly to concepts of uncertainty and confidence of the estimates.

One of the motivations for approaching the problem of resource allocation is to derive theories that predict strategies for active perception. In this context, the noise constraint of a common power source (total sum of variances) may not be appropriate. In the case of vision, for example, there may be a complex function that describes how the quality of the observation depends on gaze angle.

This study proposed a dynamic allocation method that uses a measurement of certainty to derive the noise ratio at each time step. The method was based on the idea that it is advantageous to give more resources to uncertain targets since they are harder to predict. This is clearly not the only dynamic method and further work is necessary in order to find a general solution for the optimization problem over the space of possible dynamic strategies. Furthermore, the dynamic solution described here does not take into account the cost of reassigning resources. It could be the case that switching back and forth between channels gives the lowest error, but it is so expensive (in term of energy consumption or other constraints) that it becomes a suboptimal solution.

We assumed perfect knowledge of the dynamics of the signals to be estimated. Further work is necessary to evaluate the robustness of the fixed and dynamic allocation strategies when parameters are not known accurately. Furthermore, the allocation problem should be stated in a learning framework in which estimation of the parameters is done simultaneously (perhaps at a slower time-scale) than estimation of the signals. Some researchers have investigated similar ideas in the context of competitive allocation of learning between stimuli according to their relative uncertainties (Dayan et al., 2000).

Finally, an analytical solution for the optimal noise ratio may provide further insight on the effects of the different system parameters (total system noise and relative predictability of the signals) on performance. An empirical approach, like the one used here, may miss some of these details due to the limited sampling of the parameter space.

7.4 Concluding remarks

This chapter described an empirical evaluation of different strategies for distributing noise across parallel Gaussian channels in order to minimize signal estimation error. Noise levels were constrained by a total minimum system noise, and estimation of the transmitted signals was performed by finding the most probable input given the history of noisy observations. The allocation strategies benefited from the knowledge about the dynamic properties of the signals to be estimated.

Optimal strategies with respect to the error measure defined here depended on the total amount of system noise, giving qualitatively different results for low and high SNR. For a fixed allocation case and low SNR, the best strategy is to give most resources to the least predictable channel. For high SNR, performance can be slightly improved by choosing a noise ratio close to 1. When allowing dynamic allocation at each time step, the best strategy was to rapidly switch between the channels. This strategy, compared to the fixed allocation case, significantly increased performance for low SNR cases when both channel had similar properties.

Chapter 8

Psychophysics of active perception

Summary

Perception is an active process in which sensors are oriented in a task-dependent fashion. This chapter presents experiments that evaluate overt strategies used by humans when immersed in dynamic environments.

Main contribution

We present original results from eye-movement measurements during simultaneous tracking tasks. We show that selected strategies depend on the relative uncertainty on the targets, consistent with predictions from the previous chapter.

In this chapter we measure and compare allocation strategies used by human subjects under conditions analogous to those modelled in the previous chapter. Strategies were inferred from eye-movement measurements taken while subjects performed simultaneous-tracking tasks where targets had different levels of uncertainties. The term "simultaneous" is used to emphasize that the task requires *all* targets to be estimated correctly, but the display was designed in such a way that eye-movements were required,
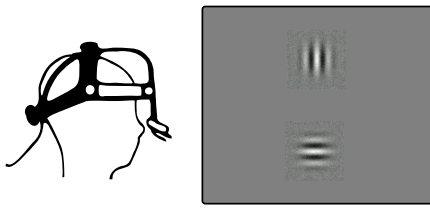
Figure 8.1: Head mounted eye-tracking system and display. Two noisy Gabor patterns were displayed at the top and bottom of the screen separated by a distance that precluded their simultaneous identification, requiring eye-movements.

and changes in covert attention alone would not be sufficient for correct identification of the targets. The uncertainty of such targets was varied by changing both, their dynamics and their noise level.

8.1 Experimental methodology

Task. Six naïve subjects, 21-31 yrs of age, with normal or corrected-to-normal vision, were required to play a video-game in which they simultaneously tracked the state of two targets. The targets did not change position during the experiment but their state, represented by their orientation, changed randomly according to Markovian dynamics. In each trial, subjects had to report the last state of both targets once they were occluded by a mask at a random time at the end of the trial. Subjects were given feedback on the correctness of their answers by scoring 20 points if both targets were estimated correctly for that trial, 3 points if only one of them was correct, and 0 points for errors in both targets. The total score was also displayed at the end of all trials in one condition. Seven conditions (Table 8.1), each one composed of 20 trials, were recorded for each subject. Subjects were allowed to practice the task for one session composed of 20 trials. No recordings were made during this session.

Targets. The target patterns were two-dimensional Gabor images oriented either vertically or horizontally, contaminated with uniformly distributed noise (Fig. 8.1). Targets were positioned at the top and bottom of the screen at a distance of 13.8 degrees of visual angle, which precluded their simultaneous identification and required the subject to make eye-movements. Two different signal-to-noise ratios (SNR) were used throughout the experiment. High SNR conditions had noise uniformly distributed in the range $\pm 10\%$ of the peak value of the Gabor image. The low SNR condition had noise in the range $\pm 60\%$. These correspond to SNR values of 10.6 dB and -5

Cond	1	2	3	4	5	6	7
P_A	0.9	0.6	0.6	0.9	0.9	0.9	0.6
P_B	0.9	0.9	0.6	0.6	0.9	0.9	0.6
SNR	10.6 dB					-5 dB	

$s_i[k] \in \{\text{Horiz}, \text{Vert}\}$
$i = A, B$
$P_i \equiv P(s_i[k+1] = s_i[k])$

Table 8.1: Seven experimental conditions were evaluated. They differed either in SNR, regulated by the amount of noise added to the target images; or in the dynamics of the targets, defined by the probability of each pattern to keep its current orientation on the next time step.

dB respectively. The target states changed independently for each channel i according to symmetric Markovian dynamics defined by the parameter P_i. This value represents the probability that the state will stay the same on the next time step (Table 8.1). Time steps had a duration of 200 ms. All subjects were given exactly the same trials but in random order and with different masking times. Before each condition, subjects were allowed to view the targets changing as they would do for that condition, and the experiment only started when subjects considered they had enough knowledge about the target dynamics.

Measurements. Eyes position was measured throughout each trial using a head-mounted EyeLink II eye-tracker system (SR Research Ltd., Ontario, Canada). This system measured the position of the center of the pupil at a sampling rate of 500Hz. Fig. 8.2 shows an example of the collected data. Full-field calibration was performed before each condition. Center calibration was performed before each trial, forcing the subject to start the trial by fixating in the middle of the display. Data were recorded for both eyes, but only data from the dominant eye of each subject were used for analysis.

The display. Stimuli were displayed in a 19″ standard computer monitor with equal brightness and contrast conditions in all sessions. Targets had a size of 1.2° and were separated (center-to-center) by 13.8°. Subjects were sitting at a distance of 74 cm, measured from their eyes to the display.

8.2 Results

The first observation from the behavioral measurements is the effect of training on the performance of the subjects. Average scores over all sub-
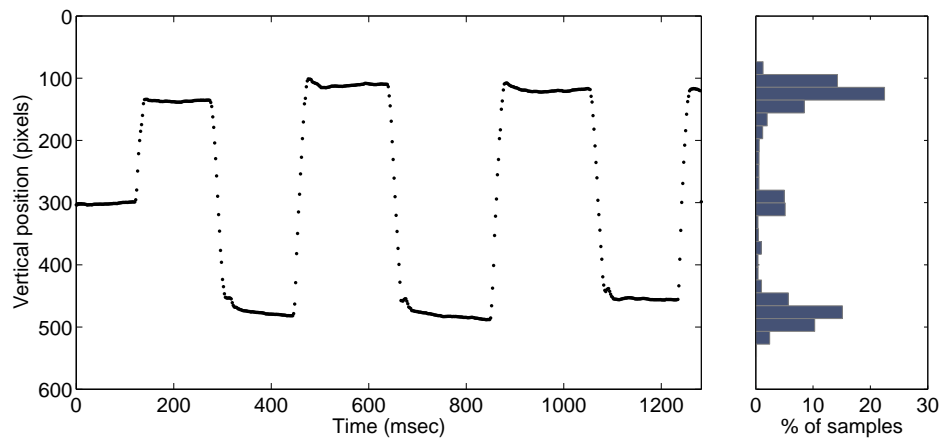
Figure 8.2: Example of eye-movements for one subject. The left panel shows eye-positions for one trial. The right panel shows the histogram of positions for all trials of condition 1 pooled together.

jects are shown in Fig. 8.3. As expected, performance was lower for those conditions in which targets had faster dynamics, *e.g.*, condition 3. When more noise was added to the targets (low SNR condition), performance decreased with respect to the last high SNR conditions, but did not change significantly when targets were made faster. This effect can be explained by the particular strategies chosen by the subjects, as described below.

8.2.1 Changing strategies according to target dynamics

To test if subjects changed allocation strategies depending on the dynamics of the targets, we compared conditions where both targets had the same probability of changing state, against conditions in which one target was more probable to change. Fig. 8.4 illustrates this analysis for conditions 1, 2, 3, and 4, 5, 6. Allocation of resources was measured as the difference between times spent in each target during all trials in a particular condition. Positive or negative values indicate that the subject spent more time looking at the top or bottom targets, respectively. The two top panels show averages over all subjects for the high SNR conditions. Both targets had the same dynamics parameters for conditions 1, 3, 5, and different for conditions 2 and 4. From these plots, we observe a clear preference for looking more often at the top target, independent of target dynamics. The first comparison, centered around condition 2, shows no effects when changing
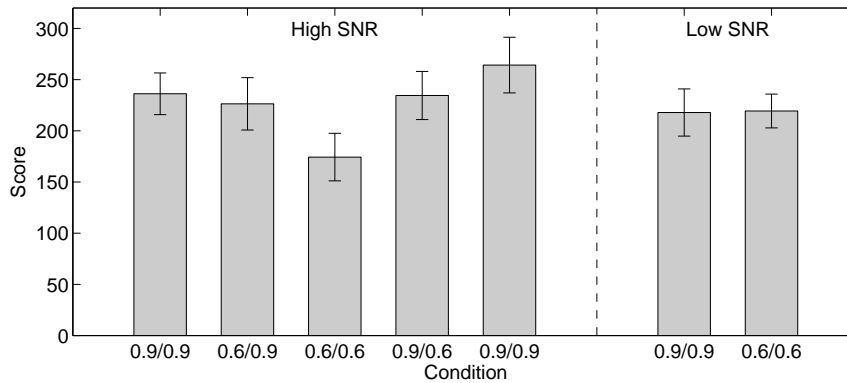
82

Figure 8.3: Average scores across subjects. Mean and standard error are shown for each condition. The maximum score possible was 400 points. Chance predicts an average score of 130, while estimating one target perfectly and guessing the other would give an average score of 230 points.

the dynamics of the targets. In contrast, the second comparison (around condition 4) shows that subjects changed strategy when the bottom target became more uncertain, giving more resources to this channels when compared to previous conditions. The strategy changed back to the previous level when both targets had again the same dynamics (condition 5). All subjects reported being aware of the targets having different dynamics for conditions 2 and 4.

8.2.2 Changing strategies according to noise level

In order to test changes in strategies depending on the level of noise added to the target images, we compared conditions that had the same target dynamics for high and low SNR. Fig. 8.5 first compares conditions 5 and 6, both of which have targets changing slowly and were performed one after the other. The figure also shows a comparison for conditions 3 and 7 where targets changed faster. The comparison differs from that in the previous section in that the index measured is not a difference in time spent between top and bottom targets, but a difference in time spent looking at one preferred target defined by the strategy on each trial. This way, if the subject spends one whole trial looking at the top target and another trial looking at the bottom target, the first index will give equal allocation whereas the second index will indicate a bias for one target, independent of which one it is. This new index was higher for lower SNR, indicating that in low SNR
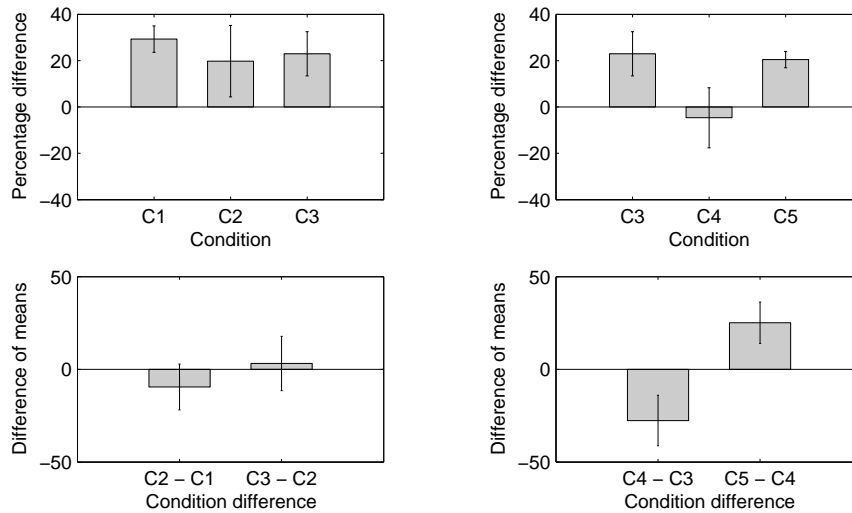
Figure 8.4: Strategies for different dynamics. The top panels show the average difference in time spent looking at the top *vs.* the bottom target. Positive values indicate that the subject spent more time on the top target. The bottom panels show the differences between these indexes for consecutive conditions. Late conditions (right panels) show an clear effect not observed in early conditions (left panels).

subjects tend to focus mostly on one target and briefly check the other one.

8.3   Discussion

A comparison of the allocation of resources under different conditions suggest that: **(1)** Learning has a significant effect on the selected strategies and performance for naïve subjects. **(2)** Selected strategies depend on the relation between dynamics of the targets to track. **(3)** Selected strategies depend on the level of noise of the targets.

The scores presented in Fig. 8.3 indicate that when targets are noisy, the dynamics of targets does not have a strong influence on performance. This can be attributed to a strategy in which the subjects has to spend most of the time following closely one target, estimating its state correctly, while assigning almost no time to the other target, and simply guessing its state. Results also indicate that subjects achieved a higher score in condition 7 (low SNR) than condition 3 (high SNR). This counter-intuitive result is explained by the effect of training in subjects as can be verified by comparing the performance in conditions 5 and 6, for which the effects of learning are minimized since they are evaluated in succession.
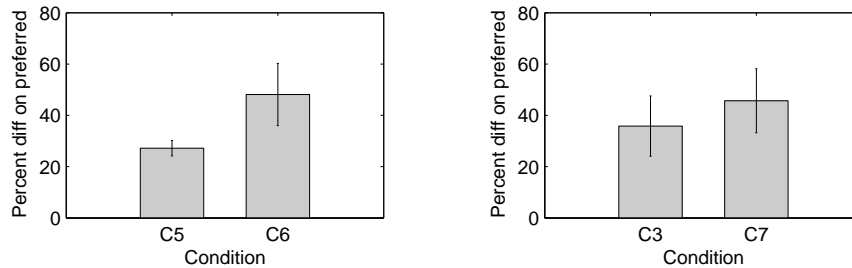
Figure 8.5: Strategies for different noise levels. The index plotted here measures the time spent on the preferred target. This index is independent of which target is preferred and was calculated for each trial before averaging. The bars indicate mean and standard error across subjects comparing conditions with different noise levels.

The comparison between allocation strategies from conditions 3, 4 and 5 show a clear change in the time allocated to each target. On average, subjects did not give more resources to the fast-changing target as expected from the model, but changed from a biased condition, preferring the top target, to a balanced condition in which time was allocated also to the fast-changing bottom target. When the same analysis was performed for conditions 1, 2 and 3, no clear changes where observed. This results may be due to the conditions being too early in the experiment. In addition, the bias towards the top target makes assignment of more resources less likely.

Changes in strategy when comparing different SNR scenarios (Fig. 8.5) are consistent with the results from the model, which suggested a tendency to increasingly focus on only one target as the noise increased (Section 7.3.1).

8.4   Concluding remarks

Researchers have previously evaluated eye-movement strategies during visual search using static displays (Liversedge and Findlay, 2000; Najemnik and Geisler, 2005). The task and measurement presented here extend these approaches by providing a simple way of evaluating allocating strategies in dynamic environments. This approach allows for the characterization of overt orienting for different levels of uncertainty in the targets to be estimated.

Chapter 9

Conclusions

Summary

In this chapter we summarize the contributions from previous chapters and evaluate their implications to our current understanding of orienting phenomena, selective attention and the neural code. In addition, we suggest future experiments in order to test the predictions derived earlier, and we propose extensions of the models with the purpose of providing additional predictions.

We have investigated orienting phenomena, not only of our peripheral sensors, but also of what we could call the *mind's senses*. The latter do not require physical movement, yet improve our performance at detecting and discriminating features of interest entering the senses. Our exploration included both a description of the phenomena at the behavioral level, and a characterization of the changes in neuronal activity associated with covert and overt orienting. We proposed general principles that can be implemented by neural architectures and that both generate novel predictions and account for experimentally observed phenomena. We went further and tested some of these predictions by measuring eye positions during simultaneous tracking tasks. These principles, derived from optimal coding and estimation from noisy observations, were presented in a framework

of resource allocation. Under this framework, we argue that the nervous system has the ability, at many levels, to reassign the function of its hardware in order to improve performance on the task at hand, at the expense of decreased processing of irrelevant features.

9.1 Contributions and implications

The work presented in this dissertation extends the characterization of orienting phenomena and suggests a relation between selective attention and the development of neuronal selectivity. This section summarizes the various contributions derived from this work and evaluates their implications on our current understanding of orienting systems.

(a) Uneven allocation of resources emerges as a natural property of a system that encodes its input efficiently under non-uniform constraints. In contrast to systems in which resources are fixed and encoding units assigned to irrelevant features are simply ignored, a system that allows for reallocation of its resources has the advantage of improving performance by using these otherwise-neglected resources. These alternatives resemble early and late selection theories, both of which have found supporting evidence. Our work suggests that if the neural hardware allows for a reallocation of the resources, early selection is likely to occur, with the amount of change in allocation being dependent on the existing mechanisms and the task itself. For instance, it is not expected that the primary visual cortex suddenly starts processing auditory stimuli when these become more relevant than their visual counterpart, but studies suggest that neuronal selectivity to features like color can change depending on relevance (Motter, 1994). In addition, as tasks become more demanding, changes are observed in both behavioral measurements and neuronal activity (Spitzer et al., 1988). In summary these results suggest that allocation of resources is a ubiquitous phenomena to be expected in any neural system wherever the mechanisms allow it, with performance improvement being a strong evolutionary force demanding the existence of these mechanisms.

(b) In a neural circuit, a set of modulatory inputs that enter the system in a fashion indistinguishable from other signals can generate an ef-

fect analogous to allocation of resources. This is one of the alternative mechanisms the nervous system may use to implement contextual modulation of sensory processing, and implies that, in some cases, local anatomical information may not be sufficient to discriminate modulatory from sensory inputs. However, learning to associate modulatory inputs to different cost functions depending on the task may require a differentiation between sensory and contextual inputs. Presumably the nervous system required both phenomena to evolve in parallel, suggesting that a mechanism that allows for both modulation and learning may be most advantageous.

(c) In a neural architecture, optimal coding under non-uniform fidelity requirements is sufficient to generate the firing rate modulation observed during selective attention experiments. This type of modulation is therefore expected in most neural systems for which the evolutionary pressure of improving performance has developed mechanisms for finding efficient coding strategies. These changes in activity can be interpreted as the signature of resource allocation at the neuronal level, and play an important role in the characterization of neuronal responses.

(d) Optimal coding principles provide a unified framework for attention and receptive field formation. Neuronal selectivity and orienting phenomena are intricately related and probably evolved together. Therefore, there must be common principles underlying their seemingly-unrelated mechanisms.

(e) An investigation of the modulation of neuronal activity by selective attention suggests that the traditional characterization of neuronal responses is incomplete and possibly inadequate. Our results suggest that the stimulus that maximally drives a neuron depends on the attentional state of the subject, and that the separability between neuronal responses and attentional modulation needs to be reevaluated. In this respect, we come to realize that characterizing a neuron for a particular attentional state does not seem to be sufficient to predict the neuron's response for a different attentional state. The appropriate characterization of neuronal responses that takes into account population coding and contextual modulation remains as an important open issue.

(f) The principles proposed above predict a clear relationship between the magnitude of attentional modulation and the capacity of a system. Simulation results suggest that stimuli of high complexity (with respect to the available system's resources) produce stronger modulations. These phenomena resemble those from experiments in which competing stimuli were necessary to observe attentional modulation (Motter, 1993). At the behavioral level, an analogous relation between perceptual load and the effect of distractors has already been demonstrated (Lavie, 2001).

(g) Lastly, using similar principles for the analysis of overt orienting in dynamic environments, we predicted that resource allocation strategies would depend on the uncertainty of the stimuli of interest. These predictions were confirmed by experiments in which humans were required to simultaneously estimate the state of multiple changing targets. The results of these experiments demonstrate that different tracking strategies are used depending on the dynamics of the targets and noise levels. In particular, in an array of features with equal relevance, those features that are less certain require more resources when the cost of estimating them erroneously is very high. In comparison to the phenomena evaluated in the first part of this dissertation, here the task itself, and the selected strategy, set the relevance of different features of the display. These examples of active perception are consistent with the hypothesis that attention serves to reduce perceptual uncertainty, and provide tools for predicting the behavior of orienting systems.

9.2 Future work

The introductory chapter presented some of the remaining questions related to orienting phenomena at both behavioral and neuronal levels. We have chosen to focus, in the first part of this dissertation, on the relation between selective attention and the neural code, suggesting various principles and models. These principles and their implementations generated a set of predictions that require experimental support. Here we propose some of these experiments. In addition, we suggest extensions of the analysis that will provide further predictions.

The resource allocation hypothesis predicts a reduced effect of distractors when attention is highly focused. The magnitude of this effect depends on the complexity of the stimulus relative to the capacity of the system. Evidence from behavioral and brain imaging studies (Lavie, 2005) exist, but there is a lack of electrophysiological measurements of single cells or small populations of cells that test this predictions (Section 5.3). Currently available techniques can be used to perform experiments in which the magnitude of the attentional modulation of neuronal activity is measured as perceptual load is varied.

Models that implement the coding principles discussed in this dissertation in a more realistic fashion will provide further predictions regarding the contextual modulation of neural processing and the characterization of neuronal response. Examples of models that take advantage of more detailed neurobiology are presented by Tiesinga (2005) and Schwabe and Obermayer (2005) for particular stages of the visual pathway. While the former study focuses on the particular mechanism of "gain modulation by inhibitory interference", the latter evaluates different mechanisms under a general coding principle. Optimal coding under non-uniform relevance could be evaluated in a similar fashion under more realistic neural architectures that include, for example, plasticity at short time scales and spike-timing information.

Some experiments have found that the stimulus eliciting maximal response in a neuron does not change with attention (McAdams and Maunsell, 1999), whereas other studies found changes in the preferred stimulus (Motter, 1993; Connor et al., 1997). New experiments must be designed to evaluate under which conditions and for which modalities these changes are observed. Analysis of these experiments should explore the separability of sensory processing and modulatory effects, by evaluating if attentional modulation of activity under some stimuli can be predicted from modulation under different stimuli. Emphasis must therefore be given not only to new measurements, but to the fashion in which responses are characterized. In visual neurons, for example, traditional methods that define classical receptive fields and tuning curves may be insufficient. New characterizations should place the activity of the neurons in a framework of population coding, and take into account the separability (or not) between modulatory effects and sensory processing.

The experiments and models discussed in the last part of the dissertation evaluated the relationship between orienting and the uncertainty of the targets. This evaluation did not take into account situations in which the task sets a different relevance to each feature, as it was assumed in the first chapters of this dissertation. Further experiments should evaluate changes in strategies when non-uniform relevance is required, and analytical solutions for the optimal strategies when these requirements are included should be derived.

9.3   Final remarks

The work presented in this dissertation demonstrates that principles of optimality account for behavioral as well as neuronal phenomena related to orienting systems and selective attention.

With the use of theoretical and computational methods to complement experimental measurements, we are approaching comprehensive answers to longstanding questions concerning the neurobiology of attention. These new explanations take us one step further towards the understanding of the neural code and how the nervous system guides behavior.

APPENDIX

A.1 Example of patterns used in simulations
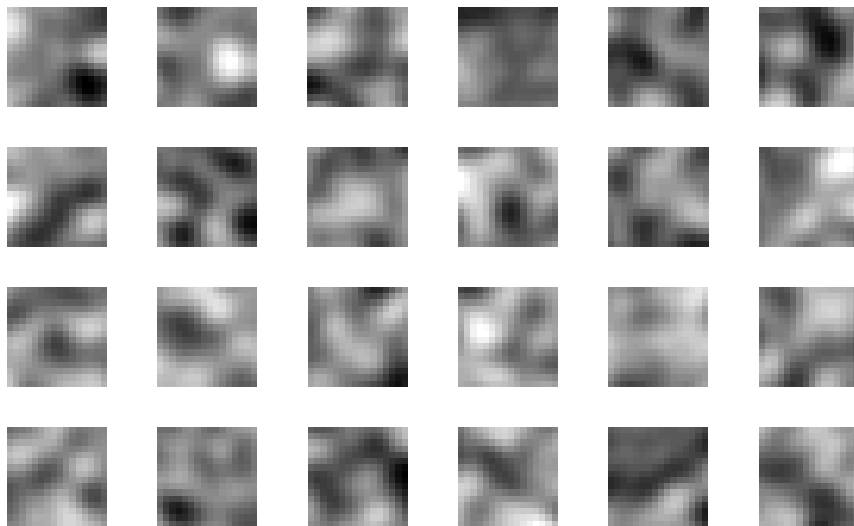
The images above correspond to 24 example patterns with the same statistics as those used for training and testing the model described in Chapters 3 and 5. Images are scaled between -1 (black) and 1 (white).

A.2 Backpropagation of modified error

x_k: input to network, unit k y_k: output of network, unit k

z_j: output of unit j w_{ji}: weight from unit i to unit j

Neuron model:
$$z_j = g(a_j) = A \tanh(B \, a_j) \tag{1}$$
$$a_j = \sum_i w_{ji} \, z_i \tag{2}$$

Error function:
$$E_\mathbf{p} = \frac{1}{2} \sum_k c_k (y_k - x_k)^2 \tag{3}$$
$$E_\text{total} = \sum_p E_p \tag{4}$$

Error gradient:
$$\frac{\partial E_\mathbf{p}}{\partial w_{ji}} = \frac{\partial E_\mathbf{p}}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} = \frac{\partial E_\mathbf{p}}{\partial a_j} z_i \tag{5}$$

Output units:
$$\delta_k \equiv \frac{\partial E_\mathbf{p}}{\partial a_k} = g'(a_k) \frac{\partial E_\mathbf{p}}{\partial y_k} \tag{6}$$
$$g'(a_k) = \frac{B}{A} [A - g(a_k)][A + g(a_k)] \tag{7}$$
$$\frac{\partial E_\mathbf{p}}{\partial y_k} = c_k (y_k - x_k) \tag{8}$$

Hidden units:
$$\delta_j \equiv \frac{\partial E_\mathbf{p}}{\partial a_j} = \frac{\partial E_\mathbf{p}}{\partial z_j} \frac{\partial z_j}{\partial a_j} = \frac{\partial E_\mathbf{p}}{\partial z_j} g'(a_j) \tag{9}$$
$$\frac{\partial E_\mathbf{p}}{\partial z_j} = \sum_k \delta_k w_{jk} \tag{10}$$

Learning rule:
$$\Delta w_{ji} = -\eta \, \delta_j z_i \tag{11}$$

A.3   Additional studies

The work included in this dissertation covers a selection of studies related to resource allocation in the nervous system, developed at the Brain & Computation Lab. Parts of this research project have been published in different forms, including conference and journal papers (Jaramillo and Pearlmutter, 2004b,a, 2006b,c,d).

During my time as a PhD candidate, I was involved in various other studies which, while not directly related to orienting systems or selective attention, represent a significant portion of my research background. In order to maintain a smooth flow in the topics covered in this manuscript, these additional studies have not been included in detail, and they are instead briefly summarize below.

A.3.1   Brightness Illusions as Optimal Percepts

This study showed that Mach bands and a number of other low-level brightness illusions can be accounted for by assuming that our perceptual system performs simple Bayesian inference using a Gaussian image prior with noisy retinal gangion cells (Pearlmutter and Jaramillo, 2003a; Jaramillo and Pearlmutter, 2006a).

A.3.2   Blind Source Separation of MEG data

In an early study, we evaluated the apparent modulation of the visual evoked response by oscillations in the alpha band. A blind source separation algorithm was applied to magnetoencephalographic signals, and resulting components that localized to the occipital lobe were analyzed. A model composed of independently generated background oscillations and evoked responses replicated the modulation observed in the latency of the responses (Pearlmutter et al., 2001).

In a later review study, we described the application of a Second-Order Blind Identification (SOBI) algorithm to magnetoencephalographic data. The manuscript focused on the advantages of this method for extracting artifacts and improving the localization of brain activity (Pearlmutter and Jaramillo, 2003b).

A.3.3 Time-frequency analysis of MEG data

The Center for Neuromagnetism (CNM) at New York University School of Medicine is a research facility dedicated to the study of cognitive processes through the use of noninvasive functional brain imaging techniques like magnetoencephalography (MEG). While working at CNM, I was involved in various projects in which Independent Component Analysis algorithms were applied to magnetoencephalographic data, and the resulting components were further localized and analyzed in the time, frequency and time-frequency domains.

One of these studies characterized the dynamics of sources associated with an audio-motor reaction-time task. Motor components showed (1) early increases in theta power in response to onset of auditory stimulus and (2) a stimulus induced suppression of alpha and beta power followed by a resychronization in alpha, beta and gamma bands. In addition, phase locking to stimulus at theta and gamma frequency bands was observed (Moran et al., 2004).

In a different study, we evaluated the location and dynamics of sources related to somotasensory and motor processes. By comparing three tasks: one mostly motor, one mostly tactile, and one involving both, we were able to distinguish features that corresponded to each of these processes. With this information, we were able to segregate single trials into motor or somatosensory given the MEG signals alone (Jaramillo et al., 2004).

Bibliography

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci*, 8:407–30. 6.3.1

Arbib, M. A., editor (2003). *The Handbook of Brain Theory and Neural Networks*. The MIT Press, second edition. 3.3.3

Aristotle (2004). *On Sense and the Sensible [350 B.C.]*. [Electronic Resource] eBooks@Adelaide. 1.3

Awh, E. and Pashler, H. (2000). Evidence for split attentional foci. *J Exp Psychol Hum Percept Perform*, 26(2):834–46. 2.2

Baccus, S. A. and Meister, M. (2002). Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–19. 1.1

Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE*, 76(8):996–1005. 7.1

Baldi, P. (2005). *Surprise: A shortcut for attention?*, chapter 5, pages 24–48. In Itti et al. (2005). 2.6

Barlow, H. (2001). Redundancy reduction revisited. *Network*, 12(3):241–53. 6.3.2

Bauer, M., Oostenveld, R., Peeters, M., and Fries, P. (2006). Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *J Neurosci*, 26(2):490–501. 4.2

Beauchamp, M. S., Cox, R. W., and DeYoe, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J Neurophysiol*, 78(1):516–20. 4.1

Bell, A. J. and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Res*, 37(23):3327–38. 6.3.2

Braun, J., Koch, C., and Davis, J. L., editors (2001). *Visual Attention and Cortical Circuits*. The MIT Press, Cambridge, MA. 1.3, A.3.3

Brefczynski, J. A. and DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci*, 2(4):370–4. 4.1

Broadbent, D. (1954). The role of auditory localization in attention and memory span. *J Exp Psychol*, 47(3):191–6. 1.3

Bundesen, C. (1990). A theory of visual attention. *Psychol Rev*, 97(4):523–47. 2.6

Bundesen, C., Habekost, T., and Kyllingsbaek, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychol Rev*, 112(2):291–328. 2.6, 2.6, 4.6

Butts, D. A. and Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biol*, 4(4):e92. 1.2

Cavanaugh, J. and Wurtz, R. (2004). Subcortical modulation of attention counters change blindness. *J Neurosci*, 24(50):11236–43. 4.5

Colby, C. L. and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu Rev Neurosci*, 22:319–49. 6.2

Connor, C. E., Gallant, J. L., Preddie, D. C., and Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol*, 75(3):1306–8. 5.2, 5.5

Connor, C. E., Preddie, D. C., Gallant, J. L., and Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *J Neurosci*, 17(9):3201–14. 4.4.2, 4.6, 5.2, 5.2, 5.5, 9.2

Constantinidis, C. and Steinmetz, M. A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. *J Neurosci*, 25(1):233–8. 4.4.4

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248(4962):1556–9. 4.1

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J Neurosci*, 11(8):2383–402. 4.1

Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. (1993). A PET study of visuospatial attention. *J Neurosci*, 13(3):1202–26. 2.1, 4.4

Corbetta, M. and Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*, 3(3):201–15. 4.1

Crowe, D. A., Averbeck, B. B., Chafee, M. V., and Georgopoulos, A. P. (2005). Dynamics of parietal neural activity during spatial cognitive processing. *Neuron*, 47(6):885–91. 1.4, 4.4.4

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience*. The MIT Press, Cambridge, MA. 1.5, 4.6, 5.4, 6.3.1

Dayan, P., Kakade, S., and Montague, P. (2000). Learning and selective attention. *Nat Neurosci*, 3 Suppl:1218–23. 7.3.4

de Boer, R. and Kuyper, P. (1968). Triggered correlation. *IEEE Trans Biomed Eng*, 15(3):169–179. 5.1

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 18:193–222. 2.6, 4.4.2, 6.2

Downing, C. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *J Exp Psychol Hum Percept Perform*, 14(2):188–202. 2.1, 2.1

Driver, J. (2001). A selective review of selective attention research from the past century. *Br J Psychol*, 92 Part 1:53–78. 1.3

Duncan, J., Ward, R., and Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, 369(6478):313–5. 2.3

Egeth, H. E. and Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu Rev Psychol*, 48:269–97. 2.3

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci*, 2(10):704–16. 4.2

Everling, S., Tinsley, C. J., Gaffan, D., and Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat Neurosci*, 5(7):671–6. 4.4.4

Everling, S., Tinsley, C. J., Gaffan, D., and Duncan, J. (2006). Selective representation of task-relevant objects and locations in the monkey prefrontal cortex. *Eur J Neurosci*, 23(8):2197–214. 4.4.4

Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu Rev Neurosci*, 23:441–71. 5.4

Field, D. J. (1994). What is the goal of sensory coding? *Neural Comput*, 6(4):559–601. **6.3.1**

Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278. **7.3**

Franconeri, S. L., Hollingworth, A., and Simons, D. J. (2005). Do new objects capture attention? *Psychol Sci*, 16(4):275–81. **1.4**

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508):1560–3. **4.2**

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*, 6(11):1216–23. **4.4.5**

Gazzaniga, M. S. (1996). *Conversations in the Cognitive Neurosciences*. The MIT Press. **1.4**

Gersho, A. and Gray, R. M. (1992). *Vector quantization and signal compression*. Kluwer Academic Publishers, Boston. **7.1**

Gilbert, C. D., Sigman, M., and Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31(5):681–97. **6.3.2**

Goldberg, M. E. and Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *J Neurophysiol*, 35(4):560–74. **1.3, 4.4**

Gottlieb, J., Kusunoki, M., and Goldberg, M. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391(6666):481–4. **4.4.4, 6.2**

Gray, C. M., Konig, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–7. **4.2**

Gruber, T., Müller, M. M., Keil, A., and Elbert, T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human EEG. *Clin Neurophysiol*, 110(12):2074–85. **4.2**

Hancock, P. J. B., Baddeley, R. J., and Smith, L. S. (1992). The principal components of natural images. *Network: Computation in Neural Systems*, 3(1):61–70. **6.3.1**

Hausser, M. and Mel, B. (2003). Dendrites: bug or feature? *Curr Opin Neurobiol*, 13(3):372–83. **6.1.2**

Hayhoe, M. and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn Sci*, 9(4):188–94. **2.5, 7.1**

Haykin, S. (1999). *Neural Networks. A Comprehensive Foundation*. Prentice-Hall, second edition. **3.2**

Heinze, H., Mangun, G., Burchert, W., Hinrichs, H., Scholz, M., Munte, T., Gos, A., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M., and Hillyard, S. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, 372(6506):543–6. **4.3**

Helmholtz, H. v. (1924). *Treatise on Physiological Optics [1866]*. The Optical Society of America, translated from the third german edition (1909). electronic edition (2001): university of pennsylvania edition. **1.3**

Hillyard, S. A. and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci U S A*, 95(3):781–7. **4.3**

Hocherman, S., Benson, D. A., Goldstein, Jr, M. H., Heffner, H. E., and Hienz, R. D. (1976). Evoked unit activity in auditory cortex of monkeys performing a selective attention task. *Brain Res*, 117(1):51–68. **1.3, 4.4.5**

Hoffman, J. E. and Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept Psychophys*, 57(6):787–95. **2.3**

Hsiao, S. S., O'Shaughnessy, D. M., and Johnson, K. O. (1993). Effects of selective attention on spatial form processing in monkey primary and secondary somatosensory cortex. *J Neurophysiol*, 70(1):444–7. **4.4.5**

Hubel, D. H., Henson, C. O., Rupert, A., and Galambos, R. (1959). Attention units in the auditory cortex. *Science*, 129(3358):1279–80. **1.3, 4.4**

Hunt, A. R. and Kingstone, A. (2003a). Covert and overt voluntary attention: linked or independent? *Brain Res Cogn Brain Res*, 18(1):102–5. **2.3**

Hunt, A. R. and Kingstone, A. (2003b). Inhibition of return: Dissociating attentional and oculomotor components. *J Exp Psychol Hum Percept Perform*, 29(5):1068–74. **2.3**

Itti, L. and Baldi, P. (2006). Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005)*, pages 1–8, Cambridge, MA. MIT Press. **1.4, 2.6**

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203. **2.6, 3.4.3, 7.1**

Itti, L., Rees, G., and Tsotsos, J. K., editors (2005). *Neurobiology of Attention*. Academic Press. 1.3, A.3.3

James, W. (1890). *The Principles of Psychology*. Electronic Resource by C. D. Green, York University, Toronto, Canada. 1.3

Jaramillo, S. and Pearlmutter, B. A. (2004a). Exploring the attentional modulation of neural codes. Computational & Systems Neuroscience, COSYNE. A.3

Jaramillo, S. and Pearlmutter, B. A. (2004b). A normative model of attention: receptive field modulation. *Neurocomputing*, 58-60:613–8. A.3

Jaramillo, S. and Pearlmutter, B. A. (2006a). Brightness Illusions as Optimal Percepts. Technical Report NUIM-CS-TR-2006-02, Dept. of Computer Science, NUIM. A.3.1

Jaramillo, S. and Pearlmutter, B. A. (2006b). Optimal coding predicts attentional modulation in neural systems. *Under review*. A.3

Jaramillo, S. and Pearlmutter, B. A. (2006c). Optimal strategies for active perception. Computational & Systems Neuroscience, COSYNE. A.3

Jaramillo, S. and Pearlmutter, B. A. (2006d). Use your powers wisely: resource allocation in parallel channels. In *IEEE Workshop on Machine Learning for Signal Processing (MLSP)*. 2.6, A.3

Jaramillo, S., Ramirez, R. R., Moran, K. A., Ribary, U., and Llinás, R. R. (2004). Can MEG tell movement from touch? In *2004 Abstract Viewer/Itinerary Planner*, number 643.11, Washington, DC. Society for Neuroscience. A.3.3

Johansen-Berg, H. and Lloyd, D. (2000). The physiology and psychology of selective attention to touch. *Front Biosci*, 5:D894–904. 2.1

Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors (2000). *Principles of Neural Science*. McGraw Hill, fourth edition. 6.1.2

Kanwisher, N. and Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nat Rev Neurosci*, 1(2):91–100. 4.1

Kastner, S. and Ungerleider, L. (2000). Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci*, 23:315–41. 4.1

Khayat, P. S., Spekreijse, H., and Roelfsema, P. R. (2006). Attention lights up new object representations before the old ones fade away. *J Neurosci*, 26(1):138–42. 4.4.1

Klein, R. and Farrell, M. (1989). Search performance without eye movements. *Percept Psychophys*, 46(5):476–82. **2.3**

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci*, 27(12):712–9. **1.2**

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–27. **2.6**

Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, 2nd, M. J., Balasubramanian, V., and Sterling, P. (2006). How much the eye tells the brain. *Curr Biol*, 16(14):1428–34. **1.1**

Körding, K. P. and König, P. (2001). Supervised and unsupervised learning with two sites of synaptic integration. *J Comput Neurosci*, 11(3):207–15. **6.1.2**

Larkum, M. E., Zhu, J. J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725):338–41. **6.1.2**

Larkum, M. E., Zhu, J. J., and Sakmann, B. (2001). Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons. *J Physiol*, 533(Pt 2):447–66. **6.1.2**

Lavie, N. (2001). *Capacity Limits in Selective Attention: Behavioral Evidence and Implications for Neural Activity*, chapter 3, pages 49–68. In Braun et al. (2001). **1.3, f**

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn Sci*, 9(2):75–82. **1.3, 9.2**

Law, C. C. and Cooper, L. N. (1994). Formation of receptive fields in realistic visual environments according to the Bienenstock, Cooper, and Munro (BCM) theory. *Proc Natl Acad Sci U S A*, 91(16):7797–801. **6.3.1**

Lee, D., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat Neurosci*, 2(4):375–81. **2.1**

Liversedge, S. P. and Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends Cogn Sci*, 4(1):6–14. **2.5, 6.2, 8.4**

London, M. and Häusser, M. (2005). Dendritic computation. *Annu Rev Neurosci*, 28:503–32. **6.1.2**

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol*, 77(1):24–42. 4.4.2

Luck, S. J., Hillyard, S. A., Mouloua, M., and Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? *J Exp Psychol Hum Percept Perform*, 22(3):725–37. 3.3.3

Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn Sci*, 4(11):432–440. 4.3

Lynch, J. C., Mountcastle, V. B., Talbot, W. H., and Yin, T. C. (1977). Parietal lobe mechanisms for directed visual attention. *J Neurophysiol*, 40(2):362–89. 1.3, 4.4

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol Bull*, 109(2):163–203. 1.1

Mazzoni, P., Andersen, R. A., and Jordan, M. I. (1991). A more biologically plausible learning rule than backpropagation applied to a network model of cortical area 7a. *Cereb Cortex*, 1(4):293–307. 3.3.3

McAdams, C. and Maunsell, J. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci*, 19(1):431–41. 4.4.1, 4.4.2, 4.6, 5.4, 9.2

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognit Psychol*, 14(2):247–79. 2.2

Moore, T. and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921):370–3. 4.5

Moore, T. and Fallah, M. (2001). Control of eye movements and spatial attention. *Proc Natl Acad Sci U S A*, 98(3):1273–6. 4.5

Moore, T. and Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *J Neurophysiol*, 91(1):152–62. 4.5

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–4. 4.4, 4.4.1, 4.4.2, 4.6, 5.5

Moran, K. A., Ramirez, R. R., Jaramillo, S., Ribary, U., and Llinás, R. R. (2004). Event related spectral perturbations and coherence in a simple audio-motor reaction time task. In *BIOMAG-2004*. A.3.3

Motter, B. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J Neurosci*, 14(4):2178–89. 4.4.2, a

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol*, 70(3):909–19. 4.4.1, 4.4.2, f, 9.2

Mozer, M. C. and Sitton, M. (1998). Computational modeling of spatial attention. In Pashler (1998), chapter 9. 2.6, 3.3.3, 3.4.3

Müller, J., Philiastides, M., and Newsome, W. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc Natl Acad Sci U S A*, 102(3):524–9. 4.4, 4.5

Müller, M. M., Malinowski, P., Gruber, T., and Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, 424(6946):309–12. 2.2

Müller, M. M., Teder-Sälejärvi, W., and Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nat Neurosci*, 1(7):631–4. 4.2

Najemnik, J. and Geisler, W. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–91. 2.6, 8.4

Nakahara, H., Wu, S., and Amari, S. (2001). Attention modulation of neural tuning through peak and base rate. *Neural Comput*, 13(9):2031–47. 4.6

Niebur, E. and Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J Comput Neurosci*, 1(1-2):141–58. 6.1.3

Niebur, E., Koch, C., and Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Res*, 33(18):2789–802. 6.1.3

Nobre, A. C. (2001). Orienting attention to instants in time. *Neuropsychologia*, 39(12):1317–28. 2.3

O'Connor, D., Fukui, M., Pinsk, M., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat Neurosci*, 5(11):1203–9. 4.1

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., and Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18(4):591–8. 4.1

Olshausen, B., Anderson, C., and Van Essen, D. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci*, 13(11):4700–19. 4.6, 5.5, 6.1.1

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–9. 5.5, 6.3.2

O'Reilly, R. C. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. The MIT Press, Cambridge, MA. 1.5

Parasuraman, R., editor (2000). *The attentive brain*. The MIT Press, Cambridge, MA. 1.3

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol Bull*, 116(2):220–44. 2.2

Pashler, H., editor (1998). *Attention*. Psychology Press. 1.3, A.3.3

Pashler, H. E. (1999). *The Psychology of Attention*. The MIT Press, Cambridge, MA. 1.3

Pearlmutter, B. A. and Jaramillo, S. (2003a). Brightness illusions and saturation. In *2003 Abstract Viewer/Itinerary Planner*, number 698.13, Washington, DC. Society for Neuroscience. A.3.1

Pearlmutter, B. A. and Jaramillo, S. (2003b). Progress in blind separation of magnetoencephalographic data. In *Proc. SPIE*, volume 5102, pages 129–134. A.3.2

Pearlmutter, B. A., Tang, A. C., Jaramillo, S., and Nolte, G. (2001). Modulation of single-trial visual evoked response onset times by alpha-band oscillation. In *2001 Abstract Viewer/Itinerary Planner*, number 821.1, Washington, DC. Society for Neuroscience. A.3.2

Pessoa, L., Kastner, S., and Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *J Neurosci*, 23(10):3990–8. 4.1

Picton, T. W. and Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr Clin Neurophysiol*, 36(2):191–9. 1.3, 4.3

Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat Neurosci*, 7(6):621–7. 6.1.2

Posner, M., Snyder, C., and Davidson, B. (1980). Attention and the detection of signals. *J Exp Psychol*, 109(2):160–74. 2.1, 2.1, 2.6

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proc Natl Acad Sci U S A*, 91(16):7398–403. 1.4

Posner, M. I., editor (2004). *Cognitive neuroscience of attention*. The Guilford Press. 1.3

Price, D. J. and Willshaw, D. J. (2000). *Mechanisms of Cortical Development*. Oxford University Press. 6.3.1

Rabiner, L. R. (1989). Tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286. 7.3

Rao, R. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, 16(16):1843–8. 4.6, 5.5

Raz, A. and Buhle, J. (2006). Typologies of attentional networks. *Nat Rev Neurosci*, 7(5):367–79. 1.1

Reynolds, J., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci*, 19(5):1736–53. 4.4.2, 4.6, 4.6, 6.1.1

Reynolds, J. H. and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annu Rev Neurosci*, 27:611–47. 1.2

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1996). *Spikes: Exploring the Neural Code*. The MIT Press. 1.2, 1.2

Roelfsema, P. R., Lamme, V. A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395(6700):376–81. 4.4.1

Romo, R., Hernandez, A., Zainos, A., and Salinas, E. (1998). Somatosensory discrimination based on cortical microstimulation. *Nature*, 392(6674):387–90. 1.2

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back–propagating errors. *Nature*, 323:533–536. 3.2

Salinas, E. and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27(1):15–21. 5

Scharf, B. (1998). *Auditory Attention: The Psychoacoustical Approach*, chapter 2, pages 75–118. In Pashler (1998). 2.1, 2.3

Schwabe, L. and Obermayer, K. (2005). Adaptivity of tuning functions in a generic recurrent network model of a cortical hypercolumn. *J Neurosci*, 25(13):3323–32. 4.6, 6.1.3, 9.2

Sheliga, B. M., Riggio, L., and Rizzolatti, G. (1994). Orienting of attention and eye movements. *Exp Brain Res*, 98(3):507–22. 2.3

Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850):338–40. 4.4.2, a

Sprague, N. and Ballard, D. (2004). Eye movements for reward maximization. In Thrun, S., Saul, L., and Schölkopf, B., editors, *Advances in Neural Information Processing Systems 16*. The MIT Press, Cambridge, MA. 2.5, 2.6

Sripati, A. P. and Johnson, K. O. (2006). Dynamic gain changes during attentional modulation. *Neural Comput*, 18(8):1847–67. 4.4.5

Stanley, G. B., Li, F. F., and Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J Neurosci*, 19(18):8036–42. 1.2

Steinmetz, P. N., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K. O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404(6774):187–90. 4.2

Styles, E. A. (1997). *The psychology of attention*. Psychology Press Ltd. 1.3

Tiesinga, P. (2005). Stimulus competition by inhibitory interference. *Neural Comput*, 17(11):2421–53. 6.1.3, 9.2

Tootell, R., Hadjikhani, N., Hall, E., Marrett, S., Vanduffel, W., Vaughan, J., and Dale, A. (1998). The retinotopy of visual spatial attention. *Neuron*, 21(6):1409–22. 4.1

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognit Psychol*, 12(1):97–136. 2.6

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychol Rev*, 76(3):282–99. 1.3

Treue, S. and Martinez Trujillo, J. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736):575–9. 4.4.3

Treue, S. and Maunsell, J. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382(6591):539–41. 4.4.3

Treue, S. and Maunsell, J. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci*, 19(17):7591–602. 4.4.3, 5.2, 5.3

van Swinderen, B. (2005). The remote roots of consciousness in fruit-fly selective attention? *Bioessays*, 27(3):321–30. 1.3

Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–6. 6.3.2

Voytko, M. L., Olton, D. S., Richardson, R. T., Gorman, L. K., Tobin, J. R., and Price, D. L. (1994). Basal forebrain lesions in monkeys disrupt attention but not learning and memory. *J Neurosci*, 14(1):167–86. 6.1.1

Winkowski, D. E. and Knudsen, E. I. (2006). Top-down gain control of the auditory space map by gaze control circuitry in the barn owl. *Nature*, 439(7074):336–9. 4.5

Wolfe, J. M. (1998). *Visual search*, chapter 1, pages 13–74. In Pashler (1998). 2.3, 2.4, 2.6

Wurtz, R. H. and Mohler, C. W. (1976). Enhancement of visual responses in monkey striate cortex and frontal eye fields. *J Neurophysiol*, 39(4):766–72. 1.3, 4.4

Yu, A. and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–92. 5.5, 6.1.1

Yu, A. J. and Dayan, P. (2005). Inference, attention, and decision in a bayesian neural architecture. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 1577–1584. The MIT Press, Cambridge, MA. 2.6, 4.6, 6.1.3

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):679–84. 5