Adaptivity of Sensory Representations
in the Adult Visual Cortex

vorgelegt von

Diplom-Informatiker

Lars Schwabe

aus Berlin

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Olaf Hellwich
Berichter: Prof. Dr. rer. nat. Klaus Obermayer
Berichter. Prof. Dr. Alessandra Angelucci

Tag der wissenschaftlichen Aussprache: 16.12.2005

Berlin 2005

D 83

Acknowledgments[1]

I express my greatest gratitude to Prof. Klaus Obermayer, the supervisor of this thesis. The scientific environment he set up at the Berlin University of Technology was highly inspiring. The numerous discussions, seminars, invited guests, workshops, and meetings with our collaborators were essential sources of motivation and triggers of new ideas. He was always critical, discussions with him were always focused, and I like to thank him in particular for shaping my scientific thinking. In this context, I also like to thank our group's Postdocs Drs. Michael Sibila, Martin Stetter, and Sepp Hochheiter, because they also contributed so much to make our group an inspiring scientific environment.

Among our collaborators, I would like to express special thanks to Profs. Alessandra Angelucci, Paul Bressloff, Jenny Lund, and Mriganka Sur. The focused scientific discussions paired with their unique openness, kindness and hospitality were unique. I learned so much by interacting with them. I also like to thank Dr. Matthias Hennig from the Humboldt University. I very much enjoined our discussions and joint work about adaptation in the cricket auditory system.

I am greatly indebted to all my colleagues at the Neural Information Processing Group, in particular Oliver Beck and Peter Wiesing – it was a pleasure to work with them in a close and fruitful collaboration –, Gregor Wenning, Roland Vollgraf, Sambu Seo, Johannes Mohr, Hendrick Purwins, Thomas Hoch, Klaus Wimmer, and Joshua Young. Interacting with all the members of the graduate school 'signaling cascades in living systems' was fun and inspiring, and without the support of this school the unforgettable 'International Neuroscience Summit 2002', which I set up together with Peter Wiesing, Gregor Wenning, and much help from our Neural Information Processing Group, would not have been possible.

I also like to thank the whole team of the 'EU Advanced Course in Computational Neuroscience', where I had the pleasure to work as a student and tutor in the summers of 2000, 2002, and 2003. In particular, I like to thank 'my students'. I hope that you learned as much as I did. Special thanks, however, go to Tanya Baker and Gordon Pipa. I am honored to had you under my tutorship, and I hope that our paths continue to cross. Of course, I also like to thank my Diploma students Jerome Rougnon-Glasson, Andre Nuppenau, Philipp Kallerhoff, and Ralf Ansorg. Interacting with them was fun and fruitful.

I dedicate this work to Anja and my parents.

Summary

The developing sensory areas of the brain are shaped by a combination of environmental signals imposing their fingerprints on them and genetically predetermined factors. Experimental studies, however, have shown that they are still surprisingly adaptive even in the adult. We investigate the representation of visual information in the visual cortex of mammals and focus on explaining their adaptivity in the adult. One insight from computer science is that the most efficient algorithms depend on the structure of the underlying machine. Therefore, we admit a "primacy of the neuronal circuits" as compared to an approach, where first an algorithmic basis for (visual) cortical function is hypothesized and then a corresponding neuronal correlate is searched.

We first investigate orientation tuning in the primary visual cortex (V1). The combination of recent measurements with modeling studies allows us to significantly constrain the actual operating regime of V1 to a recurrency-regime characterized by a co-variation of excitation and inhibition, hence almost settling the long-standing issue of how orientation tuning is computed in V1. This model does not directly relate to our main theme of "adult adaptivity". However, it is an important step for investigating adaptivity in the early visual cortex, because experiments have shown that orientation adaptation is stronger in V1 pinwheel regions compared to orientation domains (Dragoi et al., 2001). Our two-dimensional model explicitly considers these different regions, and hence could serve as an experimentally tested model for subsequent investigations.

We then re-investigate the neuronal circuitry underlying contextual effects in V1, which can be viewed as a form of stimulus-driven adaptivity of sensory representations. Again, integrating recent experimental data leads us to propose a new inter-areal model, where modulatory interactions are mediated via an inter-areal pathway, such that they could be subject to attentional modulation. The new model's specific predictions were recently confirmed experimentally. In this thesis, we view attentional top-down modulations of visual representations also as one form of adaptivity. Hence, the inter-areal model could serve as an experimentally tested model for subsequent investigations of attentional top-down control.

In a third study, we use an "optimal coding principle" to explain tuning function changes in area V4 after perceptual learning and due to visual attention. Although these two phenomena give rise to different tuning function changes, we show that they can both be understood as optimal improvements of the sensory representation, but being caused by different mechanisms. The developed approach of predicting tuning function changes for particular mechanisms is of use to investigate the mechanisms of various adaptation phenomena, which may be hard to address directly, based on the tuning functions as the "observable".

Finally, using a more abstract network model, we develop a new interpretation of task-dependent representations as caused by, for example, attentional top-down control. We argue that they are best understood as top-down modulations, which "tune" the computation of the sensory representations in a task-dependent way such that they are useful for read-out neurons, which inherit their task-dependence from the neurons in the sensory area.

Contents

Chapter 1

Introduction to this thesis

When should an organism exploit its environment in order to benefit from its previously acquired knowledge, and when should an organism continue to explore its environment to acquire even more knowledge for later exploitation? This famous and well-known *exploration-exploitation tradeoff* (Kaelbling et al., 1996) is a rather high-level problem relevant for understanding the behavior of biological organisms and for building intelligent artificial agents. If an agent never explores its environment, it will never learn something about it. On the other hand, too much exploration will also not pay off, because in this case the agent seldom exploits the acquired knowledge to obtain rewards. This problem, however, has a low-level counter part, which in turn is highly relevant for understanding biological sensory systems and, of course, for building the sensing devices of intelligent artificial agents. The so-called *plasticity-stability tradeoff* (Grossberg, 1976) refers to the problem of deciding when and how to adapt the internal sensory representation of the outside world.

On the one hand, subsequent processing stages rely on stable and reproducible sensory representations when performing their computations like, for example, the decision between exploration and exploitation. On the other hand, sensory representations need to be plastic in order to ensure a high fidelity representation of sensory stimuli whose behavioral relevance may not be known beforehand. In other words, sensory representations need to be adaptive. Indeed, adaptation is a widespread phenomenon in nervous systems, and it happens on multiple time-scales. In the context of the early visual system the activity-dependent refinement of cortical maps happens during a critical phase, which lasts weeks, perceptual learning occurs during hours and days, light adaptation in the retina and contrast-adaptation in the primary visual cortex (V1) works on the order of a few seconds, and the fast dynamics of receptive field properties on the order of a few tens or hundred milliseconds may also be viewed as an adaptive process.

Together with the apparent general applicability of the idea of adaptation comes a certain vagueness: For example, what exactly do contrast-adaptation and the development of cortical maps have in common? Is it indeed justified to think of dynamic receptive field properties or even attentional top-down modulations in the visual cortex as some sort of adaptation? Clearly, understanding adaptation only in the narrow sense of a neuronal fatigue after prolonged stimulation falls too short. Therefore, from now on we use the term 'adaptivity of

sensory representations' when referring to the adaptation of sensory representations in a wider sense, i. e. an adaptation to changing internal or external conditions at various time-scales. The term 'adaptation' will be used in the usual way as referring to the phenomenon of firing rate adaptation observed in single neurons.

It may come as a surprise, but the visual system is indeed quite adaptive even in the adult (Dragoi and Sur, 2003). This means, taking at first a rather naive view, that we seldom see the world the same way twice. This should not be confused with the phenomenon of the same visual scene being interpreted in a different way when we see it the second time, because we acquired new knowledge since we saw it first. For example, being told that the scoring of a goal in a soccer game was from an offside position alters the way the corresponding scene is interpreted. However, here we refer to phenomena, which change the representation of visual information in the first place, not only the subsequent interpretation. For example, it has been shown that practicing visual discrimination tasks alters those representations such that after this so-called perceptual learning two very similar, but actually different stimuli can be distinguished, which were indistinguishable before the learning phase.

Let us now consider some selected examples, which demonstrate that the visual cortex connects an agent to the visual world in a much more complex and "adaptive" way than simply being a mirror of the light pattern received by the two eyes.

1.1   Examples of adaptivity in the visual cortex

In this thesis we investigate the adaptivity of sensory representations in the adult visual cortex. In other words, we consider how the representation of visual information is changed in the first place as compared to altered subsequent interpretations of otherwise unchanged representations. Visual illusions have been successfully used to demonstrate aspects of the way the visual system processes visual information. Therefore, let us now consider a set of illusions selected to demonstrate the visual system's adaptivity.

Fig. 1.1a shows a variant of the so-called Zöllner illusion. Here, the long horizontal lines are indeed parallel but appear to tilt alternately. In other words, the short lines change the way the long horizontal lines are perceived. In particular, this illusion is strongest when the intersecting angle of the short and long lines is between 10° and 30°. Like almost any visual illusion, the Zöllner illusion does not reveal a malfunctioning of the visual system. Instead, using a rather artificial stimulation they demonstrate that the visual system incorporates prior knowledge about the structure of the visual world in order to compute representations, which are useful for proper perception and action during natural stimulation. For example, the Zöllner illusion might be an epiphenomenon of how depth information is computed during natural stimulation. The Zöllner illusion, however, builds upon very simple stimulus properties, and it can hardly be altered by cognitive efforts like visual attention. This suggest that the corresponding neuronal processes are rather 'hard-wired' and are likely to be found in the early stages of visual processing like area V1. Currently it is assumed that orientation tuning in area V1 also strongly depends on genetic factors (Kaschube et al., 2002). Thus, the Zöllner illusion can be viewed as an example of how the
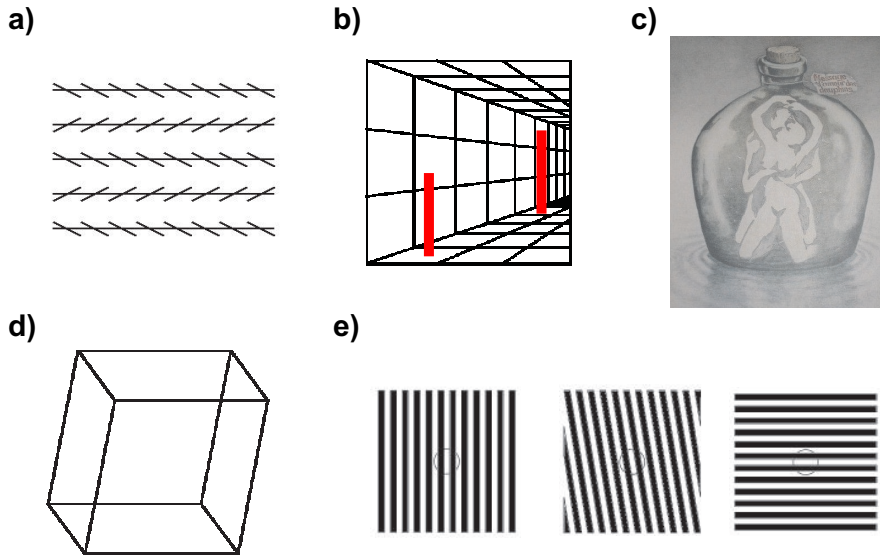
Figure 1.1: Examples illustrating adaptivity of the visual system at different time-scales. **a)** In the Zöllner illusion the long lines are indeed parallel, but due to the superimposed short lines they do not appear as parallel. **b)** In this variant of the so-called corridor illusion the two red lines have the same size, but due to the depth cue the right line appears to be longer. **c)** Adults identify the image on the bottle as a couple, whereas young children only see dolphins due the lack of prior memory associated with such scenario. **d)** In the Necker cube illusion two sides of the cube can be seen as been nearer to the observer. **e)** Stimuli to demonstrate the direct and indirect tilt aftereffect, where after prolonged viewing of the middle/right image the vertical lines appear tilted to the right/left (see text for details).

visual system adapted to the environment, but on an *evolutionary time-scale*.

An illusion explicitly exploiting the way depth information is computed by the visual system is shown in Fig. 1.1b. In this so-called corridor illusion the sets of converging lines are easily interpreted as perspective cues. As a consequence, the right bar is perceived as being longer than the left bar, although they are of the same size. Note that the impression of depth is evoked also if the corridor illusion is viewed with only one eye. In other words, the perspective cue does not depend on integrating information from both eyes. It is not clear to what extend visual experience is necessary to infer depth information from such monocular clues. It is possible that the main mechanisms underlying the corridor illusion are 'hard-wired' as it is likely the case in the Zöllner illusion, which would imply that the corridor illusion is an example of adaptivity on the evolutionary time-scale, but if visual experience is essential, then it would be an example of an adaptation to the environment at the *developmental time-scale* of the ontogenesis of the animal.

One clear example demonstrating adaptivity at the developmental time-scale is shown in Fig. 1.1c. The figure on the bottle can be identified in two different ways, namely either as a couple or a set of dolphins. It can easily be tested that

young children cannot identify the intimate couple. They have never seen such a scenario. They will only see the dolphins. In other words, prior experience can strongly influence the way a physically ambiguous stimulus is perceived. Similar to other ambiguous figures like the famous duck-rabbit or the face-vase illusion, both interpretations cannot be seen at the same time. By itself, of course, this is no direct evidence for the hypothesis that the mutual exclusive character of these illusions can be explained by assuming that the visual system only represents one interpretation at any one time.

The Necker cube shown in Fig. 1.1d is an example of a similar illusions. Two faces of this cube can appear to be nearer to the observer, but the two mutually exclusive interpretations switch after prolonged viewing or can be changed by visual attention. Assuming that visual attention can alter the way sensory representations are computed in the visual system, Figs. 1.1c,d are examples of adaptivity of sensory representations on a *perceptual time-scale* subject to cognitive top-down modulation.

The illusion shown in Fig. 1.1e, however, is largely immune against interference with cognitive processes like visual attention. Just fixate the area surrounded by the circle the middle panel of Fig. 1.1e. You can slightly move your eyes in order to avoid strong afterimages. Then, quickly fixate the left panel of Fig. 1.1e. The vertical lines should appear slightly tilted clockwise. This phenomenon is called the direct tilt aftereffect. If you fixate the horizontal lines at the right, after having fixated the area surrounded by the circle in the middle panel, then the lines should appear barely tilted counterclockwise, which is called the indirect tilt aftereffect. The indirect effect, however, is small and not always noticeable outside the laboratory. The stimulus property relevant for these effects is the orientation of the lines, and in the visual pathway of mammals, neurons selective for the orientation of a visual stimulus are first found in area V1. Therefore, the corresponding neuronal processes of this perceptual effect could be located in area V1. Interestingly, animal studies have shown that after prolonged stimulation the orientation selectivity of neurons in area V1 is slightly changed (Dragoi and Sur, 2003), which could account for these effects. These effects are examples of adaptivity on a perceptual time-scale subject to stimulus-dependent modulations.

These examples show that the visual system connects a perceiver to the world in a rather complex way. In particular, the visual system seems to be adaptive at different time-scales, and it is a challenge to explain these phenomena in terms of specific computational principles. In other words, using such an explanation one would argue, that the visual system is operating "optimal" (in a sense to be defined), and that the illusions are a natural consequence of the visual system operating according to the proposed principle (see (Schwabe and Obermayer, 2003) for a review of those approaches).

One very prominent hypothesis is to view the visual system as a statistical inference engine, which aims to infer the probable "causes" of the retinal activation patterns. In general, however, this problem is ill-posed. For example, the right bar in Fig. 1.1b could be of the same size as the left bar, if it would be located at the same distance to the perceiver as the left bar. However, it could also be larger, if it would be located at a larger distance to the perceiver. In such cases, prior knowledge has to be used in order to resolve this ambiguity. Indeed, the perspective cues suggest that the right bar is located at a larger distance than the left bar. This prior knowledge is then used by the visual system to

infer that the right bar is larger.

This view of the visual system is attractive, because it could explain its adaptivity in terms of the prior knowledge used to disambiguate the visual stimuli. The appropriate prior knowledge could vary on a case-by-case basis, but also on different time-scales. The formalism of Bayesian inference (Cox, 1961) is a prime candidate to serve as a basis for this "inference engine" perspective. However, if the (visual) brain makes use of Bayesian inference, then it has make use of proper *approximate* Bayesian learning and inference algorithms, because straight-forward implementations are only possible for almost trivial statistical models of the environment. In other words, although this perspective has gained much interest in recent years (see, e. g., (Barlow, 2001; Pouget et al., 2003; Rao, 2004; Deneve, 2005)), the specific approximate algorithms possibly used by the (visual) brain need to be spelled out in detail (but see (Yu and Dayan, 2005) for a possible and neurophysiologically plausible approach). In this thesis, however, we do not adopt this hypothesis. Instead, we will ask (see Chapter 5) as to whether this hypothesis is justified at all.

Our approach in this thesis is to admit a "primacy of the neuronal circuits". We argue that instead of postulating an algorithmic basis of visual cortical function and then searching for the corresponding neuronal correlate, focusing on the neuronal circuits with the goal of formulating simple models may turn out to be a more promising approach for developing a better understanding of visual cortical functions as well as for building "intelligent machines". Therefore, we now give a brief introduction into the architecture of the visual system before we formulate the specific questions we address (Sec. 1.3).

1.2 The central visual pathway

Here we briefly describe the central visual pathway of mammals. This pathway is certainly the best investigated signal processing pathway of higher mammals, because it is easy to access and can be stimulated in a straight forward fashion. This introduction is far from being comprehensive. Instead, it is intended as a minimal primer for the rest of this thesis. Much more in-depth reviews of this pathway can be found in textbooks like the one by Kandel et al. (2000), Zigmond et al. (2000), Palmer (1999), or Hubel (1988).

Fig. 1.2 sketches the overall anatomical structure of the early visual pathway. When a cat or monkey fixates its environment, light that falls into the eye is focused by the cornea and the lens to form two images on the *left and right retinae*. The part of the world that contributes to the image formed on the two retinae is called the *visual field* of the animal. Each retina transforms the incoming light intensity distribution into spike patterns, which are transmitted by the two *optic nerves* into the central nervous system. Each optic nerve consists of roughly 10^6 fast conducting axons, almost all of which target two structures within the thalamus, which are called *lateral geniculate nuclei* or LGN. At the *optic chiasm*, each optic nerve branches such that one half of the fibers targets the 'contralateral' LGN at the side opposite to its origin, whereas the rest contacts the 'ipsilateral' LGN at the same side as the eye of origin. The crossover of the nerve fibers occurs in a highly ordered fashion, which ensures that each LGN receives the fibers from the ipsilateral parts of both retinae. Thus, each LGN processes the signals from the contralateral hemisphere of the
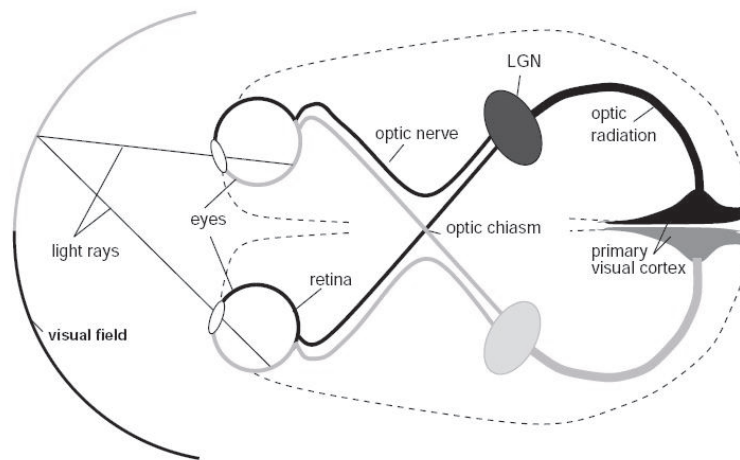
Figure 1.2: Illustration of the early visual pathway in higher mammals. The light is transformed into spikes of the ganglion cells, whose axons form the optic nerve. The signals are forwarded via the optic nerve to the optic chiasm, the LGN, and then via the optic radiation to the V1 (adapted from Bauer (1998)).

visual field.

Proceeding from the LGN, another less concentrated bundle of fibers, which is referred to as the *optic radiation*, contacts the primary visual cortex (also referred to as area V1). In contrast to the optic nerve, the optic radiation does not cross hemispheres. Hence, analogously to the LGN, each hemisphere of the primary visual cortex processes visual information from the contralateral hemisphere of the visual field. From V1, two major output streams can be divided. The first stream projects from V1 to higher visual areas, whereas the second stream projects back to the LGN and other deep structures.

The retina

The retina is the first step of information processing in the brain. Although being located at the sensory periphery, it is part of the brain itself, because it is derived during development from the neural ectoderm, which also gives rise to the brain. The retina is a layered structure composed of many different cell types. The cells transducing light into graded electrical potentials are the *photoreceptors*. The so-called "rods" are very sensitive to light and are crucial for vision in low-light environments like, for example, during the night, whereas the so-called "cones" signal color information and are crucial for vision during the day. The signals transduced by the photoreceptors are forwarded to the *ganglion cells*, which are the first to use spikes instead of graded potentials for signalling.

A single ganglion cell receives input from several photoreceptors such that beyond light adaptation of the photoreceptors a first pre-processing of the visual signals can take place. Per eye approx. 120 million photoreceptors converge onto approx. 1 million ganglion cells. In other words, this first processing can

also be characterized as a massive *data compression*. Fig. 1.3a show the two types of ganglion cells termed *ON- and OFF-center cells*. If the center region of ON-center cells is stimulated the firing rate of this ganglion cell increases (top), whereas stimulation of the surrounding region leads to a decrease in the frequency (middle) of the spike discharges. Stimulation of both the center and the surround has no effect (bottom). The response properties of the OFF-cells are inverse to those of the ON-cells.

Of course, these response properties are the results of dynamic interactions in the retinal circuitry, but they can be described in a compact way using the difference of two Gaussian functions (the so-called *DoG-model*). In Chapter 3 we will consider how the response properties of neurons in V1 are modulated by the spatial context of visual stimulation, similar to the facilitatory center and suppressive surround regions of the ON-cells. The DoG-model has also been used to describe these contextual modulations in V1, but in Chapter 3 we show that it is a poor model of V1 circuitry in the sense of failing to account for contextual modulations predicted using the model we develop in Chapter 3. In particular, it fails to account for the predicted facilitation from the far surround, which recently has been confirmed experimentally (Ichida et al., 2005). In a similar way, viewing the retina only as the location, where a linear filtering, as captured by the DoG-model, is computed may also fall too short in the sense of abstracting too much from the retinal circuits. However, so far the DoG-model has been a very accurate description of the ganglion cell's responses.

The lateral geniculate nucleus

The two LGN (see Fig. 1.2) are nuclei in the thalamus. They are often characterized as relay stages, which forward the signals received from the retinal ganglion cells via the optic nerve to V1. Fig. 1.3b shows a vertical slice through the LGN of a macaque monkey. The LGN is composed of six layers, which receive afferent input from retinal ganglion cells. The layers are innervated in an alternating fashion by ganglion cells from either the temporal region of the retina of the ipsilateral eye or the nasal region of the retina of the contralateral eye (Fig. 1.3c). The two most ventral layers are innervated by the axons of the *magnocellular* ganglion cells (for motion processing), whereas the four inner layers are innervated by axons of *parvocellular* ganglion cells (for color/form vision). For example, the right LGN receives parvo- and magnocellular inputs from the nasal region of the retina of the left eye (the left visual field in the contralateral eye), and the temporal region of the retina of the right eye (the left visual field in the ipsilateral eye). Thus, each layer in the LGN contains a representation of the contralateral visual hemield. The layers in the LGN form *topographic maps*, which are stacked on top of each other (Hubel and Wiesel, 1977). The center of the visual field, however, has a much larger representation in the LGN compared to the periphery, which derives from the fact that the density of ganglion cells in the retina is higher than at the periphery. The spatial profile of the receptive fields of a neuron in the LGN also has a circular center and surround region, and they can also be classified as ON- and OFF-cells similar to the retinal ganglion cells.
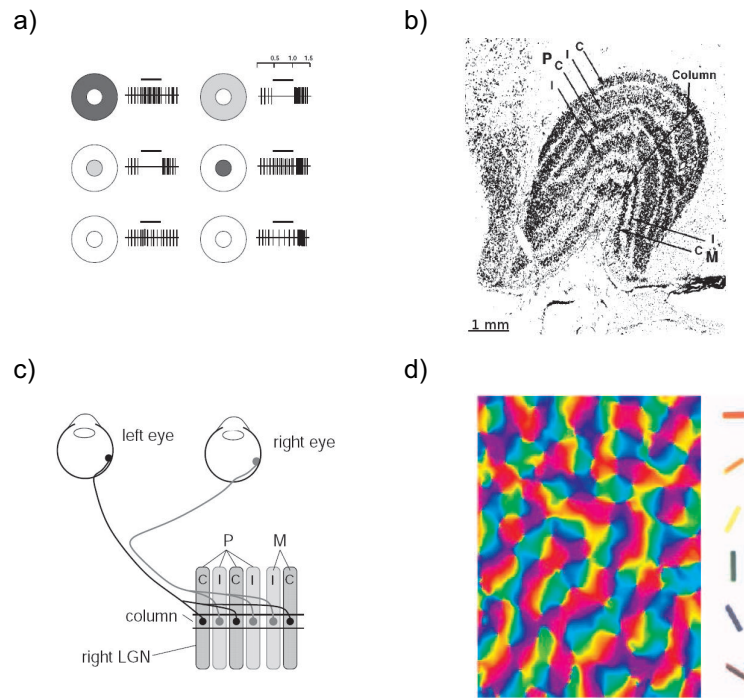
Figure 1.3: Steps in the retino-geniculocortical pathway. **a)** Illustration of ON-
and OFF-center receptive fields. The bars on top of the spike patterns indicate
the duration of stimulation (*white area*). **b)** Vertical slice of a macaque LGN
(adapted from Hubel and Wiesel (1977)) **c)** Projections of the ganglion cells
in two eyes to the layers in the right LGN. **d)** Orientation preference map of
macaque V1 with the color indicating the preferred orientation (see bars) of
neurons located at this position.

The primary visual cortex

Compared to neurons in the retina and the LGN, which have nearly circular re-
ceptive fields, cortical response characteristics show a qualitatively new feature:
Many cortical neurons respond selectively to contrast lines, bars, or gratings of
a certain orientation within the receptive field. The *orientation selective cells*
have classically been categorized as either *simple cells* or *complex cells*. The lat-
ter are selective for the orientation of a stimulus in their receptive field, but are
invariant to its relative position, whereas simple cells have a preferred phase. In
the macaque primary visual cortex, simple cells are found mainly in the input
layer 4C, whereas complex cells are found further away from layer 4C in the
layers 2, 3, 5 and 6. The prominent *hierarchical model of complex cells*, which
can explain their phase invariance, assumes a converge of the outputs of multi-
ple simple cells onto a complex cell. Although orientation selectivity has been
described the first time almost four decades ago (Hubel and Wiesel, 1962), the
way it is computed by the neuronal circuits in V1 is still not known completely.

In the context of our investigation of orientation selectivity in Chapter 2, we review models of orientation selectivity in much greater detail.

As the LGN, each V1 hemisphere represents the contralateral visual field with a much larger portion of the cortical surface being allocated to the representation of the visual field center. The projections from the LGN to V1 are in an orderly point-to-point manner. Nearby patches of V1 represent nearby patches of the visual field. In the macaque monkey, the area covered by the primary visual cortex is about $1300\,\mathrm{mm}^2$ (Hubel and Wiesel, 1977). In contrast to the massive reduction of 120 million photoreceptors to only 1 million ganglion cells in the retina, the projection from the LGN to V1 is far from being a bottleneck, because the input from 2 million LGN fibers is processes by 260 million cells in V1. This massive increase in the number of neurons processing the visual information can partly be understood by noting that certain features of the visual world are explicitly represented in V1 for each location in the visual field. For example, for each location in the visual field the full range of orientations is represented. Similar to the retinotopy of the cortical representations, where nearby patches represent nearby locations in the visual field, neurons in nearby patches have similar preferred orientations. Fig. 1.3d shows the layout of the *orientation preference map* of macaque V1 as obtained by optical imaging (Obermayer and Blasdel, 1993).

Neurons in V1, which are close to each other, communicate via *short-range lateral connections*. In addition, however, excitatory pyramidal neurons in the superficial layers 2/3 of area V1 send *long-range lateral connections* tangentially to the cortical surface within the layer to other excitatory and inhibitory neurons at distant locations. These connections are reciprocal and seem to contact predominantly neurons with a similar preferred orientation. The same kind of orientation-preference of synaptic connections seems to hold also for inter-areal *feedback connections* between area V1 and extra-striate areas. Both the intra- and inter-areal connections are only *modulatory* in the sense of only modulating, but not driving the postsynaptic neuron. A much more detailed recent review of the anatomy of V1 is given by Lund et al. (2003).

1.3   Addressed questions and outline

Schwabe and Obermayer (2003) reviewed several computational approaches, which could be used to explain adaptivity in the visual cortex. We concluded, however, that models of the kind used in these computational approaches are often too simple and far from incorporating relevant biological constraints. On the other hand, more realistic models are hard to handle mathematically. As a consequence, their (computational) properties need to be explored by means of numerical simulations.

In this thesis we mainly use numerical simulations to study visual cortical functions. We consider the following questions:

1. By means of which *mechanisms* do neuronal circuits compute experimentally characterized *feature selectivity* such as, for example, orientation tuning in the primary visual cortex?

2. Which *mechanisms* account for the *modulation of feature selectivity* such as the so-called contextual effects found in primary visual cortex and/or the changes of neuronal response properties due to attentional top-down modulation?

3. Can we explain at least some adaptivity phenomena in the visual cortex by referring to *specific computational principles* while still using biologically realistic models?

4. Which computational principle is appropriate to explain *task-dependent adaptivity* in the visual cortex?

The whole thesis is structured along these four questions, where a single chapter is devoted to each of them:

The first question is considered in *Chapter 2*, which is based on (Marino et al., 2005), where we consider the mechanisms of orientation tuning in area V1. Orientation tuning is the prime example of feature selectivity in the visual cortex. It has first been described almost four decades ago by Hubel and Wiesel (1962), but the underlying neuronal machinery is still not completely understood. If one assumes that cortical circuits are largely generic in terms of the computations they perform, general properties of the circuits in area V1 may also transfer to other areas and/or modalities. Pursuing a combination of experimental and modeling work in Chapter 2, we constrain the operating regime of area V1 to a recurrency regime with a co-variation of excitation and inhibition. It is conceivable that other sensory areas may work according to the same principle. This model does not directly focus on "adult adaptivity", but it is an important step for investigating adaptivity in the early visual cortex, because experiments suggest that orientation adaptation is stronger in V1 pinwheel regions compared to orientation domains (Dragoi et al., 2001). Our two-dimensional model explicitly considers these different regions, and hence could serve as an experimentally tested model for subsequent investigations.

In *Chapter 3*, which is based on (Schwabe et al., 2005), we re-investigate the neuronal circuits underlying the so-called contextual effects in area V1. These contextual effects are examples for an adaptivity of neuronal representations to the spatial context. Previous models assumed that these effects are realized by

computations performed solely within area V1. Recent experiments, however, have shown that other areas than V1 may be crucial for these effects. Again, pursuing a combination of experimental and modeling work, we set up a network model constrained by anatomical and physiological data. Beside accounting for a vast amount of physiological data, the model makes specific predictions for new experiments, which have recently been confirmed experimentally (Ichida et al., 2005). Since the model explains contextual effects via an inter-areal network, it can also serve as the basis for future studies concerning the computational role of inter-areal projections, which may mediate attentional top-down control of sensory processing.

In *Chapter 4*, which is based on (Schwabe and Obermayer, 2005a), we consider two specific phenomena of adaptivity in the adult visual system: the changes of tuning functions due to visual attention and perceptual learning. We seek to explain them both using the computational principle of optimally encoding sensory stimuli by a population code. Using this principle, we determine the optimal way of changing specific mechanisms in order to improve the quality of the representation of a set of stimuli. We find that the deduced tuning function changes depend on the mechanism, which highlights the importance of considering biological constraints when interpreting physiological data in the light of a functional principle. Moreover, the developed approach of deducing tuning function changes is of use to infer the mechanisms underlying other adaptation phenomena, which may be hard (or even technically not possible) to address directly. The results of this chapter are largely new explanations of already available physiological data from attention and perceptual learning experiments. However, yet to be tested predictions are also made for a new experiment in the context of visual attention to specific stimulus values.

In *Chapter 5*, which is based on (Schwabe and Obermayer, 2005b), we focus on the computational principle underlying task-dependent top-down modulations in the visual system. Here we ask the question of which principle is best suited to explain these modulations. We set up a simple model of a visual area and learn task-dependent top-down modulations in a reinforcement-learning paradigm. We propose that task-dependent top-down modulations are best understood in terms of a "tuning of a sensory pre-processor" such that the computed sensory representations remain useful for the very same read-out neurons, despite changing task-demands. We compare our view with other candidate principles and sketch an experiment, which could distinguish between our interpretation and the other candidate principles.

Each chapter begins with an abstract and a short introduction. Summaries are given at the end of each chapter. Details of the developed models and methods can be found in the Appendix.

Chapter 2

Orientation tuning in V1

Abstract. In this chapter we combine two complementary models, a single cell model and a recurrent network model, with recent intra-cellular measurements of orientation-selective neurons in cat primary visual cortex (V1) in order to determine the role the local recurrency plays in orientation tuning. It has been found experimentally that the super-threshold spike response is sharply tuned at all locations in the orientation map, but the sub-threshold membrane potential and conductance tuning depend on the map location. We demonstrate that the visually evoked excitatory synaptic conductances need to be balanced by inhibition at all map locations in order to keep the spike tuning invariant. Moreover, we show that only in such a recurrent regime the assumption of an anatomically observed isotropic local connectivity to both excitatory and inhibitory cells leads to model predictions consistent with the measured sub- and super-threshold responses. Our results suggest that at least one functional role of the local recurrency is to keep cortical computations, which are initially determined by the pattern of the afferent connections, invariant across an area. This can be realized via the mechanism of co-varying excitation and inhibition, which for orientation tuning in V1 can be induced by a spatially isotropic local connectivity.

2.1 Introduction

Neurons in the sensory cortices of the brain compute representations of the environment relevant for perception and action. Although these computations are often modulated by the spatial, temporal, and behavioral context, they are dominated by the integration of signals received via the afferent and the local recurrent connections. Since its discovery by Hubel and Wiesel (Hubel and Wiesel, 1962), orientation tuning in V1 has served as a paradigmatic example of a cortical computation, because the thalamic cells providing the afferent input to the primary visual cortex lack orientation selectivity whereas cortical cells are orientation-selective. In the last four decades, however, there has been a vivid and highly polarized discussion about the underlying cortical mechanisms. The main question is whether the recurrent connections sharpen a weakly tuned afferent input, or the afferent input is already sharply tuned, which implies that the recurrent circuitry has a different function.
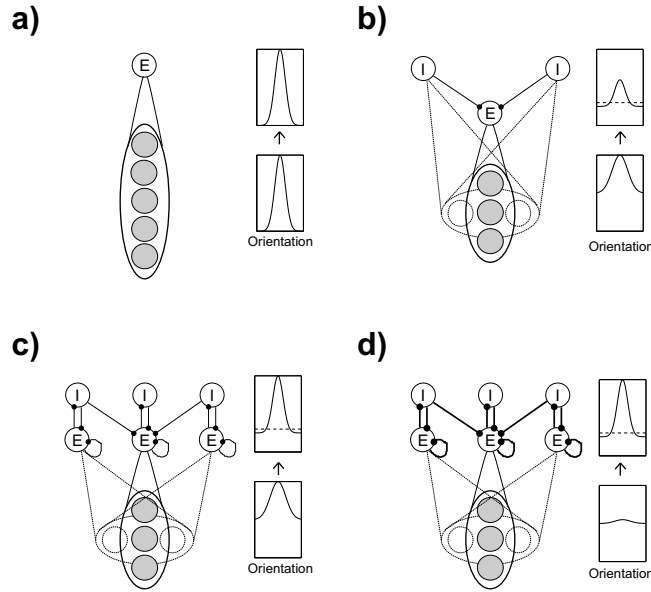
Figure 2.1: Models of orientation tuning. **a)** The *feedforward models* explain orientation tuning as resulting from the elongated spatial pattern of aligned ON- and OFF-inputs from the thalamus. **b)** The *inhibition models* assume only a weakly tuned afferent input and state that lateral inhibition suppresses non-preferred responses. **c,d)** The *recurrent models* models assume that an already orientation-tuned afferent input is amplified via intra-cortical excitation and inhibition. The operating regime of recurrent models with very strong recurrent synapses, compared to the afferent drive, is termed the "marginal phase", d), where the afferent input selects predefined activation profiles determined by the pattern of the recurrent connections. The *boxes* illustrate how the afferent input (bottom box) is transformed into an output firing rate (top box; the dashed line corresponds to the firing threshold).

Fig. 2.1 illustrates four types of models, which have been proposed to explain orientation tuning. The feedforward model (Fig. 2.1a), originally proposed by Hubel and Wiesel, states that orientation selectivity arises due to the elongated spatial pattern of appropriately aligned ON- and OFF-inputs from the thalamus. The strongest evidence for this assumption comes from experiments by Alonso et al. (1995), where monosynaptic thalamocortical connections in the cat have been characterized in terms of the receptive field properties of the pre- and postsynaptic neurons. The feedforward model, however, does not account for the invariance of the tuning with stimulus contrast (Sclar and Freeman, 1982), because it incorrectly predicts broadening of orientation tuning with increasing stimulus contrast. Moreover, it can be questioned as to whether the bias set up by the elongated projection pattern is sufficient to account for the sharp orientation selectivity of the spike response, because the aspect ratio of the afferent projection field might be too low (Chapman et al., 1991; Pei et al., 1994; Reid and Alonso, 1995). It also fails to predict the loss of orientation tuning after the localized reduction of inhibition with the $GABA_A$ antagonist

bicuculline (Nelson et al., 1994), because intra-cortical connections do not play a role in this purely "afferent model".

On the other hand, inhibition models (Fig. 2.1b) state that a strong, but weakly tuned afferent input is sharpened by intra-cortical inhibition suppressing non-preferred responses. Models of this type (Woergoetter and Koch, 1991; Sabatini, 1996; McLaughlin et al., 2000) predict contrast-invariant orientation tuning and are consistent with the observed loss of orientation selectivity after the bicuculline-induced reduction of inhibition. However, the strongest inhibitory postsynaptic potentials are observed during stimulation with the preferred orientation, and they are much weaker during stimulation with cross-oriented stimuli. Inhibitory models are also inconsistent with the observation that intracellular blockade of inhibition had almost no effect on the sharpness of the orientation-tuned spike response, which itself seems to be at odd with the finding that bicuculline-induced loss of inhibition reduces orientation selectivity.

The recurrent models (Fig. 2.1c) resolve this apparent paradox. For example, Somers at al. (1995) explain sharp orientation-tuned spike responses as the result of amplifying and sharpening a weakly orientation-tuned afferent thalamocortical input via strong iso-orientation intra-cortical excitation paired with strong iso-oriented intra-cortical inhibition. If the recurrent connections are very strong, then the activation profile is mainly determined by the pattern of the recurrent connections. In this so-called "marginal phase" (Ben-Yishai et al., 1995) the afferent input selects predefined attractor states of the network such that even a very weakly tuned afferent input can lead to a sharply tuned spike response (Fig. 2.1d).

Out of the four types of models, the recurrent models are most compatible with the currently available anatomical and physiological data. Their concrete instantiations, however, can be differentiated in terms of the assumed i) orientation-specificity of the afferent input, ii) orientation-specificity of the recurrent connections, and iii) strengths of the recurrent connections compared to the afferent inputs. Determining these characteristics experimentally is technically highly demanding, and so far no consensus has been reached. For example, we have previously argued (Adorjan et al., 2000; Adorjan et al., 2002; Schwabe, 2002) that albeit recurrent interactions can sharpen a weakly orientation-tuned input, very strong recurrent interactions might not be desirable from a functional perspective, because in such a highly 'competitive' operating regime only the most salient feature of a complex stimulus would be represented. Instead, we proposed that recurrent interactions are strong in the initial phase of a stimulus presentation, during which the most salient stimulus feature dominates the response, but become weaker in a second phase, such that the details can be represented as well.

Here we argue that recent intra-cellular measurements of the stimulus-driven response in cat V1 (Marino et al., 2005) can be used to further constrain its operating regime. It was found experimentally, that the super-threshold spike response is invariant across the cortical surface in terms of the sharpness of its tuning, but the sub-threshold membrane potential (Schummers et al., 2002) and conductance tuning (Marino et al., 2005) depend on the location in the orientation map. The recurrent circuitry, however, can vary widely within an area based on its functional architecture. In V1, the local neighborhood of neurons depends on the location in the orientation map. An experimentally measured signature of this dependence is that neurons close to pinwheel centers have more

broadly tuned sub-threshold responses compared to neurons in orientation domains.

Orientation tuning in V1 has recently been modeled mainly using simplified one-dimensional hypercolumn models. Here, however, we use two complementary models: a single cell model and a two-dimensional network model. First, using the *single cell model*, we determine the conductance tuning, which is *necessary* to produce the observed invariant spike tuning. It will turn out that a co-variation of excitation and inhibition with map location is necessary to produce location-independent spike tuning. Second, using the *network model*, we show that unspecific isotropic recurrent connections, as observed experimentally, are indeed *sufficient* to produce such a co-variation.

In addition, however, we also explicitly use the experimentally obtained data in order to further constrain our models. We use the orientation selectivity index (OSI, see Appendix B for a definition) of a set of super- and sub-threshold responses as a quantification of the responses' tuning. Then, the predictions of both the network and the single cell model for these OSIs can be compared with the experimentally measured values. In particular, models with very weak recurrency will not predict any dependence of the conductance responses on the map location, whereas models assuming a very strong recurrency will predict a strong dependence. For the single cell model, we determine those strengths of afferent drive, which are consistent with the data. For the network model we do the same for the strength of the local recurrent connections. Thus, assuming a location-independent afferent input and an isotropic local recurrency we let the experimentally measured location-dependencies constrain the possible strengths of the recurrency.

We now summarize the relevant experimental methods and findings.

2.2   Experimental methods and findings[1]

Methods

Fig. 2.2a show the experimental setup. Experiments were performed on 23 adult cats, which were anesthetized and paralyzed. First, optical imaging was used to obtain an orientation preference map of V1. Then, whole-cell recordings during visual stimulation were made and currents were injected in order to measure the evoked excitatory and inhibitory conductances during visual stimulation.

Optical imaging: Intrinsic signals were recorded in response to full field drifting square wave gratings of eight orientations (Fig. 2.2b) using a CCD camera. Single condition maps were computed by normalizing the response to each orientation (average of 24-48 trials) by the mean response to all orientations ("cocktail blank"). The orientation preference (angle) maps were computed as the pixelwise vector average of the single condition maps. High accuracy and reproducibility of the map was ensured as follows: First, reference images of the surface vasculature were obtained after every five trials, and in the event of any shift in the position of the blood vessels, the recording was aborted. Second, angle maps were computed from independent blocks of five to seven trials and

---

[1]This section is based on Marino et al. (2005) and contains only a brief summary of the methods and findings as the minimal background for this chapter. For details we refer to Marino et al. (2005).
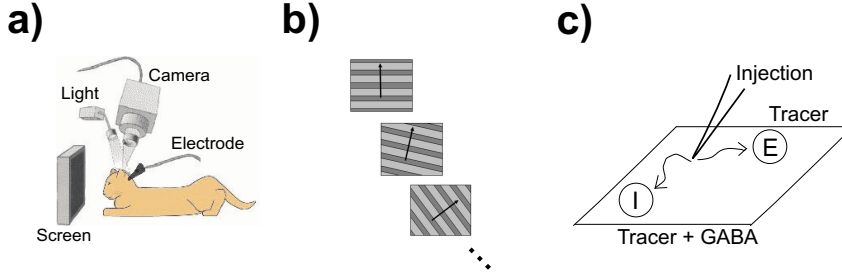
Figure 2.2: Experimental methods. **a)** Illustration of the experimental setup with intrinsic signal optical imaging and in vivo whole-cell recordings during visual stimulation. **b)** Visual stimuli were *full field drifting square wave gratings* of eight orientations. **c)** By injecting a *retrograde tracer* (from the axonal terminal back to the soma) the neurons projecting to the injection site were labeled. Those labeled neurons, which also stained positive for a *GABA antibody*, were identified as inhibitory neurons.

compared. Third, angle maps were computed from the two independent sets of four orientations contained within the stimulus set.

Furthermore, the specific sites to be targeted for patch recordings or tracer injection were chosen to be the most stable sites, assessed as above. In most cases, we were able to find pinwheel centers that shifted by no more than 1-2 pixels (13-26 μm) between maps taken from early and late trials, or from the two sets of independent stimulus orientations. An analysis of the reliability of pinwheel center localization, and the effect of variability in electrode localization on the relationship between the tuning of conductance and the local orientation distribution, is described in the Supplementary Notes of (Marino et al., 2005).

Electrophysiology: The specific sites to be targeted for the patch recordings were chosen to be the most stable sites. The glass microelectrodes (resistance 6-12 MΩ) were carefully lowered into the cortex and directed to either orientation domains or pinwheel centers, using an image of the surface vasculature aligned to the orientation map as reference. The visual stimuli were randomly generated drifting sine-wave gratings of eight orientations, moving in opposite directions, plus a blank stimulus, each of which was presented 5 times for 1 s. This protocol was repeated while injecting 3-4 different steady currents (I_{inj}), ranging from -0.2 to 0.1 nA. The entire protocol was completed in 40 cells, from which 18 were chosen for further analysis based on their stable biophysical properties, which were monitored throughout each recording by means of the I-V curves. The average series resistance was 70.5 $\pm$ 39.7 MΩ (mean $\pm$ s.d.), input resistance 27.8 $\pm$ 20.1 MΩ, time constant 17.4 $\pm$ 8 ms, and resting potential 50.4 $\pm$ 15.1 mV. Total conductance $g(t)$ at time t was estimated by regression as the inverse of the slope of a line fitted to the relation between the injected current I_{inj} and the membrane potential Vm. To calculate the inhibitory $g_i(t)$ and excitatory $g_e(t)$ conductances we assumed

$$g(t) = g_e(t) + g_i(t)$$
$$V_{rest}(t) = \frac{g_e(t)E_e + g_i(t)E_i}{g_e(t) + g_i(t)},$$

where $V_{rest}(t)$ is the membrane potential in absence of current injection, and E_e and E_i are the equilibrium potentials for $g_e(t)$ and $g_i(t)$ respectively. Then, $g_i(t)$ and $g_e(t)$ can be derived as follows:

$$g_i(t) = \frac{g(t)[V_{rest}(t) - E_e]}{E_i - E_e}$$

$$g_e(t) = \frac{g(t)[V_{rest}(t) - E_i]}{E_e - E_i}$$

We used $E_e = 0\,\text{mV}$ and $E_i = -80\,\text{mV}$.

Anatomy: In V1 of 9 cats, glass pipettes (tip diameter 10-20 μm) were used to place three distinct injections of a retrograde (from the axonal terminal back to the soma) tracer in pinwheels and domains at a depth of approximately 600 μm. Following 36-48 hrs, the cats were killed with an overdose of sodium pentobarbital. V1 was removed and sectioned (40 μm) tangential to the surface. The tissue was processed to reveal GABA positive neurons. Images of the pattern of labeled neurons near the depth of the injection and laterally were acquired. Inhibitory neurons were identified by positive staining for the GABA antibody and from retrograde tracer injection. The pattern of labeled cells was aligned to the orientation map using landmarks from three injection sites, visible penetrations from the patch pipettes, and blood vessel patterns along the cortical surface and through a depth of 600 μm. To ensure the most detailed anatomical analysis of local connectivity from our sample of 27 injections, we limited analysis to 3 pinwheel and 4 domain injections, selected because their spread of tracer uptake was confined to a diameter less than 100 μm. For these injections, the numbers of excitatory and inhibitory neurons were counted every 50 μm from the boundary of the injection site to a radius of 450 μm. This analysis indicated a sharp drop in the number of local excitatory inputs at about 250 μm from the injection, representing the limit of dense, short-range connectivity.

Findings

Fig. 2.3a-c show the average tuning curves for the membrane potential and the excitatory and inhibitory conductances for pinwheel and orientation domain neurons. For each of the three quantities, the tuning, as quantified using the OSI (see Appendix B), was less pronounced in pinwheel neurons compared to orientation domain neurons. The mean absolute change in inhibitory conductance was always larger for inhibition than for excitation, independent of orientation or map location. The map location was quantified as the OSI of the histogram of preferred orientations in a circular area with radius 250 μm. For all cells, the OSI values for g_i and g_e co-varied, indicating that regardless of location, inhibition always seemed to balance excitation. A three-way ANOVA was performed to compare OSIs for the total conductance g, g_i and g_e between map locations, recording depth and cell type (simple/complex), but only significant effect of orientation map location was found. Therefore, in the population analysis, it was not differentiated between cell type or cortical depth.

Fig. 2.3d-f show this population analysis. Fig. 2.3d shows that the tuning of the spike responses (thin line) is independent of the map location, whereas the tuning of the membrane potential (thick line) depends of the map location. The tuning of g_i and g_e is also broader (low values of the OSI) close to pinwheels
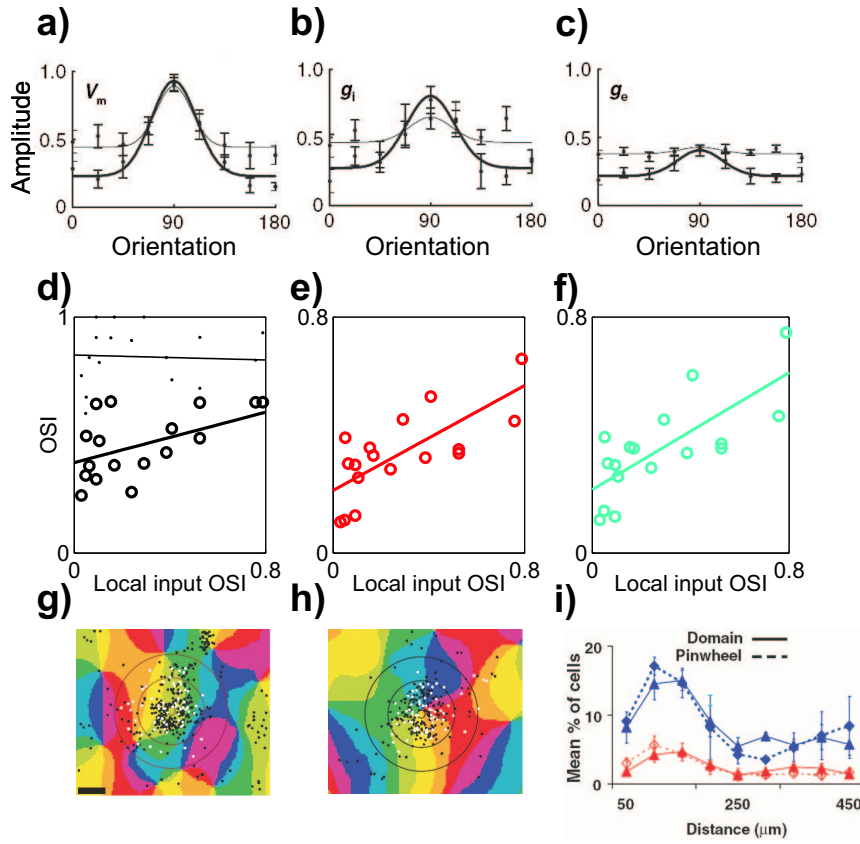
Figure 2.3: Experimental findings. **a-c)** Normalized tuning curves of the membrane potential, a), and the excitatory and inhibitory conductances, b) and c), for orientation domains (thick line) and pinwheels (thin line). **d-f)** Tuning of the membrane potential (thick line) and spike responses (thin line), d), as quantified by the OSI (see Appendix B) as a function of the map location as well as the tuning for the inhibitory and excitatory, e) and f), conductances. **g-i)** Examples for identified excitatory (*black dots*) and inhibitory (*white dots*) neurons projecting to the injection site in an orientation domain, g), and a pinwheel, h), as well as the the number of cells projecting to the injection site as a function of the distance.

(low values of the local input OSI) compared to orientation domains (high values of the local input OSI) as shown in Fig. 2.3e,f. The loss of tuning of the membrane potential, g_i and g_e close to pinwheels could potentially be explained by an isotropic pattern of local connections, in which neurons located at or near pinwheel centers would receive inputs from neurons with different orientation preferences. Orientation domain cells would be primarily driven by cells sharing the same orientation leading to a sharper tuning. The smooth transition of the tuning from broad at pinwheel centers to sharp in orientation domains suggests the existence of a single mechanism that is able to balance the different patterns of excitation and inhibition at different map locations, keeping the spike response equally selective at any site. A spatially isotropic pattern of local excitatory and inhibitory connections could provide a substrate for such a mechanisms.

To explore this possibility, the anatomical distribution of local excitatory and inhibitory inputs to neurons at different locations in the orientation map was determined. Extremely small injections (uptake zone less than 100 μm in diameter) of retrograde tracers were used to study the structure of local inputs to different sites in the orientation map, combined with labeling for GABAergic cells (see Methods above). This technique provides a fine-grained comparison of the projection patterns to orientation domains and pinwheel centers. Despite drastic differences in local orientation distributions at pinwheels and domains, no differences in the spatial distribution of either local inhibitory or excitatory cells labeled at these sites were found.

An example of the pattern of retrogradely labeled cells from a domain and a pinwheel injection is depicted in Fig. 2.3g,h. Independent of the location, the pattern of labeled cells around the injection site was always roughly circular, ignoring the distribution of orientation preferences. This distribution of cells, as shown for the two individual cases (Fig. 2.3g,h), and the population (Fig. 2.3i), indicated a local isotropic radius of influence of about 250 μm. Thus, the anatomical data provide a potential substrate for the electrophysiological measurements, suggesting that broader excitatory and inhibitory synaptic conductances at pinwheel centers arise naturally from spatially isotropic local projections. This finding of an apparently isotropic local connectivity will now serve as an assumption in the models used to explain the physiological data.

2.3   Single cell model of invariant tuning

We used the one compartment, Hodgkin-Huxley type neuron model with Na-, K- and M-currents and balanced background noise inputs given in the Appendix A. Our approach was as follows: First, the probabilities of intracortical synaptic connections with presynaptic excitatory neurons were estimated based on optically imaged orientation maps. Second, the total excitatory input conductance, given as a function of stimulus orientation, was computed for local neighborhoods of varying OSI under the assumption that the spike tuning of all excitatory neurons is the same and independent of location in the orientation map. Third, the tuning curve of the total inhibitory conductance that was necessary to obtain the observed sharp tuning curve of the firing rates was calculated.

Connection probabilities in orientation space

Using optically imaged orientation maps of cat V1 (see also the electronic material to this thesis), we first estimated the probability $P\left(\Delta\theta; x\right)$ of a neuron at location x with preferred orientation θ making a synaptic connection to a neuron with preferred orientation θ', with $\Delta\theta$ being the distance between θ and θ' on the half-circle. We assumed that the connection probability is rotationally symmetric in cortical space and depends on the distance via an alpha-function (see text Fig. 4d). Depending on the preferred orientation of the neuron at x and the orientation distribution of the local network neighborhood, this induces the orientation distribution $P\left(\Delta\theta; x\right)$. We then computed the local input OSI as the OSI of the orientation histogram compiled from all pixels with distances not larger than $250\,\mu$m and averaged $P\left(\Delta\theta; x\right)$ over all locations x with the same local input OSI (bin size 0.1) to obtain the connection probability $P\left(\Delta\theta; \text{OSI}\right)$.

Deducing the inhibition

The presynaptic activity (independent Poisson spike trains) was separated into background, feedforward, and recurrent components, which describe the ongoing activity not dependent on the stimulus (balanced noise), the afferent stimulus-driven input, and the inputs due to the activation of the local network neighborhood. The strength of the afferent drive relative to the strength of the local recurrency was parameterized with a parameter $0 \leq p \leq 1$. The model predictions were used to let the experimental data constrain the possible values of this parameter.

Since we used non-saturating "exponential synapses" (see Appendix A) the conductances induced scale linearly with the presynaptic firing rate. We adjusted the maximal conductances of the individual synapses such that mean excitatory and inhibitory synaptic conductances of size g_L and $2g_L$ are induced by presynaptic Poisson spike trains of $7000\,\text{sp/s}$ and $3000\,\text{sp/s}$. Here, g_L is the leak conductance of the cell (see Appendix A). Then, we calculated the membrane potential of the model cell by means of numerical integration of the membrane potential equation using the simulation software NEURON (fixed step size of $\Delta t = 0.25\,\text{ms}$). For every combination of input firing rates we calculated membrane potential traces of $10\,\text{s}$ simulated time. The spike response was then characterized by the time-averaged firing rate; the sub-threshold response was characterized by the mean membrane potential with the action potentials being removed, i.e., removing the membrane potential from $2\,\text{ms}$ before to $3\,\text{ms}$ after each crossing of the threshold ($-54\,\text{mV}$).

The tuning curve of the mean excitatory conductance for a given position in the map (as characterized by its local input OSI) was then calculated using the equations

$$\left\langle g^E\left(\theta; \theta^{pr}\right)\right\rangle = W_{syn}^E \cdot \left[f_{bg}^E + p f_{Aff}\left(\theta; \theta^{pr}\right) + \left(1 - p\right) f_{Rec}\left(\theta; \theta^{pr}\right)\right]$$

$$f_{Rec}\left(\theta; \theta^{pr}\right) = \int_{-90}^{+90} P\left(\Delta\theta = \theta' - \theta^{pr}; \text{OSI}\right) \cdot f_{rec}^E\left(\theta; \theta^{pr} = \theta'\right) d\theta'$$

Here, $\theta^{pr} = 0°$ is the preferred orientation of the model neuron, θ is the stimulus orientation, $P\left(\Delta\theta = \theta' - \theta^{pr}; \text{OSI}\right)$ is the estimated connection probability, and determines the excitatory conductance in the absence of visual stimulation. W_{syn}^E is a "weight", which relates the presynaptic firing rates to a postsynaptic

conductance (see Appendix B.2). We chose $f_{bg}^E = 6000\,\mathrm{sp/s}$. The feedforward input was described by

$$f_{Aff}\left(\theta;\,\theta^{pr}\right) = 3.5 \cdot f_{bg}^E \cdot \left[0.1 + 0.9\exp\left(-\frac{\left(\theta - \theta^{pr}\right)^2}{2\sigma_{Aff}^2}\right)\right],$$

with $\sigma_{Aff} = 25°$. For the recurrent excitatory input induced by a stimulus of orientation θ, we assumed

$$f_{rec}^E\left(\theta;\,\theta^{pr}\right) = 3.5 \cdot f_{bg}^E \cdot \exp\left(-\frac{\left(\theta - \theta^{pr}\right)^2}{2\sigma_{rec}^2}\right),$$

with $\sigma_{rec} = 15°$. The parameter p was set to 0.35, i. e. 65% of the stimulus induced excitation were due to local recurrence. For a pinwheel neuron this leads to an excitation induced by the stimulus of approximately $1.5g_L$ for the preferred orientation. For orientation domain neurons this leads to an excitation of approximately $2.5g_L$, also for the preferred orientation.

Given the mean excitatory conductance and our characterization of the model neuron's response, we determined for each stimulus orientation θ the *smallest* synaptic inhibitory conductance $\left\langle g^I\left(\theta;\,\theta^{pr} = 0\right)\right\rangle$ necessary to obtain an orientation tuned spike response given by

$$f\left(\theta;\,\theta^{pr}\right) = 20\,\mathrm{sp/s} \cdot \exp\left(-\frac{\left(\theta - \theta^{pr}\right)^2}{2\sigma_{rec}^2}\right).$$

In other words, we *enforce self-consistency* by requiring that the shape of the tuning curve for the spike output matches the firing rate tuning curve f_{rec}^E assumed for computing the excitatory conductance. For the fit we used the mean-squared error between the desired output firing rate and the firing rate given by our characterization of the model neuron. Note that the much higher absolute value of the f_{rec}^E is due to integration over multiple presynaptic neurons in the local neighborhood, which may also fire with individual peak responses of around $20\,\mathrm{sp/s}$.

We then calculated the total inhibitory conductance by taking the sum of the synaptic inhibitory conductance and the time-averaged conductance $g_M = \bar{g_M}n$ for the adaptation current for the corresponding output firing rate. We did this, because the latter also hyperpolarizes the neuron and - due to its reversal potential of $E_K = -90\,\mathrm{mV}$ - is likely to contribute to the experimentally measured inhibitory conductances. The resulting total mean excitatory and inhibitory conductances were normalized as follows:

$$g_e\left(\theta\right) = \frac{g_{Aff}^E\left(\theta\right) + g_{Rec}^E\left(\theta\right)}{\max\left(g_{Aff}^E\left(\theta\right) + g_{Rec}^E\left(\theta\right),\, g_M^I\left(\theta\right) + g_{Rec}^I\left(\theta\right)\right)}$$

$$g_i\left(\theta\right) = \frac{g_M^I\left(\theta\right) + g_{Rec}^I\left(\theta\right)}{\max\left(g_{Aff}^E\left(\theta\right) + g_{Rec}^E\left(\theta\right),\, g_M^I\left(\theta\right) + g_{Rec}^I\left(\theta\right)\right)}$$

Here, the $g_{Aff}^E\left(\theta\right)$, $g_{Rec}^E\left(\theta\right)$, $g_{Rec}^I\left(\theta\right)$, and $g_M^I\left(\theta\right)$ are the time-averaged conductances for the afferent and recurrent excitation, the recurrent inhibition and the non-inactivating K-current during stimulation with orientation θ.
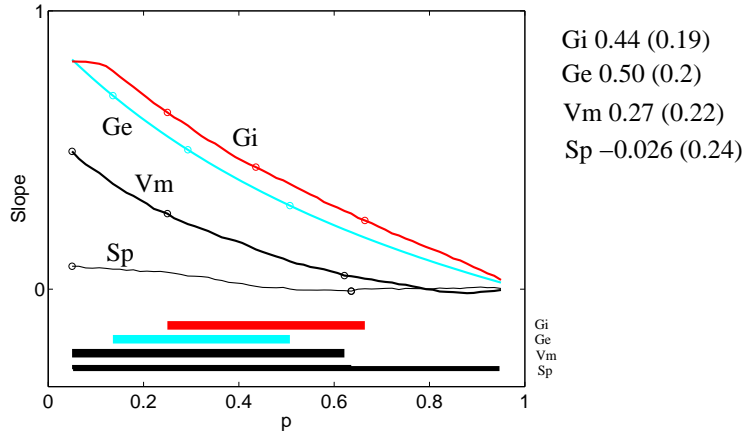
Figure 2.4: Predicted slopes for the dependence of the tuning of g_e, g_i, the membrane potential and the spike responses on the map location for different values of the free model parameter p. Shown are also the slopes and 95% confidence intervals for the linear fit of the experimental data, as well as the values of p (see bars), which lead to predicted slopes inside these confidence interval (numbers in brackets).

Model predictions and comparison with experimental data

By convolving the experimentally obtained spatial excitatory input profiles with experimentally obtained orientation maps, we calculated the tuning of the excitatory conductance g_e at locations ranging from pinwheels to domains. Then, given these g_e curves, we determined the g_i tuning curves which yield sharp spike tuning for each location. The obtained tuning curves for both g_e, g_i and the membrane potential were always broader for locations close to a pinwheel (for low values of the local input OSI) compared to orientation domains (high values of the local input OSI). In the model, this dependence of these tunings on the location was always linear, but the slope of this linear dependence was strongly affected by the value of the parameter p, which determined the strength of the afferent excitatory drive relative to the local recurrency. Therefore, we computed the slope for these dependences for different values of p and compared the predicted slopes with the slope of the linear regression of the experimental data.

Fig. 2.4 shows the predicted slopes for g_e, g_i, the membrane potential and the spike responses for different values of p. For all quantities, the slope decreases with increasing values of p. This reflects the fact that the local network neighborhood does not affect the tuning if the neuron is driven predominantly via the afferent synapses. Fig. 2.4 also shows that a range of values for p (from $p = 0.25$ for the lowest value compatible with the g_i tuning to $p = 0.5$ for the highest value compatible with the tuning of the excitatory conductance) is consistent with the experimental data in the sense of predicting slopes inside the 95% confidence interval for the slope of the linear regression of the experimental data. In principle, every value $0.25 \leq p \leq 0.5$ would result in models, which are a proper description of the data in terms of the location-dependence of the
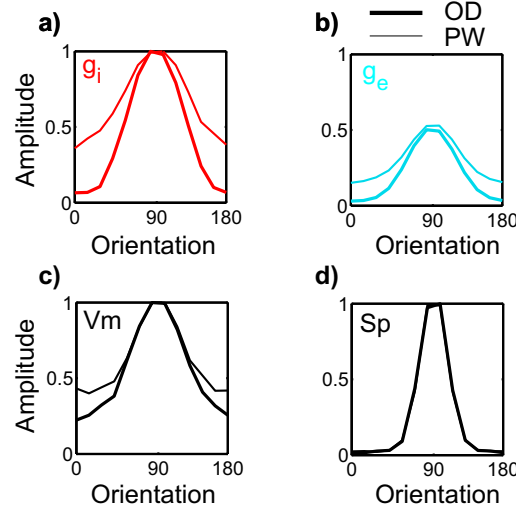
Figure 2.5: Model predictions for synaptic conductances, membrane potentials, and spike responses underlying orientation selectivity across the orientation map. **a,b)** Inhibitory and excitatory conductance tuning for a pinwheel cell and an orientation domain cell derived from the single cell model. The afferent input in orientation space is described by a Gaussian function, $\sigma = 25°$, added to an offset of 10% of the maximum value. **c)** Vm tuning at a pinwheel center and in an orientation domain with the input conductances from a,b). **d)** Spike response tuning at a pinwheel center (OSI<0.3) and in an orientation domain (OSI>0.7); the two are nearly identical but have different Vm tuning as shown in b), created by different g_i and g_e tuning as shown in a,b).

tuning. For the following simulations we used the value $p = 0.35$.

Fig. 2.5a,b shows the deduced tuning curves of g_e and g_i for neurons near pinwheel centers and in orientation domains. The shapes of these tuning curves are similar to the experimentally measured g_e and g_i tuning curves (Fig. 2.3b,c). As with the experimental data (Fig. 2.3a), the difference between pinwheel locations and orientation domains is reflected in the subthreshold membrane potential (Fig. 2.5c), but due to the appropriate inhibitory balance at orthogonal orientations, which keeps the membrane potential below threshold, not in the spike responses (Fig. 2.5d). These model results indicate that given the constraints imposed by the anatomical location within an orientation map, the inhibitory tuning that was measured resembles the tuning that is necessary, as predicted by the model, to balance excitation and yield sharp spike tuning at all locations. Fig. 2.6 shows the the predicted dependence for all map locations for the value $p = 0.35$.

The mechanism of co-varying the inhibition with the excitation is a general mechanism to account for the location-independent spike tuning independent of the absolute strength of the overall excitation. Fig. 2.7a shows two tuning curves for g_e and g_i for a pinwheel location corresponding to a low and a high value of the absolute strength of excitation. For very strong excitation (approx. $2g_L$ at the preferred orientation for a pinwheel location, OSI<0.3, solid line)
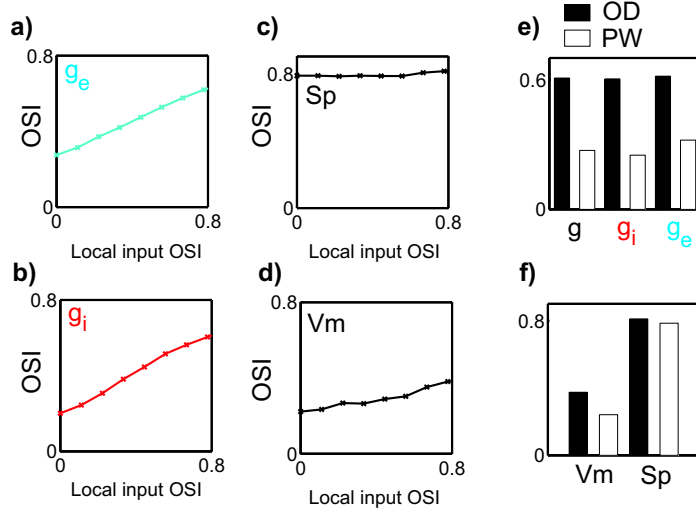
Figure 2.6: Results from the single cell model, demonstrating that inhibition balances excitation and produces sharp tuning across the orientation map. **a,b)** Predicted tuning of g_i and g_e as a function of the local input OSI. **c,d)** Tuning of the spike and Vm response as a function of the local input OSI. The Vm OSI increases as the local input OSI becomes larger, i. e. as cell locations move from a pinwheel center to an orientation domain. The OSI of the spike response, however, remains constant. **e,f)** Average OSIs for g, g_i, g_e, Vm, and the spike response for pinwheels (OSI<0.3) and orientation domains (OSI>0.7).

the inhibition needs to be increased (approx. $4g_L$ at the preferred orientation) in order to balance the excitation. However, the shape of the tuning functions remains the same. Fig. 2.7b shows the ratio of the conductance tunings for a pinwheel and an orientation domain location for different absolute strength of the excitation. The almost constant ratio indicates that above a certain strength of the recurrent excitation, a corresponding balancing by inhibition can ensure a location-independent spike tuning. However, in the orientation domain (thick line) the ratio of the g_i to the g_e tuning is approx. 1, whereas it is smaller in pinwheel locations (thin line). This means that close to pinwheels the inhibition is predicted to be slightly less tuned compared to the excitation in order to suppress non-preferred responses at orthogonal orientations. Such a difference in the deduced tuning, however, is still compatible with an isotropic local connectivity to inhibitory neurons being identical with the connectivity to excitatory neurons, as long as the inhibitory neurons have a broader orientation tuning.

We now address the issue of how the inhibitory tuning is generated within a recurrent network.

2.4 Network model of invariant tuning

In the last section we combined the pattern of the anatomical connections with the resulting strength of the excitatory synaptic drive in order to deduce the in-
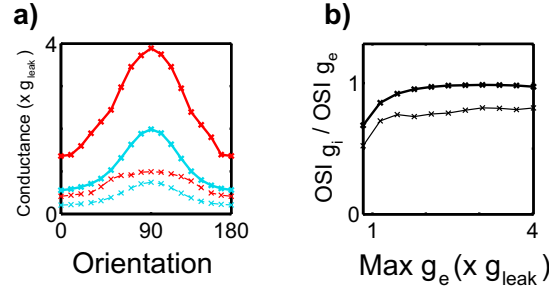
Figure 2.7: Co-varying inhibition and excitation for location-independent spike tuning for different absolute strengths of the excitation. **a)** Tuning of the excitatory conductance (g_e, blue curves) and the matched inhibition (g_i, red curves) for increasing levels of excitation (dashed and continuous lines) for a pinwheel neuron (OSI<0.3). The tuning curve for inhibition must be scaled in order to balance an increase in total excitation, but its shape remains similar. **b)** Ratio of the OSIs of g_i and g_e for an orientation domain (OSI>0.7, thick line) and a pinwheel location (OSI<0.3, thin line) as a function of the strength of excitation (peak conductance at the preferred orientation for an orientation domain neuron). Spike tuning remains constant for a broad range of excitatory strengths if it is properly balanced by inhibition. This is shown by the almost constant ratio of OSIs for the conductances.

hibition, which is necessary to yield the experimentally observed invariant spike tuning. Both the excitatory and the inhibitory conductance tuning curves were similar to the experimentally measures tuning curves. However, it is not clear a priori as to whether these tuning curves, i. e. the location-dependent conductance tuning and the location-independent spike tuning, would also emerge as the result of the recurrent interactions in a network model. Interestingly, McLaughlin et al. (2000) use a network model of a patch of V1 including pinwheels and orientation domains, but they report a dependence of the sharpness of the spike tuning on the map location. In particular, they predict that spike tuning is sharper closer to pinwheel locations. We also investigate this discrepancy between the experimental measurements of Marino et al. (2005) and the model predictions by McLaughlin et al. (2000). We set up a recurrent network model of a small patch of V1, simulate its responses to oriented visual stimuli, and characterize the resulting conductance and spike tuning curves as a function of model parameters determining the strength of the recurrent interactions.

Fig. 2.8 shows the architecture of the model (see Appendix B for a detailed description of the model). In short, the model represents a small patch of V1 with the preferred orientations of the neurons being arranged in an artificial orientation map, which contains pinwheels as well as orientation domains (Fig. 2.8a). In agreement with the anatomical findings (Fig. 2.3g-i), the spatial scale of the local recurrency was the same for excitatory and inhibitory connections (Fig. 2.8b). As in the experiments and the single cell model, for each pixel in the map, the preferred orientations of all pixels in circular region around this pixel were used to compile a histogram of preferred orientations. The OSI of this histogram was then used as a graded quantification of map location.

Small values correspond to locations close to a pinwheel, whereas large values correspond to orientation domains (Fig. 2.8c,d).
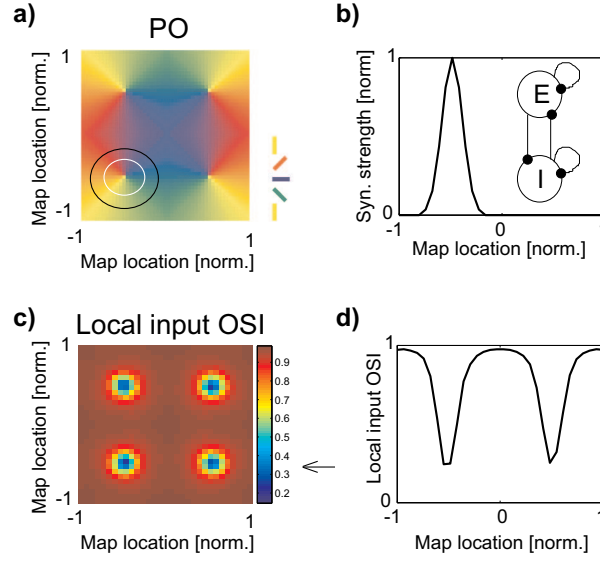
Figure 2.8: Architecture of the map model. **a)** Small patch of primary visual cortex with color-coded preferred orientations for each pair of excitatory and inhibitory neurons. The *black circle* (radius 0.4) denotes the region based on which the local input OSI was computed. The *white circle* indicates the spatial scale of the local recurrent connections ($\sigma_E = \sigma_I = 0.1$, see Appendix B for details). Those values were determined, in order to match the statistics of real orientation maps (O. Beck, personal communication). **b)** Spatial scale of all local synaptic connections. Shown is the spatial profile of the connections made by a neuron close to the pinwheel at $(-0.5, -0.5)$. **c, d)** Local input OSIs for the map shown in a) for all map locations, c), and for the neurons in the row, d), indicated by the *arrow*.

We then simulated the responses of the neurons in the network to oriented stimuli (see Appendix B) for different values of the maximal synaptic conductances G_{EE} and G_{EI}, which determine the strength of the recurrent excitation and inhibition, respectively. For every simulation, the tuning of the spike responses and the excitatory and inhibitory postsynaptic conductances for each map location was recorded. The OSIs of these tuning functions can be plotted against the local input OSI to reveal the dependence of the tuning on the map location (see also Figs. 2.3 and 2.6). The slope of the linear regression of this dependence is a single number, which here we used as one characterization of a network parameterization.

Fig. 2.9a shows these slopes for the tuning of the excitatory conductances. Model networks with strong recurrent excitation (large values of G_{EE}) predict steeper slopes than networks with less recurrent excitation. If, however, the recurrent excitation is too strong, then the network becomes unstable, but stronger recurrent inhibition (larger values of G_{EI}) can prevent this loss of stability. The model also makes predictions for the slopes of the inhibitory conductances (Fig. 2.9b) as well as for the spike tuning (Fig. 2.9c). Those can be compared with the experimental findings in order to constrain the operating

regime of V1. Similar to the comparison done for the single cell model, here we determine those parameterizations, which predict slopes falling in the 95% confidence interval of the experimentally obtained slopes. Fig. 2.9d shows those parameterizations. Interestingly, only a few parameterizations are consistent with the data in this sense.
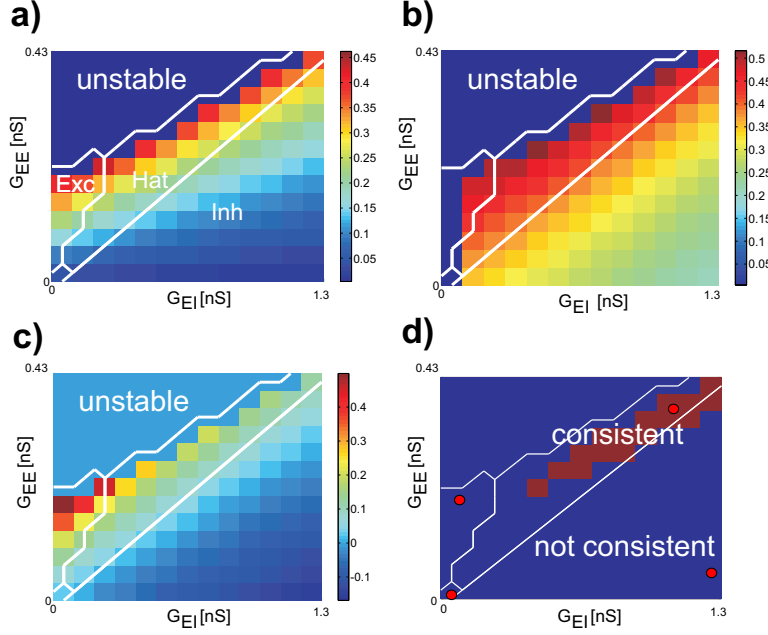
Figure 2.9: Exploration of the network model. **a)** Shown are the slopes of the linear regression for of the dependence of the excitatory conductance tuning on the map location. The regimes *Exc*, *Inh*, and *Hat* are defined w. r. t. the tuning of the currents received via the local recurrent connections (see Appendix B.4). **b,c)** Same as a), but for the inhibitory conductance and the spike tuning, respectively. **d)** For each parameterization it was determined, as to whether the slopes shown in a-c) are in the 95% confidence interval of the data. Only if this is the case for all three quantities, the corresponding parameterization was labeled as "*consistent*". The red dots correspond to the representative models shown in Fig. 2.10.

The different models may not only be distinguished in terms of their parameter values, but also in terms of more general quantitative properties. For example, Kang et al. (2003) distinguished between four different operating regimes, which were defined in terms of the effective recurrency. We also use a simple classification of our model parameterizations into a *Hat-*, *Exc-*, and *Inh-*regimes in order to structure the parameter space we explored. The regimes are defined and shortly discussed in Appendix B.4. In short, the *Hat-*regime corresponds to a regime, where both recurrent excitation and inhibition drive the neuron. In the *Exc-*regime, only the recurrent (and afferent) excitation drive the neurons while the recurrent inhibition is weak, and in the *Inh-*regime the recurrent excitation is weak, but the inhibition is rather strong. In the *Aff-*regime, the strength of all recurrent connections is zero.

Let us now consider in greater detail the responses of models representative for each regime. The *first row* in Fig. 2.10 corresponds to the model without any local recurrency. The excitatory conductance shown in the second column is due to the afferent input, which was the same for all map locations. As a consequence, the spike tuning is independent of the map location (fourth column). However, since no location-dependence of the conductances is predicted, this model fails to explain the experimental data.

The *second row* corresponds to the *Hat-*regime, where local recurrent excitation and inhibition are strong. For example, for this regime the excitatory postsynaptic conductance at the preferred orientation of an orientation domain neuron due to local recurrent excitation was twice as much as the excitation due to the afferent connections. However, due to the local inhibition, the effective current received via the recurrent connections is smaller than the current received via the afferent connections (Fig. 2.10, first column). In this regime, however, the spike tuning is almost independent of the map location, and the tuning of the excitatory and inhibitory conductances co-vary with map location (Fig. 2.10, last column), as observed in the experimental data.

The predictions of the *Inh-*regime are shown in the *third row*. The local recurrent excitation is weak, and the excitatory conductance is dominated by the afferent connections. Interestingly, in this regime a sharper spike tuning is predicted for locations close to a pinwheel (2.10, fourth and last column). Since with this parameterization the local recurrent excitation is weak at all map locations, local recurrent inhibition can sharpen a more broadly tuned afferent input. This, however, occurs only at locations close to a pinwheel, because only at those locations cross-orientation interaction takes place. Note that this regime comes closest to the network model proposed by McLaughlin et al. (2000) both in terms of the predicted sharper tuning close to pinwheels and the parameter settings they used (McLaughlin et al., 2000).

The predictions of the *Exc-*regime are shown in the *last row*. Recurrent excitation is strong and inhibition is weak. Still, however, the network operates in a stable regime (see corresponding red dot in Fig. 2.9d). This model predicts that the spike tuning depends on the map location with broader tuning close to pinwheel locations (2.10, fourth and last column). The model also predicts a co-variation of excitation and inhibition, but the recurrent inhibition is too weak to balance the recurrent excitation at orthogonal orientations, which leads to super-threshold responses close to pinwheel locations. In other words, the models in the *Exc-*regime are not consistent with the data, because they would always predict a broader spike tuning close to pinwheel locations.
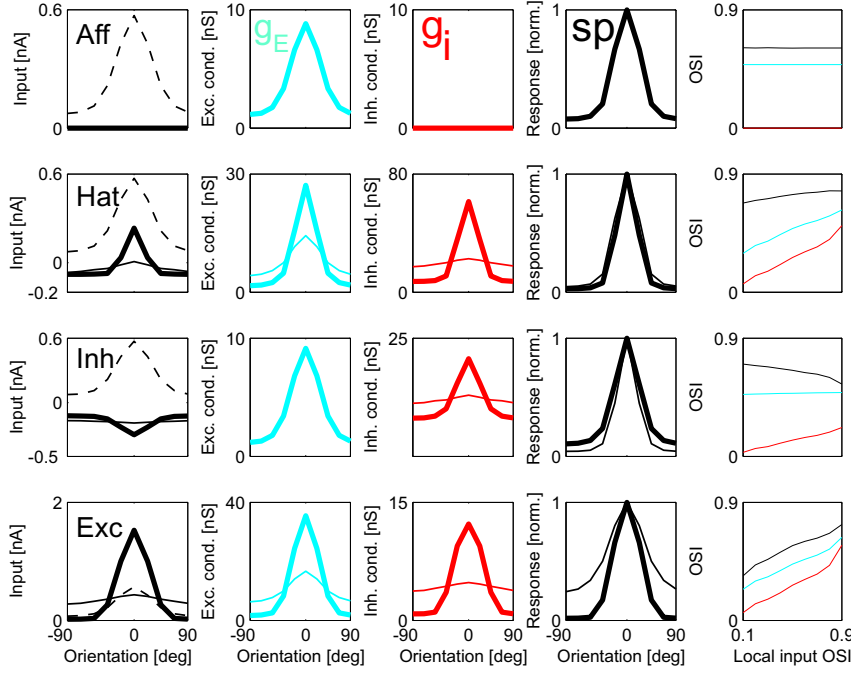
Figure 2.10: Responses of models representative for each regime. Each *row* corresponds to one regime. The solid lines in the *first column* show the sum of the input currents received via the local recurrent connections for a pinwheel (OSI=0.3, *thin line*) and orientation domain neuron (OSI=0.9, *thick line*). The *dashed line* shows the input current received via the afferent connections. The *second and third column* show the tuning of the excitatory and inhibitory conductances. The *fourth column* shows the spike tuning normalized to the maximum, and the *last column* shows how the tuning of the excitatory (*cyan*) and inhibitory (*red*) conductances and the spike responses (*black*) depend on the map location in the different parameter regimes. The regimes correspond to the following values of the connection strengths: $G_{EE} = G_{EI} = 0\,\text{nS}$ (*Aff*), $G_{EE} = 0.4\,\text{nS}$ and $G_{EI} = 1.1\,\text{nS}$ (*Hat*), $G_{EE} = 0.03\,\text{nS}$ and $G_{EI} = 1.3\,\text{nS}$ (*Inh*), and $G_{EE} = 0.2\,\text{nS}$ and $G_{EI} = 0.1\,\text{nS}$ (*Exc*).

2.5   Discussion

The experimental results have shown that visual stimulation evokes a different pattern of synaptic inputs at orientation domains compared to pinwheel centers. We demonstrated that these response patterns result from a diversity of synaptic inputs impinging on different locations in the orientation map, acting through a locally isotropic and recurrent anatomical architecture. That is, the spatial distribution of excitatory and inhibitory neurons provides the necessary anatomical inputs, and their synaptic drive provides sufficient functional balance to preserve sharp spike tuning, particularly at pinwheel centers.

We modeled these findings using a single cell and a network model. For the single cell model we constrained the free parameter p, which determines the contribution of the afferent relative to the recurrent excitation, to those values compatible with the experimentally measured data. The network model was explored by means of systematic numerical simulations. Different parameter regime were defined and characterized, and the model predictions were compared with the experimental data.

In our models, we always assumed that the orientation preference is initially set up by the pattern of the afferent connections. This is consistent with the recent experimental evidence indicating that the preferred orientation of a V1 neuron arises from the feedforward bias of its afferent inputs (Chapman et al., 1991; Reid and Alonso, 1995; Mooser et al., 2004). Orientation selectivity of the spike responses, however, appears to be narrower than afferent spread, and likely requires intracortical mechanisms for its generation as assumed by the recurrent models of orientation selectivity (Fig. 2.1c,d).

Our modeling results and the experimental data, however, suggest that the computation of orientation selectivity is best characterized as taking place in a regime, where the recurrency significantly contributes to driving the neurons[2]. For the single cell model, the smallest value of p compatible with the experimental data was $p = 0.25$. Assuming less than 25% of the average excitatory drive as originating from the afferent connections, for example, leads to predicted slopes of the g_i tuning too sttep for being compatible with the experimental data. On the other hand, assuming that the dominant part of the excitation comes from the afferent connections leads to predicted slopes of the g_e, g_i and membrane potential tuning too low for being compatible with the data.

These conclusions can be questioned, because we deduced the inhibitory tuning instead of explaining it in terms of the underlying neuronal circuits. However, our network model addressed this question. We have shown that a co-variation of excitation and inhibition indeed emerges due to the layout of the orientation map, which is most strongly in the *Hat*-regime (see Fig. B.1), and only the parameterizations in the Hat-regime are consistent with the data (see Fig. 2.9d).

Our model exploration is also suitable to resolve the apparent discrepancy between the experimental measurements of Marino et al. (2005) and the model by McLaughlin et al. (2000), which predicts sharper spike tuning close to pinwheel locations. We can reproduce their findings, as long as we operate our model in the *Inh*-regime. In this case, only the recurrent inhibition is strong,

---

[2]Here we have not elaborated on this point, but we have shown elsewhere (Wiesing et al., 2005) that the recurrency should not be as strong as assumed for recurrent models operating in the so-called "marginal phase" (Fig. 2.1c).

and it can sharpen a more broadly tuned afferent input close to pinwheel locations, because only at those locations cross-orientation interactions take place.

McLaughlin et al. (2000), however, did not assume the same spatial scale for the local excitation and inhibition. In contrast, our experimental results, consistent with a previous report by Anderson et al. (2000), demonstrate a close relationship between the spread of excitation and inhibition, which are similarly tuned regardless of map location. However, we have also shown elsewhere (Wiesing et al., 2005) that the assumption of the same spatial scales is not crucial for our conclusions. The source of most of these synaptic inputs is likely to be the local neighborhood of a neuron, though it cannot ruled out that at least excitatory inputs from iso-oriented sites that are located more distantly (Bringuier et al., 1999; Angelucci et al., 2002b) contribute some additional excitation.

In addition to location invariance, the orientation selectivity of V1 responses is also invariant with stimulus contrast. So far we have not explored as to whether the orientation tuning predicted by the models is also independent of the stimulus contrast. Given the crucial role of the threshold effect in our model to keep the spike response equally tuned at all map locations, it can be expected that the due to an iceberg-effect the tuning will not be contrast-invariant. However, it has been argued that when considering noise contrast-invariant tuning of the spike responses can be explained using a simple feedforward-model (Hansel and van Vreeswijk, 2002). Thus, the question of contrast-invariance in our models needs to be addressed in future studies.

Interestingly, a simple rule of spatial integration turns out to ensure a balance of excitation and inhibition, which produces sharp orientation tuning at all positions in the orientation map. Our results demonstrate the fundamental role of such a balance for a key emergent computation in the adult visual cortex. A similar mechanism based on the balance provided by local inputs may account for the tuning of other functional properties in visual cortex, and may be a general mechanism for generating and preserving response selectivity in sensory cortex (Wehr and Zador, 2003).

Recapitulation

The mechanism by which orientation selectivity in the primary visual cortex is computed has been discussed since it discovery by Hubel and Wiesel. One characteristic property, which is useful to categorize models of orientation selectivity, is the extend to which processing is attributed to the afferents or to the recurrency. Here we argued that the conductance measurements obtained recently (Marino et al., 2005) can greatly constrain the mechanisms by which orientation selectivity is computed. The tuning functions for the spike, membrane potential, and conductances responses were obtained as a function of the location in the orientation map, which has different local functional neighborhoods at different positions.

Assuming an isotropy of the local recurrent interactions we first determined the functional connectivity of the local recurrency for different locations in the orientation map. Using a single compartment Hodgkin-Huxley model neuron and the obtained orientation-dependent interactions we deduced the excitatory and inhibitory conductance tunings, which lead to the experimentally found location-independence of the spike tuning and are consistent with the measured dependence of the sub-threshold tunings on the map location. Furthermore, using a network model, we found that only a recurrent regime is compatible with the data.

Our findings support the following conclusions: i) Primary visual cortex is likely to operate in a regime, where local recurrent excitation and inhibition are balanced and almost cancel out each other, such that the non-uniformity of the local neighborhood due to the layout of the orientation map does not affect the spike tuning. ii) The origin of orientation selectivity is the pattern of afferent projections, but the local recurrency ensures that this tuning is preserved.

How does this study relate to our main theme of "adult adaptivity"? It has been shown (Dragoi et al., 2001) that orientation adaptation is most prominent close to pinwheel locations. Thus, our two-dimensional model is an important step for modeling those phenomena at a biological realistic level.

Chapter 3

Contextual modulation of orientation tuning in V1

Abstract. The responses of neurons in the primary visual cortex (V1) depend on the spatial context within which visual stimuli are embedded. Orientation-selective responses to stimuli in the receptive field (RF) center are suppressed by similarly oriented stimuli in the RF surround. The spatial scale and timing of surround suppression are consistent with a role for top-down feedback connections in its generation. We use an anatomically-based recurrent network model of macaque V1 to show how excitatory inter-areal feedback connections, contacting predominantly excitatory neurons, can generate suppression from the far surround. The basic mechanism involves feedback connections targeting excitatory neurons in the near surround, which then send horizontal connections to excitatory and inhibitory neurons in the RF center. A novel prediction of our model is that stimulation of the far surround does not always suppress, but can also facilitate the RF center's response, depending on the amount of excitatory input driving the local inhibitors.

3.1 Introduction

The responses of neurons in sensory cortices are not determined solely by the physical properties of the afferent stimulus, but are also modulated by the spatial context within which sensory stimuli are embedded (Moore et al., 1999; Albright and Stoner, 2002; Nelken, 2004). For example, in the primary visual cortex (V1, or striate cortex) the responses of neurons to optimally oriented stimuli within their receptive field (RF) center are usually suppressed by similarly oriented stimuli outside their RF, in the RF surround (Blakemore and Tobin, 1972; Nelson and Frost, 1978; Allman et al., 1985; Gilbert and Wiesel, 1990; DeAngelis et al., 1994; Sengpiel et al., 1997; Sceniak et al., 2001; Cavanaugh et al., 2002; Levitt and Lund, 2002). Contextual modulation from the extra-classical RF may represent the neural correlate of perceptual figure-ground segregation (Knierim and van Essen, 1992; Sillito et al., 1995; Bradley and Andersen, 1998; Nothdurft et al., 1999). Thus, identifying its anatomical substrates and mechanisms is crucial for our understanding of visual cortical processing.
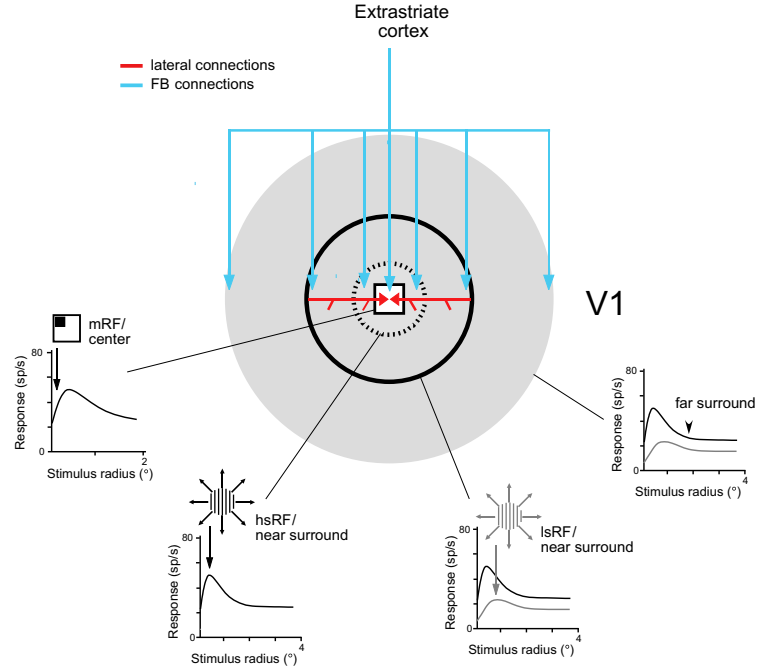
Figure 3.1: Schematic diagram comparing the spatial scale of the RF center and surround of a V1 neuron with the spatial scales of V1 horizontal connections and feedback (*FB*) connections to V1 from extrastriate cortex. The central circular area represents the various components of the RF center and surround of an example V1 neuron; this is surrounded by four identical size-tuning curves for this neuron, over which an *icon and arrow* indicate the size of the respective RF component. *Black and gray curves* are size-tuning curves measured at high and low stimulus contrast, respectively. *White square area*: minimum response field (mRF) or RF center; this is the RF region over which presentation of optimally oriented stimuli evokes spikes from the cell. Experimentally the mRF is measured using small, high contrast moving stimuli (*icon* on the top-left size-tuning curve) and delimiting the area where spikes are evoked from the cell; the *arrow under the mRF icon* indicates the size of the mRF relative to the peak of the high contrast size-tuning curve. *Dashed ring*: high contrast summation RF (hsRF); this is measured by presenting high contrast gratings of increasing radius (*icon* on the center-left size-tuning curve), and defined as the stimulus radius at the neuron's peak response (*arrow under the hsRF icon*). The region between the mRF and the hsRF is the region over which presentation of high contrast gratings at the same orientation as the center grating facilitates the cell's response to optimally oriented gratings in the center. *Continuous ring*: low contrast summation RF (lsRF); this is measured and defined as the hsRF size, but using low contrast gratings of increasing radius (*icon* on the center-right size-tuning curve). The region between the hsRF and the lsRF is the region over which presentation of gratings at the same orientation as the center grating suppresses or facilitates the cell's response to optimally oriented gratings in the center, depending on the grating's contrast. Note the shift to the right of the peak response at low contrast. *Gray area*: RF surround. We consider two

separate regions of the surround depending on their proximity to the RF center: 1) the *near surround* is the region between the mRF and the lsRF, 2) the *far surround* is the region outside the lsRF over which presentation of stimuli at the same orientation as the center stimulus usually suppress the cell's response to optimally oriented gratings in the center. Intra-areal V1 horizontal connections (*red*) are commensurate with the lsRF size of their V1 neurons of origin, while extrastriate FB (*blue*) connections to V1 are commensurate with the full spatial scale of the center and surround field of V1 neurons.

V1 cells are tuned to the size of a visual stimulus (DeAngelis et al., 1994; Sceniak et al., 2001; Cavanaugh et al., 2002; Levitt and Lund, 2002), i. e. they respond best to oriented stimuli of optimal size, with smaller and larger stimuli evoking smaller responses (Fig. 3.1). The size tuning of V1 cells is contrast-dependent (Kapadia et al., 1999; Sceniak et al., 1999), i. e. the RF region over which the cell's response increases with increasing stimulus size is larger at low stimulus contrast than when measured using high contrast stimuli (e. g. compare the peak response of black and gray size-tuning curves in Fig. 3.1). We refer to the radius of a high or low contrast stimulus evoking the largest response from the cell as the cell's high or low contrast summation RF (hsRF or lsRF) size, respectively. The summation RF (hsRF plus lsRF), to which here we refer as the near surround, is a high threshold depolarizing field immediately surrounding the minimum response field (mRF; or classical RF, cRF) of the V1 neuron (Fig. 3.1). The mRF is the low threshold spiking region of the RF center, usually mapped using small moving stimuli (Fig. 3.1) (Hubel and Wiesel, 1962; Barlow et al., 1967). Beyond the lsRF lies the suppressive region of the surround, to which we refer here as the far surround (Fig. 3.1).

Iso-orientation surround suppression in V1 has traditionally been attributed to long range intra-areal horizontal (or lateral) connections (Gilbert et al., 1996; Dragoi and Sur, 2000; Fitzpatrick, 2000; Somers et al., 2002; Stettler et al., 2002) made by excitatory neurons in layers 2/3 (Gilbert and Wiesel, 1983; Rockland and Lund, 1983), and linking neurons with similar stimulus specificity (Gilbert and Wiesel, 1989; Malach et al., 1993; Yoshioka et al., 1996; Bosking et al., 1997). As these connections terminate on both excitatory (80%) and inhibitory (20%) neurons (Martin and Whitteridge, 1984; Kisvarday et al., 1986; McGuire et al., 1991), surround suppression of the center response has been proposed to be mediated by lateral excitation of local inhibition (Lund et al., 1995; Somers et al., 1998).

Recent anatomical and physiological studies have challenged this notion that surround suppression in V1 is mediated solely by horizontal connections, and has suggested an involvement of top-down feedback (FB) signals in these effects. These studies have demonstrated that surround suppression in V1 can be evoked by stimuli at locations in the visual field far beyond the monosynaptic range of V1 horizontal connections (Sceniak et al., 2001; Angelucci et al., 2002b; Cavanaugh et al., 2002; Levitt and Lund, 2002). The spatial scale of these connections has been shown to match the spatial dimensions of V1 cells' lsRF (Fig. 3.1); this suggests that horizontal connections may, instead, play a role in shaping V1 neurons spatial summation properties at low stimulus contrast (Sceniak et al., 1999), and in the modulation of RF center responses arising from the near, but not the far, surround (Kapadia et al., 1995; Polat et al.,

1998; Chen et al., 2001). Moreover, the short temporal latency of suppression arising from the far surround (Bair et al., 2003; Muller et al., 2003) cannot be accounted for by polysynaptic chains of horizontal connections, due to the slow conduction velocity of horizontal axons (Grinvald et al., 1994; Bringuier et al., 1999; Girard et al., 2001; Slovin et al., 2002; Angelucci and Bullier, 2003). In contrast, the spatial scale and fast conduction velocity of FB projections from extrastriate cortex to V1 are commensurate with these large-scale suppressive effects (Girard et al., 2001; Angelucci et al., 2002b; Bair et al., 2003), suggesting that FB connections could represent the associated anatomical substrate (Fig. 3.1).

While reversible inactivation of extrastriate cortex had previously revealed a facilitatory influence of FB connections on the RF center response of V1 neurons (Sandell and Schiller, 1982; Mignard and Malpeli, 1991; Hupe et al., 1998; Hupe et al., 2001), recent inactivation studies have also demonstrated a role for FB connections in mediating surround suppression. Specifically, cooling of extrastriate area MT was shown to reduce the suppressive effect of surround motion stimulation in V3, V2 and V1 neurons (Hupe et al., 1998; Bullier et al., 2001). A role for FB connections in iso-orientation surround suppression is also consistent with recent evidence that FB connections are parcellated into discrete patches in V1 (Angelucci et al., 2002a; Angelucci et al., 2002b), and link cortical territories of like-orientation preference (Gilbert and Wiesel, 1989; Angelucci and Bullier, 2003; Shmuel et al., 2005). FB connections to V1 arise from excitatory neurons in layers 2/3A and 5/6 of extrastriate cortical areas V2, V3 and MT and, at least in rat area 17, target almost exclusively (97-98%) excitatory neurons (Johnson and Burkhalter, 1996). Accordingly, intracellular recordings in slices of rat visual cortex have indicated that FB pathways modulate V1 responses via synaptic mechanisms in which excitation dominates (Shao and Shao, 1996). In contrast, stimulation of horizontal connections evokes EPSPs followed by strong IPSPs (Hirsch and Gilbert, 1991; Shao and Shao, 1996). Such synaptic organization of FB connections may appear to be inconsistent with a role for this pathway in surround suppression.

In order to pinpoint the neuronal circuitry and mechanisms underlying responses within and outside the RF of macaque V1 neurons, in this study we have set up a recurrent neuronal network model whose architecture is constrained to fit the recent anatomical and physiological studies described above. We provide a solution to how excitatory extrastriate FB connections, activated by stimuli in the far surround, could exert their suppressive influence on V1 neurons, by targeting excitatory neurons in the near surround, which in turn send horizontal connections to local inhibitory neurons in the RF center. Our model can account for a wide range of physiological data regarding the static and dynamic effects of surround suppression. This includes the variation of the degree of near surround facilitation and suppression with the size and contrast of the stimulus (Sceniak et al., 1999; Cavanaugh et al., 2002; Levitt and Lund, 2002), the reduction of surround suppression following inactivation of FB from extrastriate cortex (Hupe et al., 1998; Bullier et al., 2001), and the latency of surround suppression (Bair et al., 2003). Furthermore, our model predicts that stimulation of the far surround can suppress or facilitate the RF center response, depending on the total amount of excitatory drive to the local inhibitors. This prediction is consistent with our recent physiological data (Ichida et al., 2005) demonstrating that the "suppressive surround" of V1 neurons is not always suppressive.

In order to elucidate the role of horizontal and feedback (FB) connections in generating the RF center and surround of V1 neurons, we construct a simplified recurrent network model of two interconnected visual areas, one corresponding to V1 and the other to an extrastriate area such as MT. Each area is idealized as a single layer of cells, namely, layer 2/3 of V1 and layer 6 of MT. Parameters of the model are fitted to known anatomical and physiological data, as specified below and in the Appendix C. Numerical simulations are performed for stimuli of varying size and contrast. In the figures, we represent such stimuli as gratings, in order to facilitate the link with experiments. However, in our network model, the size of a stimulus is simply taken to be the spatial extent (in degrees) of the feedforward afferents carrying nonzero currents to the V1 layer, and the contrast of the stimulus is taken to be the strength of these currents (see Appendix C). We do not model other stimulus features such as orientation and spatial frequency, since in this paper we focus on size-tuning and contrast effects.

A schematic diagram illustrating the basic network architecture is shown in Fig. 3.2a. Neurons in V1 and extrastriate cortex are identified according to the visual field position of their RF centers. Suppose that a particular V1 excitatory neuron is identified as the target neuron whose response properties we wish to determine. V1 neurons are then labeled according to the location (in degrees) of their RF centers relative to the RF of the target neuron. Thus, the center consists of neurons whose RF centers lie within the minimum response field (mRF) of the target neuron, the near surround consists of neurons whose RF centers fall within either the high (hsRF) or low (lsRF) contrast summation RF, and the far surround consists of neurons whose RF centers lie beyond the lsRF (see also Fig. 3.1). In the model, we assume that excitatory neurons in V1 receive feedforward afferent excitation from other V1 layers, local recurrent excitation and inhibition, long-range excitation via slow intra-areal horizontal connections, and FB excitation via fast inter-areal connections. Horizontal connections also target local interneurons, whereas FB connections only target excitatory neurons. This is motivated by the anatomical finding that FB axons contact predominantly excitatory neurons (Johnson and Burkhalter, 1996). The spatial extent of inter-areal FB connections is taken to be larger than that of intra-areal horizontal connections (Fig. 3.2b), following the anatomical data of Angelucci et al. (2002b). In particular, horizontal connections from the near surround, but not the far surround, connect to the RF center monosynaptically, and extend as far as the lsRF of their neurons of origin, while extrastriate FB connections extend as far as the far surround of V1 neurons.

Following previous recurrent network models (Somers et al., 1998; Dragoi and Sur, 2000), we assume that stimulation of the near surround modulates the response to a center stimulus via horizontal connections targeting both excitatory and inhibitory neurons in the center. The interneurons are assumed to have higher threshold and gain than the local excitatory neurons whose output they control ((Lund et al., 1995); Fig. 3.2c), and thus only generate suppression under sufficiently high levels of excitation, for example, when the contrast and/or size of a grating stimulus is sufficiently large. At low levels of excitation, such as for small or low contrast stimuli, the inhibitors are inactive and stimulation of the near surround facilitates the center response.
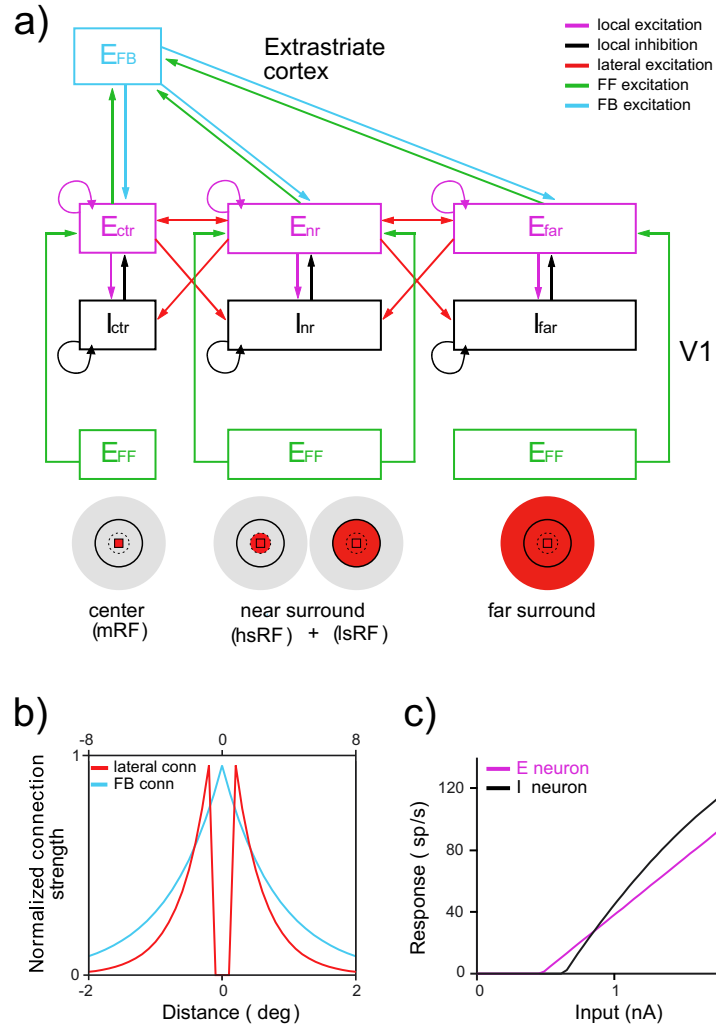
Figure 3.2: Basic architecture of the recurrent network model. **a)** Schematic diagram of the connections used in the network model. Different connection types are indicated as color-coded *arrows*. *Purple and black boxes* represent populations of excitatory (E) or inhibitory (I) V1 neurons, respectively, labeled according to the position of their RF center relative to that of the center neurons. Accordingly, *ctr* are the neurons in the RF center or minimum response field (*mRF*); *nr* are the neurons in the near surround, of which those closer to the center fall within the center neurons' high contrast summation RF (*hsRF*), while those located farther from the center fall within the center neurons' low contrast summation RF (*lsRF*); *far* are the neurons in the far surround. E_{FF}: excitatory neurons in other V1 layers sending feedforward afferents to the E neurons in V1 layers 2/3. E_{FB}: excitatory neurons in extrastriate cortex sending feedback projections to the E neurons in V1. FB connections are spatially highly divergent and convergent. Note the absence of direct FB inputs to I neurons. The latter receive monosynaptic inputs only from V1 horizontal connections

(*red arrows*) and from local E neurons via local recurrent connections (*purple arrows*). *Icons at the bottom of a)* represent the different regions of the RF center and surround (same conventions as in Figure 1), with *red areas* indicating the RF regions that are activated by each respective submodule. **b)** Different spatial scale used in the model for the feedback (*FB, blue*) and the lateral (or horizontal, *red*) connection. Plotted is the normalized strength of lateral and FB connections to a postsynaptic neuron as a function of the distance of the presynaptic neurons' RF centers from the RF center of the postsynaptic neuron. Note the different scales on the x-axes (the upper one applies to the FB connections, and the lower to the lateral connections). **c)** Firing rate of the V1 local excitatory (*E, purple*) and inhibitory (*I, black*) neurons in the model, plotted against the input current.

3.2   Minimal model

The recurrent network model shown in Fig. 3.2a illustrates the basic principles of our proposed circuit. It will serve as a guideline throughout this chapter. However, we first focus on a simplified instantiation of the model shown in Fig. 3.2a, because the basic mechanism of modulating the response of a center neuron via an excitatory inter-areal pathway can already be illustrated using such a *minimal model*. In the next section, we will use only the full recurrent network model, which explicitly takes into account the spatial extent of visual stimulation and activation in the cortex.

Model definition

The minimal model has three pairs of an excitatory and an inhibitory neuron corresponding to the RF centre, the near and far surround, and an excitatory neuron in the extrastriate area (Fig. 3.3a). Denoting with

$$r\left(t\right) = \begin{pmatrix} r_E^C\left(t\right) \\ r_E^N\left(t\right) \\ r_E^F\left(t\right) \\ r^X\left(t\right) \\ r_I^C\left(t\right) \\ r_I^N\left(t\right) \\ r_I^F\left(t\right) \end{pmatrix}$$

and

$$\mathbf{W} = \begin{pmatrix} w^{EE} & w^{EN} & 0 & w^{E_C X} & w^{EI} & 0 & 0 \\ 0 & w^{EE} & 0 & w^{E_N X} & 0 & w^{EI} & 0 \\ 0 & 0 & w^{EE} & w^{E_F X} & 0 & 0 & w^{EI} \\ w^{XC} & w^{XN} & w^{XF} & 0 & 0 & 0 & 0 \\ w^{IE} & w^{IN} & 0 & w^{I_C X} & 0 & 0 & 0 \\ 0 & w^{IE} & 0 & w^{I_N X} & 0 & 0 & 0 \\ 0 & 0 & w^{IE} & 0 & 0 & 0 & 0 \end{pmatrix}$$
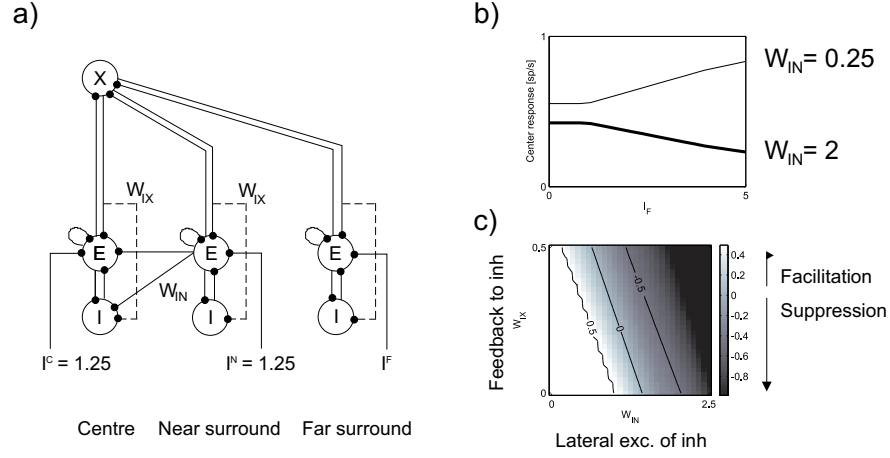
Figure 3.3: Suppression mediated via feedback excitation in the minimal model. **a)** Minimal circuit model to illustrate how stimulation of the far surround can lead to a suppression of the centre response. Excitatory feedback can mediate this suppression either indirectly via lateral excitation of a local inhibitory neuron, $w_{IX} = 0$, or directly via monosynaptic excitation of this inhibitory neuron, $w_{IX} > 0$. **b)** Responses of the centre excitatory neuron as the stimulation I_F of the far surround is increased without feedback to the inhibitory neurons, $w_{IX} = 0$, for weak (*thin line*, $w_{IN} = 0.25$) and strong (*thick line*, $w_{IN} = 0$) lateral excitation of the inhibitory neuron. **c)** Suppression/facilitation of the centre response as a function of the weights w_{IX} and w_{IN}. The suppression is computed as $-(R_C - R_{CS})/R_C$, where R_C is the response to the central stimulus alone, and R_{CS} is the response with stimulation in the far surround. Negative values correspond to a suppression.

the neuronal rates and the weight matrix, we define the dynamics of the minimal model as

$$\frac{dr_l}{dt} = -r_l + \max\left(0, \sum_m W_{lm}r_m + I_l - T_l\right), \tag{3.1}$$

where $T_l = 1$ for $l = E, I, N, F$ and $T_l = 0$ for $l = X$. Feedback to the inhibitory neurons in the far surround was excluded, because we assumed it to be weak anyway. The I_l was set to 1.25 for $l = C, N$ and changed between 0 and 5 for $l = F$. All other afferent inputs were set to 0. The values for w_{IN} and w_{IX} were varied and are given in the Caption to Fig. 3.3. The values for the other weights are $w_{EE} = w_{IE} = 0.25$, $w_{EI} = -1$, $w_{XC} = w_{XN} = 0.5$, $w_{XF} = 0.2w_{XC}$, $w_{EN} = 0.25w_{EE}$. All responses shown in Fig. 3.3 are the steady-states of Eq. 3.1.

Results

Previous models of suppressive contextual effects in V1 assumed the long-range lateral projections as the sole neuronal substrate of surround suppression (Somers et al., 1998; Dragoi and Sur, 2000). These projections, however, are

made only by excitatory neurons. Hence, by themselves they cannot have a suppressive effect. Therefore, it was postulated that they target local inhibitory neurons, which mediate the suppression. We hypothesize (see Introduction) that this strategy of decomposing a suppressive pathway into a cascade of multiple steps solves the puzzle of how the exclusively excitatory inter-areal feedback, which targets predominantly excitatory neurons (Johnson and Burkhalter, 1996), can mediate the fast suppression from the far surround.

Let us consider the minimal circuit model shown in Fig. 3.3a. The centre excitatory neuron receives feedback excitation from an extrastriate neuron, inhibition from a local interneuron, and afferent excitation. Excitation due to afferent stimulation of the far surround is relayed via the extrastiate neuron to another excitatory neuron in the near surround, which in turn excites the local inhibitory neuron. However, as to whether this feedback excitation leads to facilitation or suppression of the centre excitatory neuron depends on the strength w_{IN} of the lateral excitation of the interneuron, which itself does not receive any feedback excitation. Fig. 3.3b shows how the response of the centre excitatory neuron changes as the afferent stimulation of the far surround increases. For weak lateral excitation (thin line), the response increases, whereas it decreases for strong lateral excitation (thick line). In other words, with strong lateral excitation of the inhibitory neuron the local inhibition dominates over the excitation received via the feedback and lateral connections. Thus, suppression from the far surround, which is relayed via an extrastriate area, can be explained with solely excitatory inter-areal feedback to excitatory neurons. However, if the local interneuron also receives feedback excitation, response suppression can be evoked with weaker lateral excitation (see Fig. 3.3c).

3.3   Recurrent network model

In this section we use the recurrent network model described in detail in Appendix C. In this model, there is no monosynaptic feedback excitation of inhibitory neurons, because using the minimal model we have shown that suppression from the far surround can be evoked without monosynaptic feedback excitation of inhibitory neurons. This, however, depends on the lateral excitation of the inhibitory neurons being strong (see Fig. 3.3c).

The same dependence is also observed in the recurrent network model. Fig. 3.4 shows the suppression from the far surround analog to Fig. 3.3c for the minimal model. As in the minimal model (see Fig. 3.3c), the same amount of suppression from the far surround can be evoked with weaker lateral excitation of inhibitory neurons (G_{IE}^{SS}) if the feedback connections target the inhibitory neurons ($G_{IE}^{SX} > 0$). In the rest of this section, however, we use the network parameterization, which evokes the strongest suppression from the far surround without any feedback excitation of the inhibitory neurons (Fig. 3.4, bottom right parameterization).
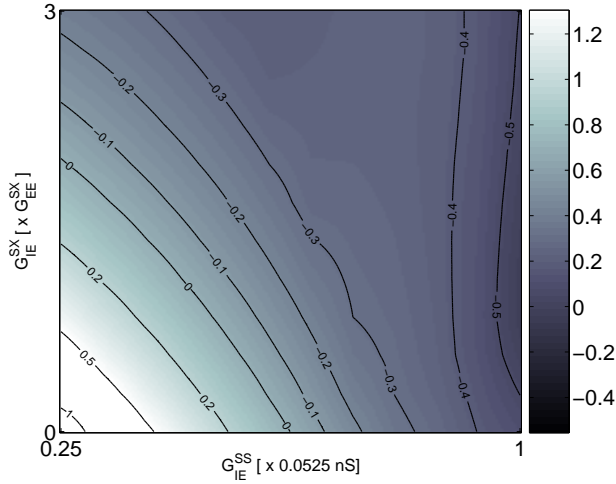

Figure 3.4: Surround suppression in the network model with different strengths of monosynaptic feedback excitation of inhibitory neurons. Shown is the facilitation/suppression (same definition as in Fig. 3.3c) as computed for a central stimulus with 0.5 deg radius and an "annulus" with inner radius of 3.2 deg and outer radius of 8 deg.

Contrast-dependent spatial summation

We now use this model to simulate experiments in which the RF size is measured as the stimulus radius at the peak response of an area-summation curve (the summation RF), using high or low contrast gratings. The summation RF size is known to depend on the stimulus contrast, with lower contrasts yielding larger summation RF sizes ((Sceniak et al., 1999); see Fig. 3.1). The results of these simulations are shown in Fig. 3.5a-c. Increasing the size of the stimulus leads to an increased response of the center excitatory neurons. This facilitation turns into suppression as the stimulus size is further increased, due to the fact that the local interneurons, now sufficiently stimulated, start firing (Fig. 3.5a). The stimulus size at which the interneurons are activated, however, depends on the stimulus contrast. At lower contrast, the firing rates of all excitatory neurons are also lower, and so the interneuron has to integrate over a larger area to reach its firing threshold. This accounts for the contrast-dependence of the RF size measured via the area-summation curves. In order for the above mechanism to be effective, in the model the center interneurons must be strongly driven by horizontal connections, while the latter provide only weak excitation to the center excitatory neurons (Fig. 3.5b,c). Fig. 3.5b,c additionally shows how, in the model, the various input currents to the excitatory and inhibitory neurons vary with stimulus size. Another interesting feature of the curves shown in Fig. 3.5a is that there is an oscillatory component to the variation in response as a function of stimulus size. This occurs when there is disinhibition of the center excitatory neurons, due to suppression of the excitatory neurons in the near surround driving the center inhibitory neuron. Such suppression in turn depends on the level of excitation of horizontal connections to near surround neurons arising from the center and far surround.

In Fig. 3.5d-f we show a sequence of diagrams for increasing stimulus size, highlighting the network components that are active and the major afferent pathways from the active regions to the center neurons. A stimulus fitted to the size of the hsRF activates the center and part of the near surround (Fig. 3.5d). The center receives excitatory inputs via feedforward afferents, via horizontal connections from the active region of the near surround, and via FB connections from extrastriate cortex. At high contrast, the local inhibitors are assumed to be close to threshold, so that a further increase in stimulus size beyond the hsRF size, as in Fig. 3.5e, leads to suppression of the center response. On the other hand, expansion of a low contrast stimulus beyond the hsRF size results in facilitation, until the full extent of the near surround is active (Fig. 3.5e). Finally, high or low contrast stimuli that extend well beyond the lsRF activate the far surround (Fig. 3.5f). This suppresses the center via FB connections exciting neurons in the near surround, which then excite local interneurons in the center. In principle, it could be possible for the far surround to suppress the center via a cascade of horizontal connections from the far, to the near surround, to the center (see Fig. 3.2a). Although such a cascade is too slow to account for the fast onset of suppression (Girard et al., 2001; Angelucci and Bullier, 2003; Bair et al., 2003), it could contribute to late suppression. However, in our simulations we find that such a contribution is negligible (see Discussion).
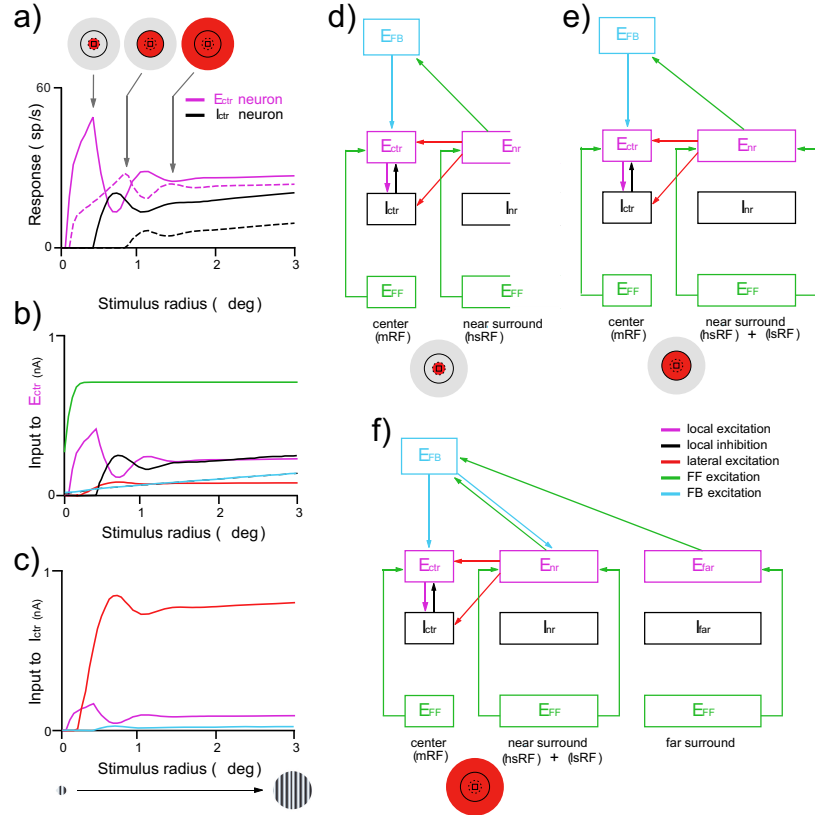
Figure 3.5: Contrast-dependent spatial summation in the network model. **a)** Response (spikes/sec) of the center excitatory (E_{ctr}, *purple*) and inhibitory (I_{ctr}, *black*) neurons as a function of the size of the afferent stimulus (in degrees), for two different stimulus contrasts (85% and 15%, *solid and dashed lines*, respectively). *Icons at the top* represent the different regions of the RF center and surround, with *red areas* indicating the RF regions that are activated at specific points in the size-tuning curve. The same icons are shown in d)-f) to indicate the respective active network components and pathways. **b,c)** Input currents to the center excitatory neurons b) and to the center inhibitory neurons c) as a function of the afferent stimulus size (in degrees). The different input types are color coded according to the legend in f). *Icons in c)*: stimulus used, and valid for a)-c). **d,f)** A sequence of diagrams highlighting, for increasing stimulus size, the network components that are active and the major afferent pathways from the active regions to the center neurons. Conventions are as in Fig. 3.2a.

Contrast-dependent suppression and facilitation from the far surround: a model prediction

To investigate more directly the contribution of FB connections to the far surround of V1 neurons, we simulate an experiment in which afferent stimulation of the region between the hsRF and lsRF is withdrawn. This allows us to minimize the afferent drive of horizontal connections in the near surround, so as to unmask the effect of FB connections from the far surround. In order to achieve this, a stimulus analogous to a high contrast annular grating is presented in the far surround together with a central high contrast stimulus fitted to the size of the hsRF (stimulus shown in Fig. 3.6b). The inner radius of the "annulus" is systematically decreased from the far surround to a size no smaller than that of the lsRF, so that near surround neurons in the lsRF but beyond the hsRF never receive afferent stimulation. As the inner radius is decreased, more neurons in the far surround receive afferent stimulation. This leads to suppression of the excitatory neurons in the center (Fig. 3.6a), as observed experimentally (Levitt and Lund, 2002). The active components in the network, and the main pathways relaying information to the center neurons in response to this stimulus configuration are illustrated in Fig. 3.6c. The mechanism leading to the response suppression is the same as that illustrated in Fig. 3.5. Specifically, as more neurons in the far surround are activated, excitatory neurons in the near surround increase their firing rate (Fig. 3.6b); this is due to increased excitation received via FB connections activated by stimulation of the far surround. The near surround neurons then activate the center interneurons via horizontal connections (Fig. 3.6a,c), resulting in suppression. However, our model also demonstrates that, when a low contrast central stimulus is presented together with a high contrast annulus in the far surround (stimulus shown in Fig. 3.6b, right panel), decreasing the inner radius of the "annulus", i.e. stimulating an increasingly larger number of neurons in the far surround, can lead to initial facilitation of the center excitatory neurons followed by suppression (Fig. 3.6a, right panel). This is due to the fact that the low contrast central stimulus is too weak to activate the center interneurons by itself (Fig. 3.6b); thus, stimulation of the far surround initially facilitates the response of the center neurons, because FB inputs to the center sum with afferent and horizontal inputs, until a critical "annulus" size is reached, beyond which the interneurons are activated (Fig. 3.6a, right panel), and suppression occurs.

One of the advantages of using a recurrent network model is that it allows us to make predictions regarding the population activity profile of neurons across V1 in response to visual stimuli. Fig. 3.7a shows the predicted activity profile of excitatory neurons as a function of position relative to the center neuron population, in the case of a small high contrast stimulus (0.4° radius, i.e. the size of the hsRF as derived from Fig. 3.5a) and a large high contrast stimulus (1.5° radius, i.e. extending into the far surround; see Fig. 3.5a). The responses of the neurons in the center and near surround decrease with increasing stimulus size (thin vs. thick line), as expected from the simulated area summation experiments (Fig. 3.5a). However, the responses of neurons with their RF centers located at the border of the large stimulus, in the far surround, are not suppressed. Fig. 3.7b shows the predicted activity profile for a small high contrast central stimulus (0.5° radius, i.e. just larger than the neurons' hsRF, but smaller than the lsRF- see Fig. 3.5a) presented with and without a high

contrast "annular" stimulus (3.2° inner radius and 8° outer radius) in the far surround. When the "annulus" is presented together with the central stimulus, the responses of the center excitatory neurons are suppressed, whereas the responses of the excitatory neurons in the near surround are facilitated. This can be explained by the same mechanism illustrated in Fig. 3.6. Namely, that activation of FB, due to stimulation of the far surround, enhances the response of excitatory neurons in both the center and near surround, which in turn activate the center inhibitors (Fig. 3.6c). It is worth emphasizing that the excitatory neurons in the near surround are not suppressed, but are facilitated, by the far surround stimulus; this occurs because their local inhibitors, in contrast to the local inhibitors in the center, do not receive sufficiently high afferent drive, due to the small size of the central stimulus.
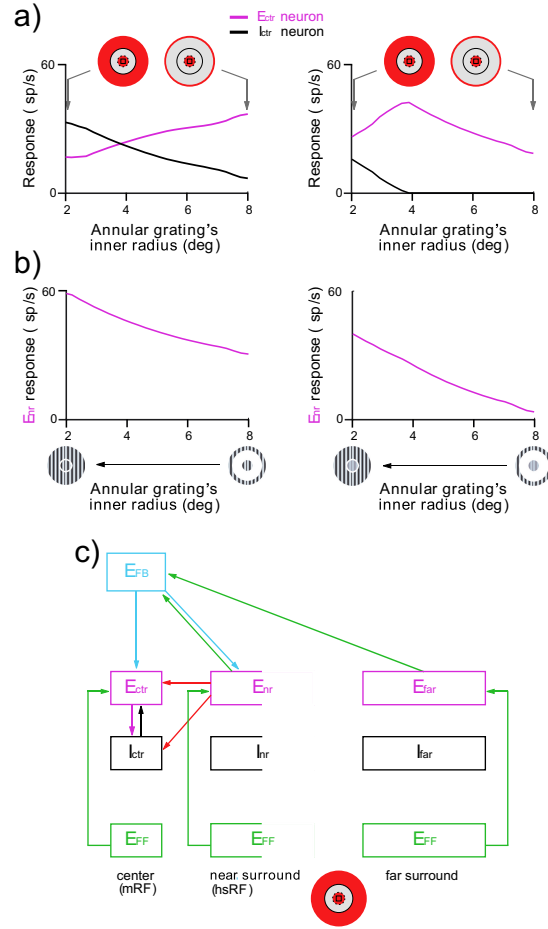
Figure 3.6: Contrast-dependent suppression and facilitation from the far surround in the model. **a)** *Left panel*: Response (spikes/sec) of the center excitatory (E_{Ctr}, purple) and inhibitory (I_{Ctr}, *black*) neurons to a high contrast (85%) central stimulus of 0.5 deg radius (i.e. just larger than the neuron's hsRF) plotted against the inner radius of a high contrast (85%) annular stimulus of 8 deg outer radius presented together with the central stimulus. *Right panel*: Same as in the left panel, but the configuration of the stimulus used consisted of a low contrast (15%) central stimulus (0.5 deg radius) presented together with a high contrast (85%) annular stimulus in the far surround. **b)** *Left panel*: Same as in a), but for excitatory neurons in the near surround (E_{nr}). *Right panel*: Same as in the left panel, but for excitatory neurons in the near surround (E_{nr}). **c)** Diagram highlighting, for the RF regions activated by the stimulus (*red areas* in the icon at the bottom), the network components that are active and the major afferent pathways from the active regions to the center neurons. Conventions are as in Fig. 3.2a, 3.5d-f.

a)
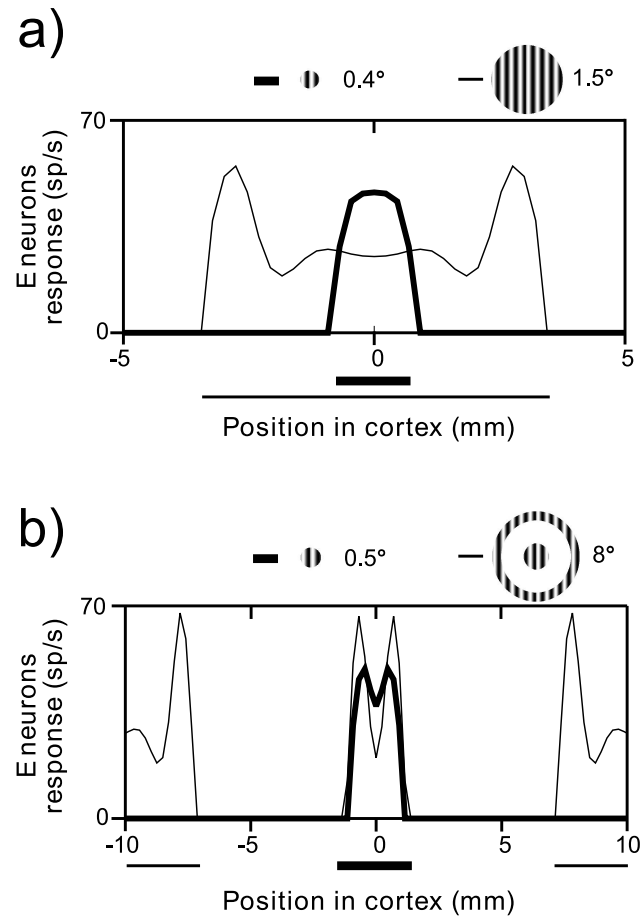
Figure 3.7: Population activity profile of the excitatory neurons across V1. **a)** Activity profile for a small (0.4 deg radius, *thick line*) and a large (1.5 deg radius, *thin line*) high contrast (85%) stimulus. Plotted are the firing rates of the excitatory neurons against their location in V1. **b)** Activity profile for a small (0.5 deg radius) high contrast (85%) central stimulus presented alone (*thick line*) or together (*think line*) with a high contrast (85%) annular stimulus (3.2 deg and 8 deg inner and outer radius, respectively). The cortical location of the neurons were computed by converting the distance in degrees between their RF centers into cortical distance, using a cortical magnification factor of 2.3 mm/deg.

Inactivation of the feedback pathway

The question of how inter-areal FB connections contribute to suppressive contextual effects was directly addressed in a series of experiments by Bullier and coworkers, using reversible inactivation of extrastriate cortex (Hupe et al., 1998; Bullier et al., 2001). Inactivation of FB connections arising from area MT greatly reduced the suppressive effect of surround stimulation (sometimes even

causing facilitation) onto the center responses of neurons in areas V1, V2 and V3. The largest reduction in surround suppression was observed for low-saliency stimuli.

In our network model, we can mimic the effect of FB inactivation simply by setting the values for the weights of the inter-areal FB connections to zero. To determine the contribution of FB connections to the center and surround of V1 neurons, we consider the change in response of the center excitatory neurons to a central stimulus presented alone or together with a surround stimulus, with and without active FB connections, and for center-surround stimuli with different saliencies. As in the experimental studies (e.g. Hupe et al., (1998)), here we define the "response change" as

$$-100\frac{R_C - R_{CS}}{R_C}$$

where R_C and R_{CS} are the responses of a neuron to the presentation of the central stimulus alone and of the central stimulus shown together with a surround stimulus, respectively. Stimulus saliency here is defined as $C_{center}/C_{surround}$, where C_{center} and $C_{surround}$ are the contrasts of the central and surround stimulus, respectively. To obtain stimuli with different saliency, the surround stimulus is always presented at low contrast (15%), while the contrast of the central stimulus (>15%) is changed. Note that the central stimulus is fit to the optimal RF size of the neurons whose response we wish to determine; this is defined as the stimulus size at peak response, derived from the neurons size tuning curve, and varies with the contrast of the stimulus (Fig. 3.5a). Thus, for stimuli of different saliency, the central stimulus ranges in size from the hsRF to the lsRF, and therefore also includes part of, or all the near surround. Fig. 3.8a,b shows the response of the center neurons to a central stimulus presented either alone or together with a surround stimulus. The response with and without FB is compared under low- (Fig. 3.8a) and high-saliency (Fig. 3.8b) conditions. In the case of the central stimulus alone, removal of FB leads to a reduction in the response of the center neurons, as in the experimental findings (Sandell and Schiller, 1982; Mignard and Malpeli, 1991; Hupe et al., 1998; Bullier et al., 2001; Hupe et al., 2001). In our model this follows from the fact that, under normal conditions, FB excites excitatory neurons in the center (see Fig. 3.2a,b). In the presence of FB connections, the response of the center neurons to the central + surround stimulus is suppressed relative to the response to the central stimulus alone, and the degree of suppression is greater for high-saliency stimuli. In the latter case, the amount of suppression is similar with and without FB (Fig. 3.8b). On the other hand, for low saliency stimuli, there is a significant reduction in the suppression when FB is removed (Fig. 3.8a). Fig. 3.8c shows the response change (as defined above) as a function of the stimulus saliency, with and without FB connections. The reduction in surround suppression when inactivating the FB is clearly strongest for low saliency stimuli. In our model, the reduction in suppression is due to the fact that the suppressive effect of the far surround is mediated by FB connections targeting excitatory neurons in the near surround (Fig. 3.2a). Removal of FB reduces excitation of the near surround, and consequently lowers the excitation of the center local inhibitors targeted by excitatory neurons in the near surround. Our results are consistent with the experimental findings (Hupe et al., 1998; Bullier et al., 2001), and suggest a crucial role for inter-areal FB connections in surround suppression and

figure-ground segregation.
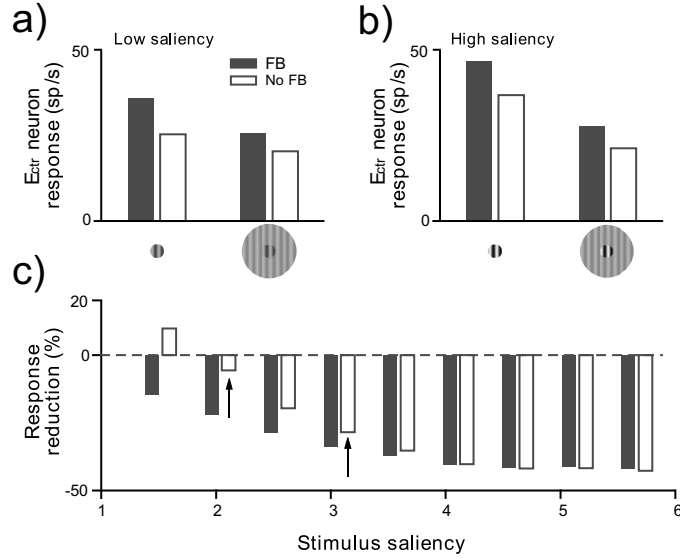
Figure 3.8: Center-surround interactions with and without feedback. **a)** Response (spikes/sec) of the center excitatory neurons (E_{Ctr}) to the central stimulus alone (*small grating*; 0.68 deg radius), and to the central stimulus presented together with a surround stimulus (*large grating*; surround stimulus of 0.68 deg inner radius and 8 deg outer radius), in the presence of feedback (*FB; gray bars*) or with the FB connections inactivated (*white bars*). The center + surround stimulus was presented at low saliency (38% vs 15% contrast; see Results for definition of saliency). **b)** same as in a), but for a high saliency center + surround stimulus (77% vs 15 % contrast). The size of the central stimulus was always fitted to the stimulus size at the peak response of the size-tuning curve (i.e. the optimal size), which varies with the stimulus contrast; thus in b) the central stimulus had a radius of 0.49 deg, which is the optimal stimulus size at 77% contrast. **c)** Response change (see text for definition) of the center excitatory neurons as a function of stimulus saliency, with and without FB. For each contrast, the central stimulus was presented at the optimal size.

Dynamics of contextual effects

Recently, the short-term dynamics of suppressive contextual effects has been investigated in macaque V1 (Bair et al., 2003). The latency of suppression induced by stimuli in the far surround was found to be almost as short as the latency of suppression induced by stimuli in the near surround. The short delay in the onset of suppression induced by far surround stimulation is not consistent with polysynaptic chains of horizontal connections mediating this effect, due to their slow conduction velocity. These results suggest a role for FB connections in the generation of suppression from the far surround.

We simulate these experiments and investigate the model's predictions for these delays. Fig. 3.9a shows the response, and Fig. 3.9b its suppression strength, of the center excitatory neurons to a stimulus protocol where first only a central stimulus, the size of the hsRF, is presented, and then a high contrast "annular stimulus" is flashed in the far surround, after the center response reaches its steady state (>200 ms). The response suppression is strongest for "annuli" with smaller inner radius, i.e. involving larger regions of the surround (Fig. 3.9d). In our model, the suppression from the far surround is mediated via relaying excitation from V1 neurons in the far surround to the extrastriate area (via feedforward connections), and from there to the excitatory neurons in the near surround (via FB connections), which in turn excite the center inhibitory neurons (via horizontal connections), as in Fig. 3.6c. The predicted responses of excitatory neurons in the near surround and of the local inhibitory neurons in the center are shown in Fig. 3.9c. The dynamics of the response of the inhibitory neurons in the center follows the response dynamics of the excitatory neurons in the near surround, which drive these inhibitory neurons. Importantly, the latency of suppression onset in the center neurons' response is almost independent of the inner radius of the annulus, i.e. it is only weakly dependent on the location of stimulation in the surround (Fig. 3.9e), as observed experimentally (Bair et al., 2003). In the model, this is due to the suppression being mediated via the fast inter-areal FB connections, which introduce a delay of only a few milliseconds.
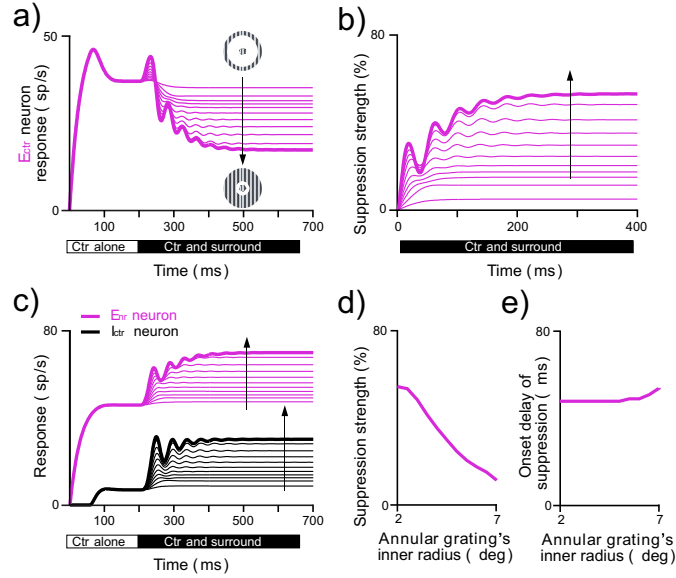
Figure 3.9: Dynamics and latency of surround suppression. **a)** Response (spikes/sec) of the center excitatory neurons to a high contrast (85%) central stimulus (0.5 deg radius) first presented alone and, after 200 ms, together with a high contrast "annular stimulus" in the far surround of varying inner radius (2.5-7 deg), plotted against the time from the onset of the central stimulus. Here, and in b) and c), the *arrow* points in the direction of decreasing inner radius of the annulus, as indicated by the *stimulus icon*. **b)** Suppression strength of

the center neurons' response induced by a stimulus in the far surround, plotted against the time from the onset of the suppression. Suppression strength is defined as $100\left(R_C - R_{CS}\right)/R_C$, where R_C and R_{CS} are the neurons' response to the presentation of the central stimulus alone and of the central stimulus shown together with the surround stimulus, respectively. **c)** Responses (spikes/sec) of inhibitory neurons in the center *(black)* and excitatory neurons in the near surround *(purple)* for annuli of varying inner radius. **d)** Suppression strength of the steady state response of the center excitatory neurons as a function of the surround annulus' inner radius. **e)** Latency of suppression onset (defined as the time from the onset of the annulus at which suppression strength reached 3%) as a function of the surround annulus' inner radius.

3.4   Discussion

We have developed an anatomically and physiologically-based recurrent neuronal network model of center-surround interactions in macaque V1. In addition to V1 horizontal connections, traditionally thought to represent the anatomical substrate for V1 cells' extra-classical RF (e.g. (Gilbert et al., 1996; Fitzpatrick, 2000; Somers et al., 2002)), we have incorporated into our model inter-areal FB projections to V1, based on recent experimental findings implicating these connections in center-surround interactions (Angelucci et al., 2002b; Angelucci et al., 2002a; Levitt and Lund, 2002; Bair et al., 2003; Angelucci and Bullier, 2003; Shapley, 2004). The model incorporates realistic spatial scales and conduction velocities for each connection type, taken from anatomical and physiological data in macaque. We have used our model to elucidate the relative roles of horizontal and FB connections in generating the spatial summation properties of V1 neurons and their dependence on stimulus contrast, as well as the timing and dynamics of center-surround interactions. Our results are consistent with the hypothesis that extrastriate FB connections underlie the modulation of V1 neurons' responses arising from the far RF surround. We propose a solution to how excitatory FB connections, targeting almost exclusively excitatory neurons in V1 (Johnson and Burkhalter, 1996), can generate surround suppression; namely by targeting excitatory neurons in the near surround sending monosynaptic horizontal connections to inhibitory neurons in the RF center. This hypothesis is also consistent with recent evidence in primates that FB connections link cortical territories of broadly similar orientation preference in V1 and extrastriate cortex (Angelucci and Bullier, 2003; Shmuel et al., 2005). Finally, our model suggests that both horizontal and FB connections contribute to the modulation arising from the near surround of V1 neurons, and both facilitate responses to stimuli within the RF center.

As in previous recurrent network models of center-surround interactions (Somers et al., 1998; Dragoi and Sur, 2000), in our model stimulation of the near surround modulates the response to a center stimulus via horizontal axons targeting both excitatory and inhibitory neurons in the center, the latter having higher threshold and gain than the local excitatory neurons whose output they control (Lund et al., 1995; Somers et al., 1998; Angelucci et al., 2002b). The specific sign of the modulatory effect, i.e. facilitation or suppression, thus depends on the overall level of excitation reaching the local inhibitory and excitatory neurons, with facilitation predominating for low levels of excitation and sup-

pression taking place for high levels of excitation. Our model, however, extends the "lateral model" (Somers et al., 1998), by adding FB inputs from extrastriate cortex as an additional source of excitation to excitatory neurons in the center and in the near surround. While the "lateral model" can account for many contextual effects, it cannot account for the experimentally observed suppression arising from locations beyond the scale of monosynaptic V1 horizontal connections (Sceniak et al., 2001; Angelucci et al., 2002b; Cavanaugh et al., 2002; Levitt and Lund, 2002). It has been argued that cascades of horizontal connections cannot account for the fast onset of suppression arising from the far surround, due to the slow conduction velocity of horizontal axons (Angelucci and Bullier, 2003; Bair et al., 2003). Instead, the spatial scale (Angelucci et al., 2002b) and conduction velocity (Girard et al., 2001) of extrastriate FB connections to V1 are commensurate with the spatial scale and fast onset of suppression from the far surround. Thus in our model, to account for far surround suppression, we have incorporated FB connections. Whereas polysynaptic chains of horizontal connections cannot account for the fast onset of suppression from the far surround, they could in principle contribute to the late phase of the suppression. In our simulations we find such a contribution to be negligible, due to surround suppression of excitatory V1 neurons at each step in the chain, which prevents horizontal propagation of signals. However, while our model excitatory neurons all show surround suppression, in real V1 some neurons do not i.e. their size tuning curves show a plateau at the peak response, or an increase in response for increasing stimulus sizes (e.g. (Levitt and Lund, 2002)). The latter kinds of neurons in our model are assumed to be the local inhibitory neurons controlling the output of local excitatory neurons (see Fig. 3a). While this assumption is supported by experimental evidence that at least some inhibitory neurons in V1 indeed do not show surround suppression (Hirsch and Gilbert, 1991), it is possible that also some excitatory neurons in V1 are not suppressed by stimuli in their RF surround. The latter neurons could in principle propagate horizontal signals from the far surround to the RF center, and thus contribute to the late phase of the suppression. Note, however, that cascading lateral connections are unlikely to contribute to the modulatory effects of the far surround in the case of the "annular grating" experiments illustrated in Fig. 3.6. In this case, excitatory neurons whose RFs lie in the visual field location of the blank stimulus do not receive afferent drive, and thus cannot effectively relay signals to their postsynaptic V1 neurons.

A novel prediction of our model is that the far surround of V1 neurons is not always suppressive, but can facilitate the RF center response under conditions in which the amount of excitatory input to the local inhibitors is relatively low, such as when the center neurons are stimulated with a low contrast stimulus in the presence of a stimulus in the far surround. Facilitation of the center response by similarly oriented stimuli in the surround has typically been reported when the surround region near the RF center is stimulated (Kapadia et al., 1995; Polat et al., 1998; Chisum et al., 2003). Iso-oriented stimuli in the RF center and far surround, instead, have typically been shown to be suppressive (DeAngelis et al., 1994; Levitt and Lund, 1997; Sengpiel et al., 1998; Sceniak et al., 2001; Levitt and Lund, 2002; Cavanaugh et al., 2002). This behavior of RF center-surround interactions has been modeled as a difference of Gaussians (DOG), i.e. as the summation of two spatially overlapping mechanisms having Gaussian sensitivity profiles: a center excitatory mechanism, corresponding to the RF center, and

a spatially larger inhibitory mechanism, corresponding to the extraclassical RF surround (reviewed in (Shapley, 2004)). Iso-orientation facilitation from the far surround of V1 neurons, as predicted by our model, has not been reported previously, and is inconsistent with traditional DOG descriptions of contextual effects in V1. However, our recent physiological studies in macaque V1 seem to confirm such prediction of our model (Ichida et al., 2005). Thus, in order to account for facilitation from the far surround, the DOG model needs to be extended.

Our model can account for several extra-classical RF phenomena previously observed experimentally. These include: 1) the expansion of the summation RF at low stimulus contrast (Kapadia et al., 1999; Sceniak et al., 1999); 2) the size tuning curves of V1 neurons (Sceniak et al., 2001; Cavanaugh et al., 2002; Levitt and Lund, 2002); 3) far surround suppression, as shown in experiments in which the afferent drive to near surround neurons sending horizontal connections to the center was partially withdrawn by interposing a blank stimulus between a grating in the RF center and a grating in the far surround (Levitt and Lund, 2002); 4) FB-mediated facilitation of responses to stimuli in the RF center, and FB mediated suppression of responses to stimuli in the RF center by stimuli in the surround, as demonstrated by experiments in which FB connections were inactivated (Hupe et al., 1998; Bullier et al., 2001); 5) the latency and dynamics of surround suppression (Bair et al., 2003). The same model can also be easily extended to account for the effects of spatial attention in V1.

Our model nicely captures the experimental observation of Bair et al. (2003) that the onset latency of surround suppression is almost independent of the distance of the surround stimulus from the RF center (Fig. 3.9e). However, it cannot account for the finding, from the same study, that suppression from the far, but not the near, surround is predominantly transient. Our model may be too simplistic to account for this transient response; however, further exploration, or an extension of the model to include mechanisms affecting the short-term dynamics of center-surround interactions, such as firing rate adaptation, could reveal transient suppression emerging as a consequence of the network dynamics.

A central assumption of our model, without which reproduction of many of the physiological observations would not occur, is that local inhibitory neurons are strongly driven by horizontal connections. This assumption is consistent with experimental observations in slices of cat and rat visual cortex (Hirsch and Gilbert, 1991; Shao and Shao, 1996). If, for example, local excitatory neurons were to be the predominant drivers of inhibitory neurons, the latter would follow the responses of the former. Strong inhibition from within the RF center would dominate, and thus mask any additional suppression arising from the surround.

One possible limitation of our model is that the extrastriate area is represented only as a one-dimensional line of neurons without intra-areal interactions, thus serving purely as a relay station. While inclusion of interactions within extrastriate cortex would certainly enrich the repertoire of response properties predictable by the model, it is remarkable that such a simple circuit model can account for a broad spectrum of V1 neuron responses.

While a role for FB connections in top-down, global-to-local processing of visual information has been previously suggested to explain phenomena, such as the responses of V1 and V2 neurons to illusory (Lee and Nguyen, 2001) or occluded (Sugita, 1999; Bakin et al., 2000) contours, and spatial attention, we

are proposing a more fundamental role for FB connections. Namely, that FB inputs from outside area V1 directly contribute to the response properties of V1 neurons, and thus V1 and extrastriate cortex act in unison to generate responses to simple visual patterns.

Recapitulation

Here we have proposed a new pathway for suppressive contextual effects in the primary visual cortex (V1). It explains suppression evoked from the far surround, which is beyond the spatial scale of the lateral connection within V1. The model is consistent with the currently available physiological and anatomical data. The key idea is that excitation from the far surround is relayed via the extra-striate cortex to excitatory neurons in the near surround, which in turn send horizontal connections to local inhibitory neurons mediating the suppression. The developed model predicts that stimulation of the far surround could also have a facilitatory effect on the centre, as long as the horizontal connections are inactive, so that the local inhibitory neurons are below their firing threshold. This prediction has recently been confirmed in physiological experiments (Ichida et al., 2005).

Our modeling results as well as corresponding experiments (Ichida et al., 2005) lead to the following conclusions: i) It is possible to mediate suppression from distal locations in the visual field, even if the mediating connections are excitatory with the only exception being the last step in this pathway, where inhibition is mediated by a local inhibitory neuron. ii) Lateral excitation of inhibitory neurons is likely to be rather strong, because this is a necessary prerequisite for explaining the observed suppression with this circuit. Our results suggest that during centre-surround interactions in V1 a near and a far excitatory mechanism are operative, which are interacting and could both turn into inhibitory mechanisms, depending on the stimulus configuration and contrast.

Assuming that our model captures the essential properties of inter-areal signaling between the striate and the extrastriate areas, then attentional top-down modulations (see, e. g., (Sharma et al., 2003)) have to use these circuits. Future modeling studies can now investigate these modulations based on an experimentally tested model.

Chapter 4

Adaptivity of tuning functions in hypercolumn models

Abstract. The representation of orientation information in the adult visual cortex is plastic as exemplified by phenomena like perceptual learning or attention. Although these phenomena operate on different time-scales and give rise to different changes in the response properties of neurons, both lead to an improvement in visual discrimination or detection tasks. If, however, optimal performance is indeed the goal, the question arises, why the changes in neuronal response properties are so different. Here we hypothesize, that these differences arise naturally if optimal performance is achieved by means of different mechanisms. In order to evaluate this hypothesis we set up a recurrent network model of a visual cortical hypercolumn and asked, how each of four different parameter sets (strength of afferent and recurrent synapses, neuronal gains, and additive background inputs) must be changed in order to optimally improve the encoding accuracy of a particular set of visual stimuli. We find that the predicted changes of the population responses and the tuning functions were different for each set of parameters, hence were strongly dependent on the plasticity mechanism which was operative. An optimal change of the strength of the recurrent connections, for example, led to changes in the response properties which are similar to the changes observed in perceptual learning experiments. An optimal change of the neuronal gains led to changes mimicking neural effects of attention. Assuming the validity of the optimal encoding hypothesis, these model predictions can be used to disentangle the mechanisms of perceptual learning, attention, and other adaptation phenomena.

4.1   Introduction

The representation of orientation information in the adult visual cortex is plastic as exemplified by phenomena like perceptual learning or attention. Although these phenomena operate on different time-scales and give rise to different changes in the response properties of neurons, both lead to an improvement in visual discrimination or detection tasks. If, however, optimal performance is indeed the goal, the question arises, why the changes in neuronal response properties are so different. Here we hypothesize, that these differences arise

naturally if optimal performance is achieved by means of different mechanisms. In order to evaluate this hypothesis we set up a recurrent network model of a visual cortical hypercolumn and asked, how each of four different parameter sets (strength of afferent and recurrent synapses, neuronal gains, and additive background inputs) must be changed in order to optimally improve the encoding accuracy of a particular set of visual stimuli. We find that the predicted changes of the population responses and the tuning functions were different for each set of parameters, hence were strongly dependent on the plasticity mechanism which was operative. An optimal change of the strength of the recurrent connections, for example, led to changes in the response properties which are similar to the changes observed in perceptual learning experiments. An optimal change of the neuronal gains led to changes mimicking neural effects of attention. Assuming the validity of the optimal encoding hypothesis, these model predictions can be used to disentangle the mechanisms of perceptual learning, attention, and other adaptation phenomena.

Orientation tuning is the paradigmatic example of stimulus selectivity in the visual cortex. It first arises in the primary visual cortex (V1), and it is preserved in higher visual areas like V2 and V4. The typically bell-shaped orientation tuning functions, however, have been shown experimentally to be highly adaptive. They depend on the temporal context of the stimulation (Muller et al., 1999; Dragoi et al., 2002), the current behavioral demands (Moran and Desimone, 1985; Treue and Trujillo, 1999; McAdams, 1999), and they also change during the course of training a perceptual task (Schoups, 2001; Ghose et al., 2002; Yang, 2004). This raises the question of why the representations of physically unchanged stimuli in the adult visual cortex are so adaptive. Wouldn't it be sufficient for an animal to act successfully if one proper representation of the environment's current 'state' were computed in the visual cortex and then forwarded to neuronal structures responsible for planning and initiating actions?

Many previous theoretical works are based on optimal coding hypotheses in order to explain the observed changes. Paradiso (1988), Seung and Sompolinsky (1993), Clifford et al. (2000), Nakahara et al. (2002; 2001), and Bethge et al. (2003), for example, used descriptive models of tuning functions and assessed the quality of the sensory representation as a function of the tuning functions' parameters. In reality, however, neuronal response properties are computed in a recurrent cortical network where architecture and plasticity mechanisms constrain the set of available tuning functions and their possible changes. Hence, the changes predicted by descriptive models may not be realizable. Teich and Qian (2003) set up a physiologically plausible model in order to explain the changes of orientation tuning functions in V1 during adaptation and perceptual learning. In this study, however, the synaptic changes were not derived from a functional principle, rather they were determined ad hoc in order to fit experimental data.

This motivated us to combine both approaches and to evaluate an optimal coding principle for a physiologically realistic model of a visual cortical hypercolumn. A recurrent neuronal network encodes a stimulus, for example the orientation θ of an oriented grating, by the activity of its output neurons. The quality of this representation can then be assessed using a hypothetical ideal observer (decoder or read-out). Within such a setting (Fig. 4.1) we address the following two questions: (i) How do the tuning functions and population responses change if the quality of representation is optimally improved for a rel-

evant set of stimuli? (ii) How are these changes affected, if plasticity is restricted to one of the four sites of plasticity: maximum conductance of the afferent or recurrent synapses, gain of the excitatory neurons, and strength of an additive (feedback) input current?

We find that optimal changes of response properties are different for different sites of plasticity and that specific changes can even go into opposite directions and still improve coding quality. This finding stresses how important it is to consider physiological constraints when interpreting data in the light of a functional principle. We also find that published experimental data on perceptual learning and on attentional modulations of tuning functions in the visual cortex are broadly consistent with the model's predictions if the recurrency (for perceptual learning) or the values of the neuronal gain (for attentional modulation) are changed. This motivates the hypothesis that seemingly unrelated phenomena may be explained by one functional principle, and that diversity emerges because different routes are taken to calibrate cortical representations according to the same goal.

In the next section we briefly motive the Fisher information objective function. The detailed description of the model (Fig. 4.1b) can be found in the Appendix D.1.

4.2   The Fisher information objective function

The average output activity of the mean-field (rate) model is then converted into a noisy spike output activity with Poisson statistics with the spikes being conditionally independent given the stimulus. The probability to count n spikes fired by neuron i in a time-interval of duration τ is given by

$$P_i(n; \theta) = \frac{[\tau \cdot f_i(\theta)]^n}{n!} \exp(-\tau \cdot f_i(\theta)),$$

where θ is the presented stimulus, and $f_i(\theta)$ is the steady-state response f_i of neuron i to the stimulus θ. We always used $\tau = 1$.

In order to quantify the quality of the representation of the stimulus θ, we consider a hypothetical ideal observer whose task is to provide the best possible estimate $\hat{\theta}$ of the stimulus given a set of spike counts and knowledge of the probability distribution $P_i(n; \theta)$. In an estimation task the Fisher information is a useful quantity to measure the quality of a representation, because (for a one-dimensional continuous stimulus θ) the Fisher information provides, via $1/J(\theta)$, a lower bound for the variance of any unbiased estimator of θ (Kay, 1993). If $P_i(n; \theta)$ is the probabilistic description of how the spike count n of neuron i relates to the stimulus value θ, then no unbiased estimate of θ based on the spike count n can have a lower variance than $1/J(\theta)$.

For a population of N neurons with their 'noise' being statistically independent given the stimulus, the population Fisher information is $J(\theta) = \sum_i J_i(\theta)$, where

$$J_i(\theta) = E\left[\left(\frac{d}{d\theta} \log P_i(n; \theta)\right)^2\right]_{P_i(n; \theta)}.$$

The Fisher information is also monotonically related to the mutual information between the stimulus θ and the whole vector $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ of the
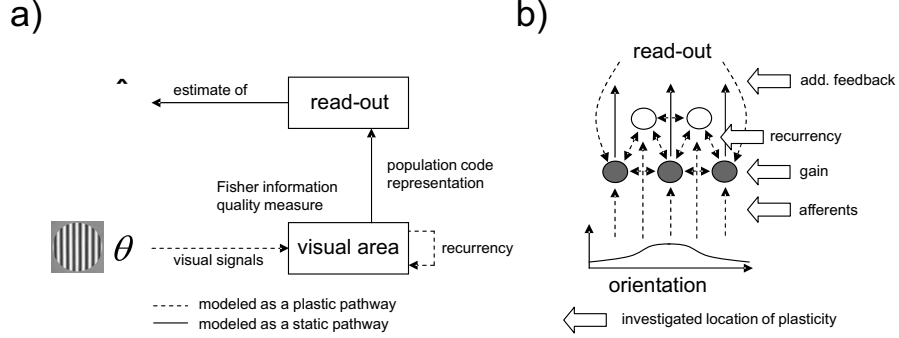
Figure 4.1: The hypercolumn model and the encoder/decoder framework for assessing the quality of sensory representations. **a)** The encoder/decoder framework. An ideal observer computes a point estimate of the stimulus θ based on the neuronal responses of the cortical hypercolumn. The variance of this estimate should be minimal, therefore the Fisher information should be maximized. **b)** Recurrent network model of a cortical hypercolumn of excitatory (filled circles) and inhibitory (open circles) neurons. The thick arrows point to the "sites of plasticity" (dashed lines) considered in this paper.

spike counts (Brunel and Nadal, 1998) as well as to the measure d' often used is the psychophysics literature (Seung and Sompolinsky, 1993). For Poisson statistics of the spike response one gets

$$J_i(\theta) = \tau \left(\frac{\partial f_i(\theta)}{\partial \theta} \right)^2 / f_i(\theta).$$

Fig. 4.2b shows $J_i(\theta = 0.5)$ as a function of neuron i (normalized to the maximum) for the initial network parameterization.

Note that here the Fisher information of a single neuron i, $J_i(\theta)$, is proportional to $[f_i'(\theta)]^2 / f_i(\theta)$, where $f_i(\theta)$ and $f_i'(\theta)$ are the tuning function of that neuron and its derivative. In order to determine how much changes in the absolute values of $f_i(\theta)$ and the changes in the slopes $f_i'(\theta)$ contributed to the total improvement of the Fisher information, we calculated the quantities $J_{add}(\theta)$ and $J_{slp}(\theta)$. For $J_{add}(\theta)$ we used the amplitudes $f_i(\theta)$ after we changed the model parameters and the derivatives $f_i'(\theta)$ before the re-parameterization. For $J_{slp}(\theta)$ we used the derivatives $f_i'(\theta)$ after but the amplitudes $f_i(\theta)$ before the re-parameterization. Thus, an increase/decrease of the encoding accuracy only due to changes of the slope is reflected by a large change of $J_{slp}(\theta)$, whereas an increase/decrease only due to changes of the response magnitude is reflected by a large change of $J_{add}(\theta)$.

Objective functions: A full optimization of the network is only a reasonable approach if additional constraints on the plausible range of values of the network parameters are imposed. Since we want our results to depend as little as possible on other constraints than the chosen network architecture and the chosen site of plasticity we consider how the objective functions change as a function of an optimal but small change of the values of the model parameters. Therefore, we slightly vary the relevant model parameters along the gradient of
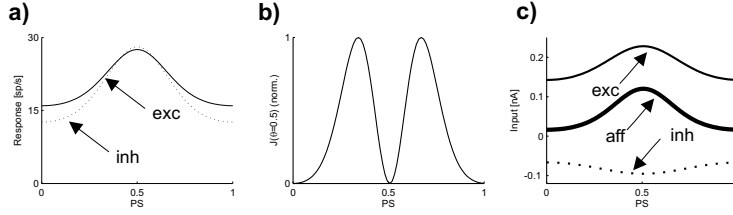
Figure 4.2: Response properties and quality of representation of the model hypercolumn for the choice of parameters given in the Methods section. **a)** Responses of excitatory (solid line) and inhibitory (dotted line) neurons to a stimulus with $\theta = 0.5$. **b)** Contribution of the excitatory neurons to the network's Fisher information for $\theta = 0.5$ normalized to the maximal value. **c)** Input currents at the steady state for an excitatory neuron with PS=0.5 as a function of the PO of the presynaptic neuron.

the objective function, similar to (Nakahara and Amari, 2002; Nakahara et al., 2001), but without further constraints.

In the following we consider the two objective functions $J(\theta = 0.5)$ and $\int J(\theta)\, d\theta$. The first objective function quantifies how well the stimulus $\theta = 0.5$ is encoded. It is an example for the task to improve coding accuracy for a small set of relevant orientations, which may happen, e. g., during a perceptual learning experiment. The second objective function quantifies how well all stimuli are encoded. It is an example for the task to overall improve coding accuracy, which may happen in a spatial attention experiment. The derivatives of $J_i(\theta)$ w. r. t. the conductance G_{rs}^{rec} at the recurrent synapses, the conductance G_r^{aff} of the afferent synapses, the gain a_r^E of an excitatory neuron r, and the additive input I_r^{add} are given in the Appendix.

4.3 Optimal changes of four selected mechanisms

In this section we report the changes of the four sets of network parameters and the resulting changes of the neuronal responses for an optimal increase of the quality of the representation. As a measure of quality we use the Fisher information for a particular stimulus value as well as the integral of the Fisher information terms over all stimulus values. The first case corresponds, for example, to the the task of improving the ability to judge orientations close to a reference orientation (e. g. vertical). The second case corresponds, for example, to the task of improving this ability for all oriented stimuli like at a particular visual field location. We choose the initial parameterization of our network such that it reproduced the "control" responses in area V4 of the macaque as reported by Yang and Maunsell (2004). We checked whether the reported response changes were different when the initial parameters were chosen differently ('Mexican-hat'-like recurrent connections with weaker and stronger values for the maximum conductances), but we found no qualitative differences.

Plasticity at the afferent synapses

We first asked how to change the maximum conductances G_i^{aff} of the afferent synapses to excitatory and inhibitory neurons in order to increase the network's Fisher information specifically for $\theta = 0.5$. We computed the gradient of the corresponding objective function w. r. t. the maximum conductances of the afferent synapses and changed their values by a small amount ΔG_i^{aff} proportional to this gradient (see Methods). Fig. 4.3a shows the ΔG_i^{aff} normalized by the "unadapted" initial values $\Delta G_{i,pre}^{aff}$ as a function of the preferred stimulus (PS) of the postsynaptic neuron i. The changes were strongest for neurons with PS differing approx. ± 0.16 from $\theta = 0.5$. For excitatory neurons with these PS the afferent synapses became stronger (thick solid line) whereas the synapses to inhibitory neurons with these PS became weaker (thick dashed line). The afferent synapses to excitatory neurons with PS very close to $\theta = 0.5$ do not change, but inhibitory neurons with these PS receive slightly more excitation via their afferents after the adjustment. We also asked how to change the afferent synapses in order to increase the Fisher information for all stimuli. The resulting changes were uniform. All synapses to excitatory neurons became stronger (thin solid line), and all synapses to inhibitory neurons became weaker (thin dashed line). Fig. 4.3b compares the population Fisher information before and after the adjustments and demonstrates that for the stimulus-specific changes the strongest increase was for the stimulus $\theta = 0.5$ (solid line). However, the performance also increased for stimuli close to $\theta = 0.5$, because the neurons which increase their contribution to the encoding of this stimulus also contribute to the encoding of the nearby stimuli. For the uniform change the Fisher information was also increased in a uniform manner (dotted line).
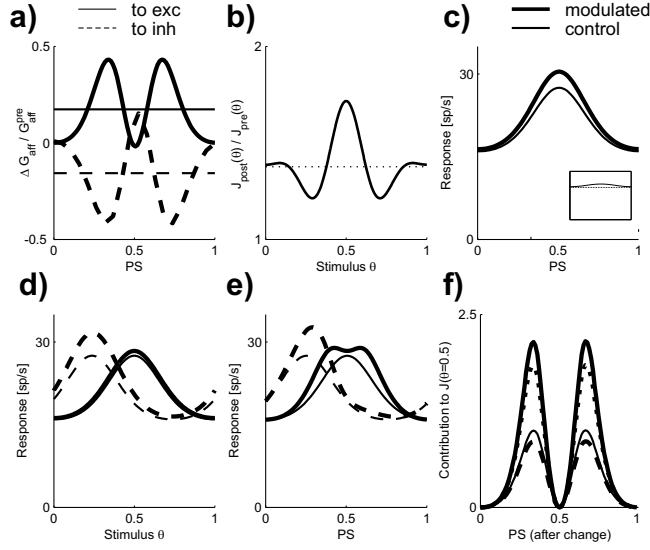
Figure 4.3: Adjusting the afferent synapses. **a)** Predicted changes of the afferent synapses to excitatory (solid line) and inhibitory (dashed line) neurons to optimally increase the objective functions $J(\theta = 0.5)$ (thick lines) and $\int J(\theta)\, d\theta$ (thin lines), see Methods. These changes were computed by following the gradient with a step size $\eta = 10^{-8}$. **b)** Relative change of the Fisher information after having adjusted the afferent synapses for the $J(\theta = 0.5)$ and $\int J(\theta)\, d\theta$ objective functions (solid and dotted lines). **c)** Population response before and after (thin and thick lines) having made the adjustment in order to increase $\int J(\theta)\, d\theta$. In this case, the shape of the population response after the adjustment is the same as the shape of the individual neurons' tuning functions. The inset shows the ratio of the responses after the adjustment to the responses with the initial parameterization (solid line, dotted line marks a ratio of 1). **d)** Tuning functions for two neurons with PSs 0.25 and 0.5 (dashed and solid lines) before and after (thin and thick lines) optimizing $J(\theta = 0.5)$. **e)** Population response to the same two stimuli $\theta = 0.25$ and $\theta = 0.5$ (thin and thick lines) for the same adjustment. **f)** Fisher information of the individual neurons for $\theta = 0.5$ before and after (thin solid and thick dotted lines) optimizing $J(\theta = 0.5)$. The solid lines shows the Fisher information of the re-parameterized network if it were only due to the changes of the response magnitude ($J_{add}(\theta = 0.5)$, thick dashed line) and tuning function slopes ($J_{slp}(\theta = 0.5)$, thick solid line) at $\theta = 0.5$ (see text for details). The normalization is w. r. t. the maximal Fisher information before the re-parameterization.

For both the stimulus-specific and the uniform changes of the afferent synapses the Fisher information increased, because the tuning functions were changed. If the adjustment was uniform, then the initially "symmetric" network parameterization (all neurons and associated connections have equal values of their model parameters) remained symmetric. For symmetric parameterizations the population response to a stimulus has the same shape as the neurons? tuning functions. However, for stimulus-specific adjustments this is no longer the case. Therefore, here as well as for the other three mechanisms we report only the population responses for uniform adjustments, but for stimulus-specific adjustments we show both the population response and the tuning functions.

Fig. 4.3c shows how the population response of the excitatory neurons to the stimulus $\theta = 0.5$ (and hence the shape of all tuning functions) changed after the uniform adjustment of the afferent synapses. All excitatory neurons received more afferent excitation, and all inhibitory neurons received less afferent excitation. Without recurrent connections this would have caused a multiplicative effect on all stimulus-driven activations, but due to the recurrent interactions shaping the activation profile, the activity increase is stronger for neurons with already high activity. The inset of Fig. 4.3c (thick line) shows the ratio of the population response after the adjustment to the initial response. The changes are not strictly multiplicative, because a multiplicative change would correspond to a horizontal line.

Figs. 4.3d,e show examples of tuning functions and population responses after the stimulus-specific adjustments. Fig. 4.3d shows the tuning functions for two neurons with PS 0.5 and 0.25 (solid and dashed lines) before and after the re-parameterization (thick and thin lines). The peak activation of the neuron with PS=0.5 increased only very little, but it showed a much stronger increase for the neuron with PS=0.25. The tuning functions gave rise to population activations shown in Fig. 4.3e. Here, two stimuli with $\theta = 0.5$ and $\theta = 0.25$ were used (solid and dashed lines). With the initial parameterization (thin lines) the population activations for every stimulus had the same shape, only the location of the peak activity was dependent on the stimulus. However, after the re-parameterization the shape of the population activation depends on the stimulus. For example, the activation profile for $\theta = 0.5$ became bimodal, because the tuning functions for neurons with PS close to 0.5 were not changed whereas the peak amplitudes of neurons with PS differing by ± 0.16 from $\theta = 0.5$ increased. For $\theta = 0.25$ the profile was again unimodal. The peak activity increased and was slightly shifted towards $\theta = 0.5$.

Let us now analyze, how the re-parameterized network achieved its increase of the quality of the representation around $\theta = 0.5$. Fig. 4.3f (thin line) shows the contribution of every neuron to the population's Fisher information for the initial parameterization (normalized to the maximal contribution, cf. Fig. 4.2b) as well as the contribution after the re-parameterization (dotted line, relative to the normalization). Neurons with an initially high contribution increased their Fisher information even more whereas the contribution of neurons with initially low Fisher information for θ remained low.

For conditionally independent Poisson spike trains the Fisher information of a single neuron i for a stimulus θ, $J_i(\theta)$, is proportional to $[f_i'(\theta)]^2 / f_i(\theta)$, where $f_i(\theta)$ and $f_i'(\theta)$ are the tuning function of that neuron and its derivative (see Methods). Fig. 3f shows how much of the overall improvement is due to changes of the response magnitudes and the slopes. For neurons with a high contribution

to the population Fisher information the value J_{add} ($\theta = 0.5$) after is below the value $J(\theta)$ before the change (thick dashed vs. thin solid line) whereas the value J_{slp} ($\theta = 0.5$) after is above the value $J(\theta = 0.5)$ before the change (thick solid vs. dotted lines). The increased response magnitudes would have caused a decrease of the encoding accuracy, but this effect was compensated by the increase of the slopes at $\theta = 0.5$, and overall the encoding accuracy for $\theta = 0.5$ was increased.

Plasticity at the recurrent synapses

Let us now consider the consequences of only adjusting the conductances at the recurrent synapses. We first adjusted the synapses in order to increase the population Fisher information specifically for the stimulus $\theta = 0.5$ (see Methods). Fig. 4.4a shows the values ΔG_{ij}^{EE} by which we changed the synapses at the recurrent connections between excitatory neurons as a function of the pre- and postsynaptic initial PS. For neurons with PS different from $\theta = 0.5$, synaptic changes depend on the postsynaptic neuron's PS. The excitation from presynaptic neurons with PS similar to the PS of the postsynaptic neuron increased, and the increase was strongest for postsynaptic neurons with PS differing approx. ± 0.16 from $\theta = 0.5$ (Fig. 4.4a, inset). The excitation from presynaptic neurons with PS different from the PS of the postsynaptic neuron decreased, and the decrease was also strongest for postsynaptic neurons with PS differing approx. ± 0.16 from $\theta = 0.5$. The changes for the other three types of recurrent connections (EI, IE and II) are complementary to the changes shown in Fig. 4.4a. Where the excitation of excitatory neurons increased, the inhibition of excitatory neurons decreased, the excitation of inhibitory neurons decreased, and the inhibition of inhibitory neurons increased.

These adjustments caused changes of the tuning functions which in turn gave rise to the population Fisher information shown in Fig. 4.4b. For the objective function $J(\theta = 0.5)$ the strongest increase was around $\theta = 0.5$ (solid line) whereas the increase was "uniform" for the uniform objective function $\int J(\theta) \, d\theta$ (dotted line). For the latter the population response to a stimulus with $\theta = 0.5$ is shown in Fig. 4.4c. The responses of all neurons in the re-parameterized network were lower compared to the responses with the initial parameterization (thick vs. thin line). The shape of the population activation (and hence of the individual tuning functions) changed as well, and overall the change was not strictly multiplicative (see inset). The reduced activity and the shape change resulted in an increased encoding accuracy for all stimuli.
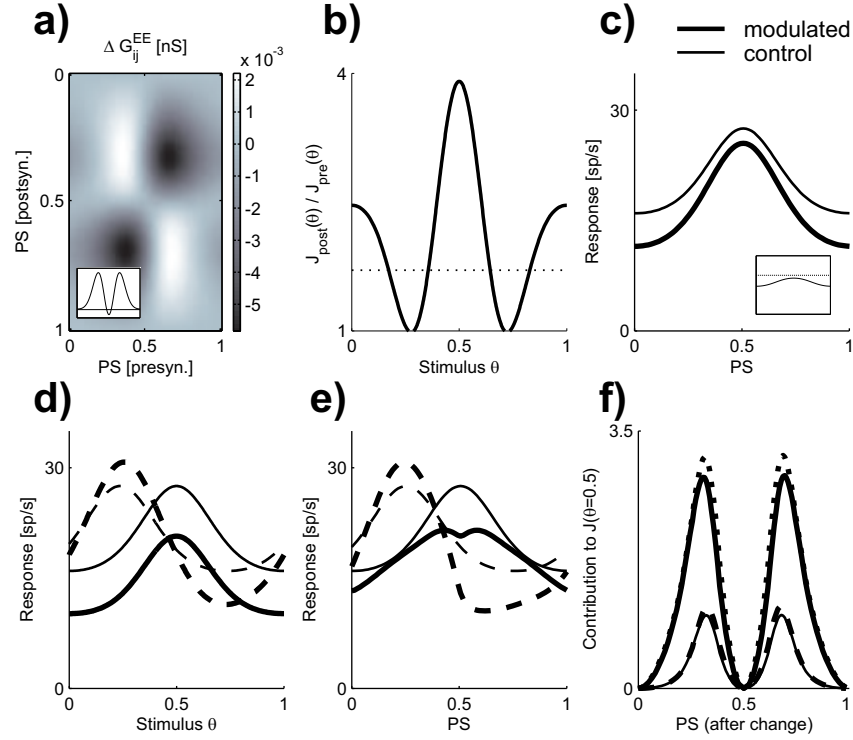
Figure 4.4: Adjusting the recurrent synapses. **a)** Predicted changes of the recurrent excitatory synapses to excitatory neurons for optimizing $J(\theta = 0.5)$. The inset shows the change of the self-excitation (the horizontal indicates no change). These changes were computed by following the gradient of the objective function with a step size $\eta = 3 \cdot 10^{-6}$. **b-f)** Same as in Fig. 4.3b-f.

Figs. 4.4d,e show examples of tuning functions and population responses after improving $J(\theta = 0.5)$. Fig. 4.4d shows the tuning functions for two neurons with PS 0.5 and 0.25 (solid and dashed lines) before and after the re-parameterization (thin and thick lines). The peak activity of the first neuron decreased, but the shape of its tuning function remained unchanged, whereas the tuning function of the second neuron became sharper and its peak activity increased. These tuning functions explain the population responses of the re-parameterized network to the two stimuli $\theta = 0.5$ and $\theta = 0.25$ shown in Fig. 4.4e. The profile of the response to $\theta = 0.5$ was bimodal and below the initial responses, whereas the response profile for $\theta = 0.25$ became sharper and its peak activity increased. Fig. 4.4f shows the two "hypothetical" Fisher information terms $J_{add}(\theta = 0.5)$ and $J_{slp}(\theta = 0.5)$ (see Methods) before as well as the Fisher information $J(\theta = 0.5)$ after the re-parameterization. In contrast to the simulations where only the afferent synapses were adjusted, the re-parameterization of the recurrency changed the tuning functions so that now the $J_{add}(\theta = 0.5)$ are above the initial $J(\theta = 0.5)$ (thick dashed vs. thin solid line). The $J(\theta = 0.5)$ for the re-parameterized network are above the $J_{slp}(\theta = 0.5)$ (thick dotted vs. thick solid line), because now both the decreased response magnitudes and the changes of the slopes contributed to increasing the encoding accuracy for $\theta = 0.5$.

Fig. 4.5a compares the shape of the tuning functions averaged over all excitatory neurons before (thin line) and after (thick line) the recurrency was adjusted. In addition to this sharpening, the peak responses were modulated depending on the PS of the neurons (see Fig. 4.5b), and the PS themselves were also changed (see Fig. 4.5c). The underlying synaptic mechanisms for the shifts of the PS are shown in Fig. 4.5d for the two neurons with maximal shifts of their PS towards and away from $\theta = 0.5$ (solid and dotted lines). After the adjustment the neuron which shifted its PS towards $\theta = 0.5$ received more excitation from neurons with PS closer to $\theta = 0.5$ (left arrow). The neuron which shifted its PS away from $\theta = 0.5$ received more excitation from neurons with PS differing even more from $\theta = 0.5$ (right arrow) after the re-parameterization.
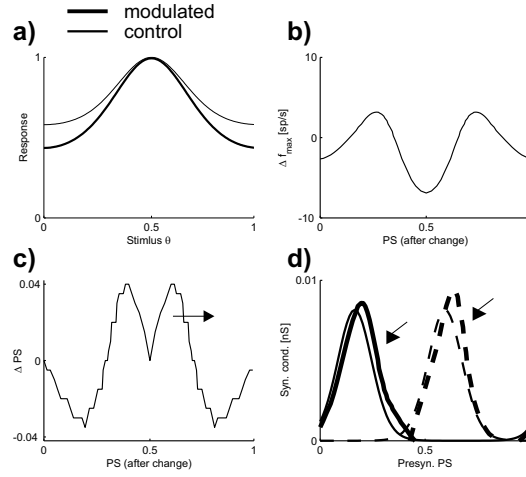
Figure 4.5: Differential changes of tuning functions after having adjusted the recurrent synapses to increase the Fisher information specifically for $\theta = 0.5$. **a)** All tuning functions, normalized to their peak value and aligned so that their POs coincide, were pooled and are shown before (thin line) and after (thick line) the adjustment of the recurrent synapses. **b)** Change of the peak response after the adjustment. Positive values correspond to an increase of the response. **c)** Changes of the PS of the excitatory neurons. Positive values correspond to a shift away from 0.5. **d)** Strength of the recurrent excitation of two excitatory neurons before (thin lines) and after (thick lines) the adjustment. The first neuron (solid lines) had the maximal shift of its PS towards 0.5, the second neuron (dashed lines) had the maximal shift of its PS away from 0.5. Note that after the adjustments to the recurrency the PS also changed. Here we refer with PS to the optimal stimulus and not to the initially "afferent" PS (see Methods) used to determine the afferent input.

Changing the gain of excitatory neurons

Another mechanism we considered is the adjustment of the gain for the excitatory neurons. We first asked how to change the gains a_i^E (see Methods) in order to increase the network's Fisher information specifically for the stimulus $\theta = 0.5$ (see Fig. 4.6a, solid line). One possible realization of this gain modulation is to change the variance of the balanced background inputs (Chance et al., 2002), which could be realized rapidly by, e. g., adjusting top-down feedback inputs. Similar to the case of changing the strength of afferent inputs to the excitatory neurons, the gains were increased mainly for neurons with PS differing approx. ± 0.16 from $\theta = 0.5$. The encoding accuracy is enhanced around $\theta = 0.5$. The changes Δa_i^E necessary to increase the network's Fisher information for all stimuli were again uniform (see Fig. 4.6a, dotted line, and Fig. 4.6b).

In contrast to the previous two mechanisms, the uniform increase of the gains for the objective function $\int J(\theta) \, d\theta$ resulted in a strictly multiplicative modulation of the population response to a stimulus ($\theta = 0.5$, see Fig. 4.6c and inset). If $J(\theta = 0.5)$ is optimized the individual tuning functions and population responses are similar, but not identical, to the changes induced when adjusting the strength of the afferent synapses. The responses for the neuron with PS=0.25 (Fig. 4.6d, dashed lines) increased for all stimuli, but the tuning function for the neuron with PS=0.5 was unaffected (solid lines). The population responses to a stimulus with $\theta = 0.5$ were also bimodal, and the peak activation of the responses to a stimulus with $\theta = 0.25$ was increased and slightly shifted towards $\theta = 0.5$ (see Fig. 4.6e). Fig. 4.6f also parallels the two-fold effects of the changed tuning functions on the encoding accuracy (cf. Fig. 4.3f). For neurons with an already high contribution to the population Fisher information the values of $J_{add}(\theta = 0.5)$ are below the values of $J(\theta = 0.5)$ for the initial parameterization whereas the values of $J_{slp}(\theta = 0.5)$ are higher than the initial values of $J(\theta = 0.5)$. Thus, the increased slopes at $\theta = 0.5$ compensated the reduced Fisher information due to the increase in activity, and encoding accuracy was enhanced for $\theta = 0.5$.
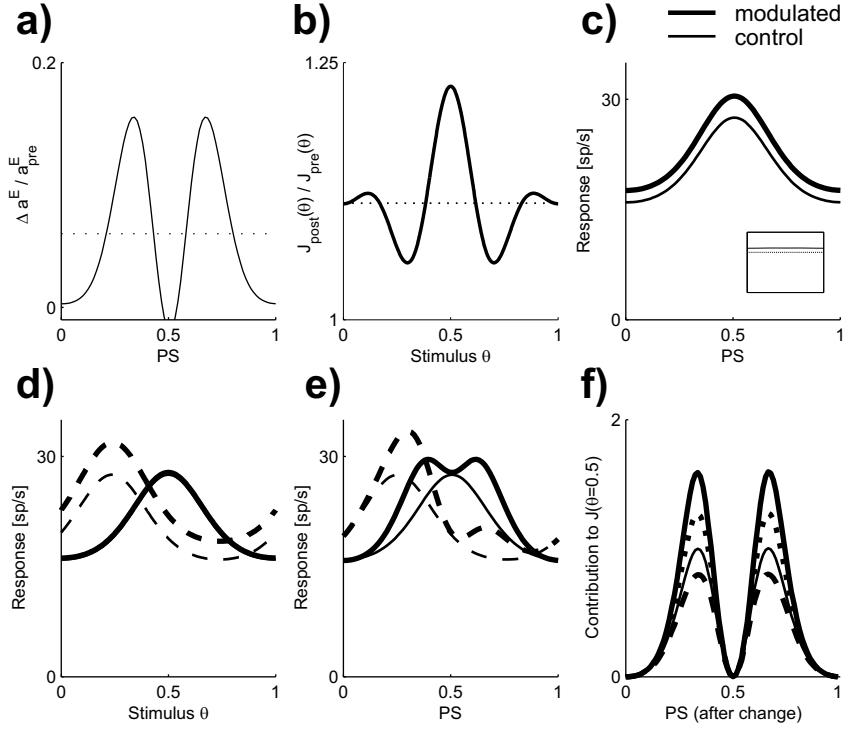
Figure 4.6: Adjusting the gain of the excitatory neurons. **a)** Predicted changes of the gains of the excitatory neurons for optimizing $J\,(\theta = 0.5)$ (solid line) and $\int J\,(\theta)\,d\theta$ (dotted line). The changes were computed by following the gradient of the objective function with a step size $\eta = 5$. **b-f)** Same as b-f in Fig. 4.3.

Changing an additive feedback input

The last mechanism we consider is the adjustment of the additive input currents I_i^{add} for both the excitatory and inhibitory neurons, as given by the gradient of the objective function w. r. t. I_i^{add}. Fig. 4.7a shows how the input to the excitatory (solid lines) and inhibitory (dashed lines) neurons were changed in order to increase the population Fisher information specifically for $\theta = 0.5$ (thick lines) as well as for all stimuli (thin lines), respectively. The additive input currents to all excitatory neurons were decreased, whereas they were increased for inhibitory neurons. For the objective function $J\,(\theta = 0.5)$ the strongest reductions of the inputs to excitatory neurons were for neurons with PS differing approx. ± 0.16 from $\theta = 0.5$. Inhibitory neurons with those PS had the strongest increase of their additive inputs. To increase the Fisher information for all stimuli, the additive inputs had to be changed in a uniform manner (thin lines). These changes resulted in an increase of the encoding accuracy around θ and for all stimuli (Fig. 4.7b).
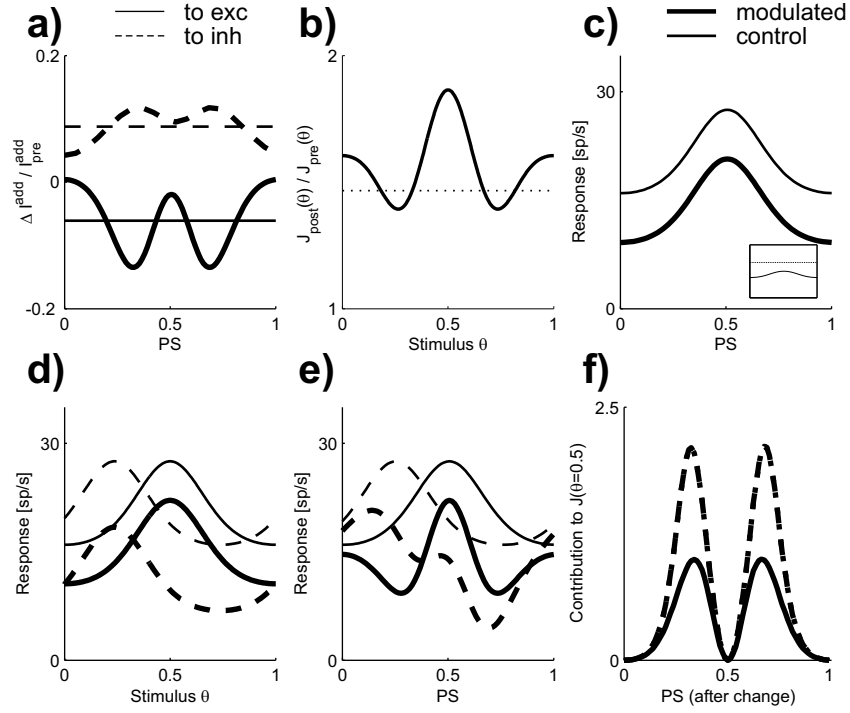
Figure 4.7: Adjusting the additive feedback input. **a)** Predicted changes of the additive feedback input to excitatory and inhibitory neurons (solid and dashed lines) to increase the objective functions $J(\theta = 0.5)$ and $\int(\theta) \, d\theta$ (thick and thin lines). These changes were computed by following the gradient of the objective function with a step size $\eta = 2 \cdot 10^{-4}$. **b-f)** Same as b-f in Fig. 4.3.

Due to the reduced excitation and increased inhibition the population response (and the tuning functions) shown in Fig. 4.7c was reduced. This reduction for the case of the objective function $\int J(\theta) \, d\theta$ was strictly subtractive. For the case of the objective function $J(\theta = 0.5)$ the tuning functions were also shifted towards lower activation levels. These stimulus-specific changes gave rise to the population responses to the two stimuli $\theta = 0.5$ and $\theta = 0.25$ as shown in Fig. 4.7e. Since the reduction of excitation and the increased inhibition were strongest for neurons with PS differing approx. ± 0.16 from $\theta = 0.5$, the strongest reduction of the responses were observed for those stimuli. Changing the values of I_i^{add} led to a change of the value of $J_{add}(\theta)$, but not of $J_{slp}(\theta)$. Fig. 4.7f shows that the $J_{slp}(\theta = 0.5)$ (after the re-parameterization) and the $J(\theta = 0.5)$ (before the re-parameterization) are identical (the thin and thick solid lines are superimposed). In other words, when adjusting only the additive inputs the improvement of the encoding accuracy around $\theta = 0.5$ is only due to the subtractive shifts of the tuning functions.

4.4 Discussion

In this section we first discuss the main finding of this paper: that an increased encoding accuracy for a continuous stimulus variable can be achieved via different mechanisms which then result in different changes of the stimulus tuning functions. Then we relate our predicted changes of neuronal responses to experimentally observed changes during attentional modulations and perceptual learning. These two phenomena happen on distinct time-scales, and they lead to different kinds of tuning function changes. We will show, however, that the observed changes are broadly consistent with the hypotheses that visual attention and perceptual learning can be explained by the common principle of optimally encoding sensory information, and that the differences observed are a result of different plasticity mechanisms being operative. Finally, we discuss the limitations of our modeling approach.

Tuning function changes and adaptation mechanisms

The Fisher information $J(\theta)$ can be increased by an increase of the slopes $f_i'(\theta)$ of the tuning functions and by a decrease of the activities $f_i(\theta)$ for these stimuli, because the contribution of the i-th neuron is proportional to $[f_i'(\theta)]^2 / f_i(\theta)$. Multiple strategies exist to adjust their values. For example, the slope for a particular value of θ could be increased by a multiplicative scaling of the tuning function or by shifting it towards lower or higher stimulus values.

Interestingly, our model predicted that not all mechanisms lead to optimal changes of both the slopes and the activation levels simultaneously, as summarized in Fig. 4.8. In two cases, the physiological constraints (the model architecture) resulted in a "compromise". When changing the gain of the excitatory neurons or the strength of the afferent synapses the increase of activity, which resulted in a decreased encoding accuracy, was compensated by an increase of the slopes (cf. Fig. 4.3f and Fig. 4.6f). However, when adjusting the recurrent connections both increase of the slopes and decrease of the activation levels contributed to the improvement of the Fisher information (Fig. 4.4f). In this case, the changes of the slopes were achieved via shifting the PS, sharpening of the tuning functions, or differentially adjusting the response amplitudes. When the additive feedback inputs were changed, the slopes remained unaffected, but the activations of all neurons were reduced (Figs. 4.7c-f) in order to increase performance.

$J(\theta) \propto$ slope^2/ value	all stimuli	only θ=0.5	
aff. synapses	activity ↑ slopes ↑	activity ↑ slopes ↑	(at θ≠0.5) (at θ≠0.5)
rec. synapses	activity ↓ slopes ↑	activity ↓↑ slopes ↑	
exc. gain	activity ↑ slopes ↑	activity ↑ slopes ↑	(at θ≠0.5) (at θ≠0.5)
add. inputs	activity ↓ slopes –	activity ↓ slopes –	

Figure 4.8: Summary of tuning function changes. The *green* arrows indicate, that the corresponding response change increased the encoding accuracy. The *red* arrows indicate that the accuracy decreased.

Model predictions and perceptual learning

Schoups et al. (2001) investigated the physiological correlate of perceptual learning in V1. They found that after monkeys were trained on an orientation discrimination task, the perceptual improvements were specific for location and orientation. The physiological correlate was an activity reduction for cells with preferred orientations (POs) around the learned orientation (Schoups et al., 1998), and an increase of the slopes of the tuning function at the learned orientation for neurons with POs differing approx. $20°$ from that orientation. When optimizing performance as measured by the objective function $J(\theta)$, our model predicts an activity reduction for neurons with PS close to the learned stimulus when assuming the recurrent connections or the additive ("feedback") inputs as the location of plasticity, but not when adjusting the afferent synapses or the neuronal gains. Furthermore, our model predicts an increase of the slopes of the tuning function for the learned stimulus when the afferent synapses, the recurrent synapses, or the single neuron gains were modified, but not when adjusting the additive inputs.

Another group performed a similar experiment (Ghose et al., 2002) and found that the perceptual improvements were orientation-specific, but transfered between retinotopic locations. Similar to Schoups et al. (2001), Ghose et al. (2002) report an activity reduction for neurons with POs close to the learned orientation. However, no increase of the tuning function slopes was found, but shifts of the POs were reported as predicted by our model when adjusting the recurrent connections. The reason for the discrepancy between the findings of the two groups in not clear.

The same group performed a similar experiment while recording neurons in area V4 (Yang, 2004). They report a sharpening of orientation tuning functions and an orientation-dependent change of the response amplitude with the largest increase of the responses for neurons with POs differing from the learned orientation, but almost no increase for neurons with a PO close to the learned orientation. When adjusting the recurrent connections, our model predicts a stimulus-dependent increase of the response amplitude for neurons with PS differing from the learned stimulus, and an increase of the slopes of the tuning

functions both of which are consistent with the data. However, model results are not fully consistent with the observed lack of change in activity at the learned orientation and better fit with the V1 data of Shoups et al. (1998) and Ghose et al. (2002). Additionally, optimally changing the recurrent connections predicts shifts of the PS. Unfortunately, shifts of the tuning functions can be addressed experimentally only indirectly, e. g., by investigating the histograms of the POs, because the time-scale of perceptual learning is too long for tracking an individual neuron's response properties. The histograms of POs shown by Yang and Maunsell (2004) are not uniform after learning, but as to whether this change is statistically significant needs to be tested.

In summary, the reported physiological correlates of perceptual learning in the visual cortex are by themselves diverse and seem to depend on the visual area. They are, however, mostly (but not completely) consistent with the model predictions, if the recurrent connections are changed in order to improve performance.

Model predictions and attentional modulations

One of the most frequently reported physiological correlates of attention in the visual cortex is an increase of the stimulus-driven neuronal activity compared to control trials. In (Treue and Trujillo, 1999) the effects of spatial and feature-based attention in area MT were disentangled and shown to contribute independently to the observed increase in activity. The effects of attention on the direction tuning curves of neurons in area MT were reported to be approximately multiplicative. Such a separation into a spatial and a feature-based component was also found in area V4 (McAdams, 2000), as well as an approximately multiplicative modulation of the whole orientation tuning function presumably mainly due to spatial attention (McAdams, 1999). In none of these studies a sharpening of stimulus tuning curves was reported.

Our model predicted that, in order to increase the encoding accuracy for all stimuli, strictly multiplicative changes of the tuning functions and the population responses are to be expected only when adjusting the neuronal gain (see inset in Fig. 4.6c). However, an approximately multiplicative change would also be compatible with the changes predicted when adjusting the afferents (see inset in Fig. 4.3c). Of course, adjusting the afferent synapses during attentional modulation via mechanisms like LTP/LTD is out of question, but a possible mechanism could be an effective increase of the impact of feedforward inputs due to synchronous activity in a lower area. Adjusting the recurrent connections or the additive ("feedback") inputs disqualify as possible mechanisms, because they lead to a decrease of neuronal activity.

So far, no study has directly tested for how attention to particular stimulus values, which would correspond to optimizing an objective function like $J(\theta = 0.5)$, changes stimulus tuning functions. The study which comes closest is the study by Treue and Martinez Trujillo (1999). There, the monkey attended to a particular stimulus direction during a direction discrimination task. The authors report an increased activity, when attention is allocated to the presented stimulus. When the afferent synapses or the gains of the excitatory neurons are assumed as the site of plasticity, our model predicts that activity increases, but that this increase is small compared to the strongest changes, which are predicted for neurons with PS differing approx. ± 0.16 from the currently rele-

vant stimulus 0.5 (see dashed lines in Figs. 4.3d, 4.4d, 4.6d, and 4.7d). Those changes, however, were not investigated by Treue and Martinez Trujillo (1999). One reason for the discrepancy could be that in our model the parameters were changed for an increase of the Fisher information for only a single value. It is conceivable that such a very specific modulation is not achievable with the neuronal circuits in the visual cortex or that the range of stimuli actually selected to be represented more accurately is broader. Changes in the recurrent connections and the additive inputs lead to a decrease of neuronal activity and are, therefore, inconsistent with the data.

In summary, one prominent physiological correlate of attentional modulations is an increase in activity for neurons responding to the attended stimulus. Approximately multiplicative modulations of tuning functions are consistent with our model predictions derived from optimizing the objective function $\int J(\theta) \, d\theta$, if the gain of the excitatory neurons or afferent synapses are adjusted in order to improve performance. When optimizing the objective function $J(\theta = 0.5)$, for all the mechanisms we investigated we predict the strongest changes for neurons with PS different from 0.5, which so far has not been tested directly in experiments. However, for this stimulus-specific attention only changing the afferent synapses or the neuronal gains are compatible with an increase in activity for the attended stimulus.

Model limitations

In our contribution we have considered changes in activity only for the mechanisms at the "different sites of plasticity" being operative individually. Since different mechanisms can lead to opposing changes in activity, but still improve performance, it is conceivable that when considering the combined action of multiple mechanisms, some of the discrepancies between model predictions and experimental data, which were mentioned in the last sections, can be resolved.

One limitation of the encoder/decoder framework used here could be the fact, that Fisher information is not always the proper quality measure of a neuronal representation. Bethge et al. (2003), for example, demonstrated that the Fisher information as a quality measure fails if the time window for decoding is very short. For very short decoding time-windows, however, the dynamics of the encoding process must be considered and our model – which was not intended to describe the activity dynamics – is no longer applicable. Xie (2002) showed that an optimal Maximum-likelihood decoding cannot be achieved if only a few neurons are available for representation. However, given that within the visual cortex larger populations of neurons are believed to encode visual information, this may not be a severe limitation (Feldman, 1984). Finally, a neural system may not be able to make optimal use of the information in its activity patterns, and neuronal structures may not be able to implement every conceivable type of optimal decoding. However, it has been shown that recurrent networks can implement Maximum-likelihood decoding procedures (Deneve et al., 1999) for the case of continuous variables and several types of statistical models.

Our modeling framework applies to the case, where the relevant property of the environment is a continuous variable and where its value has to be determined or to be discriminated from another one. Assuming the validity of the optimal encoding hypothesis, the model can then be used to disentangle the mechanisms of perceptual learning, attention, and possibly other adaptation

phenomena in the visual areas. Since the model is a generic cortex model, our predictions may transfer to other continuous stimulus domains or even to the motor cortex, which is also highly adaptive in the adult (Paz and Vaadia, 2004). When discrete stimuli are considered for the perceptual tasks, however, other optimality criteria, for example the classification error for particular classifiers or the mutual information between the stimuli and the neuronal responses, need to be considered.

Recapitulation

We set up a recurrent network model of a visual cortical hypercolumn and asked, how each of four different parameter sets (strength of afferent and recurrent synapses, neuronal gains, and additive background inputs) must be changed in order to optimally improve the encoding accuracy of a particular set of visual stimuli, where encoding accuracy was quantified using the Fisher information. We found that the predicted changes of the population responses and the tuning functions were different for each set of parameters, hence were strongly dependent on the plasticity mechanism which was operative. We compared the predicted tuning function changes with those reported in perceptual learning and attention experiments. Assuming the validity of the optimal encoding hypothesis, such a comparison can be used to infer the mechanisms underlying perceptual learning, attention, and other adaptation phenomena, which would technically be almost impossible to measure directly.

Our results lead to the following conclusions: i) An increased encoding accuracy for a continuous stimulus variable can be achieved via different mechanisms, which then result in different changes of the stimulus tuning functions. ii) Relating the predicted tuning function changes to experimental findings suggests that changing the strength of the recurrent synapses leads to changes in the response properties similar to the those observed in perceptual learning experiments, whereas optimal changes of the neuronal gains leads to changes mimicking neural effects of attention. iii) The two seemingly unrelated phenomena 'perceptual learning' and 'attention' operate on different time-scales and give rise to different changes in the response properties of neurons. They can both be explained by the common principle of 'optimal encoding'. The observed diversity in the changes of the neuronal responses, however, might be a natural consequence, because different routes are taken to calibrate cortical representations according to the same goal.

Chapter 5

Task-dependent sensory representations

Abstract. Here we propose that the effects of attentional top-down modulations observed in the visual cortex reflect the simple strategy of strengthening currently relevant pathways in a task-dependent manner. To exemplify this idea, we set up a network model of a visual area and simulate the learning of a context-dependent 'go/no-go'-task. The model learns top-down gain modulations of sensory representations based on reinforcements received from the environment. We also discuss how this idea relates to alternative interpretations like optimal coding hypotheses.

5.1 Introduction

When animals act in an environment of which they have previously acquired some knowledge, they can select appropriate actions in order to exploit this knowledge. If it turns out that the available knowledge is not sufficient for acting successfully, then an animal may further explore the environment to learn more about it. In both cases the animal perceives its environment via its sensory system which in most approaches to agent learning is assumed to be fixed and (except for sensor noise) reliable. Physiological evidence from the visual system, however, reveal that it is adaptive on a multitude of time-scales. Theoretical studies of this adaptivity are often restricted to the sensory system itself without asking, how changes in the representation affect the initiation of behavioral responses, which relies on a stable representation of an animal's environment.

It has been observed experimentally that sensory representations are adapting to an animal's internal state with attentional top-down modulations being a prime example. Here we hypothesize that these modulations are due to what we call 'architectural constraints'. More concrete, with architectural constraints we mean the 'fixed wiring' between sensory systems and their read-out structures which calls for activation-dependent processes to produce flexible behavior beyond a mere reflexive association of stimuli and behavioral responses.

The Stroop task is a well known example of a context-dependent mapping of visual information onto actions. In this task, subjects are instructed to name

the color of a word for that color (like "green", "red", etc) that is printed in
congruent or incongruent color ink. The naming of the same visual stimulus,
however, depends on a pre-defined context. For example, the word "green"
written with red ink has to be named as "green" in the context "word", whereas
it has to be named as "red" in the context "color". In their neural network
model of the Stroop task (Cohen et al., 1990) proposed that 'attentional units'
selectively strengthen particular processing pathways by disinhibiting read-out
neurons of currently relevant sensory representations, so that they can determine
the behavioral response. Here we propose that a similar principle can serve as
a functional explanation for task- and context-dependent representations in the
visual system itself.

In order to illustrate this idea, we set up a network model of a visual cortical
area, where a context-dependent feedback amplifies selected, currently relevant
stimulus dimensions so that they can dominate the responses of neurons ini-
tiating actions. If the context-dependent initiation of actions is realized by
top-down modulation of the sensory representations themselves, then the initi-
ations of actions can be more rapid compared to the case of having an inter-
mediate processing stage performing an explicit read-out. Moreover, having for
each possible context a separate read-out would also lead to a combinatorial
explosion. We hypothesize that evolution initially may have favored such direct
strategies over a strict separation between a pure sensory representation and
the subsequent action initiation.

5.2   Firing rate model and learning rules

Fig. 5.1a illustrates three types of adaptivity observable in sensory systems.
An animal is perceiving its environment via its sensors. If those are static and
do not adapt at all, learning is restricted to the selection of appropriate actions
depending on the estimated state of the environment. This is the setting usually
considered in the reinforcement learning literature, where the sensory system is
assumed to be fixed. If the sensors adapt without a read-out having explicitly
initiated this change, then these changes may be called autonomous. This is the
paradigm usually considered in most theoretical approaches to sensory adapta-
tion, where changes in the statistics of environmental signals drive the sensory
adaptation (see, e. g., (Fairhall et al., 2001; Wainwright, 1999; Adorjan et al.,
1999)). Here we consider the case where the read-out can control the state of
its sensory system and utilizes reinforcements from the environment in order to
learn how to control it.

We model the learning of a context-dependent 'go/no-go'-task shown in Fig.
5.1b. Here, the association between a simple two-dimensional stimulus and
the behavioral response depends on the context. In context 1, the selection
between the 'go'- and 'no-go'-response has to be made based on the first stimulus
dimension, whereas in context 2 it has to be based on the second dimension. In
our simulated learning scenario the currently active context is known before each
trial. The challenge, however, is to learn how to perform this task successfully
despite the architectural constraint of not having available a separate read-out
and action selection mechanism for each condition.

Independent of the concrete network architecture assumed for this task, two
ways of actually selecting the proper action a for a stimulus x in a given context
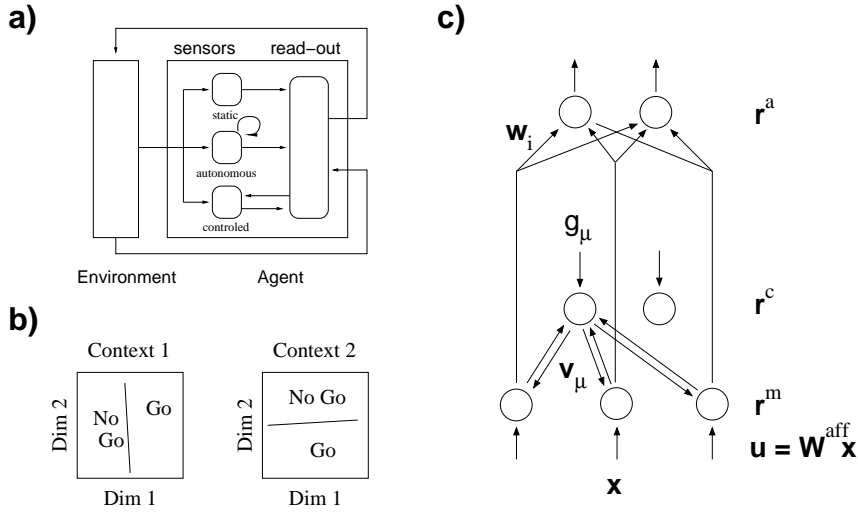
Figure 5.1: Types of adaptivity of sensory systems, task-design and model architecture. **a)** Three types of sensors: *static sensors* do not adapt, *autonomous sensors* adapt in a 'bottom-up'-way to, e. g., the statistics of the stimuli, and *controlled sensors* adapt according to the feedback received from the read-out. **b)** The simulated go/no-go task with the desired association between a two-dimensional stimulus being dependent on the currently active context. **c)** An example for a neuronal circuit, which controls the automatic initiation of actions depending on the feedforward input **u** and the context. The firing rates $\mathbf{r}^a$ of read-out neurons depend directly on the firing rates $\mathbf{r}^m$ of map neurons, which are interconnected by recurrent connections funneled through control neurons. The control neurons are subject to top-down gain modulation of their firing rates $\mathbf{r}^c$. This allows to control the effective recurrent connectivity without re-learning synaptic weights.

c are conceivable. First, a sensory representation $r = f(x)$ of a stimulus x can be computed in a context-independent way, and then the proper action $a = g(r = f(x), c)$ is computed depending on the representation r and the context c. Second, the proper action a can be computed solely based on the representation r, which itself is modulated by the context c, i. e. the action selection is best described as $a = g(r = f(x, c))$.

The linear firing rate model

In order to illustrate the above distinction between the two ways of computing context-dependent action selections, we set up a recurrent network model of a visual cortical area. The activation of neurons in this area constitute the representation of a two-dimensional stimulus x. The representation of a particular stimulus, however, depends on the recurrent connectivity as well. The network is set up such that the gain of a selected set of neurons in the sensory area can be modulated via top-down gain control, which in our model changes the effective recurrent connectivity. Hence, by modulating the gain of selected neurons

in a context-dependent way the computations of sensory representations can be modulated via changing the effective recurrent connectivity.

The model architecture is shown in Fig. 5.1c. The two ideas of modulating the effective recurrent connectivity and using top-down gain modulation to do this combined in our model have been proposed separately before by (Hahnloser et al., 1999) and (Zhang and Abbott, 2000). Each of these ideas is itself speculative, and until now it is not clear as to whether any of the proposed mechanisms is actually realized in the visual cortex. However, we chose this concrete model setup, because it reflects some prominent features of attention in sensory systems like, e. g., the gain modulation of the stimulus-driven response as reported by (Treue and Trujillo, 1999; McAdams, 1999) and the emergence of feature selectivity only during attention as reported by (McAdams, 1999) for some neurons in macaque area V4. The model is deterministic, only the action selection is modeled probabilistically. The role of this randomness is to allow for learning the read-out and the gain control in a reinforcement-based setting using the REINFORCE algorithm developed by (Williams, 1992).

The N neurons in the sensory network (the 'map neurons') receive a feed-forward input $\mathbf{u} = (u_1, u_2, \ldots, u_N)$ given by $\mathbf{u} = \mathbf{W}^{aff}\mathbf{x}$, where $\mathbf{x}$ is the two-dimensional stimulus, and $\mathbf{W}^{aff}$ is a matrix for the afferent weights. If the recurrent connections between these neurons are described by a $N \times N$ weight matrix $\mathbf{W}$ with eigenvectors $\mathbf{v}^{\mu}$ and eigenvalues λ_{μ}, $\mu = 1 \ldots N$, then the dynamics of the firing rates $\mathbf{r}^m$ and the steady-state are given by

$$\tau \frac{d}{dt}\mathbf{r}^m = -\mathbf{r}^m + \mathbf{W}\mathbf{r}^m + \mathbf{u} \qquad \text{and} \qquad r_i^m = \sum_{\mu} \frac{v_i^{\mu}}{1 - \lambda_{\mu}} \langle \mathbf{u}, \mathbf{v}^{\mu} \rangle, \qquad (5.1)$$

where τ is a time-constant, and $\langle \cdot, \cdot \rangle$ is the scalar product. In the proposed architecture, however, only the effective coupling is described by $\mathbf{W}$. The 'physical' connections between two neurons i and j with firing rates r_i^m and r_j^m are disynaptic, because they are funneled through other neurons (the 'control neurons') with their firing rates $\mathbf{r}^c$ being subject to top-down gain control. If the weight of the symmetric connection between the i-th map neuron and the μ-th control neuron is v_i^{μ}, and the gain of the μ-th control neuron is g_{μ}, then the firing rate r_i^m of the i-th map neuron is given by

$$\begin{aligned} r_i^m &= \sum_{\mu} v_i^{\mu} \cdot r_{\mu}^c = \sum_{\mu} v_i^{\mu} \cdot g_{\mu} \sum_j v_j^{\mu} r_j^m = \sum_{\mu,j} v_i^{\mu} g_{\mu} v_j^{\mu} r_j^m \\ &= \sum_j r_j \underbrace{\sum_{\mu} g_{\mu} v_i^{\mu} v_j^{\mu}}_{W_{ij}}, \end{aligned} \qquad (5.2)$$

which leads for weight vectors $\mathbf{v}_{\mu}$ to the effective weight matrix $\mathbf{W} = \sum_{\mu} g_{\mu} \mathbf{v}_{\mu} \mathbf{v}_{\mu}^T$ with eigenvectors $\mathbf{v}_{\mu}$ and eigenvalues g_{μ}.

The important property of this model we utilize for context-dependent action selections has been pointed out by (Zhang and Abbott, 2000), and is shown in Eq. 5.1. The population activity $\mathbf{r}^m$ of the map neurons, which serve as the representation of the stimulus $\mathbf{x}$, is dominated by eigenvectors $\mathbf{v}^{\mu}$ if the λ_{μ} are near to (but smaller as) 1. Thus, controlling the values of the g_{μ}, which correspond to the eigenvalues λ_{μ} of the effective recurrency $\mathbf{W}$, allows to select particular $\mathbf{v}_{\mu}$ to dominate the population activity $\mathbf{r}^m$. A possible robust mechanism for

adjusting the values of the g_μ could be changing the level balanced feedback inputs (Chance et al., 2002). Now, if each dimension of the input stimulus leads to an excitation of one particular $\mathbf{v}_\mu$, then the population activity can be selected to be dominated by one particular stimulus dimension by controlling the values of the g_μ. We will demonstrate that although the read-out neurons are connected with fixed weights to the map neurons, but the gains g_μ are adjusted in a context-dependent way, then the selectivity of the read-out neurons can be changed as needed for solving the context-dependent 'go/no-go'-task.

The final selection of actions is modeled probabilistically. Given the firing rates $\mathbf{r}^m$, the probability to select action a_i is given by

$$P\left(a = a_i \,|\, \mathbf{r}^m\right) = \frac{\exp\left(\beta \mathbf{w}_i^T \mathbf{r}^m\right)}{\sum_j \exp\left(\beta \mathbf{w}_j^T \mathbf{r}^m\right)}, \tag{5.3}$$

where the $\mathbf{w}_i$ are the weights from the map neurons to the neurons associated with the initiation of the i-th action. The parameter β determines the randomness in the action selection with low β corresponding to a more exploratory behavior. We always used $\beta = 2$.

Update rules based on a REINFORCE algorithm

Successfully performing the simulated 'go/no-go'-task means to select the action which, given the context and the sensory stimulus, yields maximal reward. In other words, we seek to adjust the free model parameters (the weights of the read-out and the feedback gain control) so that the expectation

$$E\left[R^c\left(a, \mathbf{x}\right)\right]_{P(a|\mathbf{x},c)P(\mathbf{x},c)}$$

is maximized, where c denotes the context and a is an action selected probabilistically. Optimizing this expectation directly is only tractable for very simple models, but update rules for a stochastic approximation procedure can be derived by using a REINFORCE algorithm (Williams, 1992). Given that in a trial with stimulus $\mathbf{x}$ within context c the action a_s has been selected, and the reward r has been received, then with

$$\frac{\partial}{\partial w_{ij}} \ln P\left(a = a_s \,|\, \mathbf{r}^m\right) = \beta r_j^m \left[\delta_{is} - P\left(a = a_i \,|\, \mathbf{r}^m\right)\right]$$

$$\frac{\partial}{\partial g_\mu^c} \ln P\left(a = a_s \,|\, \mathbf{r}^m\right) = \frac{\beta \langle \mathbf{u}, \mathbf{v}_\mu \rangle}{\left(1 - g_\mu^c\right)^2}$$
$$\cdot \left[\langle \mathbf{v}_\mu, \mathbf{w}_s \rangle - \frac{\sum_j \langle \mathbf{v}_\mu, \mathbf{w}_j \rangle \exp\left(\beta \mathbf{w}_j^T \mathbf{r}^m\right)}{\sum_j \exp\left(\beta \mathbf{w}_j^T \mathbf{r}^m\right)}\right]$$

we obtain the update rules

$$w_{ij} \leftarrow w_{ij} + \eta_w \cdot r \frac{\partial}{\partial w_{ij}} \ln P\left(a = a_s \,|\, \mathbf{r}^m\right) \tag{5.4}$$

$$g_\mu^c \leftarrow g_\mu^c + \eta_\mu \cdot r \frac{\partial}{\partial g_\mu^c} \ln P\left(a = a_s \,|\, \mathbf{r}^m\right), \tag{5.5}$$

where the g_μ^c are the gain factors depending on the condition c, and $\eta_w = 0.1$ and $\eta_\mu = 0.001$ are learning rates. For each simulated training sequence we randomly

selected stimuli with $-1 \leq x_1, x_2 \leq 1$ and a context c with equal probability. For the network with the context-dependent gain-modulations the gain-factors for the corresponding context were used when computing the network responses. Then, Eq. 5.3 was used to select an action. Correct 'go'-responses were rewarded with a reward of 1 for stimuli with $x_1 > 0$ (in context 1) and $x_2 > 0$ (in context 2), respectively. Moreover, for each training sequence we randomly selected orthonormal basis vectors $\mathbf{v}_\mu$ and set $\mathbf{W}^{aff} = \left(\mathbf{W}^T\right)^{-1}$ so that each stimulus dimension excites one particular $\mathbf{v}_\mu$. After each update step we enforced $0.1 \leq g_\mu^c \leq 0.95$ and ensured that $\|\mathbf{w}\| = 1$. The $\mathbf{w}_i$ were initialized randomly, and all gains were set to $g_\mu^c = 0.3$ before each training sequence.

5.3   Simulations of the firing rate model

We now apply the update equations 5.4 and 5.5 for simulating the learning of the context-dependent 'go/no-go'-task. We compare the performance of the network model, which learns context-independent weights $\mathbf{w}_1$ and $\mathbf{w}_2$ (for the 'go' and 'no-go' response) of the action initiation and context-dependent feedback controls g_1^c and g_2^c with two other variants of the model. First, we compare it with a network without context-dependent feedback control (all $g_\mu^c = 0.3$), but a separate action initiation for each context. Second, we compare it with a network without context-dependent feedback control (all $g_\mu^c = 0.3$), but also with only a single action initiation used in both contexts.

Fig. 5.2a shows the probabilities after learning for making a 'go'- and 'no-go'-response as a function of the two-dimensional stimulus in each of the two contexts for the three variants of the network. The probabilities for making 'go'- and 'no-go'-responses learned by the network with context-dependent feedback control ("Gain-modulated") correspond to the context-dependent classification, i. e. a 'go'-response only has a high probability if $x_1 > 0$ in context 1 and $x_2 > 0$ in context 2. Complementary to this, a 'no-go'-response has a high probability if $x_1 < 0$ in context 1 and $x_2 < 0$ in context 2. The variant of the network with a separate read-out for each task ("Separate read-outs") learns similar response probabilities like the network with the context-dependent feedback gain control. However, the response probabilities learned by the network with only a single read-out for both contexts ("Single read-out") neither correspond to context 1 nor context 2. In contrast, they correspond to a mixture between the two contexts, because the best performance in terms of the average reward with a single action initiation is achieved if the context-independent decision for a 'go' or 'no-go' response is made for a separation line, which lies between the two context-dependent ones.

Fig. 5.2b shows how the average reward changes during the learning. If context-dependent feedback can be utilized, then the average reward is highest. This network is even superior to the network with a separate read-out for each context, because close to the mid-lines the finally learned action initiation is less random compared to the network with separate read-outs. If for the latter we increase β after having learned the weights for the read-outs, then these borders become sharper and the average reward approaches the maximum value 1 as well. The boxes in Fig. 5.2b show the learned gain control g_μ^c after one training sequence for the two contexts reflecting that different stimulus dimensions are amplified in order to dominate the action initiation. However, the mechanism of
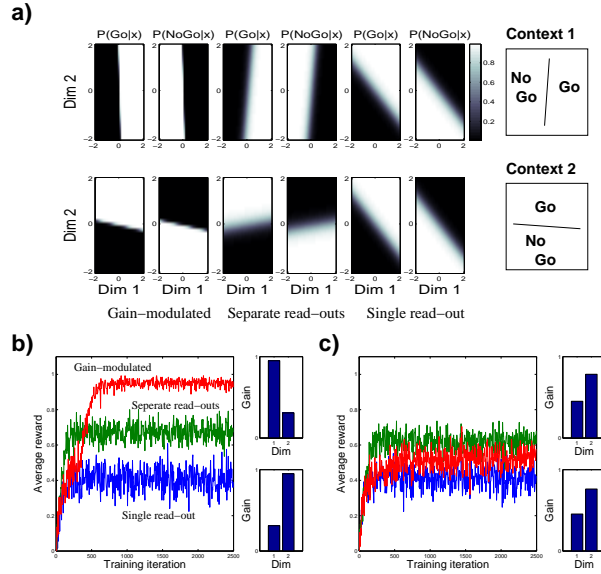
Figure 5.2: Results of model optimization for two-dimensional stimuli and two-dimensional networks. **a)** Probabilities to select a 'go' and 'no-go' response, respectively, depending on the currently active context for the three network variants: a single read-out for both contexts without gain-modulation, separate read-outs for each context without gain-modulation, and a single read-out with gain-modulation. **b)** Average reward during learning with the separation between 'go' and 'no-go' stimuli (see Fig. 5.1b) parallel to the the basis vectors $\mathbf{v}^\mu$. **c)** Same as in b), but with the separation rotated by 45°, i. e. not parallel to the basis vectors $\mathbf{v}^\mu$. The boxes in b) and d) show the learned feedback gain-modulation for context 1 (upper box) and context 2 (lower box). Rewards were averaged over 250 simulations.

amplifying a relevant stimulus dimension so that it can dominate the response of read-out neurons is restricted to cases, where stimuli along these 'relevant' stimulus dimensions excite selected eigenvectors $\mathbf{v}_\mu$ of the sensory network. Fig. 5.2c shows the average reward during learning, where we rotated the separation between the 'go'- and 'no-go'-regions in the input space by 45°. Now, stimuli in the 'go'- and 'no-go'-regions equally excite both eigenvectors $\mathbf{v}_\mu$, so that distinguishing between 'go'- and 'no-go'-stimuli on the basis of a single stimulus dimension is no longer possible. The results for the networks without feedback control do not differ compared to Fig. 5.2b, but the usefulness of the feedback gain control is diminished. The learned feedback control also does not differ between the two contexts (see right boxes).

The mechanism underlying the context-dependent selection of 'go'- and 'no-go'-responses in the network with feedback gain control is further illustrated in Fig. 5.3. In context 1 (upper row) only the response r_1^m of the first map neuron, which changes its firing rate orthogonal to the separation between the 'go'- and 'no-go'-regions in context 1, has a high response amplitude. It dominates the responses of the two read-out neurons, so that the first read-out neuron has high
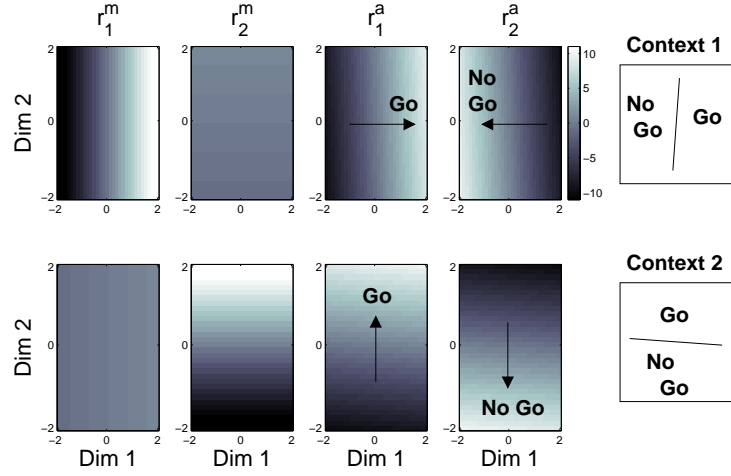
Figure 5.3: Responses of the map and read-out neurons for a two-dimensional stimulus space, where the separation between the 'go' and 'no-go' stimuli is parallel to the basis vectors $\mathbf{v}^\mu$. The responses r_1^m and r_2^m of the two map neurons are independent of the context in terms of their tuning, but their response amplitudes are modulated in a context-dependent way. In contrast, the responses r_1^a and r_2^a of the two read-out neurons change their tuning depending on the context.

response r_1^a for stimuli in 'go'-region of the input space, and the second neuron has a high response for stimuli in the 'no-go'-region. In context 2 (lower row) only the response r_2^m of the second map neuron has a high amplitude. Since the second map neuron changes its firing rate orthogonal to the separation between the 'go'- and 'no-go'-regions in context 2, and it dominates the responses of the two read-out neurons, they are still selective for the correct 'go'- and 'no-go'-responses. The stimulus selectivity of the map neurons does not differ between the two contexts, only their response amplitude is modulated. In contrast, the selectivity of the read-out neurons w. r. t. the two-dimensional input stimuli changes, but it remains invariant in terms of the context-dependent 'go'- and 'no-go'-responses.

Certainly, the situation that both the afferent and the recurrent weights are set such that feedback gain control can lead to an amplification of currently relevant stimulus dimensions is highly idealized and speculative. Therefore, we also explored as to whether the idea of amplifying relevant stimulus dimensions carries over to a more realistic scenario. We considered the case, where a two-dimensional stimulus is embedded into a high-dimensional space. Here, the afferent weights are not matched to the recurrent weights in the sense that an afferent stimulus excites selected eigenvectors (see the definition of $\mathbf{W}^{aff}$ in Sec. 5.2). Now, each afferent input u_i is computed as the value a corresponding Gaussian basis function yields for a particular two-dimensional stimulus. The basis function centers were equally spaced between -1 and 1 on each axis, and amplitudes and widths were set to 1 and $\sigma = 0.3$ corresponding to an idealized afferent input in a two-dimensional retinotopic map. The separation between
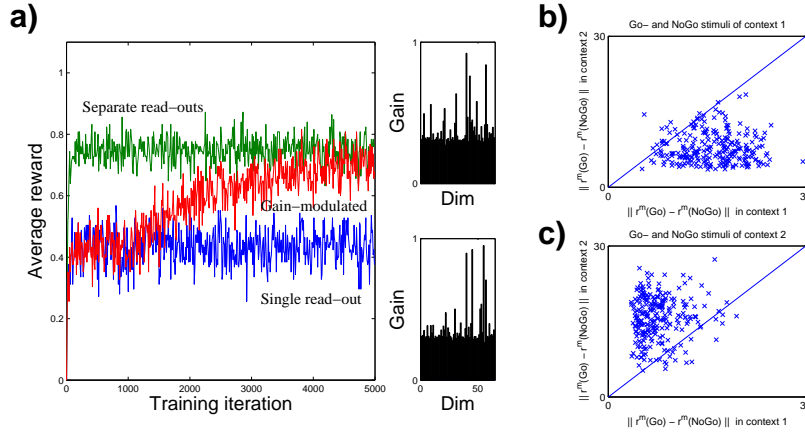
Figure 5.4: Results of model optimization for an embedding of the two-dimensional stimulus space into a 64-dimensional space. **a)** The gain-modulation (right boxes) is learned in a way which amplifies the 'relevant' directions in the 64-dimensional representation space. **b,c)** Distances between the population vector representations of the two prototypical 'go' and 'no-go' stimuli of context 1 and context 2 during the learned gain-modulation for context 1 and context 2.

the 'go'- and 'no-go'-regions was rotated by 45° . Again, we randomly selected a set of now $N = 64$ orthonormal eigenvectors $\mathbf{v}^\mu$ and simulated the learning of the three network variants. Since it is not known beforehand, which of the 64 randomly selected eigenvectors $\mathbf{v}^\mu$ will be the relevant ones, we learned the gain factors g_μ^c for all 64 dimensions.

Fig. 5.4a shows the average reward during learning as well as an example of the finally learned gain modulations for each context (right boxes). Similar to the two-dimensional case (cf. Fig 5.2b) the performance of the network with the context-dependent gain modulation is higher than the one with a single read-out and no gain modulation. However, is this improvement indeed achieved via selectively amplifying the 'relevant' eigenvectors $\mathbf{v}^\mu$? We tested this by investigating the distance between the population vector representations of the prototypical 'go'- and 'no-go' stimuli, i. e. the mean stimulus of each class, of each context.

Fig. 5.4b shows that the distance between the population vector representations of the prototypical 'go'- and 'no-go'-stimuli of context 1 is larger than the distance between the population vector representations of the prototypical 'go'- and 'no-go'-stimuli of context 2 when the gain modulation learned for the first context is used. Complementary to this, Fig. 5.4c shows that the distance between the two representations of the 'go'- and 'no-go'-stimuli of context 2 is larger when the learned gain modulation for the second context is used. In other words, the gain modulations were always learned such that the distance between the representations of the currently relevant prototypical stimuli was increased. If the learned selective amplifications were not along the 'relevant' directions, which are those connecting the centers of the two stimulus classes, then we would have obtained a change in the distances independent of the context. In

other words, the feedback gain-modulation increased the distance between the
population vector representations of the currently relevant prototypical 'go' and
'no-go' stimuli. Also note, that we have not optimized the embedding of the
low-dimensional stimulus, but it is conceivable that biological systems may have
learned those embeddings which are suitable for a number of these classification
tasks.

5.4   Discussion

In summary, we proposed as an explanation of task-dependent representations
in the visual system the selective strengthening of sensory pathways so that cur-
rently relevant stimulus properties can dominate the responses of downstream
neurons. In order to exemplify this idea, we set up a network model and learned
a context-dependent top-down gain control using reinforcements received from
the environment.

Relation to optimal coding hypotheses

Our explanation of task-dependent representations via cortical feedback is mainly
'mechanistic'. This rather simple-minded approach, however, has the distinct
advantage of making explicit the problem for the need for a co-adapting read-
out of adaptive representations in the sensory systems. If representations in
sensory systems are adapting to either external or internal conditions like, e. g.,
changing statistics of environmental signals or changing task demands, then the
corresponding read-out structures have to adapt as well.

As long as the statistics of environmental signals change on a slow time-
scale, it is conceivable that the corresponding read-out structures can track
these changes. In this case, the changes of the sensory representations are in-
duced in a 'bottom-up' manner (cf. adaptation of "autonomous sensors" in Fig.
5.1a). However, if rapidly changing task demands call for an adaptation of sen-
sory representations, then they could only be induced in a 'top-down' manner.
The prominent optimal coding hypothesis (Atick, 1992) would explain these
task-dependent changes as a reallocation of limited representational resources
in order to ensure a high fidelity representation of the currently relevant stim-
ulus aspects. As a consequence of such an altered neuronal code, however, the
read-out structures would have to adapt as well. In our model the need for task-
dependent representations derives from the need to adjust them so that they
become useful for the very same read-out structures without assuming limited
representational resources.

Sensory representations as probabilities

Another possible approach to explain task-dependent representations in sensory
systems is to interpret them as the signatures of probabilistic computations per-
formed on previously learned statistical models of the sensory data. Although
the involved models of the sensory data are usually probabilistic, learning and
inference are in principle deterministic. This makes them attractive as a pure
computational approach, which is not spoiled by the additional problems in-
duced when computations are done with unreliable 'noisy' neurons leading to

limited representational resources. For example, in (Rao and Ballard, 1999) the feedback within the visual cortex can be viewed as part of a hierarchical model of the sensory data which explains the afferent input by predicting it. A related interpretation is to view sensory systems as performing Bayesian inference, where the activations represent probabilities (see (Pouget et al., 2003; Barlow, 2001) for an introduction and (Rao, 2004; Deneve, 2005; Sahani and Dayan, 2003) for particular approaches). A possibility to explain task-dependence within this framework could be to assume that different tasks correspond to different computations with the represented probabilities. For example, the operation of 'integrating out' currently irrelevant variables could be reflected in the neuronal activity (Sahani, personal communication).

The basic assumption of these probabilistic interpretations, however, can easily be tested experimentally. By varying independently the qualities 'probability' and 'value' of stimuli in different contexts, one can examine as to whether responses in the visual cortex are related only to the subjectively estimated probabilities of stimuli, or also to their subjective value. If the latter is the case, then this interpretation disqualifies as the sole interpretations of activations in sensory cortices, whereas our interpretation would still be applicable.

Context-dependent sensory-motor transformations

Our work is certainly not the first to address the problem of how to transform sensory information in a context-dependent way into motor outputs. For example, in (Cohen et al., 1990; Gilbert and Shallice, 2002) this issue has been addressed within a 'parallel distributed processing model', and recently circuit models for context-dependent sensory-motor transformations have been suggested (Salinas, 2004b; Salinas, 2004a). Our approach differs from these previous ones in two ways. First, we used a different model architecture, where an effective context-dependent recurrent connectivity was modulated by top-down gain control. Second, we motivated our approach by comparing it with alternative interpretations for adaptivity in the visual system. We proposed that top-down modulations of sensory representations are best understood as the signature of a transformation of sensory signals into currently appropriate behavioral responses. In other words, we hypothesize that parts of those (often context-dependent) transformations are already realized within the visual cortex an not exclusively in downstream frontal areas.

In this paper we suggested that the functional role of *feedback into a sensory system*, for example from prefrontal regions into the ventral visual pathway, is to 'tune' the sensory processing in a context-dependent way. Of course, for each possible context the proper top-down modulations have to be memorized, and storing those comes with a cost. It remains to be determined as to whether the particular top-down modulation scheme we suggest is robust in the sense that storing only a few gain-factors is sufficient, because otherwise the controlling network would be more complex than the controlled sensory network. However, our results indicate that when sensory representations are embedded into a high-dimensional representation space, selectively amplifying only a few directions is indeed sufficient to significantly improve the performance. Finally, it remains to be determined as to whether our functional interpretation of feedback signals carries over to *feedback within the visual system*, for example between the striate and extra-striate cortex.

Recapitulation

We proposed as an explanation of task-dependent representations in the visual system the selective strengthening of sensory pathways so that currently relevant stimulus properties can dominate the responses of downstream neurons. In order to exemplify this idea, we set up a network model and learned a context-dependent top-down gain control using reinforcements received from the environment. We also discussed how our interpretation of task-dependent representations relates to alternative views.

Compared to the other studies of this thesis, which were related in a close way to particular experimental findings, the work presented in this chapter is more speculative. However, although we focused on functional "interpretations" of neuronal activity in the visual cortex, we suggested a new line of experiments, which could distinguish between our interpretation and other alternatives like, for example, the view of visual cortex as being a Bayesian inference engine (see Sec. 1.1). In particular, we suggested to selectively manipulate the two variables "probability" and "value" in experimental settings. Thus, future experimental studies could determine, which interpretation is most appropriate.

Appendix A

Hodgkin-Huxley and synapse model

A.1 Single cell model

The single cell model is the Hodgkin-Huxley-type neuron given in (Destexhe et al., 2001). The dynamics of the membrane potential of a neuron i is described by

$$C_i^m \frac{dV_i}{dt} = -g_i^L \left(V_i - E_i^L \right) - \sum_{\text{int}} I_{\text{int}} - I_i,$$

where the I_{int} denote the intrinsic voltage-dependent currents, g_i^L and E_i^L are the leak conductances ($g_i^L = g_E^L = 22.74\,\text{nS}$ for excitatory and $g_i^L = g_I^L = 2 \cdot g_E^L$ for inhibitory neurons) and reversal potentials ($E_i^L = E_E^L = E_I^L = -80\,\text{mV}$), C_i^m denotes the membrane capacitance ($C_i^m = C_E^m = C_I^m = 0.5\,\text{nF}$), and t is the time. Each current I_{int} is described by a Hodgkin-Huxley equation

$$I_{\text{int}} \left(t \right) = \bar{g} \cdot m^M \left(t \right) \cdot h^N \left(t \right) \cdot \left(V \left(t \right) - E \right),$$

where $\bar{g}$ is the peak conductance, E is the reversal potential, and $m \left(t \right)$ and $h \left(t \right)$ are the activation and inactivation variables. Three voltage dependent currents are included: a fast Na$^+$ current and a delayed-rectifier K$^+$ current for the generation of action potentials, and a slow non-inactivating K$^+$ current responsible for spike frequency adaptation.

A.2 Ionic currents

For the Na$^+$ current we used

$$I_{Na} = \bar{g}_{Na} m^3 h \left(V - E_{Na} \right)$$

with

$$\frac{dm}{dt} = \alpha_m\left(V\right)\left(1 - m\right) - \beta_m\left(V\right)m$$

$$\frac{dh}{dt} = \alpha_h\left(V\right)\left(1 - h\right) - \beta_h\left(V\right)h$$

$$\alpha_m\left(V\right) := \frac{-0.32\left(V - V_T - 13\right)}{\exp\left(-\left(V - V_T - 13\right)/4\right) - 1}$$

$$\beta_m\left(V\right) := \frac{0.28\left(V - V_T - 40\right)}{\exp\left(\left(V - V_T - 40\right)/5\right) - 1}$$

$$\alpha_h\left(V\right) := 0.128\exp\left(-\left(V - V_T - V_S - 17\right)/18\right)$$

$$\beta_h\left(V\right) := \frac{4}{1 + \exp\left(-\left(V - V_T - V_S - 40\right)/5\right)}$$

with parameter values $V_T = -58\,\mathrm{mV}$, $V_S = -10\,\mathrm{mV}$, and $\bar{g}_{Na} = 17.87\,\mu\mathrm{S}$ and $E_{\mathrm{Na}} = 50\,\mathrm{mV}$.

For the "delayed-rectifier" K$^+$ current we used

$$I_{Kd} = \bar{g}_{Kd}n^4\left(V - E_K\right)$$

with

$$\frac{dn}{dt} = \alpha_n\left(V\right)\left(1 - n\right) - \beta_n\left(V\right)n$$

$$\alpha_n\left(V\right) := \frac{-0.032\left(V - V_T - 15\right)}{\exp\left(-\left(V - V_T - 15\right)/5\right) - 1}$$

$$\beta_n\left(V\right) := 0.5\exp\left(-\left(V - V_T - 10\right)/40\right)$$

with parameter values $E_K = -90\,\mathrm{mV}$, $\bar{g}_{Kd} = 3.46\,\mu\mathrm{S}$.

For the non-inactivating K$^+$ current we used

$$I_M = \bar{g}_M n\left(V - E_K\right)$$

with

$$\frac{dn}{dt} = \alpha_n\left(V\right)\left(1 - n\right) - \beta_n\left(V\right)n$$

$$\alpha_n\left(V\right) := \frac{0.0001\left(V + 30\right)}{1 - \exp\left(-\left(V + 30\right)/9\right)}$$

$$\beta_n\left(V\right) := \frac{-0.0001\left(V + 30\right)}{1 - \exp\left(\left(V + 30\right)/9\right)}$$

with $\bar{g}_M = 0.28\,\mu\mathrm{S}$ for excitatory and $\bar{g}_M = 0.028\,\mu\mathrm{S}$ for inhibitory neurons.

A.3   Noise model

The model neurons additionally receive balanced excitatory and inhibitory synaptic background inputs. The corresponding conductances are described by a

stochastic process similar to an Ornstein-Uhlenbeck process with the following update rule (Destexhe et al., 2001):

$$g_{bg}\left(t + \Delta t\right) = g_{bg}^0 + \left[g_{bg}\left(t\right) - g_{bg}^0\right] \exp\left(-\Delta t/\tau\right) + A \cdot N\left(0, 1\right),$$

where g_{bg}^0 is the average conductance, τ is a synaptic time constant, A is the amplitude coefficient, and $N\left(0, 1\right)$ is a normally distributed random number with zero mean and unit standard deviation. The amplitude coefficient has the following analytic expression

$$A = \sqrt{\frac{D\tau}{2}\left[1 - \exp\left(-2\frac{\Delta t}{\tau}\right)\right]},$$

where $D = 2\sigma^2\tau^{-1}$ is the diffusion coefficient. Numerical values for the background conductances are $\sigma = 3\,\text{nS}$ and $\sigma = 6.6\,\text{nS}$ for the variances of the excitatory and inhibitory conductance, $\tau = 2.7\,\text{ms}$ for the excitatory and $\tau = 10.5\,\text{ms}$ for the inhibitory time constant, and $g_{bg}^0 = 12.1\,\text{nS}$ for the mean excitatory and $g_{bg}^0 = 57.3\,\text{nS}$ for the mean inhibitory conductance. The reversal potentials were $0\,\text{mV}$ and $-75\,\text{mV}$ for the excitatory and inhibitory conductances.

A.4 Firing rate functions

We simulated the spike responses of this model neuron to current injections for different values of the leak conductance in the presence of the fluctuating background conductance. We considered only the already adapted responses, which were then best fitted with thresholded polynomials. These fits were done by minimizing the mean squared error between the simulated firing rate and the one predicted by the thresholded polynomials. We obtained

$$
\begin{aligned}
F_i^E\left(I\right) &= \max\left(0, a_i^E\left(I - \Delta I\left(g_L\right)\right)\right) \\
F_i^I\left(I\right) &= \max\left(0, a^I\left(I - \Delta I\left(g_L\right)\right) + b^I\left(I - \Delta I\left(g_L\right)\right)^2\right)
\end{aligned}
$$

with $a_i^E = a^E = 71.9$, $I_E^c = 0.13\,\text{nA}$, and $V_E^c = 15.2$ for excitatory and $a^I = 133$, $b^I = -28$, $I_I^c = -0.02\,\text{nA}$, $V_I^c = 14.6\,\text{mV}$ for inhibitory neurons. These firing rate functions are then used in the mean-field network models in the Chapters 2, 3, and 4 (see Appendix B.2). The parameters V_I^c and I_E^c are the parameters of the linear function ΔI, which determines the shift of the firing rate function depending as a function of the leak conductance (see Appendix B.2 for a definition of ΔI).

A.5 Synapse model

A synaptic current $I_{syn}\left(t\right)$ due to activity of AMPA- and GABA$_\text{A}$-synapses was computed as

$$I_{syn}\left(t\right) = \sum_j \bar{g}_j g_j\left(t\right)\left(V\left(t\right) - E_j\right),$$

where g_j and E_j are the time-dependent conductance for the j-th synapse and the corresponding reversal potential, and $\bar{g}_j$ scales the strength of the synaptic

connection. We always used $E_j = E_E = 0\,\mathrm{mV}$ for the excitatory AMPA- and $E_j = E_I = -80\,\mathrm{mV}$ for the inhibitory GABA$_\mathrm{A}$-synapses. The dynamics of the conductance variable g_j is given by

$$\frac{d}{dt}g_j\left(t\right) = -\frac{g_j\left(t\right)}{\tau_j} + \sum_k \delta\left(t - t_j^k\right),$$

where τ_j is the time-constant of the j-th synapse, and t_j^k denotes the time of the k-th spike of neuron j. Note that only in Section 2.3 we explicitly simulated individual synapses. However, this synapse model is also the basis for the mean-field network model we used in the Chapters 2, 3, and 4.

Appendix B

The V1 map model

B.1 The orientation selectivity index (OSI)

The orientation selectivity index (OSI) for a set of N values $\{x_i\}_{i=1}^{N}$ with corresponding orientations θ_i is given by:

$$OSI\left(\{x_i\}_{i=1}^{N}\right) = \frac{\sqrt{\left(\sum_i x_i \cos\left(2\theta_i\right)\right)^2 + \left(\sum_i x_i \sin\left(2\theta_i\right)\right)^2}}{\sum_i x_i} \tag{B.1}$$

In Chapter 2 we use this index to quantify the degree of tuning for the conductances, the membrane potential, and the spike responses. High values of the OSI correspond to well tuned responses, whereas low values correspond to weakly tuned responses.

We also use the OSI as a graded measure for the location in an orientation map. Considering a pixel in an orientation map, the preferred orientations of all pixels in a circular region around this pixel can be used to compile a histogram of preferred orientations. The OSI of this histogram quantifies the map location. Small values correspond to locations close to a pinwheel, whereas large values correspond to orientation domains.

B.2 Firing rate model of coupled neurons

The network models we use in the Chapters 2, 3, and 4 are based on the single cell model described in Appendix A. They are simplified rate models for the dynamics of neural populations, which can be derived from a more detailed biophysical model, as described elsewhere (Shriki et al., 2003).

However, let us briefly summarize the main ideas developed by Shriki et al. (2003): Consider a large network of synaptically coupled neurons which is in an asynchronous state, that is, there is a low probability that neurons fire synchronously. Let $r_i = F_i\left(I_i\right)$ denote the firing-rate response of neuron i to a total input current I_i, where r_i is a state-variable denoting the firing rate of neuron i, and F_i is the current-frequency function of neuron i. It can be shown that asynchronous synaptic inputs generate a stationary additive contribution to applied currents, and increase the passive leak conductance of the postsynaptic cell. In many cortical cells, increasing the leak conductance does not alter the

current-frequency function, rather it increases the threshold current. Under these conditions, the steady-state firing rate takes the form (Shriki et al., 2003)

$$r_i = F_i \left(I_i^{aff} + I_i^{bg} + \sum_j W_{ij} r_j - \Delta I_i \left(g_i^L \right) \right)$$

for all i where $W_{ij} = G_{ij} \left(E_j - E_i^L - V_i^c \right)$ is the "synaptic weight" of the connection between postsynaptic neuron i and presynaptic neuron j with associated synaptic conductance G_{ij}. In addition, I_i^{aff} is an afferent input current, I_i^{bg} is an additional additive background input, E_j is the reversal potential for the synaptic connections of neuron j (see Appendix A.5), and g_i^L and E_i^L are the leak conductance and reversal potential of neuron i, respectively. The shift of the firing rate function is given by $\Delta I_i \left(g_L \right) = I_i^c + V_i^c g_L$, where the voltage V_i^c determines the rate of increase in the shift with the increase of the leak conductance.

Having determined the steady-state firing-rates (see Appendix A.4), one can incorporate the effects of synaptic dynamics by considering the first-order rate equation

$$\tau_j \frac{d}{dt} r_j \left(t \right) = -r_j \left(t \right) + F_j \left(I_j \left(t \right) \right)$$

where τ_j is an effective synaptic time constant.

B.3   Model architecture

Here we describe the specificities of the map model. The parameters are given in Tab. B.1.

Spatial layout

For the map model we used a *two-dimensional grid* with $N = 32 \cdot 32 = 1024$ locations spaced equally between -1 and 1 on both x- and y-axis. With each two-dimensional location $\mathbf{x}_i$ a preferred orientation was associated as given by the *artificial orientation map* shown in Fig. 2.8a.

Recurrent connectivity

The maximal conductance $G_{ij}^{\alpha\beta}$ for the connection between the presynaptic neuron j of type $\beta \in \{E, I\}$ and the postsynaptic neuron i of type $\alpha \in \{E, I\}$ was given by

$$G_{ij}^{\alpha\beta} = \bar{G}_{ij}^{\alpha\beta} \cdot \frac{1}{Z_i} \exp \left(-\frac{d \left(\mathbf{x}_i, \mathbf{x}_j \right)^2}{2\sigma_\beta^2} \right),$$

where σ_β determines the spatial scale of the connections made by neurons of type β, and Z is a normalization constant such that $\sum_{j=1}^{N} G_{ij}^{\alpha\beta} = \bar{G}_{ij}^{\alpha\beta}$ for all neurons i. The distance $d \left(\mathbf{x}_i, \mathbf{x}_j \right)$ between the locations of neuron i and neuron j is defined to implement *circular boundary conditions*.

Parameter	Description	Value
	Network architecture	
$N_E = N_I$	Number of exc. and inh. neurons	$32 \cdot 32 = 1024$
	Connectivity	
$\sigma_E = \sigma_I$	Spatial scale of local connections	0.1
M	Scaling of afferent input	$2000\,\mathrm{sp/s}$
κ	Tuning of afferent input	1.7
	Strengths	
G_E^{aff}	Strength of afferent input to all exc. neurons	$0.0044\,\mathrm{nS}$
G_I^{aff}	Strength of afferent input to all inh. neurons	$0\,\mathrm{nS}$
G_{EE}	Exc. of exc. neurons	$0..0.43\,\mathrm{nS}$
G_{EI}	Inh. of exc. neurons	$0..1.3\,\mathrm{nS}$
G_{IE}	Exc. of inh. neurons	$0.15\,\mathrm{nS}$
G_{II}	Inh. of inh. neurons	$0\,\mathrm{nS}$
	Background inputs	
I_E^{bg}	Background input to exc. neurons	$0.38\,\mathrm{nA}$
I_I^{bg}	Background input to inh. neurons	$0.74\,\mathrm{nA}$

Table B.1: Parameters specific for the map model.

Afferent input

In the model, only the excitatory neurons receive an afferent input. The input was a function only of the circular distance $\Delta\theta$ between the preferred orientation of the postsynaptic neuron and the stimulus orientation θ, but not of the location $\mathbf{x}_i$ of neuron i. When a visual stimulus of orientation θ is presented to the network, it leads to the afferent input current

$$I_i^{aff}(\theta) = M \cdot W_i^{aff}\left[0.1 + 0.9 \cdot \exp\left(\kappa\left[\cos\left(\frac{\pi}{90} \cdot d_i(\theta)\right) - 1\right]\right)\right]$$

with

$$W_i^{aff} = G_i^{aff}\left(E_E - E_i^L - V_i^c\right)$$
$$d_i(\theta) = \min\left(|\theta - \theta_i|, \, 180 - |\theta - \theta_i|\right).$$

B.4 Regimes of the map model

Kang et al. (2003) distinguished in their analysis between four different operating regimes, which were defined in terms of the effective recurrency, and in (Wiesing et al., 2005) we adopted these definitions. For technical reasons, however, their definitions are not applicable in our case. In particular, the model used by Kang et al. (2003) assumes i) linear firing rate functions and ii) all neurons above threshold. In our model, the firing rate function of the inhibitory neurons is non-linear. Moreover, due to the thresholding, which in our model accounts for the location-independent spike tuning, not all excitatory neurons are above threshold.

Therefore, we use the following definitions: If the sum of the input currents received (by a neuron in an orientation domain) via the local recurrent excitatory and inhibitory connections has a positive maximum and a negative minimum, we refer to those models as models in the *Hat*-regime. If both maximum and minimum are non-negative, the local recurrency of the corresponding models is dominated by excitation (*Exc*-regime), and if both maximum and minimum are negative, the recurrency is dominated by inhibition (*Inh*-regime). Those are the regimes indicated in Fig. 2.9. Finally, with *Aff* we refer to the single model parameterization without any local recurrency ($G_{EE} = G_{EI} = 0$), which corresponds to the lower left pixel in Fig. 2.9a-d.

The definitions we use are not intended to correspond one-to-one to particular properties of the corresponding model networks. They are intended only as a simple classification of the parameter space we explore. However, one could also ask, as to whether the classification we use is specific to neurons in the orientation domain. Therefore, we applied the same definition to neurons close to a pinwheel. In this case, the classification borders indeed change, but not completely (Fig. B.1a).

For pinwheel neurons, the *Hat*-regime becomes smaller while the *Exc*-regime becomes larger. In particular, some parameterizations classified as belonging to the *Hat*-regime using the original definition are now labeled as belonging to the *Exc*-regime. One property, however, seems to correlate well with our definition of the *Hat*-regime defined either for an orientation domain or pinwheel location. The highest values of the co-variation between excitation and inhibition are found in the *Hat*-regime (see Fig. B.1b).
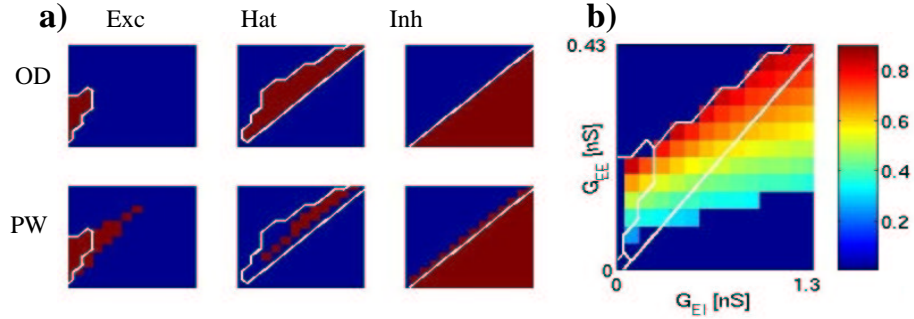
Figure B.1: Regimes of the map model. **a)** Comparison of applying the definition for the regimes (see text) to orientation domain and pinwheel neurons. The *white lines* in the top and bottom row indicate the classification as obtained for an orientation domain neuron. The *red regions* mark the corresponding regime. **b)** The co-variation of excitation and inhibition as quantified by the ration of the slopes for g_e to g_i. A value of 1 would correspond to a "perfect" co-variation.

Appendix C

The inter-areal network model

We describe here the construction of our cortical network model and how the parameters of the model are fitted to anatomical and physiological data. The values of all model parameters are summarized in Tab. C.1.

Parameter	Description	Value
	Network architecture	
$N_E^S = N_I^S$	No. of exc. and inh. neurons in striate cortex	161
N_E^X	No. of exc. neurons in extrastriate cortex	33
	Size of simulated visual field representation	$16\,\mathrm{deg}$
	Connectivity	
λ^{SS}	Scale of lateral connections	$2.3\,deg^{-1}$
$\lambda^{XS} = \lambda^{SX}$	Scale of inter-areal connections	$0.3\,deg^{-1}$
$v_{S\leftrightarrow S}$	Conduction velocity of lateral connections	$200\,\mathrm{mm/s}$
$v_{S\leftrightarrow X}$	Conduction velocity of inter-areal connections	$4000\,\mathrm{mm/s}$
$x_{S\leftrightarrow X}$	Distance between V1 and extrastriate area MT	$7\,\mathrm{mm}$
σ_{aff}	Spatial spread of aff. input	$0.1\,\mathrm{deg}$
	Strengths	
G_{EE}^{SS}	Lateral excitation of exc. neurons	$0.005\,\mathrm{nS}$
G_{IE}^{SS}	Lateral excitation of inh. neurons	$0.053\,\mathrm{nS}$
$G_{EE}^{XS} = G_{EE}^{SX}$	Inter-areal excitation	$0.007\,\mathrm{nS}$
G_{EE}^{S}	Local excitation of excitatory neurons	$0.13\,\mathrm{nS}$
G_{IE}^{S}	Local excitation of inhibitory neurons	$0.05\,\mathrm{nS}$
G_{EI}^{S}	Local inhibition of exc. neurons	$0.79\,\mathrm{nS}$
G_{II}^{S}	Local inhibition of inh. neurons	$0.08\,\mathrm{nS}$

Table C.1: Parameter values specific for the inter-areal model.

C.1   Cortical network model

In our recurrent network model we consider two areas of an idealized cortex with each area being composed of a single layer of cells. The neurons in the model

cortex are identified with the visual field position of their receptive field (RF) centers, and the patterns of the intra-areal lateral (or horizontal) and inter-areal feedback (FB) connections are defined as translation-invariant in visual field coordinates.

For simplicity, we do not model stimulus features such as orientation and spatial frequency, since we focus on size-tuning and contrast effects in this paper. Thus, we model experimental paradigms in which two-dimensional gratings are presented at the optimal spatial frequency and orientation of the center target neuron. In real life, such stimuli activate discrete patches across cortex corresponding to those cortical columns that have sufficiently similar orientations and spatial frequencies to that of the stimulus. We idealize this patchwork structure as a one-dimensional network of cells along the collinear axis of the preferred orientation, in agreement with recent data showing that horizontal (Bosking et al., 1997; Sincich and Blasdel, 2001) and FB (Angelucci and Bullier, 2003; Shmuel et al., 2005) connections in V1 link regions of similar orientation preference along an axis collinear with the preferred orientation of their neurons of origin.

Let $r_l^\alpha(x_i, t)$ denote the local activity of excitatory ($l = E$) and inhibitory ($l = I$) neurons with their RF center at in striate ($\alpha = S$) and extrastriate ($\alpha = X$) cortex. The RF centers were equally spaced between $-8°$ and $+8°$. For the striate and extrastriate area we used 161 and 33 model neurons, respectively. The activity r_l^α evolves according to the rate equation (see also Appendix B.2)

$$\tau_l \frac{dr^\alpha(x_i, t)}{dt} = -r_l^\alpha(x_i, t) + F_l^\alpha(I_l^\alpha(x_i, t))$$

where $F_l^\alpha(I)$ is the current-frequency function as given in Appendix A.4 and $I_l^\alpha(x_i, t)$ is the total synaptic current into the neuron. In the striate cortex, the synaptic current is given by

$$\begin{aligned} I_l^S(x_i, t) &= \sum_{m=E,I} W_{lm}^S r_m^S(x_i, t) \\ &+ \sum_{\beta=S,X} \sum_j W_{lE}^{S\beta}(x_i|x_j) \cdot r_E^\beta(x_j, t - \Delta t^{S\beta}(x_i|x_j)) \\ &+ h(x_i, t) \delta_{l,E} \end{aligned}$$

where the first term corresponds to the local inputs, the second corresponds to the lateral and feedback inputs from the extrastriate cortex, and the last term denotes the afferent inputs. For the extrastriate cortex it has the form

$$\begin{aligned} I_l^X(x_i, t) &= \sum_{m=E,I} W_{lm}^X r_m^X(x_i, t) \\ &+ \sum_j W_{lE}^{XS}(x_i|x_j) \cdot r_E^S(x_j, t - \Delta t^{XS}(x_i|x_j)), \end{aligned}$$

where the first term corresponds to local inputs, and the second corresponds to afferent excitation from the striate cortex. Lateral interactions in the extrastriate area were not part of the model. The weights for the non-local intra-(lateral) and inter-areal (FB) connections depend on the distance between the pre- and postsynaptic neurons' RF centers x_i and x_j and are given by

$$W_{lE}^{\alpha\beta}(x_i|x_j) = \bar{W}_{lE}^{\alpha\beta} \cdot \exp\left(-\lambda^{\alpha\beta}|x_i - x_j|\right)$$

where $\lambda^{\alpha\beta}$ determines the spatial scale of the connections. Lateral connections in V1 were made only between neurons with the distance between RF centers larger than 0.1° (equivalent to 230 μm in cortical distance, at 5° retinal eccentricity-see below). For example, self-excitation via lateral connections was excluded. The absolute values for the conductances associated with the weight coefficients $\bar{W}_{lE}^{\alpha\beta}$ for the modulatory connections and for conductances for the coefficients W_{lm}^{S} for the local connections are given in the Tab. C.1.

For simplicity, all recurrent connections in extrastriate cortex were set to zero ($\bar{W}_{lE}^{XX} = W_{lm}^{X} = 0$) so that it effectively served as a relay station without further processing.

The time-delays are given by

$$\Delta t^{\alpha\beta}\left(x_i \,|\, x_j\right) = \begin{cases} \frac{|x_i - x_j|}{v_{S \to S}} & \text{for lateral connections in striate cortex} \\ \frac{\Delta x_{S \to X}}{v_{S \to X}} & \text{for inter-areal connections} \end{cases}$$

where $v_{S \to S}$ and $v_{S \to X}$ are the conduction velocities for signaling within the striate cortex and between the striate and the extrastriate area, respectively, and $\Delta x_{S \to X}$ is the distance between the two areas. The values for the $\lambda^{\alpha\beta}$ and the conduction velocities are given in Tab. C.1.

C.2 Afferent input

The model neurons in the recurrent network correspond to neurons in layer 2/3 of area V1. Therefore, the afferent input to these neurons arises from layer 4C. Here we model the input to the i-th neuron in the network with RF center at x_i as a current $I^{aff}(x_i)$. Let $c(y)$ denote the contrast of the stimulus at the visual field location y, then the input current is given as

$$I^{aff}\left(x_i\right) = \int G\left(y;\, x_i\right) \cdot I\left(c\left(y\right)\right)\, dy,$$

where $G\left(y;\, x_i\right)$ and $I\left(c\right)$ account for the spatial summation over the afferent inputs and the contrast-dependence of the input, respectively. These functions are defined as

$$\begin{aligned} G\left(y;\, x_i\right) &= \left(\sqrt{2\pi}\sigma_{aff}\right)^{-1} \cdot \exp\left(-\left(y - x_i\right)^2 / \left(2\sigma_{aff}^2\right)\right) \\ I\left(c\right) &= \begin{cases} 0 & \text{for } 0 \leq c < c^{th} \\ \frac{I^{low}}{c^{low} - c^{th}}\left(c - c^{th}\right) & \text{for } c^{th} \leq c < c^{low} \\ \frac{I^{high} - I^{low}}{c^{high} - c^{low}}\left(c - c^{low}\right) + I^{low} & \text{for } c^{low} \leq c \end{cases} \end{aligned} \quad \text{(C.1)}$$

Here, the parameter values for the contrast-dependence of the input currents were set to $c^{th} = 0.1$, $c^{low} = 0.15$, $c^{high} = 0.85$, $I^{low} = 0.58\,\text{nA}$, $I^{high} = 0.71\,\text{nA}$, and $\sigma_{aff} = 0.1$ (Angelucci et al., 2002a; Bauer et al., 1999).

C.3 Parameterization of spatial scales and delays

The parameters $\lambda^{\alpha\beta}$ determine the spatial scale of the connections between the areas α and β. We used $\lambda^{SS} = 2.3\,\text{deg}^{-1}$ and $\lambda^{XS} = \lambda^{SX} = 0.3\,\text{deg}^{-1}$. These values were computed as follows. Defining the scale of the connections

between the areas α and β as the value Δx at which $\exp\left(-\lambda^{\alpha\beta}\Delta x\right) = 0.05$, we have $\lambda^{\alpha\beta} = -(\Delta x)^{-1}\ln 0.05$. For the lateral connections we used $\Delta x = 0.5 \cdot D_{(mm)} \cdot MF^{-1} = 1.3\,\text{deg}$ and obtained $\lambda^{SS} = 2.3\,deg^{-1}$, where $D_{(mm)} = 6\,\text{mm}$ is the average length of the lateral connections in macaque V1 (Angelucci et al., 2002b), and $MF = 2.3\,\text{mm}/^\circ$ is the cortical magnification factor in macaque V1 at $5\,\text{deg}$ eccentricity along the iso-polar lines of the visual field representation (Essen et al., 1984).

In a previous study (Angelucci et al., 2002b), we found that the average visuotopic extent of the lateral connections closely matches the low contrast summation field (lsRF). For example, at $5\,\text{deg}$ eccentricity the physiologically determined average size of the lsRF is approx. $2.6\,\text{deg}$ in diameter, which is compatible with $2\Delta x = 2.6\,\text{deg}$. For the inter-areal connections we used $\Delta x = 0.5 \cdot D_{(mm)} \cdot MF^{-1} = 9.89\,\text{deg}$, and obtained $\lambda^{XS} = \lambda^{SX} = 0.3\,\text{deg}$, where $D_{(mm)} = 8.9\,\text{mm}$ is the average cortical extent of the neuronal fields of FB connections to V1 within area MT in the macaque (Angelucci et al., 2002b), and $MF = 0.45\,\text{mm}/\text{deg}$ is the cortical magnification factor in macaque MT at $5\,\text{deg}$ eccentricity (Albright and Desimone, 1987).

For the conduction velocities used to compute the time-delays $\Delta t^{\alpha\beta}$ we used $v_{S\rightarrow X} = 4000\,\text{mm}/\text{s}$ for the conduction velocity of the inter-areal connections between V1 and MT in the macaque, and $\Delta x_{S\rightarrow X} = 7\,\text{mm}$ for the distance between the areas V1 and MT (Hupe et al., 1998). For the intra-areal connections in the striate cortex we used $v_{S\rightarrow X} = 200\,\text{mm}/\text{s} \cdot MF^{-1} = 86.9\,\text{deg}/s$ (Bringuier et al., 1999), where $MF = 2.3\,\text{mm}/\text{deg}$ is the cortical magnification factor for the striate cortex at $5\,\text{deg}$ eccentricity. These values lead, e. g., to time-delays of $\Delta t = 5^\circ/(86.9^\circ/s) = 57.5\,\text{ms}$ for signals traveling a distance of $5\,\text{deg}$ within the model V1 compared to a delay of only $\Delta t = (2 \cdot 7\,\text{mm})/(4000\,\text{mm}/s) = 3.5\,\text{ms}$.

Appendix D

Generic hypercolumn model and Fisher information

D.1   The generic hypercolumn model

Here we describe in detail the network model we used in Chapter 4. The architecture of the recurrent network model is shown in Fig. 4.2b. Excitatory and inhibitory neurons receive already tuned afferent inputs from a lower visual area as well as additive and background (feedback) inputs. The latter are not described explicitly, but their overall effects are summarized by a fluctuating background conductance and an additive input current ("feedback"). Parameterizations were chosen such that the model reproduced the "average response" of a V4 cell in the control trial of a perceptual learning experiment (Yang, 2004). Here, we also used the firing rate functions from Appendix A.4. The parameters for this model are summarized in Tab.

We separate the overall input current $I_i = I_i^{aff} + I_i^{rec} + I_i^{add}$ into an afferent, a recurrent, and an additive background component. A large fraction of the latter is assumed to be a result of direct feedback received from a downstream area. For excitatory and inhibitory neurons, we initially set I_i^{add} to 0.6 and 0.64 nA, respectively. This gives rise to a background activity of 3.6 spikes/s (sp/s) for the excitatory and inhibitory neurons. When a stimulus θ is presented to the network, it leads to an afferent input to neuron i, which is calculated using a bell-shaped input tuning function with a peak at θ_i^{pr}, which is the preferred stimulus of neuron i:

$$
\begin{aligned}
I_i^{aff}(\theta) &= W_i^{aff} M \exp\left(|\cos\left(2\pi d_i(\theta)\right) - 1|\right) \quad \text{with} \\
W_i^{aff} &= G_i^{aff}\left(E_E - E_i^L - V_i^c\right),
\end{aligned}
$$

where

$$
d_i(\theta) = \min\left(|\theta - \theta_i^{pr}|,\, 1 - |\theta - \theta_i^{pr}|\right)
$$

is the circular distance between stimulus θ and the "afferent" preferred stimulus $\theta_i^{pr} = i/N$ of neuron i, N is the number of excitatory or inhibitory neurons, depending on the type of neuron i, and $M = 2000\,\mathrm{sp/s}$. For convenience, we consider dimensionless one-dimensional "circular" stimuli with $0 \leq \theta \leq 1$. To

obtain numerical values for stimulus domains such as "orientation" or "direction", θ needs to be multiplied by 180° and 360°, respectively.

The recurrent input is a weighted sum of the output activities of the neuron and is given by the following equations:

$$
\begin{aligned}
I_i^{rec} &= \sum_j W_{ij}^{rec} r_j \\
W_{ij}^{rec} &= G_{ij}^{rec} \left(E_j - E_i^L - V_i^c \right) \\
G_{ij}^{rec} &= Z_i \exp \left(\kappa_i \cos \left(2\pi d_i \left(\theta_j^{pr} \right) \right) \right),
\end{aligned}
$$

where κ_i ($\kappa_i = \kappa_I = 1$ for inhibitory neurons and $\kappa_i = \kappa_E = 4$ for excitatory neurons) and Z_i determine the specificity and the strength of the recurrent connections. r_j denotes the firing rate of the presynaptic neuron j. The strengths Z_i are set so that $\sum_j G_{ij}^{rec} = 0.135\,\text{nS}$ for the excitatory and $0.2813\,\text{nS}$ for the inhibitory neurons j.

Parameter	Description	Value
	Network architecture	
$N_E^S = N_I^S$	No. of exc. and inh. neurons in striate cortex	32
	Afferent input	
M	Scaling of afferent drive	$2000\,\text{sp/s}$
$G_i^{aff} = G_E^{aff}$	Strength of afferent input to exc. neurons	$9.3 \cdot 10^{-4}\,\text{nS}$
$G_i^{aff} = G_I^{aff}$	Strength of afferent input to inh. neurons	$5.8 \cdot 10^{-4}\,\text{nS}$
	Background inputs	
$I_i^{bg} = I_E^{bg}$	Background input to exc. neurons	$0.6\,\text{nA}$
$I_i^{bg} = I_I^{bg}$	Background input to inh. neurons	$0.64\,\text{nA}$
	Recurrent connections	
κ_E	Specificity of recurrent exc. connections	4
κ_I	Specificity of recurrent inh. connections	1
$\sum_j G_{ij}^{rec}$	Sum of all presynaptic exc. weights	$0.135\,\text{nS}$
$\sum_j G_{ij}^{rec}$	Sum of all presynaptic exc. weights	$0.2813\,\text{nS}$

Table D.1: Parameter values specific for the generic hypercolumn model.

D.2 The Cramer-Rao bound

The Fisher information used in Chapter 4 is rooted in statistics, where it determines, via the so-called Cramer-Rao bound, a lower bound on the variance of any possible unbiased estimator. Here we give some background on how this quantity is derived.

Consider the problem of estimating a single scalar parameter θ from the knowledge of some measurements $y_1, y_2, ..., y_N \equiv \mathbf{y}$. (Note that the measurements are scalar values. They are referred to as the components of a

vector only to shorten the notation.) The corresponding system of N measurements of the underlying parameter θ is described by the liklihood $p(\mathbf{y}|\theta)$. The observed datapoints are subject to some (un)known additive noise: $\mathbf{y} \equiv (\theta + x_1, \theta + x_2, ..., \theta + x_N)$. The *Cramer-Rao bound* gives a lower bound for the expected squared error assuming an unbiased estimator:

$$E\left[\left(\hat{\theta}\left(\mathbf{y}\right) - \theta\right)^2\right] \cdot I\left(\theta\right) \geq 1,$$

where

$$I\left(\theta\right) = \int \left(\frac{\partial}{\partial\theta} \ln p\left(\mathbf{y}|\theta\right)\right)^2 p\left(\mathbf{y}|\theta\right) d\mathbf{y}$$

is the Fisher information with respect to θ. The Cramer-Rao bound can be derived as follows: Since an unbiased estimator assumes

$$E\left[\hat{\theta}\left(\mathbf{y}\right) - \theta\right] = \int \left(\hat{\theta}\left(\mathbf{y}\right) - \theta\right) p\left(\mathbf{y}|\theta\right) d\mathbf{y} = 0,$$

the goal is to fulfill this constraint. This is performed by deriving with respect to θ:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\theta} \int (\hat{\theta}(\mathbf{y}) - \theta) p(\mathbf{y}|\theta) d\mathbf{y} \\
0 &= \int \frac{\partial}{\partial\theta} (\hat{\theta}(\mathbf{y}) p(\mathbf{y}|\theta) - \theta p(\mathbf{y}|\theta)) d\mathbf{y} \\
0 &= \int p(\mathbf{y}|\theta) \underbrace{\frac{\partial}{\partial\theta}\hat{\theta}(\mathbf{y})}_{0} + \hat{\theta}(\mathbf{y}) \frac{\partial}{\partial\theta} p(\mathbf{y}|\theta) - p(\mathbf{y}|\theta) \underbrace{\frac{\partial}{\partial\theta}\theta}_{1} - \theta \frac{\partial}{\partial\theta} p(\mathbf{y}|\theta) d\mathbf{y} \\
0 &= \int \hat{\theta}(\mathbf{y}) \frac{\partial}{\partial\theta} p(\mathbf{y}|\theta) - p(\mathbf{y}|\theta) - \theta \frac{\partial}{\partial\theta} p(\mathbf{y}|\theta) d\mathbf{y} \\
0 &= \int (\hat{\theta}(\mathbf{y}) - \theta) \frac{\partial}{\partial\theta} p(\mathbf{y}|\theta) d\mathbf{y} - \int p(\mathbf{y}|\theta) d\mathbf{y}
\end{aligned}
$$

Using the identity $\frac{\partial}{\partial\theta} p = p \frac{\partial}{\partial\theta} \ln p$ and the fact that $p(\mathbf{y}|\theta)$ is normalized gives

$$\int \left(\hat{\theta}\left(\mathbf{y}\right) - \theta\right) p\left(\mathbf{y}|\theta\right) \frac{\partial}{\partial\theta} \ln p\left(\mathbf{y}|\theta\right) d\mathbf{y} = 1.$$

Factorizing the integrand and then squaring the equation yields

$$
\int \underbrace{\left[(\hat{\theta}(\mathbf{y}) - \theta)\sqrt{p(\mathbf{y}|\theta)}\right]}_{a} \underbrace{\left[\sqrt{p(\mathbf{y}|\theta)}\frac{\partial}{\partial\theta} \ln p(\mathbf{y}|\theta)\right]}_{b} d\mathbf{y} = 1
$$

$$
\left[\int (ab) d\mathbf{y}\right]^2 = 1
$$

Here, $\int (ab)\, d\mathbf{y}$ should be realized as representing the scalar product $\langle a, b \rangle$ between two functions in a function space. This is used when applying the Schwarz inequality

$$\|a\| \cdot \|b\| \geq |\langle a, b \rangle| \equiv \|a\|^2 \cdot \|b\|^2 \geq |\langle a, b \rangle|^2 \equiv \langle a, a \rangle \cdot \langle b, b \rangle \geq |\langle a, b \rangle|^2,$$

to give

$$\int \left[(\hat{\theta}(\mathbf{y}) - \theta) \sqrt{p(\mathbf{y}|\theta)} \right]^2 d\mathbf{y} \cdot \int \left[\sqrt{p(\mathbf{y}|\theta)} \frac{\partial}{\partial \theta} \ln p(\mathbf{y}|\theta) \right]^2 d\mathbf{y} \geq 1$$

$$\int (\hat{\theta}(\mathbf{y}) - \theta)^2 p(\mathbf{y}|\theta) d\mathbf{y} \cdot \int \left[\frac{\partial}{\partial \theta} \ln p(\mathbf{y}|\theta) \right]^2 p(\mathbf{y}|\theta) d\mathbf{y} \geq 1$$

$$E[(\hat{\theta}(\mathbf{y}) - \theta)^2] \cdot I(\theta) \geq 1$$

which is the Cramer-Rao inequality.

D.3   Gradients of the Fisher information

Let $f_i(\theta; w)$ be the tuning function of a neuron i, where w is a parameter of this tuning function. Then, the derivative $\partial J_i / \partial w$ of the Fisher information J_i for a given value of the stimulus θ w. r. t . the parameter w is given by

$$\frac{\partial J_i}{\partial w} = \tau \frac{\partial f_i}{\partial \theta} \left(2 f_i \frac{\partial^2 f_i}{\partial w \partial \theta} - \frac{\partial f_i}{\partial \theta} \frac{\partial f_i}{\partial w} \right) / (f_i)^2 .$$

In order to determine how small changes of the network parameters (recurrency and gain of excitatory neurons) change the objective function, one needs the derivatives of the tuning functions. But how to compute these derivatives for tuning functions given only implicitly as the attractors of the recurrent network? First, note that at a steady-state

$$r_i \left(\theta, \mathbf{G}^{aff}, \mathbf{G}^{rec}, \mathbf{a}^E, \mathbf{I}^{add} \right) = F_i \left(I_i \left(\theta, \mathbf{G}^{aff}, \mathbf{G}^{rec}, \mathbf{a}^E, \mathbf{I}^{add} \right) \right) \quad \text{(D.1)}$$

$$I \left(\theta, \mathbf{G}^{aff}, \mathbf{G}^{rec}, \mathbf{a}^E, \mathbf{I}^{add} \right) = I_i^{aff} + \sum_{j=1}^{N} W_{ij}^{rec} r_j + I_i^{add} \quad \text{(D.2)}$$

must hold for all i. We calculated the derivatives necessary to evaluate the gradient of the objective function w. r. t. the network parameters by first simulating the network until a steady-state is reached. Then, we derived both sides of Eq. D.1 and re-arranged the resulting set of equations using a matrix notation.

Sensitivity to the stimulus parameter: In order to obtain an expression for $\partial r_i / \partial \theta$ we derive both sides of Eq. D.1 w. r. t. θ and obtain

$$\frac{\partial r_i}{\partial \theta} = \frac{\partial F_i}{\partial I_i} \frac{\partial I_i^{aff}}{\partial \theta} + \frac{\partial F_i}{\partial I_i} \sum_{j=1}^{N} W_{ij}^{rec} \frac{\partial r_j}{\partial \theta} \quad \text{with} \quad \text{(D.3)}$$

$$\frac{\partial I_i^{aff}}{\partial \theta} = \frac{\partial d}{\partial \theta} k_i \exp \left(\kappa \left[\cos \left(2\pi d \left(\theta, \theta_i^{pr} \right) \right) - 1 \right] \right), \quad \text{(D.4)}$$

where $k_i = -2\pi\kappa W_i^{aff} M \sin \left(2\pi d \left(\theta_j, \theta_i^{pr} \right) \right)$, $\partial d / \partial \theta = \text{sgn} \left(\theta - \theta_i^{pr} \right)$ for $0 < |\theta - \theta_i^{pr}| \leq 0.5$, and $\partial d / \partial \theta = -\text{sgn} \left(\theta - \theta_i^{pr} \right)$ for $0.5 < |\theta - \theta_i^{pr}|$. Here, $\text{sgn}(x)$ is a signum function with $\text{sgn}(x) = 1$ for $u > 0$, $\text{sgn}(x) = -1$ for $x < 0$, and $\text{sgn}(x) = 0$ for $x = 0$. Let $h_i = \partial I_i^{aff} / \partial \theta$, $r_i^\theta = \partial r_i / \partial \theta$ and $F_{ij} = \delta_{ij} \partial F_i / \partial I_i$ and be the components of $\mathbf{h}$, $\mathbf{g}^\theta$ and $\mathbf{F}$. Then, we can write the Eqns. D.3 as

$$\mathbf{r}^\theta = \mathbf{F}\mathbf{h} + \mathbf{F}\mathbf{W}^{rec}\mathbf{r}^\theta = [\mathbf{1} - \mathbf{F}\mathbf{W}^{rec}]^{-1} \mathbf{F}\mathbf{h}. \quad \text{(D.5)}$$

Recurrent connections: In order to obtain an expression for $\partial^2 r_i/\partial G_{rs}^{rec}\partial\theta$ we derive both sides of Eq. D.3 w. r. t. G_{rs}^{rec} and get

$$
\begin{aligned}
\frac{\partial^2 r_i}{\partial G_{rs}^{rec}\partial\theta} &= \frac{\partial^2 F_i}{\partial G_{rs}^{rec}\partial I_i}\left[\frac{\partial I_i^{aff}}{\partial\theta}+\sum_{j=1}^N W_{ij}^{rec}\frac{\partial r_j}{\partial\theta}\right]\\
&\quad +\delta_{ri}V_{rs}\frac{\partial F_i}{\partial I_i}\frac{\partial r_s}{\partial\theta}+\frac{\partial F_i}{\partial I_i}\sum_{j=1}^N W_{ij}^{rec}\frac{\partial^2 r_j}{\partial G_{rs}^{rec}\partial\theta}
\end{aligned}
\tag{D.6}
$$

Let $F_{ij}^{rs}=\delta_{ij}\partial^2 F_i/\partial G_{rs}^{rec}\partial I_i$ and $h_i^{\theta,rs}=\delta_{ri}V_{rs}\partial F_i/\partial I_i\cdot\partial r_s/\partial\theta$ be the components of $\mathbf{h}^{\theta,rs}$ and $\mathbf{F}^{rs}$ with $V_{rs}=E_s-E_L-V_r^c$. Then, we can write

$$
\mathbf{r}^{\theta,rs} = [\mathbf{1}-\mathbf{F}\mathbf{W}^{rec}]^{-1}\left[\mathbf{F}^{rs}\mathbf{h}+\mathbf{F}^{rs}\mathbf{W}^{rec}\mathbf{r}^\theta+\mathbf{h}^{\theta,rs}\right]
\tag{D.7}
$$

and get $r_i^{\theta,rs}=\partial^2 r_i/\partial G_{rs}^{rec}\partial\theta$. To obtain $r_i^{rs}=\partial r_i/\partial G_{rs}^{rec}$ we derive both sides of Eq. D.1 w. r. t. G_{rs}^{rec} and get

$$
\frac{\partial r_i}{\partial G_{rs}^{rec}} = \frac{\partial F_i}{\partial I_i}\left[\delta_{ri}V_{rs}r_s+\sum_{j=1}^N W_{ij}^{rec}\frac{\partial r_j}{\partial G_{rs}^{rec}}\right]\quad\text{and hence}\tag{D.8}
$$

$$
\mathbf{r}^{rs} = [\mathbf{1}-\mathbf{F}\mathbf{W}^{rec}]^{-1}\mathbf{F}\mathbf{h}^{rs}\quad\text{with}\quad h_i^{rs}=\delta_{ri}V_{rs}r_s^\infty\tag{D.9}
$$

Gain-modulation of excitatory neurons: In order to obtain $m_i^{\theta,r}=\partial^2 r_i/\partial a_r^E\partial\theta$ we derive both sides of Eq. D.3 w. r. t. a_r^E. Denoting with $A_{ij}^{\theta,r}=\delta_{ij}\partial^2 F_i/\partial a_r^E\partial I_i$ the elements of $\mathbf{A}^{\theta,r}$, we get

$$
\mathbf{m}^{\theta,r} = [\mathbf{1}-\mathbf{F}\mathbf{W}^{rec}]^{-1}\mathbf{A}^{\theta,r}\left[\mathbf{h}+\mathbf{W}^{rec}\mathbf{r}^\theta\right].
\tag{D.10}
$$

To obtain $m_i^r=\partial r_i/\partial a_r^E$ we derive both sides of Eq. D.1 w. r. t. a_r^E and get

$$
\frac{\partial r_i}{\partial a_r^E} = \delta_{ri}\left[I_i-\Delta I_i\left(g_L\right)\right]+a_i^E\sum_{j=1}^N W_{ij}^{rec}\frac{\partial r_j}{\partial a_r^E}
\tag{D.11}
$$

$$
\frac{\partial r_i}{\partial a_r^E} = \left[a^I-2b^I\left(I_i-\Delta I_i\left(g_L\right)\right)\right]\cdot\sum_{j=1}^N W_{ij}^{rec}\frac{\partial r_j}{\partial a_r^E}
\tag{D.12}
$$

for excitatory and inhibitory neurons i, respectively. Let $I_i^r=\delta_{ri}\left[I_i-\Delta I_i\left(g_L\right)\right]$, $A_{ij}^r=\delta_{ij}a_i^E$ and $I_i^r=0$, $A_{ij}^r=\delta_{ij}\left[a_i^I-2b_i^I\left(I_i-\Delta I_i\left(g_L\right)\right)\right]$ for excitatory and inhibitory neurons i, respectively, then we get

$$
\mathbf{m}^r = \mathbf{I}^r+\mathbf{A}^r\mathbf{W}^{rec}\mathbf{m}^r = [\mathbf{1}-\mathbf{A}^r\mathbf{W}^{rec}]^{-1}\mathbf{I}^r.
$$

Derivatives of gain functions: For the gain functions the derivatives below the threshold are 0, above they are $dF_i^E/dI_i=a_i^E$ and $dF_i^I/dI_i=a^I+2b^I\left(I_i-\Delta I\left(g_L\right)\right)$, and due to the discontinuity they are not defined at the threshold. However, for all computations we never needed to evaluate the derivative at the threshold 0. For $\partial^2 F/\partial G_{rs}^{rec}\partial I$ we get $\partial^2 F_i^E/\partial G_{rs}^{rec}\partial I_i=0$ and $\partial^2 F_i^I/\partial G_{rs}^{rec}\partial I_i=2b^I\left[\delta_{ri}V_{rs}r_s+\sum_{j=1}^N W_{ij}^{rec}\partial r_j/\partial G_{rs}^{rec}\right]$, or in the shorthand notation $\mathbf{F}^{I,rs}=2b^I\left[\mathbf{h}^{rs}+\mathbf{W}^{rec}\mathbf{r}^{rs}\right]$. For the gain-modulation of the excitatory neurons we get $\partial^2 F_i^E/\partial a_r^E\partial I_i=\delta_{ri}$ and $\partial^2 F_i^I/\partial a_r^E\partial I_i=2b^I\sum_{j=1}^N W_{ij}^{rec}\partial r_j/\partial a_r^E$.

Bibliography

Adorjan, P., Piepenbrock, C., and Obermayer, K.: 1999, *Rev Neurosci* **10(3-4)**, 181

Adorjan, P., Schwabe, L., Piepenbrock, C., and Obermayer, K.: 2000, in *Advances in Neural Information Processing Systems 12*, Vol. 12, pp 89–95, MIT Press

Adorjan, P., Schwabe, L., Wenning, G., and Obermayer, K.: 2002, *Neuroreport* **13(3)**, 337

Albright, T. and Desimone, R.: 1987, *Exp Brain Res* **65(3)**, 582

Albright, T. D. and Stoner, G. R.: 2002, *Annu Rev Neurosci* **25**, 339

Allman, J., Miezin, F., and McGuinness, E.: 1985, *Annu Rev Neurosci* **8**, 407

Anderson, J., Carandini, M., and Ferster, D.: 2000, *J Neurophysiol* **84(2)**, 909

Angelucci, A. and Bullier, J.: 2003, *J Physiol Paris* **97(2-3)**, 141

Angelucci, A., Levitt, J. B., and Lund, J. S.: 2002a, *Prog Brain Res* **136**, 373

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., and Lund, J. S.: 2002b, *J Neurosci* **22(19)**, 8633

Atick, J. J.: 1992, *Network* **(3)**, 213

Bair, W., Cavanaugh, J. R., and Movshon, J. A.: 2003, *J Neurosci* **23(20)**, 7690

Bakin, J. S., Nakayama, K., and Gilbert, C. D.: 2000, *J Neurosci* **20(21)**, 8188

Barlow, H.: 2001, *Network* **12(3)**, 241

Barlow, H., Blakemore, C., and Pettigrew, J.: 1967, *J Physiol* **193(2)**, 327

Bauer, U.: 1998, *Ph.D. thesis*, Technische Fakultät, Universität Bielefeld

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H.: 1995, *Proc Natl Acad Sci U S A* **92(9)**, 3844

Bethge, M., Rotermund, D., and Pawelzik, K.: 2003, *Network* **14(2)**, 303

Blakemore, C. and Tobin, E.: 1972, *Exp Brain Res* **15(4)**, 439

Bosking, W., Zhang, Y., Schofield, B., and Fitzpatrick, D.: 1997, *J Neurosci* **17(6)**, 2112

Bradley, D. and Andersen, R.: 1998, *J Neurosci* **18(18)**, 7552

Bringuier, V., Chavane, F., Glaeser, L., and Fregnac, Y.: 1999, *Scie* **283(5402)**, 695

Brunel, N. and Nadal, J. P.: 1998, *Neural Computation* **10(7)**, 1731

Bullier, J., Hupe, J., James, A., and Girard, P.: 2001, *Prog Brain Res* **134**, 193

Cavanaugh, J. R., Bair, W., and Movshon, J. A.: 2002, *J Neurophysiol* **88(5)**, 2530

Chance, F. S., Abbott, L. F., and Reyes, A. D.: 2002, *Neuron* **35(4)**, 773

Chapman, B., Zahs, K., and Stryker, M.: 1991, *J Neurosci* **11(5)**, 1347

Chen, C. C., Kasamatsu, T., Polat, U., and Norcia, A. M.: 2001, *Neuroreport*

12(4), 655

Chisum, H. J., Mooser, F., and Fitzpatrick, D.: 2003, *J Neurosci* **23(7)**, 2947

Clifford, C. W., Wenderoth, P., and Spehar, B.: 2000, *Proc R Soc Lond B Biol Sci* **267(1454)**, 1705

Cohen, J., Dunbar, K., and McClelland, J.: 1990, *Psychol Rev* **97(3)**, 332

Cox, R. T.: 1961, *The Algebra of Probable Inference*, The Johns Hopkins Press

DeAngelis, G., Freeman, R., and Ohzawa, I.: 1994, *J Neurophysiol* **71(1)**, 347

Deneve, S.: 2005, in L. K. Saul, Y. Weiss, and L. Bottou (eds.), *Advances in Neural Information Processing Systems 17*, pp 353–360, MIT Press, Cambridge, MA

Deneve, S., Latham, P. E., and Pouget, A.: 1999, *Nat Neurosci* **2(8)**, 740

Destexhe, A., Rudolph, M., Fellous, J. M., and Sejnowski, T. J.: 2001, *Neuroscience* **107(1)**, 13

Dragoi, V., Rivadulla, C., and Sur, M.: 2001, *Nature* **411(6833)**, 80

Dragoi, V., Sharma, J., Miller, E. K., and Sur, M.: 2002, *Nat Neurosci* **5(9)**, 883

Dragoi, V. and Sur, M.: 2000, *J Neurophysiol* **83(2)**, 1019

Dragoi, V. and Sur, M.: 2003, *The Visual Neurosciences*, Chapt. Plasticity of orientation processing in adult visual cortex, pp 1654–1664, MIT Press

Essen, D. V., Newsome, W., and Maunsell, J.: 1984, *Vision Res* **24(5)**, 429

Fairhall, A., Lewen, G., Bialek, W., and de Ruyter Van Steveninckck, R.: 2001, *Nature* **412(6849)**, 787

Feldman, M. L.: 1984, *Cerebral Cortex*, Vol. 1, Chapt. Morphology of the Neocortical Pyramidal Neuron, pp 123–200, Plenum, New York

Fitzpatrick, D.: 2000, *Curr Opin Neurobiol* **10(4)**, 438

Ghose, G. M., Yang, T., and Maunsell, J. H.: 2002, *J Neurophysiol* **87(4)**, 1867

Gilbert, C., Das, A., Ito, M., Kapadia, M., and Westheimer, G.: 1996, *Proc Natl Acad Sci U S A* **93(2)**, 615

Gilbert, C. and Wiesel, T.: 1983, *J Neurosci* **3(5)**, 1116

Gilbert, C. and Wiesel, T.: 1989, *J Neurosci* **9(7)**, 2432

Gilbert, C. and Wiesel, T.: 1990, *Vision Res* **30(11)**, 1689

Gilbert, S. J. and Shallice, T.: 2002, *Cognit Psychol* **44(3)**, 297

Girard, P., Hupe, J. M., and Bullier, J.: 2001, *J Neurophysiol* **85(3)**, 1328

Grinvald, A., Lieke, E. E., Frostig, R. D., and Hildesheim, R.: 1994, *J Neurosci* **14(5)**, 2545

Grossberg, S.: 1976, *Biol Cybern* **23**, 121

Hahnloser, R., Douglas, R., Mahowald, M., and Hepp, K.: 1999, *Nat Neurosci* **2(8)**, 746

Hansel, D. and van Vreeswijk, C.: 2002, *J Neurosci* **22(12)**, 5118

Hirsch, J. A. and Gilbert, C. D.: 1991, *J Neurosci* **11(6)**, 1800

Hubel, D. H.: 1988, *Eye, brain, and vision*, Scientific American Library

Hubel, D. H. and Wiesel, T. N.: 1962, *J. Physiol.* **165**, 559

Hubel, D. H. and Wiesel, T. N.: 1977, *Proc R Soc Lond B Biol Sci* **198(1130)**, 1

Hupe, J., James, A., Girard, P., and Bullier, J.: 2001, *J Neurophysiol* **85(1)**, 146

Hupe, J., James, A., Payne, B., Lomber, S., Girard, P., and Bullier, J.: 1998, *Nature* **394(6695)**, 784

Ichida, J. M., Schwabe, L., Bressloff, P. C., and Angelucci, A.: 2005, in *SfN Abstracts*, SfN

Johnson, R. and Burkhalter, A.: 1996, *J Comp Neurol* **368(3)**, 383

Kaelbling, L. P., Littman, M. L., and Moore, A. W.: 1996, *Journal of Artificial Intelligance Research* **4**, 237

Kandel, E. R., Schwarz, J. H., and Jessell, T. M.: 2000, *Principles of Neural Science*, Edward Arnold

Kang, K., Shelley, M., and Sompolinsky, H.: 2003, *Proc Natl Acad Sci U S A* **100(5)**, 2848

Kapadia, M., Westheimer, G., and Gilbert, C.: 1999, *Proc Natl Acad Sci U S A* **96(21)**, 12073

Kapadia, M. K., Ito, M., Gilbert, C. D., and Westheimer, G.: 1995, *Neuron* **15(4)**, 843

Kaschube, M., Wolf, F., Geisel, T., and Loewel, S.: 2002, *J Neurosci* **22(16)**, 7206

Kay, S. M.: 1993, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall

Kisvarday, Z. F., Martin, K. A., Freund, T. F., Magloczky, Z., Whitteridge, D., and Somogyi, P.: 1986, *Exp Brain Res* **64(3)**, 541

Knierim, J. and van Essen, D.: 1992, *J Neurophysiol* **67(4)**, 961

Lee, T. S. and Nguyen, M.: 2001, *Proc Natl Acad Sci U S A* **98(4)**, 1907

Levitt, J. and Lund, J.: 1997, *Nature* **387(6628)**, 73

Levitt, J. B. and Lund, J. S.: 2002, *Vis Neurosci* **19(4)**, 439

Lund, J. S.; Angelucci, A. B. P. C.: 2003, *Cerebral Cortex* **13(1)**, 15

Lund, J. S., Wu, Q., Hadingham, P. T., and Levitt, J. B.: 1995, *J Anat* **187 (3)**, 563

Malach, R., Amir, Y., Harel, M., and Grinvald, A.: 1993, *Proc Natl Acad Sci U S A* **90(22)**, 10469

Marino, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M.: 2005, *Nat Neurosci* **8(2)**, 194

Martin, K. A. and Whitteridge, D.: 1984, *J Physiol* **353**, 463

McAdams, C. J.; Maunsell, J. H.: 1999, *J Neurosci* **19(1)**, 431

McAdams, C. J.; Maunsell, J. H.: 2000, **83(3)**, 1751

McGuire, B. A., Gilbert, C. D., Rivlin, P. K., and Wiesel, T. N.: 1991, *J Comp Neurol* **305(3)**, 370

McLaughlin, D., Shapley, R., Shelley, M., and Wielaard, D.: 2000, *Proc Natl Acad Sci U S A* **97(14)**, 8087

Mignard, M. and Malpeli, J. G.: 1991, *Science* **251(4998)**, 1249

Moore, C., Nelson, S., and Sur, M.: 1999, *Trends Neurosci* **22(11)**, 513

Mooser, F., Bosking, W. H., and Fitzpatrick, D.: 2004, *Nat Neurosci* **7(8)**, 872

Moran, J. and Desimone, R.: 1985, **229(4715)**, 782

Muller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P.: 1999, **285(5432)**, 1405

Muller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P.: 2003, *J Neurophysiol* **90(2)**, 822

Nakahara, H. and Amari, S.: 2002, *Neural Netw* **15(1)**, 41

Nakahara, H., Wu, S., and Amari, S.: 2001, *Neural Comput* **13(9)**, 2031

Nelken, I.: 2004, *Curr Opin Neurobiol* **14(4)**, 474

Nelson, J. and Frost, B.: 1978, *Brain Res* **139(2)**, 359

Nelson, S., Toth, L., Sheth, B., and Sur, M.: 1994, *Science* **265(5173)**, 774

Nothdurft, H., Gallant, J., and Essen, D. V.: 1999, *Vis Neurosci* **16(1)**, 15

Obermayer, K. and Blasdel, G.: 1993, *J Neurosci* **13(10)**, 4114

Palmer, S.: 1999, *Vision Science: Photons to Phenomenology*, Bradford

Paradiso, M. A.: 1988, *Biol Cybern* **58(1)**, 35

Paz, R. and Vaadia, E.: 2004, *PLoS Biol* 2(2)

Pei, X., Vidyasagar, T., Volgushev, M., and Creutzfeldt, O.: 1994, *J Neurosci* **14(11)**, 7130

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M.: 1998, *Nature* **391(6667)**, 580

Pouget, A., Dayan, P., and Zemel, R. S.: 2003, *Annu Rev Neurosci* **26**, 381

Rao, R. and Ballard, D.: 1999, *Nat Neurosci* **2(1)**, 79

Rao, R. P. N.: 2004, *Neural Comput* **16(1)**, 1

Reid, R. C. and Alonso, J. M.: 1995, *Nature* **378(6554)**, 281

Rockland, K. and Lund, J.: 1983, *J Comp Neurol* **216(3)**, 303

Sabatini, S. P.: 1996, *Biol Cybern* **74(3)**, 189

Sahani, M. and Dayan, P.: 2003, *Neural Computation* **15(10)**, 2255

Salinas, E.: 2004a, *BMC Neurosci* **5(1)**, 47

Salinas, E.: 2004b, *J Neurosci* **24(5)**, 1113

Sandell, J. H. and Schiller, P. H.: 1982, *J Neurophysiol* **48(1)**, 38

Sceniak, M., Hawken, M., and Shapley, R.: 2001, *J Neurophysiol* **85(5)**, 1873

Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R.: 1999, *Nat Neurosci* **2(8)**, 733

Schoups, A., Vogels, R., and Orban, R.: 1998, in *Invest Ophthalmol Vis Sci Suppl*, No. 39, p. 684

Schoups, A.; Vogels, R. Q. N. O. G.: 2001, *Nature* **412(6846)**, 549

Schummers, J., Marino, J., and Sur, M.: 2002, *Neuron* **36(5)**, 969

Schwabe, L.; Obermayer, K.: 2002, *Biosystems* **67(1-3)**, 239

Schwabe, L., Angelucci, A., Obermayer, K., and Bressloff, P. C.: 2005, *submitted*

Schwabe, L. and Obermayer, K.: 2003, *Neural Networks* **16(9)**, 1353

Schwabe, L. and Obermayer, K.: 2005a, *J Neurosci* **25(13)**, 3323

Schwabe, L. and Obermayer, K.: 2005b, *Vision Res*

Sclar, G. and Freeman, R.: 1982, *Exp Brain Res* **46(3)**, 457

Sengpiel, F., Baddeley, R. J., Freeman, T. C., Harrad, R., and Blakemore, C.: 1998, *Vision Res* **38(14)**, 2067

Sengpiel, F., Sen, A., and Blakemore, C.: 1997, *Exp Brain Res* **116(2)**, 216

Seung, H. S. and Sompolinsky, H.: 1993, *Proc Natl Acad Sci U S A* **90(22)**, 10749

Shao, Z. and Shao, A.: 1996, *J Neurosci* **16(22)**, 7353

Shapley, R.: 2004, *Neural Netw* **17(5-6)**, 615

Sharma, J., Dragoi, V., Tenenbaum, J. B., and Sur, E. K. M. M.: 2003, *Science* **300(5626)**, 1758

Shmuel, A., Korman, M., Sterkin, A., Harel, M., Ullman, S., Malach, R., and Grinvald, A.: 2005, *J Neurosci* **25(8)**, 2117

Shriki, O., Hansel, D., and Sompolinsky, H.: 2003, *Neural Computation* **15(8)**, 1809

Sillito, A., Grieve, K., Jones, H., Cudeiro, J., and Davis, J.: 1995, *Nature* **378(6556)**, 492

Sincich, L. C. and Blasdel, G. G.: 2001, *J Neurosci* **21(12)**, 4416

Slovin, H., Arieli, A., Hildesheim, R., and Grinvald, A.: 2002, *J Neurophysiol* **88(6)**, 3421

Somers, D., Dragoi, V., and Sur, M.: 2002, *The Cat Primary Visual Cortex*, Chapt. Orientation selectivity and its modulation by local and long-range

connections in visual cortex, pp 471–520, Academic Press

Somers, D., Nelson, S., and Sur, M.: 1995, *J Neurosci* **15(8)**, 5448

Somers, D. C., Todorov, E. V., Siapas, A. G., Toth, L. J., Kim, D. S., and Sur, M.: 1998, *Cereb Cortex* **8(3)**, 204

Stettler, D. D., Das, A., Bennett, J., and Gilbert, C. D.: 2002, *Neuron* **36(4)**, 739

Sugita, Y.: 1999, *Nature* **401(6750)**, 269

Teich, A. F. and Qian, N.: 2003, *J Neurophysiol* **89(4)**, 2086

Treue, S. and Trujillo, J. C. M.: 1999, *Nature* **399(6736)**, 575

Wainwright, M. J.: 1999, *Vision Research* **39(23)**, 3960

Wehr, M. and Zador, A. M.: 2003, *Nature* **426(6965)**, 442

Wiesing, P., Beck, O., Schwabe, L., Marino, J., Schummers, J., Lyon, D. C., Sur, M., and Obermayer, K.: 2005, *submitted*

Williams, R. J.: 1992, *Machine Learning* **8**, 229

Woergoetter, F. and Koch, C.: 1991, *J Neurosci* **11(7)**, 1959

Xie, X.: 2002, *Network* **13(4)**, 447

Yang, T.; Maunsell, J. H.: 2004, *J Neurosci* **24(7)**, 1617

Yoshioka, T., Blasdel, G., Levitt, J., and Lund, J.: 1996, *Cereb Cortex* **6(2)**, 297

Yu, A. J. and Dayan, P.: 2005, *Neuron* **46(4)**, 681

Zhang, J. and Abbott, L. F.: 2000, *Neurocomputing* **32-33**, 623

Zigmond, M. J., Bloom, F. E., and Landis, S. C.: 2000, *Fundamental Neuroscience*, Academic Press