Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex

James A. Bednar

Report   AI97-259   January 1997

jbednar@cs.utexas.edu
http://www.cs.utexas.edu/users/jbednar/

Artificial Intelligence Laboratory
The University of Texas at Austin
Austin, TX 78712

Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex

by

James Albert Bednar, B.S., B.A.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 1997

Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex

Approved by
Supervising Committee:

RISTO MIIKKULAINEN

WILSON S. GEISLER

Acknowledgments

I would like to express my sincerest gratitude to Dr. Risto Miikkulainen for his invaluable advice and guidance throughout this work. He contributed many valuable insights and criticisms and also greatly improved the quality of the writing. I would also like to thank Dr. Wilson S. Geisler for providing background information, and for reading the final draft and related papers. I am grateful to Joseph Sirosh for his assistance on the RF-LISSOM code.

JAMES A. BEDNAR

The University of Texas at Austin
May 1997

Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex

James Albert Bednar, M.A.
The University of Texas at Austin, 1997

Supervisor: Risto Miikkulainen

The psychological phenomenon known as the tilt aftereffect was used to demonstrate the functional properties of RF-LISSOM, a self-organizing model of laterally connected orientation maps in the primary visual cortex. The same self-organizing processes that are responsible for the development of the map and its lateral connections are shown to result in tilt aftereffects as well. The model allows analysis of data that are difficult to measure in humans, thus providing a view of the cortex that is otherwise not available. The results give computational support for the idea that tilt aftereffects arise from lateral interactions between adapting feature detectors, as has long been surmised. They also suggest that indirect tilt aftereffects could result from the conservation of synaptic resources. The model thus provides a unified computational explanation of self-organization and both direct and indirect tilt aftereffects in the primary visual cortex.

Contents

List of Figures

Chapter 1

Introduction

The tilt aftereffect (TAE, Gibson and Radner 1937) is a simple but intriguing visual phenomenon. After staring at a pattern of tilted lines or gratings, subsequent lines appear to have a slight tilt in the opposite direction. Figure 1.1 demonstrates the effect. The effect resembles an afterimage from staring at a bright light, but it causes changes in orientation perception rather than color or brightness perception.

In general, the visual system provides an accurate means of measuring the orientation of visual contours such as lines and edges (see Howard and Templeton 1966,p.179–183 for a review). However, contours presented close together or one after the other in the same location can interact, causing distortions in their apparent orientations. When the lines are presented simultaneously, this effect is known as the *tilt illusion*, when they are presented successively, it is known as the *tilt aftereffect*. This thesis will focus on the tilt aftereffect, but since the tilt illusion and aftereffect are widely held to have closely related causes, the tilt illusion will be discussed as well.

The prevailing theory for these effects attributes them to lateral interactions between orientation-specific feature-detectors in the primary visual cortex (Tolhurst and Thompson 1975). The inhibitory connection strengths between activated neurons are believed to increase temporarily while the subject focuses on an input pattern, causing changes in the perception of subsequent orientations. This occurs because the detectors are broadly tuned, and detectors for neighboring orientations also adapt somewhat (Hubel and Wiesel 1968). When a subsequent line of a slightly different orientation is presented, the most strongly responding units are now the ones with orientation preferences further from the adapting line, resulting in a change in the angle perceived.

Although the fundamentals of the theory were proposed in the 1970s, it has only recently become computationally feasible to test it in a detailed model of cortical function. A Hebbian[1] self-organizing process (the Receptive-Field Laterally Interconnected Synerget-

---

[1] In Hebbian learning, connections between two neurons are strengthened when both neurons are active
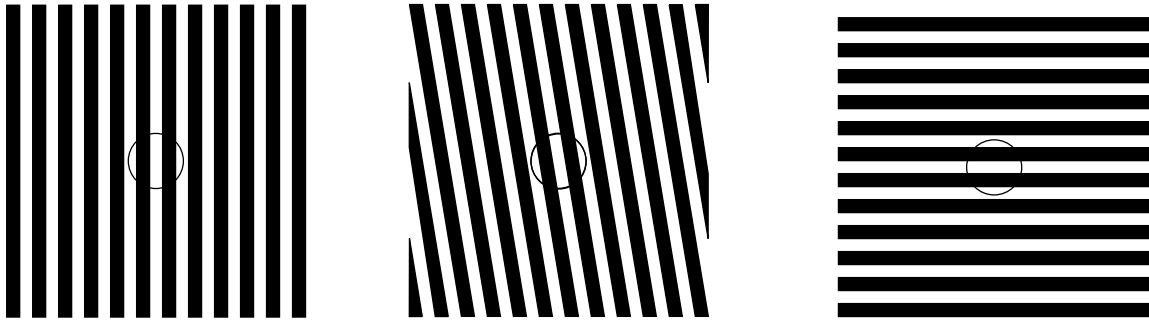
Figure 1.1: **Tilt aftereffect patterns.** Fixate your gaze upon the circle inside the square at the center for at least thirty seconds, moving your eye slightly inside the circle to avoid developing strong afterimages. Now fixate upon the figure at the left. The vertical lines should appear slightly tilted to the right; this phenomenon is called the direct tilt aftereffect. If you fixate upon the horizontal lines at the right, they may appear slightly tilted counterclockwise, though not every observer reports this indirect tilt aftereffect. (Adapted from Campbell and Maffei 1971.)

ically Self-Organizing Map, or RF-LISSOM; Miikkulainen et al. 1997; Sirosh and Miikkulainen 1997; Sirosh et al. 1996) has been shown to develop feature detectors and specific lateral connections that could produce such illusions and aftereffects. The RF-LISSOM model gives rise to anatomical and functional characteristics of the cortex such as topographic maps, ocular dominance, orientation, and size preference columns, and patterned lateral connections between them. Although other models exist that explain how the feature-detectors and afferent connections could develop by input-driven self-organization, RF-LISSOM is the only model that also shows how the lateral connections can self-organize as an integral part of the process. The laterally connected model has also been shown to account for many of the dynamic aspects of the visual cortex, such as reorganization following retinal and cortical lesions (Miikkulainen et al. 1997; Sirosh and Miikkulainen 1994b; Sirosh et al. 1996).

The current work is a first study of the *functional* behavior of the model, specifically the response to stimuli similar to those known to cause the TAE in humans. Because RF-LISSOM is a computational model, it can demonstrate phenomena in high detail that are difficult to measure experimentally, thus presenting a view of the cortex that is otherwise not available. The results suggest that tilt aftereffects are not flaws in an otherwise well-designed system, but an unavoidable result of a self-organizing process that aims at producing an efficient, sparse encoding of the input through decorrelation (as proposed by Barlow 1990; see also Dong 1995, 1996; Field 1994; Sirosh et al. 1996).

The rest of this thesis is organized as follows. Chapter 2 is a survey of related and previous work in the study of tilt aftereffects and early vision in general. Chapter 3

---

simultaneously.

explains the RF-LISSOM system in detail, including the network architecture, activity calculation, and connection weight learning mechanisms. It also presents an overview of previous results with the RF-LISSOM model, and evaluates the biological plausibility of the model. Chapter 4 describes how the realistic cortical orientation map used in the aftereffect simulations was trained. Detailed explanations of the experimental settings such as the training schedule and training parameter values are given in this chapter. The resulting map is compared to anatomical and physiological data from humans and other mammals. Chapter 5 describes the aftereffect experiments and results using this orientation map, and demonstrates that the model closely reproduces the psychophysical data for the tilt aftereffect in humans. Chapter 6 further relates the results of the model to the human data, and speculates on the details of the biological mechanisms causing the observed effects. Some directions for future work are suggested, including an examination of tilt illusions, extensions that may be needed for the model to account for behavior at low contrasts, and studies of aftereffects in other modalities. Chapter 7 summarizes the major conclusions from this study. It is argued that this thesis presents the first detailed and convincing computational explanation for the tilt aftereffect, and that it does so within a very general and biologically-plausible self-organizing model of the afferent and lateral connections within the cortex.

Chapter 2

Related Work

This chapter will present an overview of previous experimental and theoretical work on the tilt aftereffect. First, a brief review of the visual system and visual processing areas of the brain will be given. Next, an outline of the research and theories on tilt aftereffects to date will be presented, followed by a discussion of previous mathematical and computational models of the TAE and related aspects of early visual processing.

2.1   Anatomy and physiology of the visual system

Figure 2.1 shows a diagram of the main feedforward pathway in the visual system of humans (see Wandell 1995 and Kandel et al. 1991 for an overview). Very similar structures are present in other mammals, including such well-studied laboratory animals as monkeys, cats, and rats. For each species, light entering the eye is detected by the *retina*, an array of photoreceptors on the inside of the rear surface of the eye. The photoreceptors and related circuitry in the retina encode light levels as electrical signals. These signals leave the eye through nerve cells called retinal ganglion cells. From there, they travel to the *lateral geniculate nucleus* (LGN) of the thalamus, at the base of each side of the brain. From the LGN, the signals travel to the *primary visual cortex* (V1) at the rear of the brain, the first stage of cortical processing of vision. The output from this area goes to many different higher areas of the brain, eventually contributing to visual perception in ways that are only beginning to be understood (Merigan and Maunsell 1993; Van Essen et al. 1992).

There is a consensus that significant information processing occurs in the retina and V1 (Kandel et al. 1991). Although they are often modeled as simple filters, the earliest stages of visual perception are perhaps better described by a *feature detector* model (Hubel and Wiesel 1959, 1965; Van Essen et al. 1992). Neurons at each level detect certain features, sending their outputs to higher levels for the detection of progressively more complex stimuli.

For instance, retinal ganglion cells in the eye typically respond most strongly to
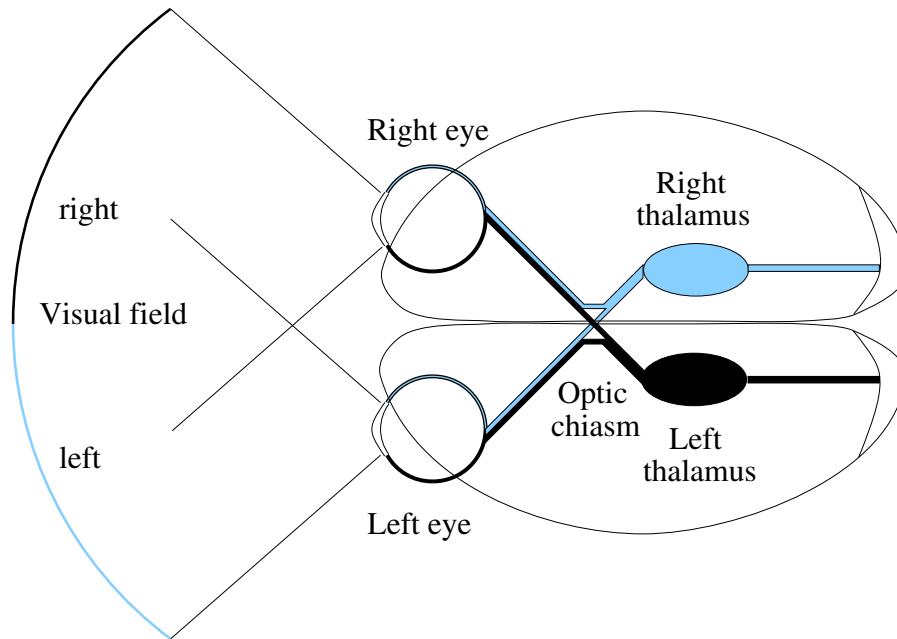
Figure 2.1: **Human visual sensory pathways (*top view*).** Visual information travels in separate pathways for each half of the visual field. Light entering the eye from the right hemifield hits the left half of the retina, on the rear surface of each eye. The inputs from each eye are combined at the optic chiasm, and travel to the left lateral geniculate nucleus (LGN) of the thalamus, then to the primary visual cortex (also called V1, visual area 1, striate cortex, or Brodmann area 17) of the left hemisphere. Light entering the eye from the left hemifield strikes the right half of each retina, which send signals to the right LGN and then to the right half of V1.

a circular light or dark spot in a particular area of the retina (its *receptive field*) (Dacey 1994; Kaplan 1989). Neurons in the LGN behave similarly (Casagrande and Norton 1989). Beginning in the primary visual cortex, most neurons are found instead to prefer inputs elongated in some particular direction (Hubel and Wiesel 1968). Such neurons respond most strongly to an oriented stimulus such as a line or an edge close to a preferred orientation. They respond less strongly the more the orientation differs from the preference, do not generally respond to unoriented inputs, and respond equally well over a wide range of contrasts. Thus V1 neurons are considered detectors for their preferred orientation, and neural explanations of orientation perception usually begin at this level of processing. Higher visual processing areas in the brain all use the output from this stage, eventually detecting more complex features such as human faces (Van Essen et al. 1992).

The primary visual cortex, like the other parts of the cortex, is composed of a two-dimensional, slightly folded sheet of neurons and connections between them. If flattened, human V1 would cover an area of nearly four square inches (Wandell 1995). It contains at least 150 million neurons, each making many hundreds of specific connections with other

neurons (Wandell 1995). The neurons are arranged in six layers with different anatomical characteristics (using Brodmann's scheme for numbering laminations in human V1; see Henry 1989 for more details). Input from the thalamus (the *afferent* input) typically goes to layer 4 (Casagrande and Norton 1989; Henry 1989). Neurons in the other layers form local connections within V1 or connect to other visual processing areas such as V2 (adjacent to V1).

At a given location on the cortical sheet, the neurons in a vertical section through the cortex generally have the same preference for the eye of origin, stimulus orientation, stimulus size, etc. It is customary to refer to such a section as a *column* (Gilbert and Wiesel 1989). The RF-LISSOM model discussed in this thesis will treat each column as a single unit, thus representing the cortex as a purely two-dimensional surface. This is only an approximation, but it is a valuable one because it greatly simplifies the analysis while retaining the basic functional features of the cortex.

Nearby columns generally have similar, but not identical, preferences, and slightly more distant columns generally have more dissimilar preferences. Preferences repeat at regular intervals (approximately 1–2 mm) in every direction. For orientation preferences, this arrangement of detectors forms a smoothly varying *orientation map* of the retinal input (Blasdel 1992a; Blasdel and Salama 1986; Grinvald et al. 1994; Ts'o et al. 1990; Weliky et al. 1995). (See figure 4.3*b* in chapter 4 for an example of an orientation map.) Each location on the retina is mapped to a region on the orientation map, with each possible orientation at that location represented by different orientation detectors. The global layout of the orientation map, and consequently the orientation preferences of the individual neurons in the map, is formed during the early development of the animal based on its visual experience (Blakemore and Cooper 1970; Blakemore and van Sluyters 1975; Hubel and Wiesel 1962, 1968; Movshon and van Sluyters 1981).

Extensive, long-range *lateral connections* are present between neurons in neighboring columns with similar preferences (figure 2.2; Gilbert and Wiesel 1983; Gilbert et al. 1990). The lateral connectivity is not uniform or genetically determined, but develops based on visual experience (Burkhalter et al. 1993; Dalva and Katz 1994; Fisken et al. 1975; Gilbert 1992; Katz and Callaway 1992; Löwel and Singer 1992). The connections are initially widespread, but develop into clustered patches at approximately the same time as the orientation maps form. The lateral connections are far more numerous than the afferents and they are believed to have a substantial influence on cortical activity. However, it is not well understood what mechanisms underlie the development of lateral connections, why their patterns are related to the response properties of cortical cells, or what their function is in visual processing.

Although the afferent structures and lateral connections are influenced by visual experience during development, after a critical period early in development they are not as
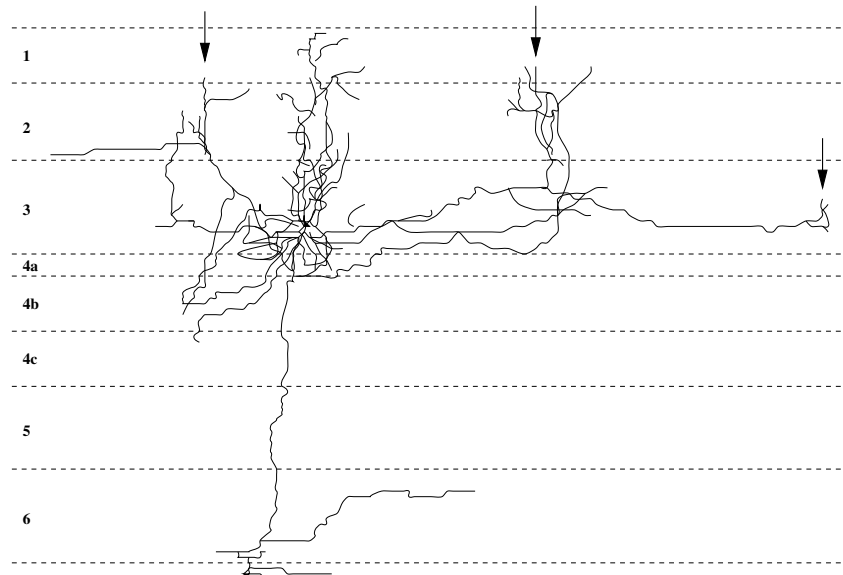
Figure 2.2: **Long-range lateral connections of a typical neuron.** The connections run over distances covering several degrees of the visual field, and sprout branches at intervals (marked by arrows). The branches form a local cluster of connections to other cells in the region. Such clusters occur only in regions with similar functional properties as the parent cell. Adapted from Gilbert and Wiesel (1989).

easily modified (Movshon and van Sluyters 1981). However, recent results show that the adult cortex can undergo significant, often reversible, reorganization in response to various sensory and cortical manipulations such as lesions in the receptive surface and the cortex (Gilbert 1992; Gilbert et al. 1996; Kaas 1991; Kapadia et al. 1994; Merzenich et al. 1990; Pettet and Gilbert 1992).

2.2 Experimental data on the TAE

Although much is known about the anatomy of the early visual areas such as V1, the mechanisms of visual perception are not yet fully understood. Studies of visual illusions and aftereffects can help provide a window into this processing, giving clues about how the underlying systems must be operating. The tilt aftereffect is a particularly important phenomenon to study in a model of vision, since it is generally thought to arise in V1, the earliest cortical processing stage. Accordingly, there have been many studies of the TAE since the first experiments published by Gibson and Radner in 1937.

7

(a) TAE adaptation procedure

(b) TAE test procedure
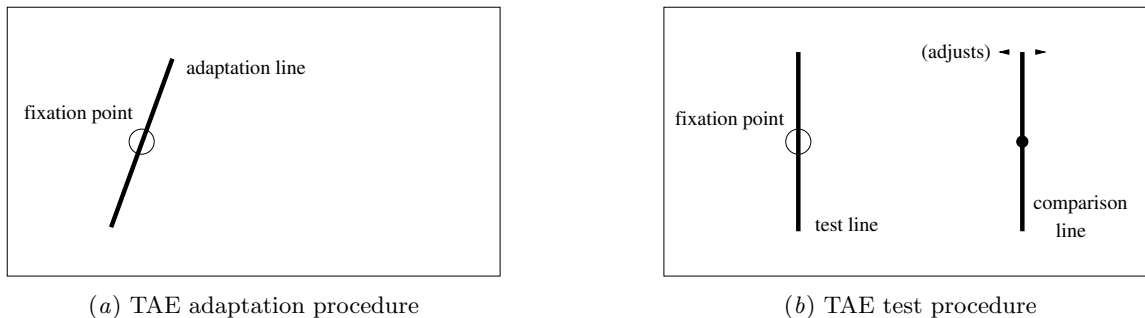
Figure 2.3: **Psychological paradigm for tilt aftereffect experiments.** The subject first adapts to a figure at a particular orientation presented for a fixed amount of time (a). The amount of TAE is determined by having the subject adjust the comparison figure to match the apparent orientation of the test figure (b). Some studies use lines or bars, as shown here; others show similar results for sinusoidal or square-wave gratings in this or similar configurations.

2.2.1 Measuring the TAE in humans

A variety of experimental paradigms have been used to measure tilt aftereffects in humans, but most are variations on the following procedure (figure 2.3). Subjects are tested in a dark room without any clues to orientation except gravity. The subject's head is immobilized, and the subject is directed to look at some fixation point in the visual field. Oriented geometric figures are used in most cases, generally lines, bars, or gratings (i.e., multiple parallel bars, as in figure 1.1). The figures are presented in various locations relative to the fixation point. First, a baseline is determined by presenting a *test* figure at a given angle. The perceived orientation of that figure for the subject is determined by various methods, such as:

Directly estimating the numerical angle. This is generally considered unreliable since most subjects are poor judges of angles other than vertical and horizontal relative to the retina (Mansfield 1974; Mitchell and Muir 1976).

Setting the figure to apparent vertical. This can be done with accuracy within a degree by most unadapted subjects (Howard and Templeton 1966,p.179).

Setting a *comparison* figure parallel to the test figure. This can be done with fairly high consistency, although it is not necessarily an accurate measure of the actual orientation. It is particularly useful because it can measure the effect for oblique testing angles.

After the baseline is computed, the subject views an *adaptation* figure of a different orientation for a fixed amount of time; this period is called tilt adaptation. Finally, the test figure is presented again, and its orientation measured once more. The difference between

8
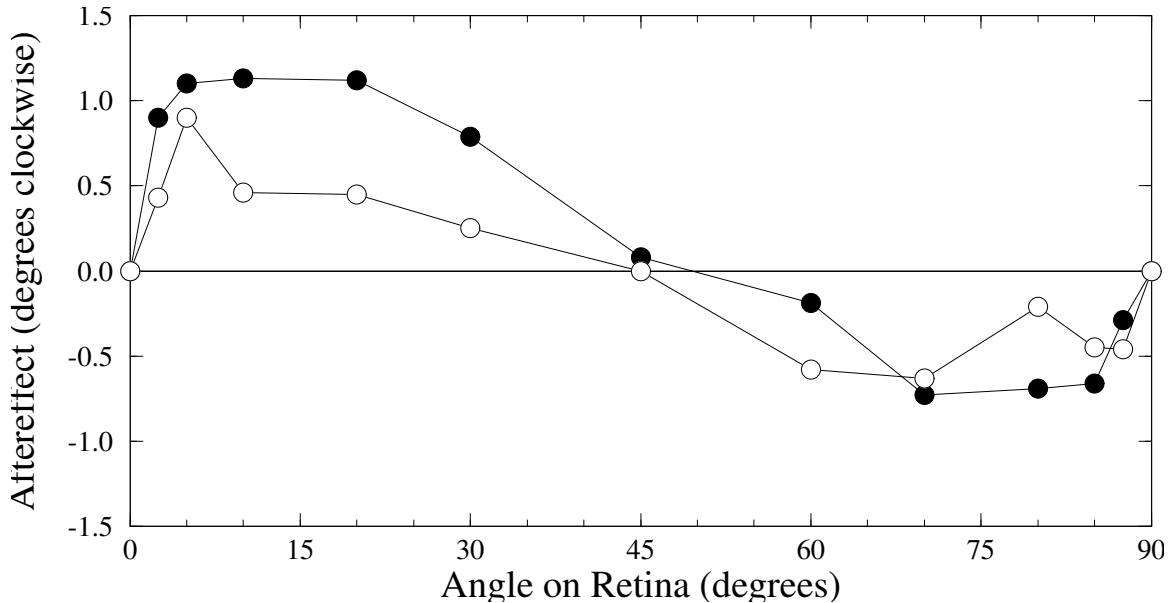
Figure 2.4: **Original study of tilt aftereffect versus angle separation on retina.** The two lines shown indicate the tilt aftereffect for the two different observers in Gibson and Radner (1937). Each point represents an average of 6 trials. Each trial consisted of a 45-second adaptation to a tilted line, then an adjustment of a nearly vertical test line until it appeared vertical. The difference between the perceived orientation and the actual orientation of the test line was taken as the tilt aftereffect at that angle. The graph shows both direct and indirect effects, with expansion of small angles and contraction of large angles.

the final and initial readings of the perceived orientation of the test figure is the value of the tilt aftereffect for that particular angle separation between test and adaptation figures (Campbell and Maffei 1971; Gibson and Radner 1937; Mitchell and Muir 1976; Muir and Over 1970).

2.2.2   Human psychophysical data

The TAE was first documented in detail by Gibson and Radner (1937), who measured the angular function (figure 2.4) and time course of the aftereffect. The precise shape of the TAE versus angle curve varies widely for different subjects and different measurement paradigms, but generally retains the S-shape seen in figure 2.4.

Gibson and Radner introduced terminology that describes the aftereffect as composed of a *direct* effect and an *indirect* effect (figure 2.5). The direct effect consists of a perceived expansion of small angles, i.e. the positive half of the 'S' in figure 2.5. A maximum overestimation of a few degrees is experienced at an angular separation between 5° and 20° (Howard and Templeton 1966,p.216). Larger angles begin to show less of an after-
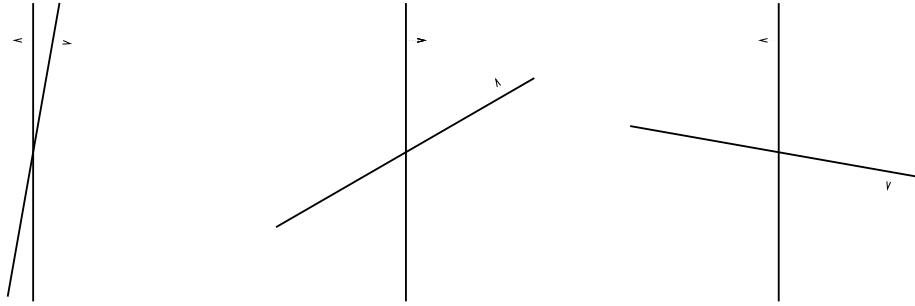
Figure 2.5: **Angle expansion and contraction effects.** Arrows indicate the direction of change of the perceived orientation of that line after adaptation on the other; the effects are equal in both directions marked (Mitchell and Muir 1976). The direct effect (left) consists of a perceived expansion of small angles, such as the one shown. The indirect effect (center) consists of a perceived contraction of larger angles. The maximum angle separation is 90°, since if two endpoints are separated by larger angles than 90°, another two form an angle smaller than 90° (right).

effect, eventually reaching zero somewhere between 25° and 50° (Campbell and Maffei 1971; Mitchell and Muir 1976; Muir and Over 1970). Even larger angles (up to 90°) generally result in the indirect effect, which consists of a smaller perceived *contraction* of the angle, peaking between 60° and 85° (Campbell and Maffei 1971; Mitchell and Muir 1976; Muir and Over 1970). Note that for overlapping lines, the largest possible separation is 90° due to symmetry (figure 2.5).

Early studies found tilt aftereffects only for horizontal or vertical test lines (Gibson and Radner 1937), as did some later studies (Campbell and Maffei 1971). However, it has since been demonstrated that although subjects do not spontaneously report any difference in the perceived orientation of an oblique line (i.e., neither horizontal nor vertical) after adaptation, when asked to set a comparison line parallel to the test line they make systematic errors identical to those made on vertical or horizontal test lines (Mitchell and Muir 1976). The variance of the settings was significantly larger for the oblique case, which contributed to the difficulty of establishing the effect for oblique lines. The current consensus is that the angular separation between the test and adaptation figures is the important parameter, not the obliqueness of the line with respect to the horizontal or vertical axis.

The magnitude of the TAE increases logarithmically with increasing adaptation time and decreases logarithmically with time elapsed since the adaptation period (Gibson and Radner 1937). However, the magnitude of the effect saturates at approximately 4° (Campbell and Maffei 1971; Greenlee and Magnussen 1987; Magnussen and Johnsen 1986; Mitchell and Muir 1976). The maximum indirect effect documented in central (foveal) vision is approximately 2.5°, and reaches up to about 60% of the magnitude of the direct effect for a given subject and paradigm (Campbell and Maffei 1971; Mitchell and Muir 1976; Muir and Over 1970). A preliminary report indicated that only a direct effect is seen for

all angles 0° to 90° when the test and adaptation figures are in peripheral vision (Muir and Over 1970), but this data has not yet been replicated by other laboratories.

The TAE is localized to what appear to be a specific set of orientation detectors. It is spatially localized on the retina (Gibson and Radner 1937); adaptation for a figure in one location has no measurable effect on test figures in other locations sufficiently distant. It is spatially selective for frequency (Ware and Mitchell 1974); adapting on a figure with narrow bars has no measurable effect upon a figure with wide bars, and vice versa. Finally, it transfers completely from one eye to the other (Campbell and Maffei 1971; Gibson and Radner 1937); adapting one eye causes equal effects upon test lines in the same location in the visual field for either eye.

2.3   Proposed theories

2.3.1   Gibson normalization

Gibson and Radner (1937) hypothesized that the TAE results from a process they called *normalization*, wherein the subjective vertical and horizontal norms for a subject are modified by visual experience. After prolonged inspection of a nearly-vertical stimulus, the subjective vertical was thought to be shifted slightly towards the fixation stimulus. If an objectively vertical line were presented subsequently, it would be perceived as slightly tilted in the opposite direction relative to the new subjective vertical. Indirect effects would be due to a constraint that the horizontal and vertical axes remain perpendicular, so that lines close to the horizontal axis would appear shifted relative to that axis. The smaller magnitude of indirect effects was attributed to some degree of "play" between the axes, i.e. that the linkage was imperfect.

One consequence of the normalization theory is that no effects should be seen when vertical or horizontal adaptation lines are used, since such lines cannot change the supposed vertical and horizontal norms in either direction. More recent experiments have shown that this assumption is invalid, since similar effects occur regardless of the adaptation angle (Mitchell and Muir 1976). Another likely consequence is that tilt effects should be zero at precisely 45° (the midpoint between the two axes), as Gibson and Radner (1937) found for their subjects and setup. However, subsequent studies have found that the crossover point between direct and indirect effects varies significantly from 45°, over a range of 25° to 50° (Campbell and Maffei 1971; Mitchell and Muir 1976; Muir and Over 1970).

2.3.2   Satiation

Later researchers were unsatisfied with the clearly *ad hoc* explanation of Gibson's normalization theory, and proposed alternatives (see Howard and Templeton 1966 for a review).

One of the most prominent theories (Köhler and Wallach 1944) postulated that the TAE is an example of a larger class of figural aftereffects. For each of these, activation of cortical sensory areas was presumed to result in local electrical fields that spread electrotonically to nearby areas. Activated areas experienced satiation, which would result in displacement of subsequent perceptions to unsatiated areas. This metaphor was apparently inspired by the beginnings of electronic technology at the time; no physiological evidence for fields of this nature has ever been found (Barlow 1990). In addition, this theory entirely fails to account for the indirect effect, since satiation would result in expansion only.

2.3.3   Feature-detector fatigue

The *feature-detector* model of the visual cortex based on the pioneering work of Hubel and Wiesel (1959, 1962, 1965, 1968) led to theories formulated in terms of broadly-tuned orientation-specific detectors. One intriguing finding of this work was that individual orientation detectors become more difficult to excite during repeated presentation of appropriately oriented stimuli, and this desensitization persists for some time afterwards. This apparently supports a model like that of Köhler and Wallach (1944): if neurons with orientation preferences close to the adaptation figure become *fatigued* as a result of activation, a different set of neurons will be activated for the test figure. Assuming the perceived orientation is some sort of weighted average over the orientation preferences of the activated neurons, the perceived orientation would thus show the direct TAE (Coltheart 1971).

Sutherland (1961) proposed such an explanation originally, and Coltheart (1971) extended it to account for indirect effects. Coltheart noted that some neurons (here called cross-neurons) had been discovered that had *two* preferred orientations, each orthogonal to the other (Hubel and Wiesel 1965, 1968). While an adaptation figure is being shown, some cross-neurons (i.e., those having one of their axes matching the figure) would desensitize in the same way as the simpler neurons whose orientation preference matched. However, subsequent figures having orientations nearly 90° from the adaptation figure would activate fewer cross-neurons than otherwise. This would effectively shift the perceived orientation of the test figure *towards* that of the adaptation figure, since it will result in less activation of neurons having orientation preferences even further from the adaptation figure than the test figure was.

Muir and Over (1970) proposed an alternative theory of indirect effects within this framework. It had been observed that when an oriented figure is presented, neurons with orientation preferences orthogonal to that of the figure actually have *less* activity than their resting level (Hubel and Wiesel 1967). Muir and Over (1970) proposed that these neurons would not only *not* be fatigued after adaptation, they would be actively facilitated, i.e. somehow become more susceptible to future stimulation. This could cause the indirect effect for lines near their orientation preference.

12

The feature-detector fatigue theory could apparently account for much of the observed phenomena with either extension for indirect effects. However, it is now discounted for a number of reasons, including the following:

- The proposed fatigue mechanism has not been found to occur in single cells *in vitro* — individual cortical neurons can be activated repeatedly with direct electrical simulation, without decreasing their output or showing less input sensitivity (Thomson and Deuchars 1994).

- Adaptation still occurs when local application of the inhibitory transmitter GABA prevents the cell from firing (Vidyasagar 1990), indicating that firing of the target cell is not necessary for adaptation to occur.

- Adaptation does not occur when a cell is made to fire by local application of the excitatory transmitter glutamate (Vidyasagar 1990), indicating that firing is not sufficient to cause adaptation.

- Adaptation to two orientations simultaneously, each of which would cause adaptation separately, can actually reduce the amount of the TAE, rather than increase it as the fatigue theory would predict (Magnussen and Kurtenbach 1980a). This effect is known as disinhibition, and is also discussed in section 2.3.4.

Together, the above factors imply that the observed desensitization must somehow depend on the activation of multiple neurons, including inhibitory interneurons, not merely on changes within the neuron itself.

2.3.4 Modifiable lateral inhibition

Direct effect

Other researchers have noted that an increase in lateral and local inhibition via the connections *between* neurons could have very similar results to a decrease in the inherent excitability of the neuron (Blakemore et al. 1970; Carpenter and Blakemore 1973). Since typical test figures activate a large number of nearby neurons, increasing the strength of the inhibitory connections between them would also produce the observed desensitization. Such an explanation also accounts for similar effects between nearby stimuli presented simultaneously, known as the tilt illusion (Carpenter and Blakemore 1973). The two tilt effects have very similar characteristics along a number of different dimensions, so related explanations have generally been expected (Magnussen and Kurtenbach 1980b; Tolhurst and Thompson 1975).

The inhibition theory also correctly accounts for the phenomenon of *disinhibition* documented for the tilt illusion (Carpenter and Blakemore 1973) and the tilt aftereffect

13

(Magnussen and Kurtenbach 1980a). This phenomenon arises for the TAE when multiple adaptation figures with different orientations are presented simultaneously. For a given test line, two figures which each cause direct tilt aftereffects would be expected to cause an even stronger TAE when presented together during adaptation, if the fatigue theory were correct (Blakemore and Carpenter 1971)). This result follows simply from having more detectors activated, and thus more detectors fatigued. However, such adaptation actually *reduces* the amount of the TAE (Magnussen and Kurtenbach 1980a). The reduction might be explained by the inhibition theory — the two adapting figures would inhibit the response to each other during simultaneous presentation, and thus the total activation of the orientation detectors near the test line would be smaller than it would be when either line is presented alone. This would cause less adaptation than before, so the TAE will be lower as observed.

Furthermore, there is a wealth of evidence for extensive, modifiable lateral connections (Gilbert et al. 1990; Gilbert and Wiesel 1990; Hirsch and Gilbert 1991; McGuire et al. 1991; Weliky et al. 1995). However, as discussed in the next chapter, the detailed behavior of these lateral connections is still a matter of some dispute, particularly with respect to whether they are predominantly inhibitory or excitatory. In any case, the currently accepted explanation of the direct TAE is in terms of the inhibitory effects.

Current theories of the TAE are typically vague about exactly what type of mechanism is causing prolonged inhibition to result in the direct TAE. Most seem to assume that some sort of intracellular buildup of inhibitory transmitters occurs in the target cell (Gelbtuch et al. 1986; Masini et al. 1990; Tolhurst and Thompson 1975). However, the experiments by Vidyasagar (1990) suggest that this simple explanation will not suffice. Vidyasagar was able to excite or inhibit cells in the visual cortex of cats by locally applying excitatory or inhibitory transmitters, but regardless of the duration of the excitation or inhibition the cells did not show adaptation effects when tested with a visual pattern. Yet the cells showed clear signs of adaptation when a visual pattern was used instead, which prompted the neighboring cells to deliver those same transmitters to the target cell. Thus, unless there is a large gap in our knowledge of the neurophysiology of these cells, the adaptation must be occurring elsewhere, not within the target cell. As suggested by Barlow (1990), this thesis will argue that the changes that occur during adaptation are changes in the strength of connections between neurons. These effects will thus only be seen when multiple nearby neurons are activated simultaneously, as they would be for the patterns typically used (cf. Vidyasagar 1990; Wilson and Humanski 1993).

Indirect effect

Unlike the direct effect, the indirect effect has not been examined by very many researchers within the context of the inhibition theory. The theory itself does not account for indirect effects in any obvious way, since increases in inhibition around the training orientation can

only repel other lines, other things being equal. This difficulty appears to have discouraged research on the topic so far, but as this thesis will show, indirect effects can be explained entirely within the inhibition theory.

So far, the most prominent explanation is from Wenderoth and Johnstone (1988); Wenderoth et al. (1989), who have proposed that indirect effects arise from lateral inhibitory repulsion from a *virtual axis* of the testing line. According to this view, the axes of symmetry of an object can contribute to the perception of it, even if they are not visible. A line (or grating) is symmetrical about two axes, one of which is the line itself, the other of which is a line perpendicular to it. The axis nearest the testing figure is assumed to have the greatest repulsive effect on the figure. Thus direct effects would result (as described above) from repulsion from the actual line, and indirect effects would result from repulsion from the virtual axis. The smaller magnitude of the indirect effects would presumably result from the lower perceptual saliency of virtual axes (Wenderoth et al. 1989). Other differences between the indirect and direct effects have been found in a long series of experiments (Wenderoth and Johnstone 1988; Wenderoth et al. 1989), most of which indicate that indirect effects develop later in time than the direct effects. Wenderoth et al. interpret these differences as evidence that the processing of indirect effects occurs higher in the visual hierarchy, i.e. that the neural substrate for the virtual axis is in areas higher than V1.

The virtual axis explanation does not appear to have been examined critically by other researchers, perhaps because it appears untestable and nearly as *ad hoc* as that originally proposed by Gibson and Radner (1937). To make it somewhat more concrete, Spivey-Knowlton (1993) has proposed that the virtual axis results from the cross-neurons described by Coltheart (1971). However, as Spivey-Knowlton himself notes, those neurons are present in V1 as well as in the other early areas. Thus if those neurons represent the virtual axis, they would have to be inherently slower to activate than other V1 neurons, since the indirect effects have been shown to have a later onset (Wenderoth and Johnstone 1988; Wenderoth et al. 1989). Yet De Valois et al. (1982) found no significant differences in the time behavior of cross-neurons relative to other V1 neurons, so these neurons must not represent the substrate for the virtual axis.

Thus the virtual axis remains an unspecified higher-level process, essentially untestable although possible in theory. In contrast, the explanation for the indirect effect that will be described in this thesis is based on entirely local synaptic resource conservation mechanisms within V1. These mechanisms have already been found elsewhere in the cortex, and they follow from general computational principles, while the virtual axis appears to have been invented purely to explain indirect tilt effects.

2.4 Computational and mathematical models

Computational models can play a fundamental role in understanding the development and function of complex systems such as the cortex. With the introduction of massively parallel computers in the last five years, it has become possible to simulate large numbers of neural units and their connections. At the same time, neurobiological techniques for mapping the response properties and connectivity of neurons have become sophisticated enough to constrain and validate such models. This provides a timely opportunity to test the inhibition theory of the TAE through large-scale computational experiments. By iteratively making and testing hypotheses about which cortical features are responsible for which behaviors, one can extend and revise a computational model so that it can predict cortical behavior in detail. At each step of the way, the model itself can provide predictions for verification or refutation in human and animal subjects. The overall goal is to represent the essential computational and organizational principles of the cortex in as simple a model as possible.

2.4.1 Previous models of orientation maps

To test the inhibition theory, a model of the orientation detectors is needed. Several computational models have shown how receptive fields (such as selectivity to different orientations) and their global organization in the cortical network can develop through Hebbian self-organization of afferent synapses (Erwin et al. 1995; Goodhill 1993; Kohonen 1982; Miller 1994; Miller et al. 1989; Obermayer et al. 1990; von der Malsburg 1973). These models have not taken the lateral interactions between cells into account, or have assumed that they are preset and fixed and have a regular profile. Thus they are unsuitable for testing the inhibitory theory of tilt aftereffects, since it depends upon modifiable connections between specific orientation detectors.

2.4.2 Previous models of tilt illusions and aftereffects

Of the models that have been suitable for examining tilt aftereffects, none have successfully accounted for both direct and indirect effects in a biologically realistic manner. The virtual axis/lateral inhibition theory of Wenderoth et al. (1989) has been implemented in a very simple computational model by Spivey-Knowlton (1993). He found that the model could produce an S-shaped curve for the angular function of the tilt illusion; the curve was a reasonably good match to human data. Presumably a similar model could be used for the tilt aftereffect, but this has not yet been implemented. However, since the virtual axis theory was explicitly developed as an explanation for the indirect tilt illusion data, demonstrating it in a computational model only serves to verify that the theory is internally consistent, rather than supporting or explaining it. The model to be described in this thesis, in contrast,

was developed primarily as a model of self-organization of cortical structures, and exhibits tilt aftereffects only as an emergent phenomenon.

Another mathematical model for neural development by Dong (1995, 1996) has also been shown to exhibit the tilt illusion and tilt aftereffect. The model is a formal expression of the information processing principle of *decorrelation*, that is, the reduction of redundancy in an image or other sensory data. All natural images are redundant to some extent, and the nervous system appears to make wide use of this fact (Barlow 1990). The amount of redundancy may be expressed in terms of correlations: two pieces of information that are highly correlated with each other could in theory be represented as a single entity, and thus they contain redundant information. A network or algorithm that compacts the image by removing such redundancies is said to *decorrelate*. After full decorrelation, an input image would consist of purely white noise, where every bit of information is independent of each of the others. Dong shows that a network that satisfies a simple approximation to full decorrelation will exhibit both direct and indirect tilt aftereffects for Gaussian-shaped inputs. The model predicts an S-shaped curve that is a reasonably good approximation to the average human TAE data from Campbell and Maffei (1971). Dong's theory may be seen as a mathematical expression of some of the principles also operating in the RF-LISSOM model; it is complementary to the more detailed column-level description of actual neural behavior presented in later chapters of this thesis.

Finally, Wilson and Humanski (1993) proposed a linear differential equation model of a cortical gain control mechanism that exhibits direct tilt aftereffects. Their model is a proposed circuit for achieving contrast independence via a divisive inhibitory gain control. The model had been shown to predict contrast sensitivity before and after adaptation, and with the same parameters it was found to exhibit direct tilt effects that have an angular function somewhat similar to the measured direct effects for humans. The same type of contrast gain control is exhibited by RF-LISSOM (Sirosh 1995), so one can consider their model to represent one part of the cortical processes modeled by RF-LISSOM. However, they do not discuss indirect effects, and it is difficult to see how those would occur in their model. Thus this model, like the others, does not fully account for both direct and indirect tilt aftereffects in terms of actual known cortical structure.

2.5   Conclusion

Although the tilt aftereffect has been studied for many years, a fully satisfying explanation has yet to be established. The prevailing theory is based on lateral inhibitory interactions between orientation detectors in the primary visual cortex. This theory can account for the angle repulsion effects found at small angle differences between test and adaptation figures. However, the angle attraction effects found for larger angle differences have not

yet been explained satisfactorily. A detailed neural model that includes modifiable lateral connections, such as that presented in the next chapter, allows the lateral inhibitory theory to be studied in more detail. Subsequent chapters will show that tilt aftereffects, both direct and indirect, arise out of the same processes responsible for self-organization of such a model of orientation detectors in the primary visual cortex. Unlike TAE models which are merely a mathematical fit to the data, the model in this thesis is a general explanation of the organization and function of the primary visual cortex.

Chapter 3

The RF-LISSOM Model

3.1 Introduction

This thesis is a first functional study of the RF-LISSOM model (Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map) developed by Sirosh and Miikkulainen (1994a, 1996, 1997). This chapter describes the architecture of the model in detail, summarizes previous results using RF-LISSOM to model the cortex, and examines the biological underpinnings of the model. Later chapters will show how this basic architecture can account for the psychological phenomenon of the tilt aftereffect.

The RF-LISSOM model is a descendent of computational models developed by von der Malsburg (1973; see also Amari 1980 and Grossberg 1976). These early models demonstrated that simple computational rules could account for the development of oriented receptive fields from visual input. Since these models, new discoveries about intra-cortical connectivity have fundamentally altered our understanding of the primary visual cortex (Gilbert and Wiesel 1983; Gilbert et al. 1990). In addition to the afferent input leading from the eye to cortical areas, there are highly-specific patterns of lateral connectivity that develop in response to visual experience (Burkhalter et al. 1993; Dalva and Katz 1994; Fisken et al. 1975; Gilbert 1992; Katz and Callaway 1992; Löwel and Singer 1992). RF-LISSOM incorporates this new evidence into a computational model of structure and function.

In previous work with the RF-LISSOM model, Sirosh and Miikkulainen (1994a, 1996, 1997) showed that Hebbian self-organization in a large recurrent network of simple, laterally-connected neural elements can provide a unified account of self-organization and plasticity in the visual cortex. They demonstrated computationally (1) how receptive fields develop selectivity to orientation, ocular dominance, and size, (2) how such receptive fields organize into global structures of intertwined columnar areas, (3) how the lateral connections develop synergetically with the afferent connections and follow their global organization, and (4)
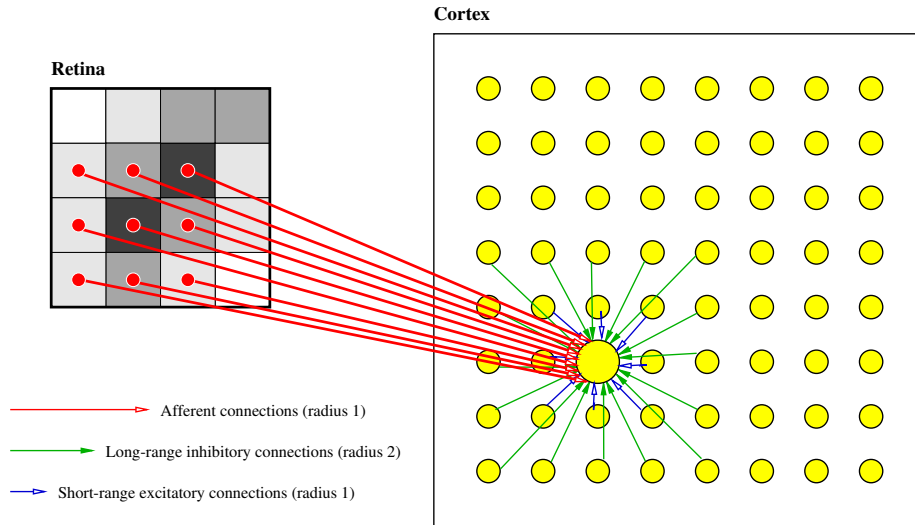
Figure 3.1: **Architecture of the RF-LISSOM network (*color figure*).** A tiny RF-LISSOM network and retina are illustrated, along with connections to a single neuron (shown as a large circle). Other neurons have similar connections to other areas of the retina and to the neurons surrounding them. The afferent connections form a local anatomical receptive field on the simulated retina. Neighboring neurons have different but highly overlapping RFs. Each neuron computes an initial response as a dot product of its receptive field and its weight vector. The responses then repeatedly propagate within the cortex through the lateral connections (only a very small connection radius is shown) and evolve into an activity "bubble". After the activity stabilizes, weights of the active neurons are adapted.

how such structures are maintained in a dynamic equilibrium with the input, resulting in reorganization after retinal and cortical lesions. The model also suggests a functional role for the lateral connections: during development, they learn the activity correlations between cortical neurons, and during visual processing, filter out these correlations from cortical activity to form a redundancy-reduced sparse coding of the visual input.

3.2   The RF-LISSOM Architecture

The RF-LISSOM model is based on a simulated network of neurons with afferent connections from the external world and recurrent lateral connections between neurons. Connections adapt based on correlated activity between neurons. The result is a self-organized structure where afferent connection weights form a map of the input space, and lateral connections store long-term correlations in neuronal activity.

In RF-LISSOM, the cortical architecture has been simplified and reduced to the minimum necessary configuration to account for the observed phenomena. Because the focus is on the two-dimensional organization of the cortex, each "neuron" in the model

20

cortex corresponds to a vertical column of cells through the six layers of the human cortex. This columnar organization helps make the problem of simulating such a large number of neurons tractable, and is viable because the cells in a column generally fire in response to the same inputs (chapter 2). Thus RF-LISSOM models biological mechanisms at an aggregate level, so it is important to keep in mind that RF-LISSOM "neurons" are *not* strictly identifiable with single cells in the human cortex.

The cortical network is modeled with a sheet of interconnected neurons and the retina with a sheet of retinal ganglion cells (figure 3.1). Neurons receive afferent connections from broad overlapping patches on the retina. The $N \times N$ network is projected on to the retina of $R \times R$ ganglion cells, and each neuron is connected to ganglion cells in a circular area of radius r around the projections. Thus, neurons at a particular cortical location receive afferents from the corresponding location on the retina. Depending on its location, the number of afferents to a neuron varies from roughly $r \times r$ (at the corners) to $2r \times 2r$ (at the center).

In addition, each neuron has reciprocal excitatory and inhibitory lateral connections with itself and other neurons. Lateral excitatory connections are short-range, connecting each neuron and its close neighbors. Lateral inhibitory connections run for comparatively long distances, but also include connections from the neuron and its immediate neighbors to itself. Thus the "lateral" connections in the model are not exclusively from neurons located laterally, since they include self-recurrent connections.

The input to the model consists of 2-D patterns of activity representing retinal ganglion cell activations. Each ganglion cell is modeled only by its activation levels, not by its receptive field, so the input pattern is equivalent to an image after it has been processed by the retina. In addition, the transformation of retinal activation patterns by the LGN has been bypassed for simplicity, since LGN neurons do not change the shape of the receptive fields of the retina (chapter 2). Thus the "retina" of the model could equivalently be considered to represent a pattern of activity across neurons in the retinotopic map of the LGN.

The RF-LISSOM network will self-organize to represent the most common features present in the input images it has seen (Sirosh 1995). Since tilt aftereffects appear to arise in the areas processing oriented inputs (chapter 2), simple oriented inputs (two-dimensional Gaussians) were used in the experiments presented in this thesis, as described in more detail in chapter 4.

The weights are initially set to random values or a smooth distribution (such as a Gaussian profile), and are organized through an unsupervised learning process. At each training step, neurons start out with zero activity. The initial response η_{ij} of neuron (i, j)

activity level

1.0

0

threshold   ceiling

net input

0
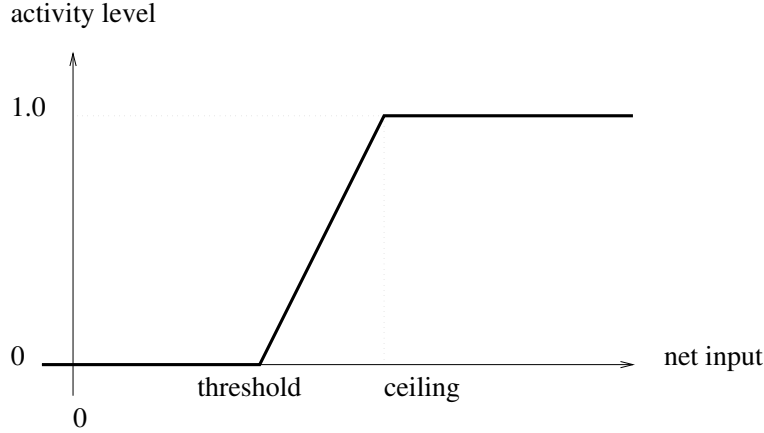
Figure 3.2: **The RF-LISSOM neuron activation function.** The neuron requires an input as large as the threshold δ before responding, and saturates at the ceiling β. The output activation values are limited to $[0, 1]$. The activation function is an easy-to-compute approximation of the sigmoid function.

is calculated as a weighted sum of the retinal activations:

$$\eta_{ij} = \sigma \left(\sum_{a,b} \xi_{ab} \mu_{ij,ab} \right), \tag{3.1}$$

where ξ_{ab} is the activation of retinal ganglion (a, b) within the anatomical RF of the neuron, $\mu_{ij,ab}$ is the corresponding afferent weight, and σ is a piecewise linear approximation of the sigmoid activation function (figure 3.2):

$$\sigma(x) = \begin{cases} 0 & x \le \delta \\ (x - \delta)/(\beta - \delta) & \delta < x < \beta \\ 1 & x \ge \beta \end{cases} \tag{3.2}$$

The weighted sum in equation 3.1 is a measure of similarity of the afferent input to the weight vector. The sigmoid introduces a nonlinearity into the response, and makes the neuron selective to a small range of input vectors that are close to the afferent weight vector.

Lateral interaction in the cortex sharpens neuronal response by repeated exchange of activation (Kohonen 1989; Mountcastle 1968). In RF-LISSOM, the initial response evolves over a very short time scale through lateral interaction. At each time step, the neuron combines the above afferent activation $\sum \xi \mu$ with lateral excitation and inhibition:

$$\eta_{ij}(t) = \sigma \left(\sum \xi \mu + \gamma_e \sum_{k,l} E_{ij,kl} \eta_{kl}(t - 1) - \gamma_i \sum_{k,l} I_{ij,kl} \eta_{kl}(t - 1) \right), \tag{3.3}$$

22

where $E_{ij,kl}$ is the excitatory lateral connection weight on the connection from neuron (k, l) to neuron (i, j), $I_{ij,kl}$ is the inhibitory connection weight, and $\eta_{kl}(t - 1)$ is the activity of neuron (k, l) during the previous time step. All connection weights are positive. The scaling factors γ_e and γ_i determine the relative strengths of excitatory and inhibitory lateral interactions. While the cortical response is settling, the retinal activity remains constant.

The activity pattern starts out diffuse and spread over a substantial part of the map, but within a few iterations of equation 3.3, converges into a small number of stable focused patches of activity, or activity bubbles. This sharpens the contrast between areas of high and low activity, helping it become focused around the maximally responding area.

After the activity has settled, the connection weights of each neuron are modified. Both afferent and lateral weights adapt according to the same mechanism: the Hebb rule, normalized so that the sum of the weights is constant:

$$w_{ij,mn}(t + \delta t) = \frac{w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}}{\sum_{mn} [w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}]}, \tag{3.4}$$

where η_{ij} stands for the activity of neuron (i, j) in the final activity bubble, $w_{ij,mn}$ is the afferent or lateral connection weight (μ, E or I), α is the learning rate for each type of connection (α_A for afferent weights, α_E for excitatory, and α_I for inhibitory) and X_{mn} is the presynaptic activity (ξ for afferent, η for lateral). Afferent inputs, lateral excitatory inputs, and lateral inhibitory inputs are normalized separately.

Following the Hebbian principle, the larger the product of the pre- and post-synaptic activity $\eta_{ij} X_{mn}$, the larger the weight change. Therefore, when the pre- and post-synaptic neurons fire together frequently, the connection becomes stronger. Both excitatory and inhibitory connections strengthen by correlated activity; normalization then redistributes the changes so that the sum of each weight type for each neuron remains constant. In effect, such a rule distributes the weights of each neuron η_{ij} in the proportion of its activity correlations with other neurons η_{kl}, $k, l = 1..N$.

At long distances, very few neurons have correlated activity and therefore most long-range connections eventually become weak. The weak connections can be eliminated periodically by the researcher; through the weight normalization, this will concentrate the inhibition in a closer neighborhood of each neuron. The radius of the lateral excitatory interactions starts out large, but as self-organization progresses, it is decreased (according to a schedule set by the researcher) until it covers only the nearest neighbors. Such a decrease is necessary for global topographic order to develop and for the receptive fields to become well-tuned at the same time (for theoretical motivation for this process, see Kohonen 1982, 1989, 1993; Obermayer et al. 1992; Sirosh and Miikkulainen 1997; for neurobiological evidence, see Dalva and Katz 1994; Hata et al. 1993.) Together the pruning of lateral connections and decreasing excitation range produce activity bubbles that are gradually more focused and local. As a result, weights change in smaller neighborhoods, and receptive fields become

better tuned to local areas of the retina.

3.3   Previous work with the RF-LISSOM model

The RF-LISSOM model has been used to examine a number of cortical phenomena. The experiments show how the observed organization of feature detectors and lateral connections in the primary visual cortex could form based on activity-dependent self-organization, driven by the regularities in the input (Miikkulainen et al. 1997; Sirosh and Miikkulainen 1995, 1996; Sirosh et al. 1996). As will be described in more detail in chapter 4, when the model is trained on retinal input patterns consisting of elongated Gaussian spots, it develops orientation columns organized into realistic orientation maps. When trained on Gaussian spots at slightly different positions on two separate receptive surfaces, the model develops realistic ocular dominance columns (areas favoring each eye). When trained on Gaussian spots of different sizes, size-selective columns develop much like those starting to be found in the cortex. In all these cases, the lateral connectivity patterns are found to follow the receptive field properties, as found in the cortex. When computational resources permit, future simulations will examine self-organization of all these parameters simultaneously. Such simulations would be trained on Gaussians varying along all of these dimensions, or on natural images processed by a model of the retina.

In addition to these developmental simulations, RF-LISSOM has been used to model cortical plasticity in the adult brain. The fundamental hypothesis is that the cortex is a continuously adapting structure in a dynamic equilibrium with both the external and intrinsic input. This equilibrium is maintained by cooperative and competitive lateral interactions within the cortex, mediated by lateral connections. As a test of this hypothesis, simulated cortical and retinal lesions were made in the model and they were shown to result in reorganization similar to that seen in monkey cortex (Miikkulainen et al. 1997; Sirosh and Miikkulainen 1994b; Sirosh et al. 1996). This demonstrated that the self-organizing principles may still be operating in the adult brain during recovery from trauma. This thesis extends those results to show that self-organization may even influence the behavior of the intact adult brain, during normal visual processing.

3.4   Sparseness and decorrelation

Since the model has shown that it organizes like the cortex, one can examine in more detail what the goal of this organization might be. Field (1994, 1987) and Barlow (1972) have suggested that the RFs in the primary visual cortex act as filters that form a sparse coding of the visual input. Sparse codes minimize the number of active neurons in the cortex, and are well suited for the detection of suspicious coincidences, pattern recognition,

associative memory and feature grouping (Field 1994, Barlow 1972, 1985, 1990). Prior work has demonstrated that the coding of visual input produced by the RF-LISSOM model is sparse in this sense (Miikkulainen et al. 1997; Sirosh and Miikkulainen 1996, 1997; Sirosh et al. 1996). The self-organized lateral connections have proven to be crucial for reducing redundancies to achieve this coding.

By Hebbian self-organization, the lateral connections in the model learn correlations between the feature-selective cells. The stronger the correlation between two cells' activity has been in the past, the larger the connection strength between them. Because the long-range connections are inhibitory, strongly correlated regions of the network inhibit each other. At the same time, the short-range lateral excitation locally amplifies the responses of active neurons. As will be seen in chapter 4, the recurrent excitation and inhibition focuses the activity to the neurons best tuned to the features of the input stimulus, thereby producing a sparse coding of the input. This same process manifests itself in the model as tilt illusions and aftereffects, as will be shown in chapter 5.

3.5 Biological basis of the model

RF-LISSOM models the interactions of small groups of neurons and the connections between them rather than individual neurons and synapses. This is appropriate because detailed low-level knowledge of the cortex is still very patchy, so its function cannot simply be extrapolated from the fragments of anatomy and physiology which have been established. Instead, a more promising approach at present is to work backward from observed psychophysical and other aggregate phenomena to describe the basic computations that are being performed. When unequivocal low-level physiological or anatomical data is available, it is used to constrain the possible models.

Many of the fundamental assumptions of the model, such as the computation of the input activity as a weighted sum and the sigmoidal activation function, are common to most neural network models. Their biological validity has been examined in detail previously by other researchers. Representing activity as a scalar value rather than a spike train is a common abstraction made for computational convenience; RF-LISSOM can be extended to model the low-level time-dependent behavior of neurons when studying phenomena that depend upon it (Choe and Miikkulainen 1996; Miikkulainen et al. 1997).

However, there are two key aspects of the model that remain particularly controversial: whether the long-range horizontal connections are primarily inhibitory in effect, and the balance between genetic factors (i.e., hardwiring) and environmental factors (i.e., due to visual input) in the organization of the cortex. These topics will be discussed in detail in the remaining sections to support the claim that RF-LISSOM represents a biologically realistic model of cortical phenomena, including the tilt aftereffect. Furthermore, recent

experimental data has clarified many of the issues involved since the discussion by Sirosh (1995), and thus it is worthwhile to reexamine these issues.

3.6   Long-range inhibitory connections

The effect of the long-range lateral connections in the cortex has been controversial for some time, with some studies indicating that they must be inhibitory, and some indicating that they must be excitatory. All of the long-range lateral connections in the RF-LISSOM model are inhibitory, as has been found to be required to produce self-organization (Sirosh 1995). However, a consensus appears to be emerging that they cannot always be considered strictly excitatory *or* inhibitory, but instead have different effects in different circumstances (Weliky et al. 1995).

Anatomical surveys show that 80% of the synapses of long-range lateral connections connect directly between pyramidal cells, which are thought to make excitatory synapses only (Gilbert et al. 1990; McGuire et al. 1991). The other 20% of the connections target inhibitory interneurons which in turn contact the pyramidal cells, and thus represent inhibitory connections. Even though the inhibitory connections are outnumbered, the net effect at the columnar level has been difficult to establish conclusively with anatomical studies. For instance, the interneurons often synapse at regions such as the soma where their effects may be larger than those of excitatory neurons, which synapse farther out on the dendrites (Gilbert et al. 1990; McGuire et al. 1991).

Physiological and psychophysical evidence now indicates that the balance between these two types of connections is actually contrast-dependent: the influence of the lateral connections impinging upon a neuron is mildly excitatory when the surrounding area is activated by a low-contrast stimulus, and strongly inhibitory when the surround is activated by a high-contrast stimulus. This has been demonstrated conclusively at the cellular level in tissue slices (Hirsch and Gilbert 1991), and more recently *in vivo* at the level of cortical columns (Weliky et al. 1995). As discussed below, it has been hypothesized that these complex connections help enhance the ability to detect low-contrast inputs while suppressing redundant activation for high-contrast inputs.

However, it remains unclear how exactly this contrast dependence is implemented in the cortex. An early proposal was that the inhibitory interneurons are inherently more effective than the direct excitatory connections, but have a higher threshold for activation (Sillito 1979). At very low stimulus levels, the excitatory effects would predominate, but at high levels the inhibitory interneurons would become progressively more active and eventually would suppress the response of the target cell.

Douglas et al. (1995) have presented a more detailed proposal which takes into account the recurrent circuit provided by the type of short-range excitatory lateral connections

26

found in LISSOM. That paper shows how the relatively few lateral inhibitory connections could be effective enough to dominate the response even though they are fewer in number. Simplified versions of such circuits have been modeled by Stemmler et al. (1995) and Somers et al. (1996). They propose that these complex connections help enhance the ability to detect weak, large-area stimuli while suppressing spatially redundant activation for strong stimuli.

The current RF-LISSOM model emphasizes the suppressive effects only. Since all inputs used for self-organization have been high-contrast, the assumption that the effects are primarily inhibitory is well-founded. Because the amount of synaptic change due to the learning mechanisms is very small for low-contrast inputs, the behavior for such inputs does not significantly affect self-organization. Furthermore, since all the inputs used in this thesis are high-contrast, the results presented for the tilt aftereffect should not depend upon this simplifying assumption.

Future versions of RF-LISSOM may be extended so that the effect of the long-range connections varies according to circumstances. However, it is not yet clear what type of extension is necessary to account for the phenomena. One possibility is to multiply each lateral connection impinging upon a cell by a factor that depends upon the activation of the cell. This would treat any lateral input as a mild excitation if the cell is inactive, and as strong inhibition if the cell is strongly active (Sillito 1979; Somers et al. 1996). Another alternative extension is to use the net lateral input to the cell to compute this scaling factor. Future work may help to clarify which of these two, or others not yet formulated, is most appropriate.

3.7   Environmental versus genetic factors in development

The RF-LISSOM model allows researchers to explore the relationship between environmental and genetic factors in development. The relative weights of these factors have been debated for hundreds of years (for review see Diamond 1974), but recent evidence is beginning to clarify how development actually occurs in the nervous system. This section will first sketch the very general constraints that apply to a developing organism and the tradeoffs involved in satisfying them. It will then examine how these tradeoffs might be optimized in the mammalian visual system in particular, and how they relate to RF-LISSOM. The intent is to show how a self-organizing model can explain aspects of both environmentally- and genetically-driven development. This helps to clarify the significance of the RF-LISSOM model, and to suggest the role that it can play in an explanation of the cortex.[1]

---

[1]This section is optional; the rest of the thesis does not rely heavily upon the arguments presented here.

3.7.1 Evolutionary and developmental constraints

Every organism develops from a single cell. As such, it can make use of material within the cell, specifically the genome, and information in its environment, measured in some way. Different organisms appear to use these sources of information in very different ways, appropriate to their ecological niches and overall degree of complexity.

For instance, an animal's development could be specified in precise detail by its genome (Purves 1988). This would be feasible if the important factors of an animal's environment remain essentially constant for millennia, and if the adult animal is not particularly complex. Each individual of the species would express exactly the same pattern, and evolutionary selection processes would ensure that the pattern is appropriate for the environment of the organism.

This developmental program appears to apply to the nervous system of the nematode worm *C. elegans*. Under ordinary circumstances, this organism has 302 neurons in the same configuration in every individual (Sulston and Horvitz 1977). The advantage of such a scheme is that each individual will need only a fairly short developmental period, after which it will be fully prepared for life. A competing organism which requires a long maturation period would be at a disadvantage. While the competitor is still developing, the hard-wired organism will already have appropriate default behaviors even for circumstances that it has not previously encountered.

However, if the environment changes unpredictably over time scales shorter than those required for evolution to change the genome significantly, then a hard-wired organism will often encounter novel situations for which its behavior is inappropriate. An adaptive organism that can make use of clues found in the environment could be much more successful in those situations than a hard-wired ecological competitor.

It would be most efficient for the adaptive organism to extract all the relevant information from the environment at birth, so that it could subsequently devote its resources to other tasks. However, this would risk optimizing the organism for the particular circumstances surrounding that short time period. Optimizations made at that time might be wholly inappropriate for different situations encountered later, particularly since birth is an inherently anomalous event in an organism's lifetime. Thus it is desirable to integrate the information over as long a time as possible, to maximize the chances that the learned behavior is appropriate for the environment.

Yet the longer an organism takes to reach full competence, the more it is susceptible to accidents and predators. In the limit, the finite lifespan of an organism provides an upper bound upon how much environmental information can be integrated into the animal's repertoire. But since a single lifetime will only span a certain range of experiences, it cannot provide as much information as the millions of years that have presumably contributed to the formation of the genome.

If an organism became entirely adapted to the environment as it has encountered it, it would eventually encounter an important new event for which it was unprepared. In this case its hard-wired competitor would have the advantage. The genome of the hard-wired organism will presumably provide some appropriate behavior if the "novel" events actually recur quite often on an evolutionary timescale. So an organism cannot rely solely upon environmental cues, and to be fully suited to its environment it must (paradoxically) make use of some genetic information. The long-term information represented in the genome thus complements the direct environmental influences upon the organism after birth.

As discussed below, one might want the genetic influences to act as a constraint upon the amount of learning that can occur. This would prevent the organism from becoming too specifically adapted to its particular circumstances. The goal would be to help ensure that the organism remains capable of detecting unusual situations; e.g. those that are fatal but rare.

Thus there are a number of important tradeoffs between environmental and genetic control of development. For simple organisms with well-defined ecological roles, hard-wiring may suffice. However, more complex animals with more variable environments will optimally use some cues from the environment when developing. Fully adapting to the environment would not be entirely desirable, since it would amount to ignoring information only available in the genome.

Mammals, being quite complex and capable of handling a wide variety of different environments, appear to represent a particularly successful balance of these developmental and evolutionary constraints. Although significant genetic components are present, an extensive array of early environmental influences on visual system development have been documented in mammals (for review see Movshon and van Sluyters 1981). The following sections will argue that mammals make use of a simple technique that allows them to make use of both environmental and genetic information efficiently. This process will also be related to RF-LISSOM as a hypothesis for what the self-organization process represents.

3.7.2 Two-stage model for development

Since each individual begins from a single cell, yet some types of learning require multi-cellular sensory systems, genetic factors must determine the very first stages of development. Thus one simple way to combine genetic and environmental factors is for entirely gene-based development to progress to a certain point, e.g. birth or shortly thereafter, and then transfer developmental control to a learning mechanism (Blakemore and van Sluyters 1975). This seems particularly appropriate in organisms such as mammals that have a long gestation period, during which their ability to measure the environment is severely limited. This approach would help minimize the time spent in learning, yet allow the environment to have some influence on the structures developed.

Once the developing systems are nearly complete, it would be desirable to limit their adaptability. Otherwise, eventually all traces of the initial genetic order would disappear (Jouvet 1980). Allowing the organism to significantly adapt further would thus amount to ignoring valuable information from the genome (as described in the previous section).

In a computer model of sensory development such as RF-LISSOM, the hard-wired aspects would correspond to the initial state of the model, before any self-organization occurs. For instance, one could assume that a roughly retinotopic map on the LGN or visual cortex is formed through genetic means. Subsequent sharpening of the map could occur through visual experience, thus ensuring that the map is appropriate for the particular environment found.

However, this simple two-stage approach does not adequately account for all of the data on innate and environmental factors in mammals. Firstly, the genetic process cannot fully specify the initial phase of development in detail. There is simply not enough space available in the genome of a mammal (on the order of 10^5 genes) to specify every connection in its nervous system (as many as 10^{15}; Kandel et al. 1991,p.885). Thus the genetic developmental mechanisms must express many highly repetitive structures or they must specify the structures only approximately (Purves 1988; Shatz 1996).

Secondly, even during embryonic stages, the developmental process appears quite flexible (Purves 1988). Developing organisms adapt systematically to a number of rather drastic modifications, such as the implantation of a third eye in a frog (Reh and Constantine-Paton 1985) or the removal of large areas of the brain of a primate (Goldman-Rakic 1980). These adaptations presumably arise out of a repertoire of possible responses to damage or malfunction in the developing embryo. The embryonic adaptive mechanisms would be even more important for mammals than for simpler creatures, since greater complexity increases the likelihood of isolated malfunctions (Purves 1988). Thus if each connection were hardwired, an explosive number of backup connections would need to be specified as well, or else some general adaptation procedure would be needed (Shatz 1996). If there were such a general procedure, then the initial precise specification would be superfluous anyway.

Finally, evolution seems unlikely to have developed such a wastefully modular system, with two complex and non-overlapping systems to perform similar tasks. The first system would express genetic information according to a fixed plan. Since it must start from a single cell, it would need a set of bootstrapping mechanisms to construct the "preset" pattern determined by the genome. These mechanisms would have to be quite complex to account for the complex structures seen even at birth in mammals (Blakemore and van Sluyters 1975). At some point determined by the genome, control would shift to the learning mechanism, and the cellular or intracellular structures used for bootstrapping would need to be deactivated or perhaps even dismantled. Otherwise, the system would not be

able to accept subsequent cues from the environment. Instead of such a clean separation between genetic and environmental factors, it seems more likely that the initial mechanisms or structures are reused to some extent in later activity-dependent developmental processes (Shatz 1996).

Cumulatively, the above constraints strongly discourage large-scale fully-specific hardwiring, yet somehow most organisms (including humans) develop with a large degree of structure already present at birth (Blakemore and van Sluyters 1975). The next section explores an alternative, complementary hypothesis for combining genetic and environmental factors that addresses these considerations, and the following section will review evidence that it is actually occurring in several systems of mammals. This type of developmental strategy makes widespread use of activity-dependent self-organization, and so it helps to show how and why animals may implement the processes modeled in RF-LISSOM.

3.7.3   Overlapping model for development using pattern generation

True environmental information is only available to the visual system after birth (and often even later, depending on the time of eye-opening). However, coherent inputs arising *at any point along the pathway* from the receptor to the cortical areas could activate cortical areas. They could thus have effects similar to genuine visual experience, yet under genetic control (Constantine-Paton et al. 1990; Maffei and Galli-Resta 1990; Marks et al. 1995; Roffwarg et al. 1966; Shatz 1990, 1996). So once the basic, general connectivity and cellular mechanisms are in place, particularly the learning mechanisms modeled by RF-LISSOM, the system can self-organize based upon *any* input, not just those in the environment.

If the organism itself generates those signals with some form of internal pattern generator, the organism could direct its own development. Neural mechanisms to generate coherent patterns are well-documented in a number of different systems (Marder and Calabrese 1996). Such a pattern generator would allow genetic factors to be expressed through the same mechanisms also used later when adapting to the visual environment (Jouvet 1980).

As simulations with RF-LISSOM have demonstrated, even very simple patterns can drive the development of extraordinarily complex structures, such as the orientation map in V1 (described in chapter 4). For the orientation map, simple two-dimensional oriented Gaussian inputs, described by a single equation with six parameters, were presented to the general RF-LISSOM model. The model did not previously contain any representation of orientation. These inputs prompted the model to develop an organized set of thousands of orientation-specific feature-detectors, with many millions of specific connections between them. Thus quite detailed structures could be specified by the genome merely by specifying what type of pattern should be generated internally. Obviously, the nervous system might use any number of different mechanisms to generate appropriate patterns, not all of

which would be as simple as the equation mentioned above, but in general much simpler mechanisms would suffice for the generation of training patterns than to specify the final structures.

Since the specification for the pattern generator could be encoded into a very small amount of genetic information (Jouvet 1980), very different structures could be specified by changing only a small part of the genome. Random mutations in that portion of the genetic code would be likely to cause development of slightly different patterns, which might lead to quite different cortical structures. This would greatly facilitate the processes of evolution, since it increases the chance that a mutation is useful rather than merely debilitating.

To compare, imagine that the genome of a phylogenetically more primitive animal, such as a mouse, specified its nervous system in complete detail. The chance that it would eventually evolve into an animal with a more complex nervous system, such as a monkey, would be quite small. Any single change would be likely to damage the nervous system, yet an enormous number of coordinated changes would be needed to transform a fully-specific mouse genome into a fully-specific monkey genome (Purves 1988). But even though much of the DNA is shared between mammalian species, an extraordinary variety of mammals exist, with quite different cortical organizations (Purves 1988). Internally generated patterns may help to explain how evolution leapt over the gaps between successive species in phylogeny. This interpretation is supported by the observation that learning mechanisms appear to be highly conserved between species and brain areas in mammals (Kirkwood and Bear 1994), while presumably pattern generators would vary significantly between different species.

The patterns of fur coloring on different individuals of the same species, e.g. domestic cats, might represent a graphic example of internal pattern generation in development. Swindale (1980) proposed a similar analogy between zebra stripes, fingerprint whorls, and ocular dominance stripes. Different kittens in the same litter often have very different fur patterns, some quite complex, yet each individual is genetically quite similar to the others. But purebred cats generally have the same overall fur patterns for every individual of that subspecies. The lack of variation in purebreds indicates that the patterns are controlled genetically, but the variety seen in mixed breeds suggests that the patterns are controlled by a small region of the genome, which can express quite different patterns with only small changes. This same type of pattern generation may be occurring in the mammalian nervous system, as explored in the next section.

3.7.4   Pattern-generated development in mammals

Retinal waves

It is not presently known what role internal pattern generation plays in mammalian development. However, evidence is rapidly accumulating that such processes are occurring in at

32

least two pathways leading to mammalian visual cortex: the developing retina and the brain stem. In the retina, the patterns take the form of intermittent spatially-coherent waves of activity across groups of ganglion cells (Meister et al. 1991; Sirosh 1995; Wong et al. 1993). They appear to arise from as-yet-unknown events in networks of developing amacrine cells that provide input to the ganglion cells (Catsicas and Mobbs 1995; Feller et al. 1996; Shatz 1996). These waves may represent training inputs for the prenatally developing LGN and cortex, providing a simple explanation of how the system could be activity-dependent, yet already organized at birth.

The waves begin before photoreceptors have even developed (Maffei and Galli-Resta 1990), so they do not result from visual input of any sort. They are presumably entirely unsynchronized between the two eyes, since they occur spontaneously and locally, with long pauses in between (Shatz 1990). Blocking this activity in the retina prevents the segregation of the LGN into eye-specific layers before birth (Shatz 1990, 1996), so the activity must be playing a role in the segregation. Blocking the activity in the retina, but substituting certain patterns of artificial electrical stimulation, results in dramatically different cellular properties than if other patterns are used (Mooney et al. 1993). This suggests that the patterns of electrical activity are the important feature of the waves (Shatz 1990, 1996). As the waves subside gradually in strength and frequency due to developmental processes in the retina, the connectivity in the LGN stabilizes (Wong et al. 1993). At approximately the same time, visual input becomes functional, but it is not yet known whether visual input causes the waves to cease or whether they cease in anticipation of visual input. Manipulations of the visual environment have been shown to have dramatic effects on the organization of the visual cortex (Movshon and van Sluyters 1981), the next step in the visual pathway. Thus internally-generated retinal waves appear to be instrumental in pre-visual development, at least of the LGN, and they appear to play a role similar to that which visual input plays for higher areas. It is not yet known what role the retinal waves play in the development of the cortex, but they may represent the prenatal portion of the activity seen by the cortex, which later receives visual input that is crucial for its development.

It has further been proposed that the spontaneous activity is responsible for maintaining topographic maps during the growth of the connecting pathways from the retina to the LGN (Bunt et al. 1979). That is, when the ganglion cells of the retina extend axons forming the optic nerve, they would keep a retinotopic order as a result of activity-dependent processes driven by these retinal waves. Throughout the developmental process, they would remain in this retinotopic order until they reached the LGN. This proposal has not been definitively established, but there is some evidence for the proposed mechanisms and results (Bunt et al. 1979). If it turns out to be well-supported, it would take still more of the burden off of the genetic factors in development. That is, since the retina is inherently "retinotopic", if the fibers remain in that organization throughout the journey

to the cortex, no genetic specification of the retinotopic ordering is needed.

Finally, similar waves have been documented in early areas of other sensory systems, such as the auditory systems of birds (Lippe 1994). Such activity may be a general feature of the earliest sensory areas in more complex organisms such as birds and mammals, allowing genetic factors to be take effect via the same mechanisms which later incorporate environmental influences (Shatz 1990). This would represent a developmental mechanism which is quite efficient and effective on a number of different levels, as described above.

PGO waves

Internal pattern generation may continue at somewhat reduced levels, even after development. For instance, the generation of appropriate training patterns has been proposed to be one function of the sleep stage known as rapid eye movement sleep (REM; Jouvet 1980; Roffwarg et al. 1966). Infants exhibit very large durations of REM sleep during the ages when their nervous systems are the most highly plastic, and the amounts of REM sleep and plasticity decrease similarly with age (Roffwarg et al. 1966). Furthermore, birds and mammals, which are much more adaptable to different environments than other animals, are the only organisms known to exhibit REM sleep in the adult (Jouvet 1980).

During and just before REM sleep, internally generated phasic waves called ponto-geniculo-occipital (PGO) waves can be measured in the LGN, V1, and many other cortical areas (Jouvet 1980; Steriade et al. 1989). These apparently genetically-programmed waves consist of large, slow increases in field potential when measured with an EEG (electroen-cephalogram). They appear to cause specific eye movements which have been statistically correlated with reports of the direction of gaze in dream imagery (Jouvet 1980). PGO waves originate in the pons of the brain stem (hence ponto-) and travel via direct pathways to the lateral geniculate nucleus (hence -geniculo) and to visual cortex (in the occipital lobe, hence -occipital; Steriade et al. 1989). They appear to be relayed from the LGN and visual cortex to many other areas of the cortex (Jouvet 1980).

Jouvet (1980) has proposed that these waves help direct the course of brain maturation in early life, and specifically that they allow genetic differences among individuals to be expressed. Tentative support for this hypothesis was obtained by Marks et al. (1995), who found that depriving kittens of REM sleep during the critical period significantly *enhanced* the effects of visual experience during that time. The heightened effects of experience were interpreted as a *weakening* of genetic control over development.

Ordinarily, blocking signals from one eye of a kitten for even a few days during the critical period for visual development (4–6 weeks after birth) causes dramatic strengthening of connections to visual cortex from the non-deprived eye, and a loss of connections from the deprived eye (Movshon and van Sluyters 1981). With shorter deprivation times, the effects are less pronounced. If both eyes are deprived of input, no detectable loss or strengthening

34

occurs (Movshon and van Sluyters 1981). Thus a monocular deprivation paradigm offers an opportunity to test the role of environmental and genetic factors in development, both of which appear to be operating in this system.

Marks et al. (1995) blocked signals from one eye of kittens for a portion of the critical period. It was found that blocking the input had a much larger effect on animals deprived of REM sleep than for control animals which had normal REM sleep. Similarly heightened sensitivity to deprivation was found when the PGO waves were blocked directly by lesions in the pons. Thus some mechanism in which the PGO waves participate appears to limit the amount of plasticity in this system in the normal individual.

Unlike the retinal waves discussed in the previous section, PGO waves are correlated between the two eyes. The patterns sent to each of the eye-specific layers of the LGN are generated separately, but match quite closely (Jouvet 1980). The correlation presumably represents the pattern of similar, but not identical, eye movement directions for the two eyes. This type of correlated input would ordinarily counteract the imbalance of inputs from the deprived and non-deprived eyes. Thus the monocular deprivation protocol should have had greater effect in the absence of PGO waves, which was in fact seen (Marks et al. 1995).

Jouvet (1980) speculates that PGO waves are a very general signal that activates local circuits in each brain area. Each local circuit would then generate training inputs appropriate for its area. He further proposes candidate substrates for such circuitry, which are beyond the scope of this discussion. However, if Jouvet's proposal is correct, then to the extent that the pattern evoked by PGO waves differs from the typical response to the environment, REM sleep would serve to ensure that each brain area develops and is maintained in readiness for its genetically-determined function. The genetically-determined structure would be fine-tuned based on information from the environment, but it would persist regardless of how much learning occurs. This would help ensure that the structure and function of each area of the brain would be suitable for the species of the organism, and thus for its ecological niche. It would also help differentiate different brain areas intended for different tasks yet which receive similar inputs from the environment; without such differentiation or some form of competition the areas would eventually become identical as they adapt to visual input.

As argued above, it is *not* desirable for an organism to completely adapt to the environment it experiences. The amount of time available for adaptation is very small compared to the time over which genetic information has been compiled, and thus not all the relevant information is available from the environment. Thus the PGO waves may also serve to limit such unwarranted adaptation. Since adults exhibit significant amounts of REM sleep as well, these pattern generators may also be operating in the developed animal (Jouvet 1980). They may help maintain a balance between specific visual correlations

learned during the day, such as a preponderance of certain orientations of lines, and circuitry capable of handling a wide variety of processing tasks, such as detecting all possible lines. Roffwarg et al. (1966) and Steriade et al. (1989) speculate that these generated patterns constitute some of the vivid imagery experienced during REM sleep.

Jouvet (1980) further speculates that the function of REM sleep is to "program" the nervous system with genetically encoded behaviors and capabilities. In preliminary experiments, he tested the effect of REM sleep deprivation on two strains of genotypically similar laboratory mice. These two strains ordinarily exhibit a small number of well-defined differences in maze-learning behavior even when raised in identical environments. When deprived of REM sleep, the difference in behavior *decreased*, i.e. the mice became even more similar. Although the experiment was not conclusive, it suggests that the REM sleep deprivation inhibited the expression of genetic differences between the two strains. In general, Jouvet hypothesizes that individuals would exhibit far less variation if deprived of the effects of REM sleep, since their behavior would be determined primarily by the environment. Thus "genetic programming" during REM sleep may help to ensure that different species, individuals of the same species, and even different areas of a single brain retain diversity, thus increasing the likelihood that some member of the group will be appropriate for a given task.

As a cautionary note, the REM sleep effects are not nearly as well characterized as those in the retina, primarily because sleep deprivation has a number of side effects. For instance, both behavioral and pharmacological methods of REM deprivation cause significant stress, and the pharmacological methods involve drugs that are not particularly specific in their actions (Jouvet 1980). This makes REM sleep results difficult to interpret. The PGO waves themselves are well-documented, however, and since they occur in many different areas of the brain, they could represent a general feature of mammalian neural systems. They may be a part of the mechanism by which genetic and environmental influences are combined by activity-dependent learning processes.

Although research into the effects of pattern generation on the developing brain is just beginning, it is already clear that the distinction between "genetic" and "environmental" origins of brain structures is blurring. Many alternatives in between can be determined by the ratio of pre-programmed "experience" to actual sensory experience. The use of pattern generators to direct experience-dependent learning processes represents an effective solution to the general constraints faced by a complex organism with a limited genome and a limited time for development.

3.7.5 Pattern-generated development in self-organizing models

Self-organizing models like RF-LISSOM allow hypotheses about the contribution of genetic and environmental factors to be tested in detail. The training inputs for self-organizing mod-

els can represent either internally generated patterns or environmentally realistic stimuli, or any combination thereof. RF-LISSOM also incorporates hard-wired genetic information, such as the basic learning and activity computation algorithms, which are assumed to be essentially the same between different individuals.

The initial connectivity parameters, such as the extent of excitatory and inhibitory connections or the amount of initial order, can be set by the researcher to represent different possible starting states. The RF-LISSOM research so far has tried to begin from as much initial randomness as is possible for self-organization. This shows how RF-LISSOM can explain the most difficult cases. Adding further initial order, representing some innate bias towards the appropriate structure (as often found in biological systems), just makes the learning process easier.

Similarly, simulations are run using the simplest training inputs that will account for the structures seen in the cortex. These inputs can then be compared with known internal pattern generators, and with the typical features of the visual environment. Finding that the required inputs match with those available from either of those two sources helps confirm that the model is appropriate. Using such techniques, the RF-LISSOM model can help test hypotheses about the relative importance of environmental and genetic factors in development.

3.8   Conclusion

The RF-LISSOM model demonstrates that the afferent and lateral connections in the primary visual cortex can self-organize simultaneously and synergetically based on a single Hebbian adaptation process. The model is well supported by anatomical and physiological evidence for high contrast stimuli. It has already been very successful at modeling structural development and adult plasticity on the cortex.

The self-organization process stores long-range activity correlations between units into the lateral connections. During visual processing, this information is used to eliminate redundancies and to form an efficient sparse coding of the input. Training inputs can come from the visual environment or from internally generated sources. If the inputs have features that are oriented (i.e., are not radially symmetrical), the model will self-organize to represent that aspect of the inputs. Chapter 4 will examine how this self-organization leads to the development of orientation preferences. Later chapters will show that the same decorrelating process that forms a sparse coding of orientation and position also results in psychological artifacts known as the tilt aftereffect.

Chapter 4

Training the Orientation Map

As discussed in chapter 2, neurons in the primary visual cortex are selective for the orientation and position of stimuli. Furthermore, the neurons are arranged retinotopically across the cortex, forming an *orientation map* (Blasdel 1992a; Blasdel and Salama 1986; Grinvald et al. 1994; Ts'o et al. 1990; Weliky et al. 1995). Each local area in the map contains neurons with each possible orientation preference, all responding to the same location on the retina. The neurons tuned to a particular orientation are activated when an input of that orientation is present at their preferred location on the retina.

It has been shown previously that, given oriented inputs, the RF-LISSOM model develops orientation maps similar to those seen in the primary visual cortex (Miikkulainen et al. 1997; Sirosh and Miikkulainen 1997; Sirosh et al. 1996). In this thesis, the functional aspects of such self-organized orientation maps will be studied. This chapter describes the process of organizing and characterizing an orientation map, and chapter 5 will examine how the behavior of the map leads to tilt aftereffects.

4.1   Training inputs

To develop an orientation map in the RF-LISSOM model, one of the simplest oriented training patterns was used: a 2-dimensional elongated Gaussian. For these inputs, the activity at each retinal ganglion cell is calculated according to the equation

$$\xi_{r_1,r_2} = exp(\quad -\frac{((r_1 - x_i)cos(\alpha) - (r_2 - y_i)sin(\alpha))^2}{a^2}$$
$$-\frac{((r_1 - x_i)sin(\alpha) + (r_2 - y_i)cos(\alpha))^2}{b^2}) \qquad (4.1)$$

where a^2 and b^2 specify the length along the major and minor axes of the Gaussian, (x_i, y_i) specifies its center, and α its orientation. The x and y coordinates of the centers are each

(a) +45° Gaussian at (14,13)   (b) +88° Gaussian at (6,4)   (c) -5° Gaussian at (8,14)
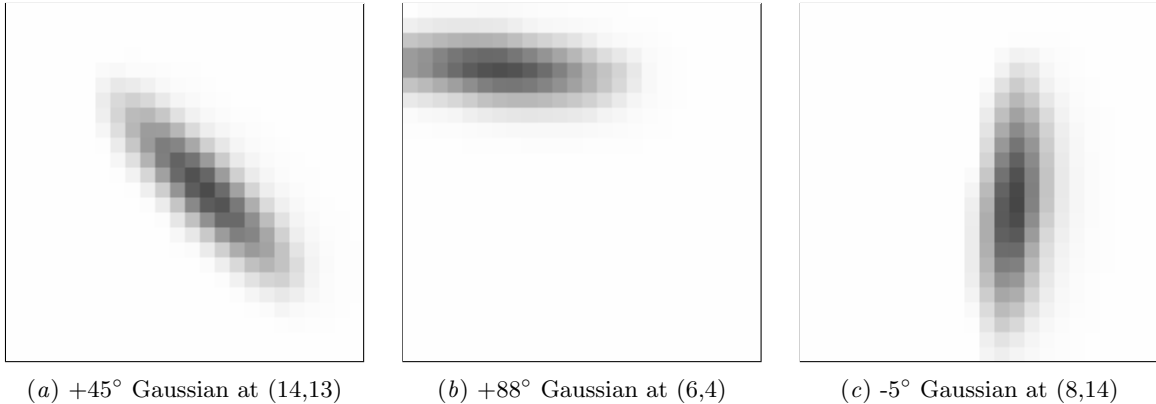
Figure 4.1: **Sample training inputs plotted on the retina.** The retina consisted of 24×24 units, with an origin at the top left corner. Angles cited are relative to vertical, taken to be 0°, and they increase counterclockwise. For all the Gaussian inputs in this thesis, the major axis half-width was $a = 7.5$ units, and the minor axis half-width was $b = 1.5$.

chosen randomly within the retinal area, and the orientation is chosen randomly from the uniform distribution in the range $0° \leq \alpha < 180°$. Figure 4.1 shows several sample training inputs.

These patterns approximate natural visual stimuli after the edge detection and enhancement mechanisms in the retina. Such edge-enhanced images have a predominance of elongated features. Similar features may also be found in the intrinsic retinal activity waves that occur in late prenatal development in mammals, and they are believed to drive the initial organization of the visual cortex (as discussed in section 3.7.4; Catsicas and Mobbs 1995; Meister et al. 1991; Wong et al. 1993). The RF-LISSOM network models the self-organization of the visual cortex based on these natural sources of elongated features.

4.2 Training parameters

The model consisted of an array of 192×192 neurons, and a retina of 24×24 ganglion cells. The circular anatomical receptive field of each neuron was centered in the portion of the retina corresponding to the location of the neuron in the cortex. The RF consisted of random-strength connections to all ganglion cells less than 6 units away from the RF center. For example, the neuron at the center of the cortex was connected to the ganglion cells inside a circle of radius 6 at the center of the retina. The top left neuron was connected to the top left retinal ganglion, and to the other ganglion cells in the top left corner. Sample initial receptive fields are shown in figure 4.2.

The cortex was self-organized for 30,000 iterations on oriented Gaussians which each had a major axis of half-width $a = 7.5$ units and a minor axis of half-width $b = 1.5$.

The initial lateral excitation radius was 19 and was gradually decreased to 1. The lateral inhibitory radius of each neuron was 47, and inhibitory connections whose strength was below 0.00025 were pruned away at 30,000 iterations. The lateral inhibitory connections were preset to a Gaussian profile with $\sigma = 100$, and the lateral excitatory connections to a Gaussian with $\sigma = 15$. The lateral excitation γ_e and inhibition strength γ_i were both 0.9. The learning rate α_A was decreased from 0.007 to 0.0015, α_E from 0.002 to 0.001 and α_I was a constant 0.00025. The lower and upper thresholds of the sigmoid was increased from 0.1 to 0.24 and from 0.65 to 0.88, respectively. The number of iterations for which the lateral connections were allowed to settle at each training iteration was initially 9, and was gradually increased to 13 over the course of training.

These parameters were chosen by Sirosh (1995) in order to develop a biologically realistic orientation map, prior to any of the experiments done on tilt illusions or aftereffects for this thesis. Small variations of these parameters produce roughly equivalent results. The training took 8 hours on 64 processors of a Cray T3D at the Pittsburgh Supercomputing Center. The model requires more than three gigabytes of physical memory to represent the more than 400 million connections in this small section of the cortex.

Although there is significant order in the model even before self-organization, with training the RFs will sharpen into smooth profiles selective for orientation, the topographical organization will be refined, and the lateral connections will become patchy and focused. The initial order represents part of the genetically-determined development of the visual cortex, and the self-organization to be described in the following sections represents the activity-dependent developmental processes, as outlined in section 3.7.

4.3   Receptive fields and orientation maps

The self-organization of afferent weights results in oriented synaptic weight patterns forming afferent receptive fields. Figure 4.2 shows examples of these weight patterns plotted on the retinal surface for several neurons. A variety of such RFs are produced in the self-organizing process, most highly selective to inputs of a particular orientation, others unselective.

To show how the orientation preferences are located across the network, an orientation map was computed by labeling each neuron with its orientation preference, as determined from its afferent weights. The afferent weights were fitted to an ellipsoidal Gaussian (equation 4.1) using the nonlinear programming package NPSOL (Gill et al. 1986). The orientation of the fitted function was taken to be the orientation preference of that neuron; it is the orientation of the Gaussian that maximally excites the afferent weights. Because the receptive fields were themselves developed based on Gaussian inputs, their orientation preferences were generally unambiguous, and other methods of computing the orientation map would show very similar results.

40

(a) RF tuned to +45°   (b) RF tuned to -30°   (c) Unselective RF
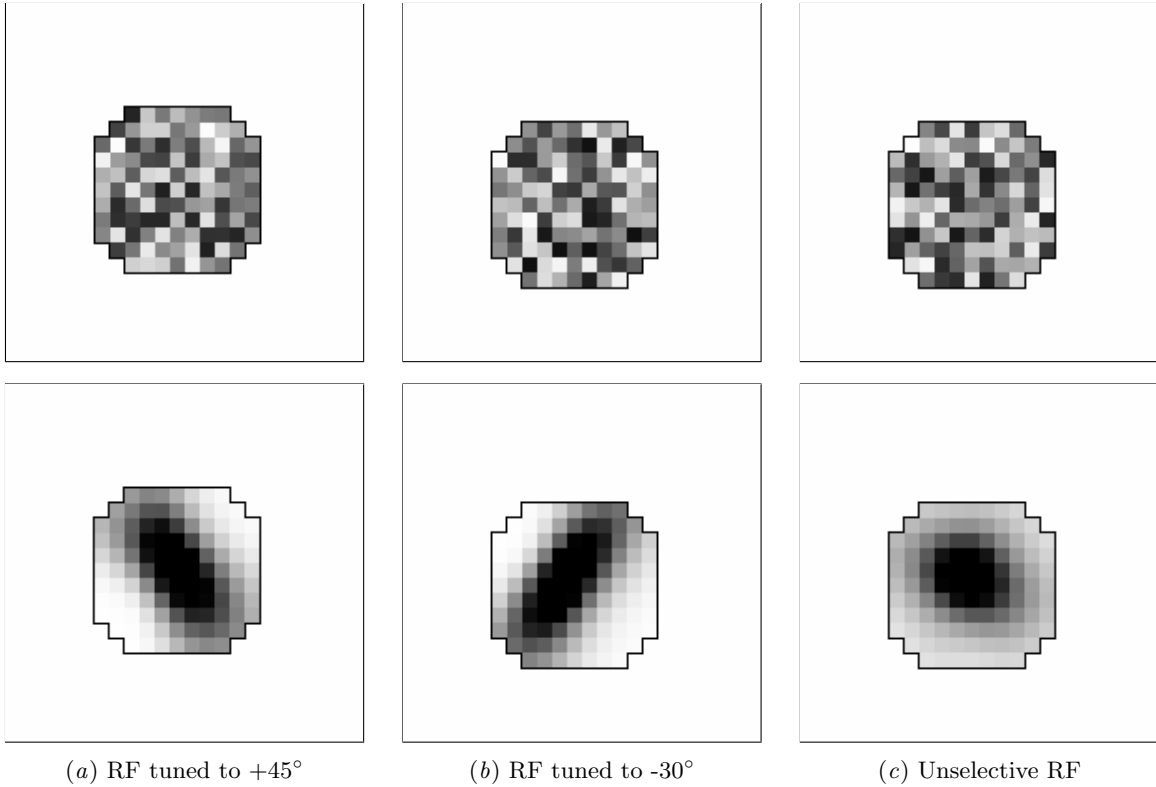
Figure 4.2: **Self-organization of afferent receptive fields.** The afferent weights to three neurons near the center of the cortex are shown, plotted on the retinal surface. The top row shows the initial afferent weights for each neuron, and the bottom shows the weights after self-organization (at iteration 30,000). Initially, all of the neurons had random afferent weights across the anatomical receptive field area, and were not selective for orientation. Through self-organization, the first two neurons became highly selective for a particular orientation. The third became selective for retinal position within its RF, but remained unselective for orientation; this type of RF is relatively rare in V1 as well as in the model. When a Gaussian input of a particular orientation is presented at the location of these receptive fields, the first neuron will fire strongly only if the Gaussian is oriented near +45°, since otherwise very little of its receptive field will be activated. Similarly, the second neuron will fire for a Gaussian oriented near -30°, while the third will have approximately equal response to all orientations. These types of receptive fields have also been found in the cortex, and serve to encode the local orientation at that position on the retina.

Figure 4.3 shows the global organization of the receptive fields in the network before and after training. The color of each neuron indicates its orientation preference as indicated in the key. Initially, all the afferent weights are random. As a result, the orientation preferences of the RFs are random and most RFs do not have a strong preference for any particular orientation. As self-organization progresses and afferent weights develop oriented receptive fields like those in figure 4.2, a complex organization of orientation preferences

41

(a) Initial unordered map (b) Final orientation map

+90°   +45°   0°   -45°   -90°

Figure 4.3: **Self-organization of the orientation map (*color figure*).** Each neuron in these plots of the cortex is colored according to the orientation preference of its afferent weights (as shown on the key.) (*a*) Initially, the afferent weights are random within a fixed receptive field. However, near the borders the receptive fie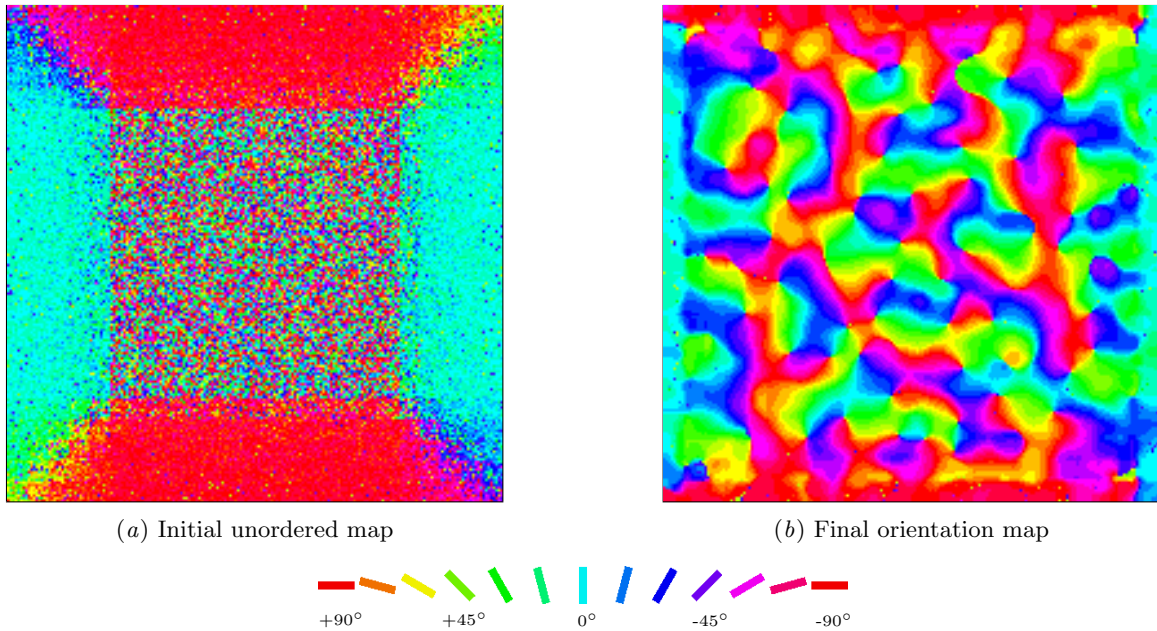lds are elongated in shape because they are cut by the edge of the retina. This gives neurons near the edge an initially oriented receptive field, even though the weights themselves are random. (*b*) After 30,000 input presentations, the receptive fields organize into continuous and highly selective bands of orientation columns. Near the borders, the patterns have seamlessly self-organized in a way that takes advantage of the bias towards receptive fields parallel to the borders. It is not known whether such edge effects occur in biological systems, where only central regions have been studied in detail. The central section of the map is both qualitatively and quantitatively similar to those found in the macaque monkey (Sirosh 1995). Both include (1) *pinwheels*, points around which the orientation preferences change continuously (e.g. just below and to the left of the center of the cortex) (2) *linear zones*, bands where the orientation preferences change continuously, like a rainbow (e.g. from just below the center, down to the left at a 30° angle from vertical), and (3) *fractures*, regions where the orientation preference changes abruptly between two distant orientations (e.g. almost half-way over from the center to the left, between the red and green areas).

develops. The map is remarkably similar in structure to those observed in the primary visual cortex by recent imaging techniques (Blasdel 1992b; Blasdel and Salama 1986), and contains complicated structures such as pinwheels, fractures and linear zones.[1] The results suggest that Hebbian self-organization of afferent weights, based on recurrent lateral interactions,

---

[1] The similarity of the model and experimental maps was measured with Fourier transforms, autocorrelation functions, and correlation angle histograms. See Erwin et al. (1995) for a discussion of these methods; see Sirosh (1995); Sirosh and Miikkulainen (1996, 1997) for those measurements for RF-LISSOM. An MPEG animation of the self-organizing process is found in Sirosh et al. (1996).

could underlie the development of orientation maps in the cortex.

4.4   Self-organization of lateral connections

The lateral connection weights self-organize at the same time as the orientation map forms. Initially, the connections are spread over long distances and connect to neurons of all types. As the lateral weights self-organize, the connections between uncorrelated regions grow weaker, and after pruning, only the strongest connections remain (figures 4.4 and 4.5). The strongest lateral connections of highly-tuned cells (figure 4.4c) link areas of similar orientation preference, and avoid neurons with the orthogonal orientation preference, as found in the cortex. Other neurons remain unselective (for instance, at points surrounded by continuously changing orientation preferences), and they connect to cells of all orientations equally (figure 4.5b,c). The connections of unselective neurons have not yet been studied, so this result represents a prediction of the RF-LISSOM model.

Furthermore, the connection patterns of highly oriented neurons are typically elongated along the direction in the map that corresponds to the neuron's preferred stimulus orientation (as verified subsequently for monkey cortex; Fitzpatrick et al. 1994). This organization reflects the activity correlations caused by the elongated Gaussian input pattern: such a stimulus activates primarily those neurons that are tuned to the same orientation as the stimulus, and located along its length (see Sirosh et al. 1996 for details).

4.5   Orientation encoding

The previous two sections showed how the afferent and lateral connections self-organize into a highly structured map with very specific internal connections, as found in the cortex. As one might expect from the properties of the individual neurons in the map, the response of the network to an input varies systematically depending upon the orientation and position of that input. This section will show examples of the network's actual response to different orientations, and will examine possible methods for determining what orientation is perceived by the cortex for that activity pattern. Calculating the perceived orientation is an essential prerequisite to measuring tilt aftereffects and illusions, since those are manifested as differences in the perceived orientation in different circumstances.

At any point in time the visual system is processing an image that has only a small number of oriented features in any local area, so only a few portions of the orientation map will be active at a given time. Figure 4.6 shows the sparse activity that results for Gaussian inputs of various orientations at the center of the retina. Initially, each input activates a wide range of neurons with orientation preferences somewhat similar to the orientation of the input pattern (figure 4.6b). Through the highly-specific lateral interactions, the

(a) Afferent weights   (b) Excitatory weights   (c) Inhibitory weights

+90°   +45°   0°   -45°   -90°

Figure 4.4: **Self-organization of lateral connections to a highly selective neuron (*color figure*).** The plots show each type of weights going to the neuron at (88,111), both before (top row) and after self-organization (bottom row, after 30,000 iterations). A border is drawn around the active weight area in each plot, and the neuron itself is marked with a tiny box in (*c*). Bright colors at a location on the cortex in (*b*) and (*c*) indicate a strong lateral connection from that neuron to neuron (88,111); the hue encodes the orientation preference of the neuron. Initially, each neuron had lateral inhibitory connections to every surrounding neuron, so each connection was very weak. After organization, the neuron developed a preference of -30° (blue), and its connections come primarily from other blue neurons. The connections follow the twists and turns of each blue iso-orientation column (compare to figure 4.3*b*), and are elongated upon the orientation they encode.

network then settles into a stable pattern with sharply-defined regions of active neurons. The neurons that remain active are those that prefer orientations close to that of the input pattern. In other words, the patterns of activity clearly encode the orientation of the input.

However, it is unknown precisely how the higher visual areas extract the encoding of orientation from figure 4.6 and arrive at the perception of an individual oriented line. It has been suggested that the perceived orientation is either the orientation preference of the unit with the highest activation (Carpenter and Blakemore 1973), or a weighted average

44

(a) Afferent weights   (b) Excitatory weights   (c) Inhibitory weights

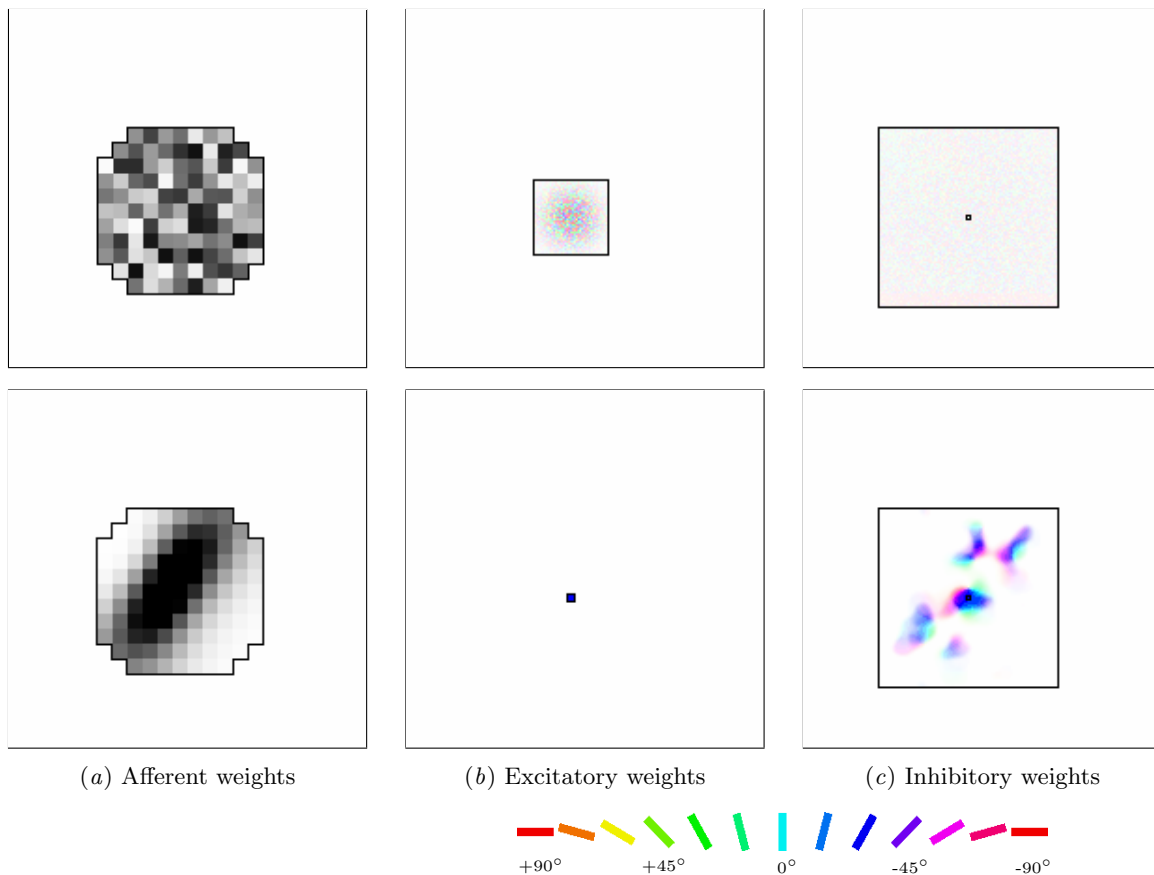+90°   +45°   0°   -45°   -90°
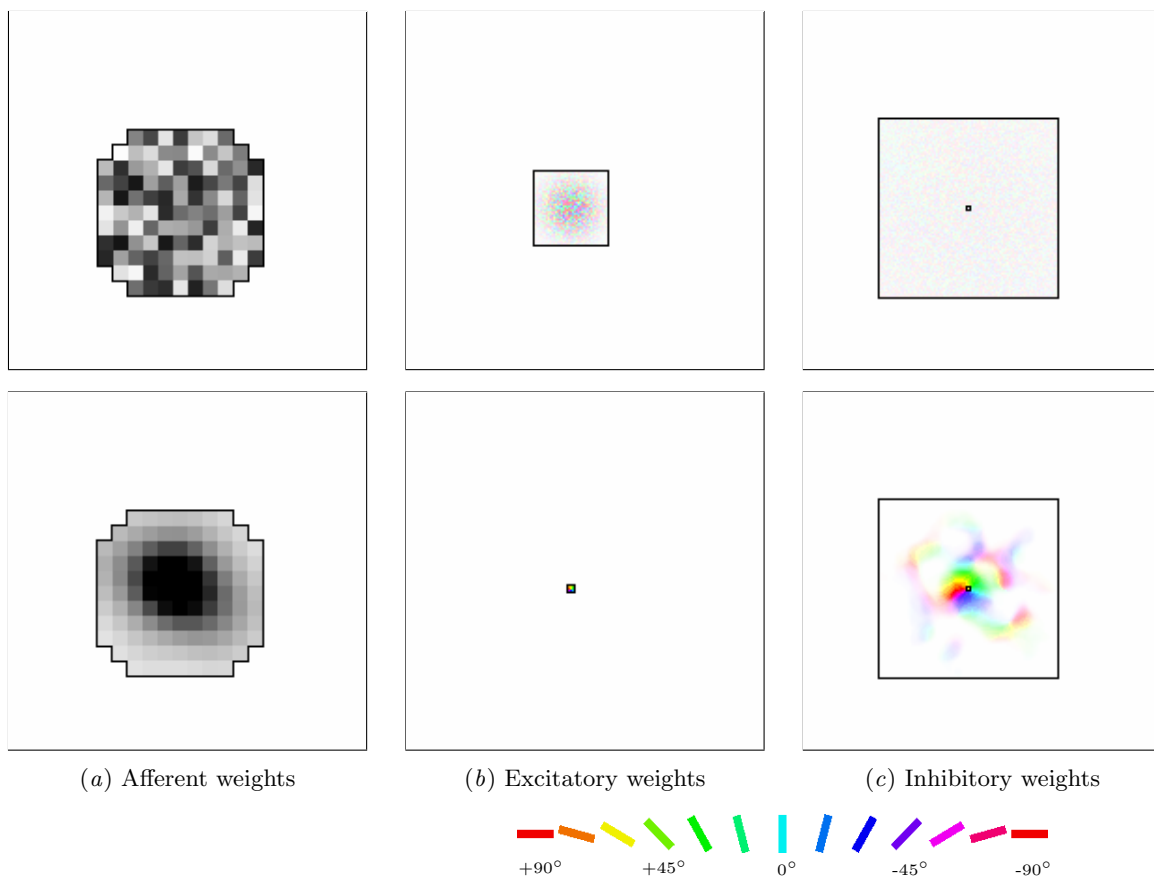
Figure 4.5: **Self-organization of lateral connections to an unselective neuron (*color figure*).** The plots show each type of weights going to the neuron at (88,105), both before (top row) and after self-organization (bottom row, after 30,000 iterations). The weights are colored as in figure 4.4. After self-organization, this neuron is in the center of a pinwheel, around which orientation preference changes continuously. The neuron is colored yellow because it has a very slight preference for +30°, but neighboring neurons of all orientation preferences connect to it.

of the preferences of all active units (Coltheart 1971). An intermediate method could also be used, computing an average of all the units having activity greater than some arbitrary activity level. In order to determine if a choice between these methods is crucial, the two extreme options were tested for the RF-LISSOM model.

The weighted average of orientation preferences must be computed as a vector sum, since angles repeat every 180°. Two nearly horizontal lines (e.g. -85° and +85°) should average to represent a horizontal line (±90°). However, the arithmetic average of -85° and +85° is 0°, which is a vertical line and is clearly incorrect as an estimated perception. Instead, each neuron is represented by a vector. The vector must represent adjacent orientations as adjacent vector angles. Thus each of the 180° possible orientations must be
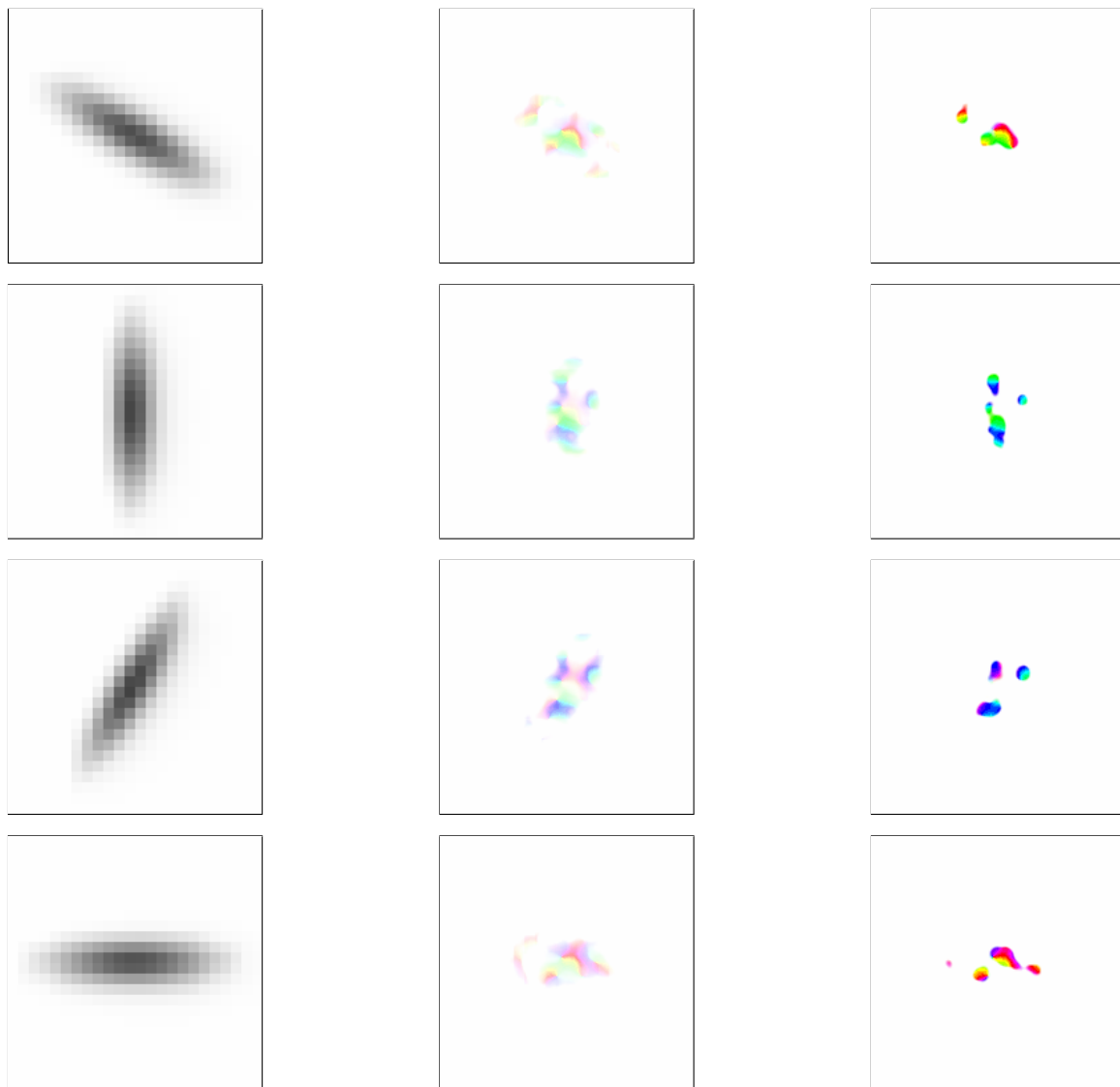
(*a*) Input pattern   (*b*) Initial activity   (*c*) Settled activity
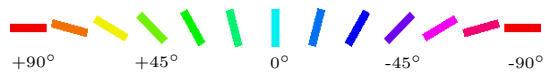
Figure 4.6: **Cortical responses to various oriented inputs (*color figure*).** The first column (*a*) shows sample oriented Gaussians, at +60°, 0°, -30°, and -90° from vertical (top to bottom, respectively.) The second column (*b*) shows the initial response of the trained map to that input, based on the afferent weights only (before the lateral interactions are allowed to settle). The third column (*c*) shows the settled activity for that input. For (*b*) and (*c*), colors indicate the orientation preference of each activated neuron, as in figure 4.3. Each input activates neurons which prefer orientations like it, within the cortical region corresponding to the active area of the retina. The initial response is wide and diffuse, like the input pattern, but the settled response is focused and sharp.

46
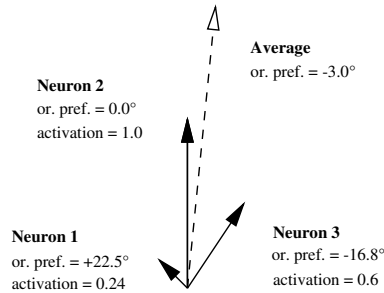
Figure 4.7: **Computing the weighted average orientation.** The activations and orientations of three neurons are shown as vectors (solid lines). The angle of each vector is twice the orientation preference of that neuron, in order to map the 180° of orientation preferences onto the full 360°, making the vector orientation for 0° and 180° identical. The magnitude of each vector is the activation level, to ensure that the encoding primarily reflects those neurons that are most active. The sum of these vectors is shown as a dashed line. The average orientation of the three neurons shown is half of the angle of the dashed vector.

scaled by two to get the angle of the vector, which ranges over 360°. Since each neuron is to contribute only to the extent that it is active, the magnitude of the vector is taken to be the activation level of the neuron. Once the neurons have been represented in this fashion, the average orientation can be computed from the orientation of the vector sum. Figure 4.7 illustrates these calculations.

Using the vector sum method of computing the average, the perceived orientation was calculated for the trained network using both the average value and the maximum value methods. For each each angle, an oriented Gaussian (of the same shape as in training) was presented at the center of the retina, and the estimated orientation was computed using each method. The results are presented in figure 4.8.

Both methods appear to be reasonably accurate representations of orientation. Small deviations from the true angle are present, however, because the activated cortical area is comparable to the size of an orientation column. The number of neurons with RFs receptive to the peak of the input Gaussian is relatively small, and thus the distribution of orientation preferences will not necessarily be uniform within that population. Thus, at different locations on the retina, the estimated orientation will differ slightly. If much larger inputs are used, activating a large cortical area, these effects should cancel out, allowing arbitrarily accurate orientation encoding. However, this would be computationally prohibitive to verify at present since much larger cortex and retina sizes would need to be simulated.

As one might expect, utilizing all active units in the calculation of the perceived orientation is slightly more accurate than only using those at maximum activation, but the difference is not significant because many neurons are at maximum and thus most
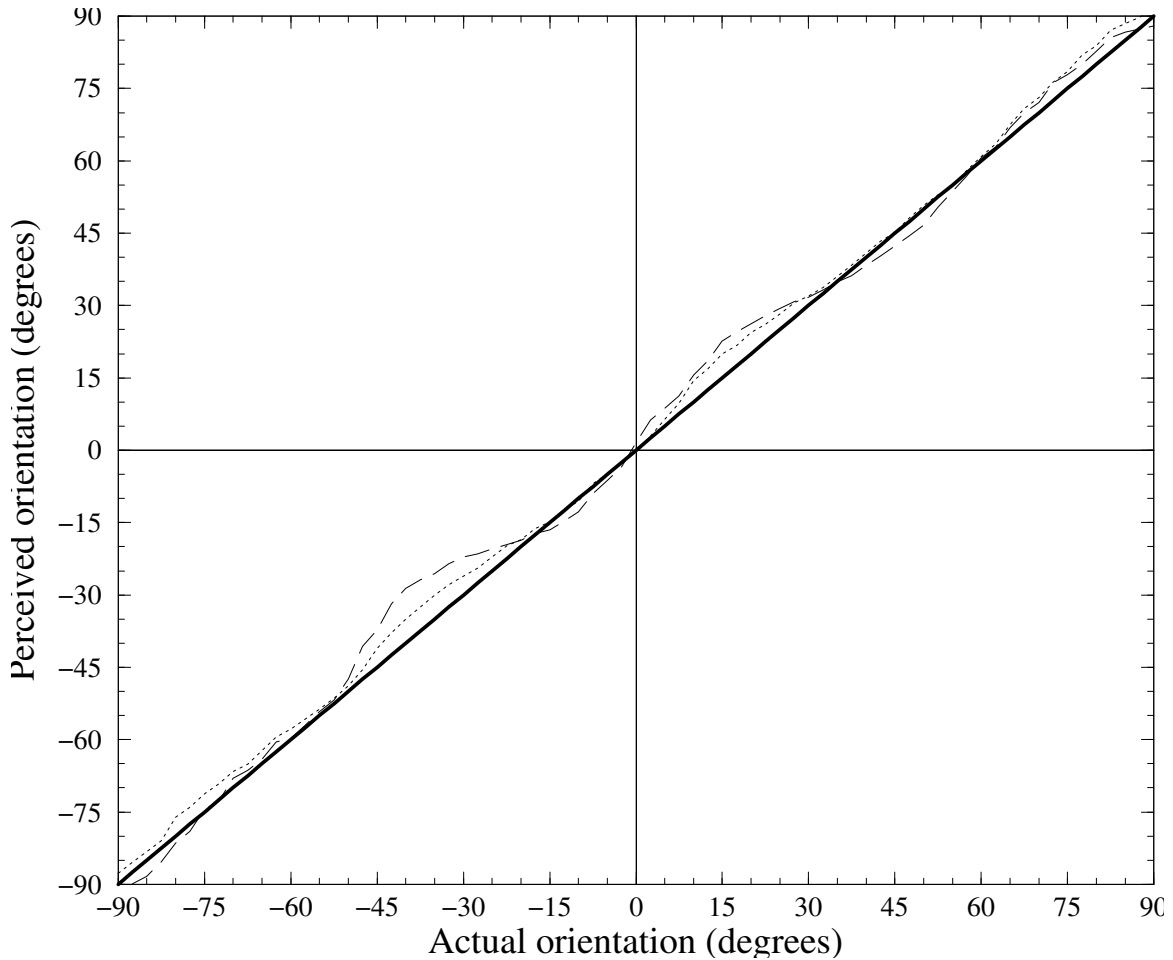
Figure 4.8: **Estimates of perceived orientation.** For each possible orientation of a Gaussian at the center of the retina, the heavy line shows the orientation that would be perceived if perception were veridical (i.e., we would perceive the actual orientation of the Gaussian.) The other lines show the perceived orientation of a Gaussian at the center of the RF-LISSOM retina estimated using two different algorithms. The dotted line shows the estimate obtained by computing an average of the orientation of each activated neuron weighted by the activation level of that neuron. The dashed line shows the estimate obtained by performing a similar computation on only those units that have reached maximum activation. Both methods are reasonably accurate estimates, although using all activated units (dotted line) is slightly more accurate.

48

of the neurons averaged in each case are identical. For simplicity, I have chosen in later experiments to use only the method of averaging all units, but the results should be the same for either method, or for any similar method. Note that there is insufficient biological evidence to support a particular method, and I do not claim that orientation perception actually need be occurring in precisely this way. In any case, to ensure that the biases shown in figure 4.8 do not distort the results, all perceived orientation measurements in this thesis are stated in terms of differences in perceived angles, rather than in terms of the actual orientation on the retina.

4.6   Conclusion

Starting from unoriented connections, neurons in the RF-LISSOM model develop oriented receptive fields and patterned lateral connections cooperatively and simultaneously. This input-driven self-organization represents the salient features of the training examples: the afferent connections develop feature detectors to distinguish between the inputs seen, and the inhibitory connections represent the long-range activity correlations between feature-selective cells. During visual processing, this information is used to eliminate redundant information, and enhance the selectivity of cortical cells. The self-organized map forms an accurate representation of the orientation and position of the input lines, suitable for use at higher levels and for further processing. If these self-organizing processes remain active even in the adult animal, they will continually act to adapt cortical response properties to match the visual environment. As discussed in the next chapter, this can result in phenomena such as tilt illusions and aftereffects.

Chapter 5

Aftereffect Experiments

Using an orientation map apparently much like those in chapter 4, humans experience tilt aftereffects that vary systematically with the angle between test and adaptation lines. The longer one adapts to an orientation, the further away similar orientations seem to be, and the closer distant orientations seem to be. These perceptions have been measured in some detail in humans, and since they can be measured in the RF-LISSOM model as well, they offer an opportunity to put the inhibition theory of the TAE to the test.

This chapter will describe how the orientation map from the previous chapter was set up to test for the TAE, and it will show that the model exhibits quite realistic tilt aftereffects. It will also show precisely how those effects arise in the model, with detail as yet unavailable in the cortex. This analysis provides testable predictions for future biological and psychophysical studies. Chapter 6 will discuss the significance of these results and propose other areas for investigation.

5.1 Experimental setup

The trained orientation map network from chapter 4 was used to examine the tilt aftereffect by measuring the activity of the cortex with different inputs and learning rates.[1] The state of the map at 30,000 iterations was taken as a starting point for each independent experiment in this chapter, so the results roughly correspond to testing a single human subject under different conditions. To simulate adapting to an oriented stimulus as in the psychophysical experiments, the position and angle of the inputs were fixed to a single value for a number of iterations (figure 5.1a). Previously, during self-organization, a uniform, random distribution of all angles and positions was used, so that the cortex adapted to a wide variety of stimuli. To see significant tilt aftereffects, the inputs must be restricted to a small range of orientations, though they need not remain perfectly fixed as was done for

---

[1]This section may be skipped unless you are interested in the details of the TAE testing setup.

(a) TAE adaptation procedure   (b) TAE test procedure
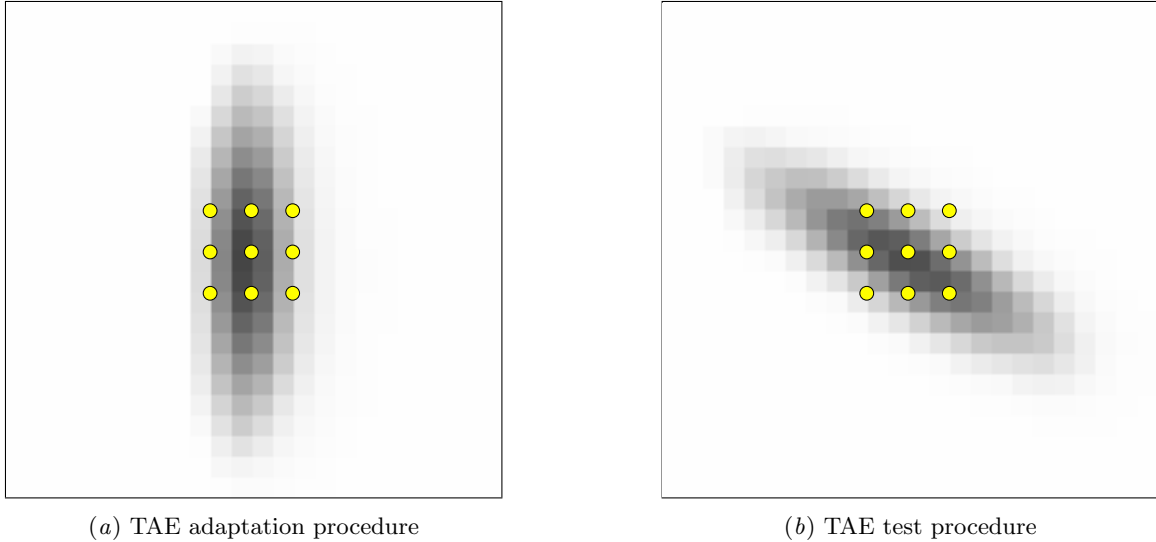
Figure 5.1: **RF-LISSOM TAE measurement procedure.** To simulate a single trial of adaptation to a fixed stimulus, a fixed location for an input was chosen from one of the nine positions on the retina indicated by the small circles. Then a vertical Gaussian was presented at that location (as illustrated in (a) for the center location), activity was propagated, the lateral interactions were allowed to settle, and weights were modified according to the usual RF-LISSOM algorithms. This procedure was repeated for a number of iterations. Afterwards, the net effect of the adaptation was evaluated for test lines of all different orientations presented at the same position chosen for adaptation, as illustrated (b). The spacing of the positions was chosen so that the entire stimulus remained within the central region of the cortex, yet covered at least a full set of orientation preferences on the map.

the results presented here. The position of the adaptation inputs is not important as long as some of them appear in the area that will later be tested for the TAE; each input causes adaptation in the area corresponding to it in the cortex. Finally, to illustrate the separate contributions of adapting the afferent, lateral excitatory, and lateral inhibitory weights, the learning rate of each (α_A, α_E, and α_I, respectively) was varied relative to the others. All other parameters remained as in chapter 4, including the size and shape of the oriented Gaussian inputs.

As adaptation progressed, test lines at various orientations were presented without modifying any weights (figure 5.1b). For each test line, the perceived orientation was measured as described in section 4.5. The magnitude of the tilt aftereffect was defined as the perceived orientation of the test line after adaptation minus the perceived orientation before adaptation.

This procedure is similar to the procedure used for human subjects described in section 2.2.1. For computational efficiency, it takes advantage of the fact that learning can be turned off in the model. Since experiments with humans cannot turn off learning for testing,

they cannot evaluate the effects of a single adaptation episode on multiple test orientations equally. That is, the data for the second test figure presented after adaptation would be affected by the presentation of the first test figure. Instead, psychophysical experiments use a single test figure. They evaluate the effect on its perceived orientation by adaptation at different angles relative to the test figure (see section 2.2.1). This alternative procedure is computationally prohibitive to simulate (not to mention extraordinarily time-consuming to measure in humans) because one must ensure that the cortex has returned to equilibrium between the adaptation episodes. Since the angular function of the tilt aftereffect has been demonstrated to be similar for all orientations (Mitchell and Muir 1976), the two procedures should give an equivalent measure of the amount of aftereffect at each angle. The procedure used here allows much more comprehensive data to be collected for the model, as shown in section 5.5.

Because the TAE curves differ substantially between individuals, particularly in the zero crossing between direct and indirect effects, it can be misleading to average results from different subjects or testing paradigms. For instance, if the zero crossings vary over some range, a null area will show up in the graph around that region, even though no individual exhibited a null area. However, since the data are generally too erratic to interpret from a single run, multiple runs from a single individual are usually averaged (e.g. Mitchell and Muir 1976).

To obtain similar measurements for the RF-LISSOM model, a single orientation map was tested separately at 9 different positions forming a 3×3 grid covering a retinal area 6×6 units around the center of the retina (figure 5.1). This range of positions covers many different orientation preferences. Averaging over these positions reduces the random fluctuations in TAE magnitude (see section 4.5), but it does not change the basic shape of the curve presented in the following section.

5.2   Angular function of the tilt aftereffect

The TAE testing procedure described in the previous section was used to determine the amount of aftereffect for each difference between adaptation and testing orientations in the model. Figure 5.2 plots the aftereffects after adaptation for 90 iterations of the RF-LISSOM algorithm. For comparison, figure 5.2 also shows the most detailed data available for the TAE in human foveal vision (Mitchell and Muir 1976).

The results from the RF-LISSOM simulation are strikingly similar to the psychophysical results. For the range $5°$ to $40°$, all subjects in the human study (including the one shown) exhibited angle repulsion effects nearly identical to those found in the RF-LISSOM model. The magnitude of the TAE increases very rapidly to a maximum angle repulsion at approximately $10°$, falling off somewhat more gradually to zero as the angular separa-
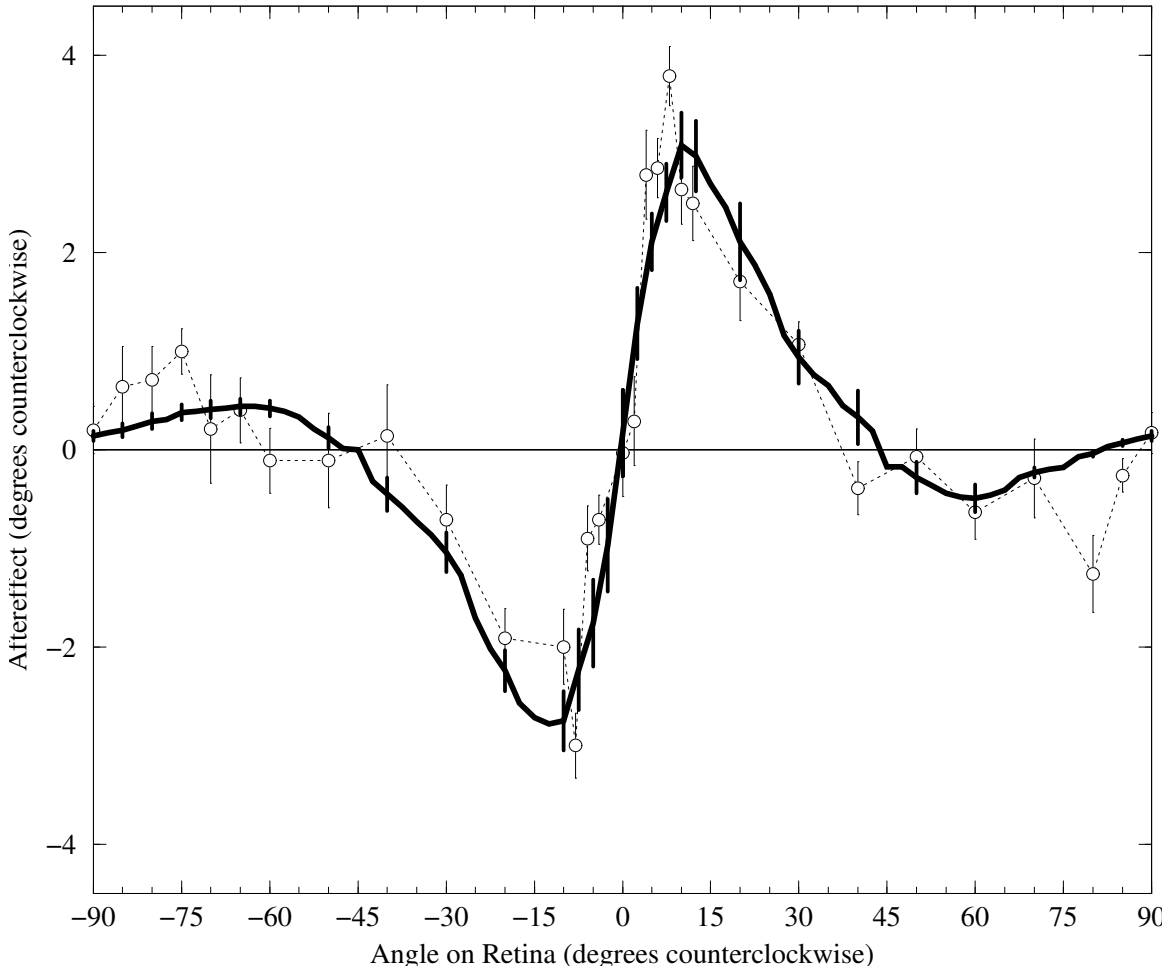
Figure 5.2: **Tilt aftereffect versus retinal angle.** The open circles represent the average tilt aftereffect for a single human subject (DEM) from Mitchell and Muir (1976) over ten trials. For each angle in each trial, the subject adapted for three minutes on a sinusoidal grating of a given angle, then was tested for the effect on a horizontal grating. Error bars indicate ± 1 standard error of measurement. The subject shown had the most complete data of the four in the study. All four showed very similar effects in the range $\pm 40°$; the indirect TAE for the larger angles varied widely in the range $\pm 2.5°$. The graph is roughly anti-symmetric around $0°$, so the TAE is essentially the same in both directions relative to the adaptation line. For comparison, the heavy line shows the average magnitude of the tilt aftereffect in the RF-LISSOM model over nine trials, as described in section 5.1. Error bars indicate ± 1 standard error of measurement. The network adapted to a vertical adaptation line at a particular position for 90 iterations, then the TAE was measured for test lines oriented at each angle. Positive values of aftereffect denote a counterclockwise change in the perceived orientation of the test line. The duration of adaptation was chosen so that the magnitude of the TAE matches the human data; as discussed in section 5.5 the shape of the curve is nearly constant, but the magnitude increases with adaptation. Learning rates were $\alpha_A = \alpha_E = \alpha_I = 0.00005$. The result from the model closely resembles the curve for humans at all angles, showing both direct and indirect tilt aftereffects.

tion increases. Section 5.4.2 examines the mechanisms responsible for these *direct* effects in RF-LISSOM.

The results for larger angular separations (from 45° to 85°) show a greater inter-subject variability in the psychophysical literature, but those found for the RF-LISSOM model are well within the range seen for human subjects. The indirect effects for the subject shown were typical for that study, although some subjects showed effects up to 2.5°. Section 5.4.3 examines the mechanisms responsible for these *indirect* effects in the RF-LISSOM model, and section 6.2.4 of the discussion proposes explanations for the variety of effects seen in human subjects.

The TAE seen in figure 5.2 must result from changes in the connection strengths between neurons, since there is no other component of the model which changes as adaptation progresses. In particular, there is no lasting change in the neuron's inherent excitability or sustained activation level. Thus there is nothing that could correspond to the concept of whole-cell neural fatigue.

Three sets of weights adapt synergetically: the afferent weights, the lateral excitatory weights, and the lateral inhibitory weights. The lateral inhibitory theory would predict that the inhibitory weights are primarily responsible for the TAE magnitude at each angle. To determine whether this is the case, the contribution of each of the weight types was evaluated independently of the others (figure 5.3). The small component of the TAE resulting from adaptation of either type of excitatory weights is almost precisely opposite the total effect. Without inhibitory learning, Hebbian adaptation of the excitatory weights causes the network to have a greater response to the input after every iteration, and boosts the response to neighboring orientations as well. If there were no inhibitory learning, this would result in a contraction of small angles, with the test line tending to be perceived as closer to the adapting line than it really is. Such contraction effects have not been documented in psychophysical experiments.

While the excitatory connections are adapting, the inhibitory connections adapt as well. Each inhibitory connection adapts with the same learning rate as the excitatory connections ($\alpha_I = \alpha_A = \alpha_E = 0.00005$), but there are many more inhibitory connections than excitatory connections. The combined strength of all the small inhibitory changes outweighs the excitatory changes, and results in a curve with a sign opposite that of the components from the excitatory weights. If excitatory learning is turned off altogether, the magnitude of the TAE increases slightly on average, but the shape of the curve does not change significantly (figure 5.3). Similar results should be obtained if there are not as many inhibitory connections as used here, but the ones present change more rapidly than afferent connections; this would represent an alternative interpretation of the results.

Thus the inhibitory connections are clearly responsible for both the direct and indirect tilt aftereffects observed in the RF-LISSOM model. The following sections will examine
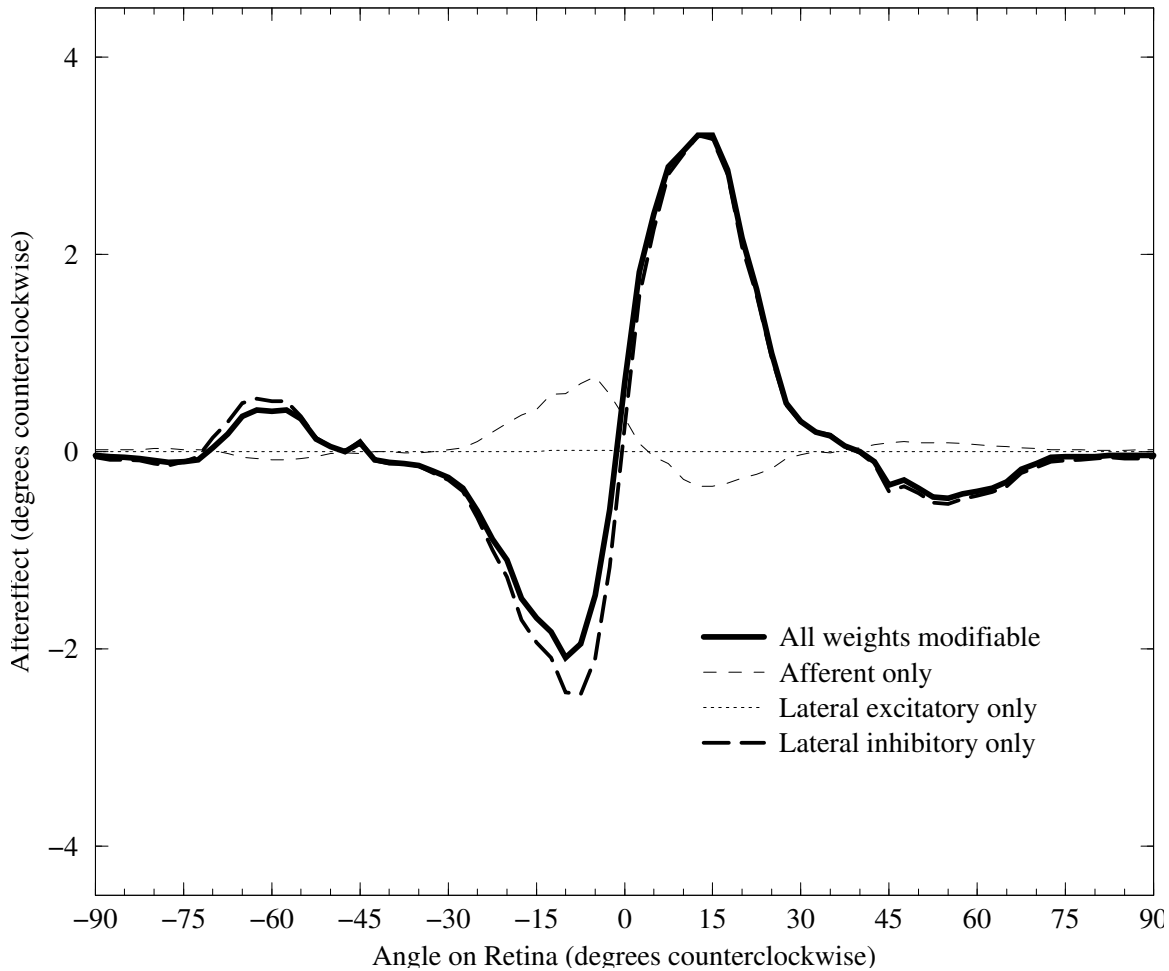
Figure 5.3: **Components of the TAE due to each weight type.** The heavy line shows the magnitude of the TAE for a single trial from the average shown in figure 5.2. This trial was at the center of the retina, and is typical of the effect seen at the other eight locations. The other curves in this figure show the contribution of adapting each weight type separately; other than the learning rates the parameters for each were identical. The thin dashed line represents the contribution from the afferent weights ($\alpha_A = 0.00005$; $\alpha_E = \alpha_I = 0$). Adaptation of the afferent weights contributes in a direction opposite to that of the overall TAE curve. The dotted line represents the contribution from the lateral excitatory weights ($\alpha_E = 0.00005$; $\alpha_A = \alpha_I = 0$). Note that the x-axis is not shown because it would have covered up this line. Adapting the lateral excitatory weights results in a curve with a similar shape as for afferent weights, but it is so small in magnitude that it is insignificant. The thick dashed line represents the contribution from the inhibitory weights ($\alpha_I = 0.00005$; $\alpha_A = \alpha_E = 0$). The adaptation of the lateral inhibitory weights clearly determines the shape of the overall curve, though it is somewhat reduced in magnitude by the afferent contribution.

exactly how changing the inhibitory weight strength produces the effects seen. To make the analysis clear and unambiguous, the simplest case that shows a realistic TAE was chosen. The discussion will center upon a single trial, using a Gaussian at the exact center of the retina, with only the inhibitory weights adapting. The TAE curve for that case is shown in figure 5.3. The curve is still quite similar to the average result when all weights were adapting at the same rate, as well as to the human data (figure 5.2). The analysis will show that Hebbian modification of the lateral inhibitory connection weights, followed by normalization of the total connection strength, systematically alters the response in a way that results in the tilt aftereffect.

5.3 Changes in the patterns of connection strength

The TAE magnitudes shown in figure 5.3 result from changes in the response to the test line after adaptation. Each point on the graph represents the change in the overall activity distribution for a test line, so it is an aggregate measure of many small changes. To illustrate the changes in more detail, this section examines how the individual weights change with adaptation. For clarity, only inhibitory weights were allowed to be modified ($\alpha_I = 0.00005$ and $\alpha_A = \alpha_E = 0$). Figure 5.4 shows how the connections of a typical neuron in the central region of the model cortex changed during adaptation.

Adaptation has redistributed the inhibitory weights. Connections to neurons that have orientation preferences similar to the adaptation line became stronger (the blue areas in figure 5.4d). The strengthening results in the direct effect: inhibition for orientations close to the adaptation orientation is increased. However, since total weight strength for all inhibitory weights is constant (equation 3.4), the connections to other orientations (the yellow and red areas) must simultaneously *decrease*. This causes the indirect effect: inhibition for distant orientations is reduced. Chapter 6 discusses possible biological mechanisms for this weight normalization.

Since the changes for each neuron are relatively small, the difference may not be readily apparent from figure 5.4. Some changes are visible on close inspection in the central blue and red-orange areas; compare figure 5.4 (c) and (d). To highlight the changes, figure 5.5 shows the result of subtracting the weight plots from before and after adaptation. Subtracting the *before* plot from the *after* plot shows which connections have increased in strength, and subtracting the *after* plot from the *before* plot shows which ones have decreased. Clearly, inhibitory connections to orientations near the adaptation line have increased in strength, while those to distant orientations have decreased. These systematic changes in the inhibitory weights result in both direct and indirect tilt after-effects, as seen in the next section.

(a) Afferent weights

(b) Lateral excitatory weights

(c) Inhibitory weights before adaptation

(d) Inhibitory weights after adaptation

+90°   +45°   0°   -45°   -90°
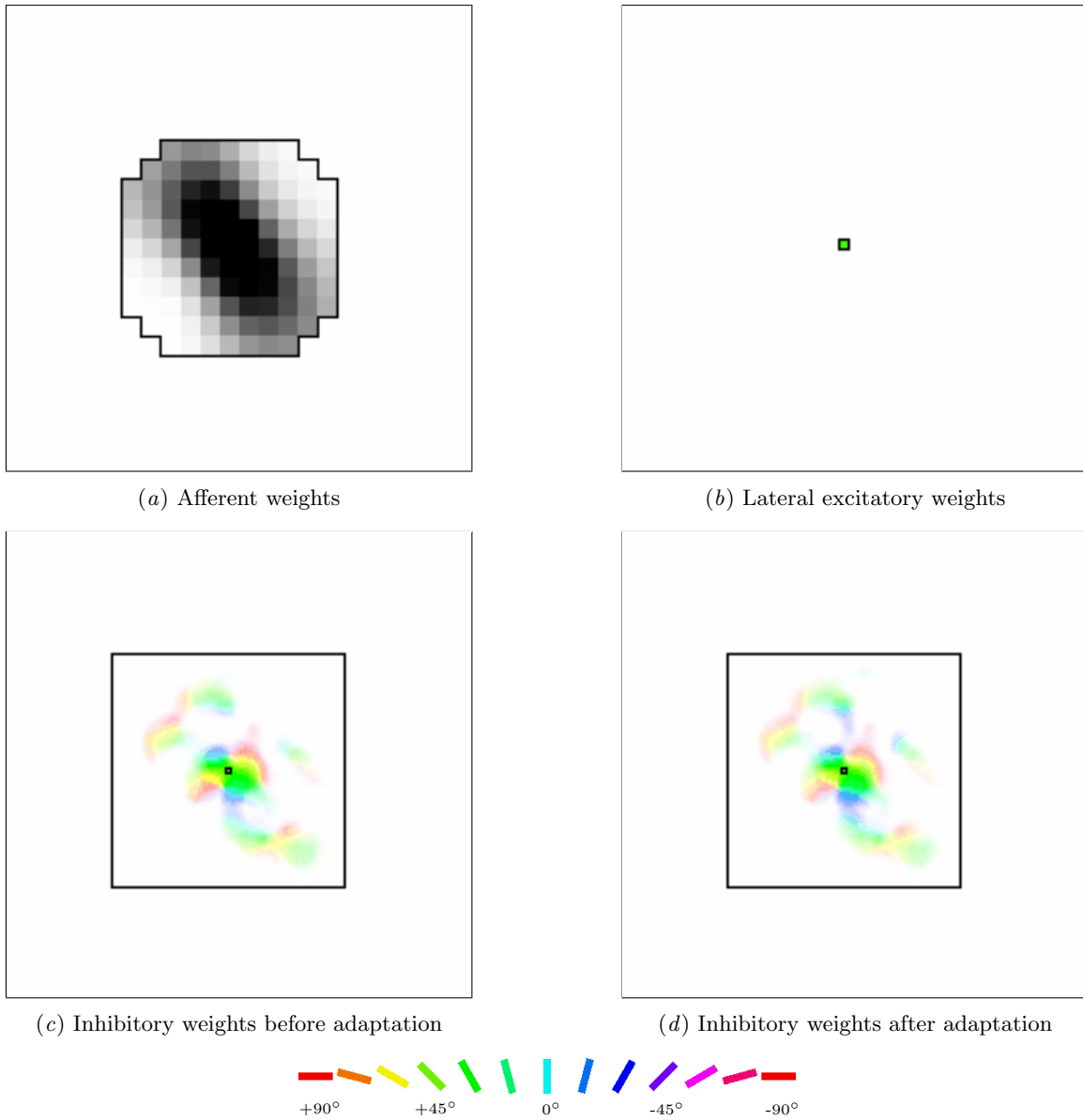
Figure 5.4: **Weights of a central neuron before and after adaptation (*color figure*).** These plots show the strength of the weights of each type going to the neuron at (91,98), which is marked with a tiny box in the cortex plots of (c) and (d). The afferent weights (a) of this neuron have an orientation preference of +45°, and were not modifiable in this experiment. The fixed excitatory weights in (b) connect only to the nearest neighbors in the cortex. Those neurons all have similar orientation preferences, so even if these weights had been modifiable they would not have changed the response of the neuron significantly. The inhibitory connections span a wide range of locations and orientation preferences. They undergo large and orientation-specific changes with adaptation. Comparing (c) with (d) shows that the blue areas along the vertical line down the center have increased in strength as a result of adapting to the vertical test line.

57

(a) Inhibitory weights that increased

(b) Inhibitory weights that decreased

+90°   +45°   0°   -45°   -90°
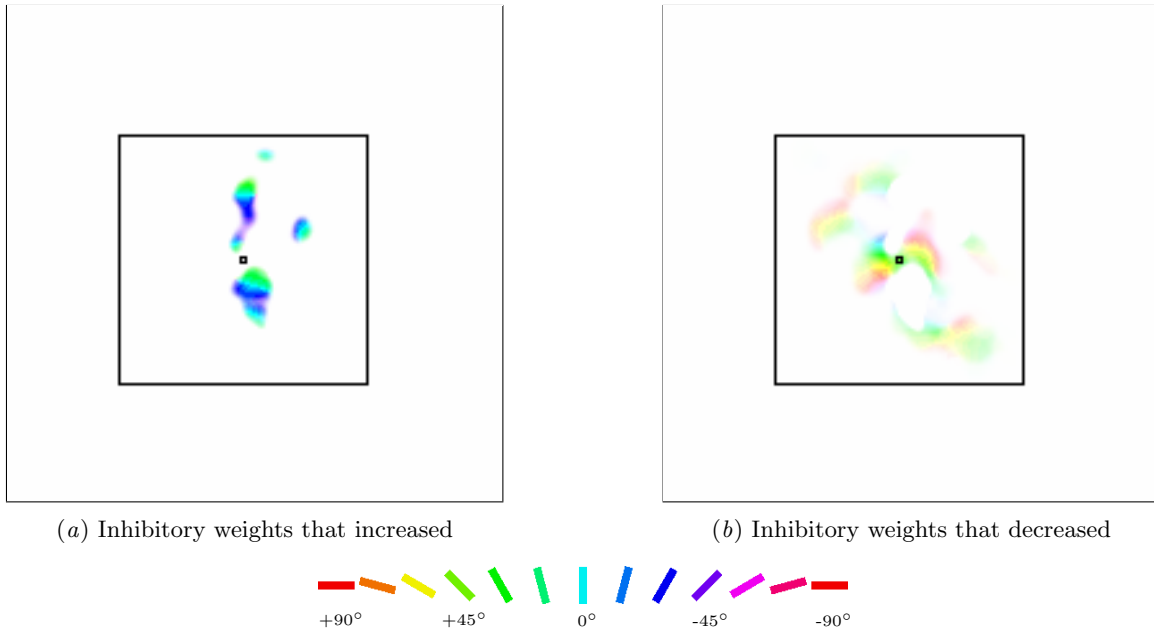
Figure 5.5: **Weight changes for a central neuron with adaptation (*color figure*).** These plots show the weight changes for the neuron in figure 5.4 in more detail. Colored areas in (*a*) represent inhibitory connections that increased in strength. The plot was computed by subtracting the weights in figure 5.4*c* from those in figure 5.4*d*, scaling the value up to a visible level, then labeling each connected neuron with the color corresponding to its orientation preference. Only connections within the cortical area corresponding to the input Gaussian have been strengthened, and only those whose orientation preferences are close to the adaptation angle (i.e., neurons colored blue, cyan, or green). These changes cause the direct tilt aftereffect, as explained in section 5.4.2. While those connections strengthened, the rest of the connections to that neuron had to weaken (due to equation 3.4). The colored areas in (*b*) represent inhibitory connections that *decreased* in strength. The plot was computed by subtracting figure 5.4*d* from figure 5.4*c*, scaling it by the same factor as in (*a*), and labeling each point with the orientation preference of the connected neuron. The weakened connections are all those present for this neuron other than those that were strengthened; each was weakened in proportion to its current strength. These changes cause the indirect tilt aftereffect, as explained in section 5.4.3.

5.4   Changes in the response to test lines

The weight changes described in the previous section result in changes in the pattern of the response to a test line. The differences in the responses were summarized as a single number (the perceived orientation) for each point on the curve in figure 5.3. This section will examine exactly how the activity patterns themselves change. First, it will look at how the cortical response to the adaptation stimulus changes, even though the perceived orientation does not change. It will then show the change in the response to test lines at 10° (where the peak direct effect was found) and 60° (where the peak indirect effect was

found). The activity patterns are each plotted onto the orientation map from figure 4.3 so that the orientation preference of the activated neurons will be evident. These plots constitute predictions of what should be found in the cortex if the TAE is occurring as it does in the RF-LISSOM model.

5.4.1 Sharpening of the response to the adaptation line

Figure 5.6 shows the response to the adaptation line before and after adaptation. The perceived orientation was approximately 0° both before and after adaptation (figure 5.3). However, there were significant changes in the settled activity (i.e., the activity remaining after lateral interactions have been allowed to settle). More neurons near the center of the cortex, where the input was strongest, were activated after adaptation. Fewer were activated in the outlying areas. Thus adaptation has sharpened the activity pattern.

The sharpening is caused by the strengthening of inhibition between activated neurons — only the areas with the strongest activation remain active after settling. The rest of the neurons initially activated in figure 5.6b are silenced by the active ones in the center. The sharpening effect is present even before adaptation, which is why the settled activity differs from the initial activity, but it gets stronger with adaptation. The inhibitory weight changes and the activity changes can be related to each other by comparing the areas showing changes between figure 5.6c and 5.6d with the areas showing increased inhibition in figure 5.5a. The increased inhibition has effectively created a group of competitive neurons where only a central region of strongly activated neurons is allowed to continue responding.

None of these changes show up in the TAE curve in figure 5.3 at 0° because the distribution of activated units contains approximately the same proportion of each orientation preference both before and after adaptation. Compared to figure 5.6c, there are more green units (orientation preference less than 0°) activated in the center of figure 5.6d. At the same time, there are more blue units (orientation preference greater than 0°) activated just outside that area. These changes cancel out for the adaptation orientation, since they are centered upon it. As discussed in the next section, the direct TAE occurs for test lines of slightly different orientations, where the changes do *not* cancel out.

5.4.2 Expansion of small angles (direct effect)

For the peak of the direct effect, the perceived orientation of a 10° test line shifted away from 0° by several degrees (figure 5.3). This change is quite subtle but evident in the activity patterns for the +10° test line in figure 5.7. After adaptation, fewer of the units encoding orientations less than +10° are activated (the cyan and blue areas). Slightly more of the units encoding orientations greater than +10° are activated (the yellow, yellow-green, and green areas). The net effect is a shift of the perceived orientation *away* from 0°, resulting

(a) Input pattern

(b) Initial activity

(c) Settled activity before adaptation

(d) Settled activity after adaptation

+90°   +45°   0°   -45°   -90°
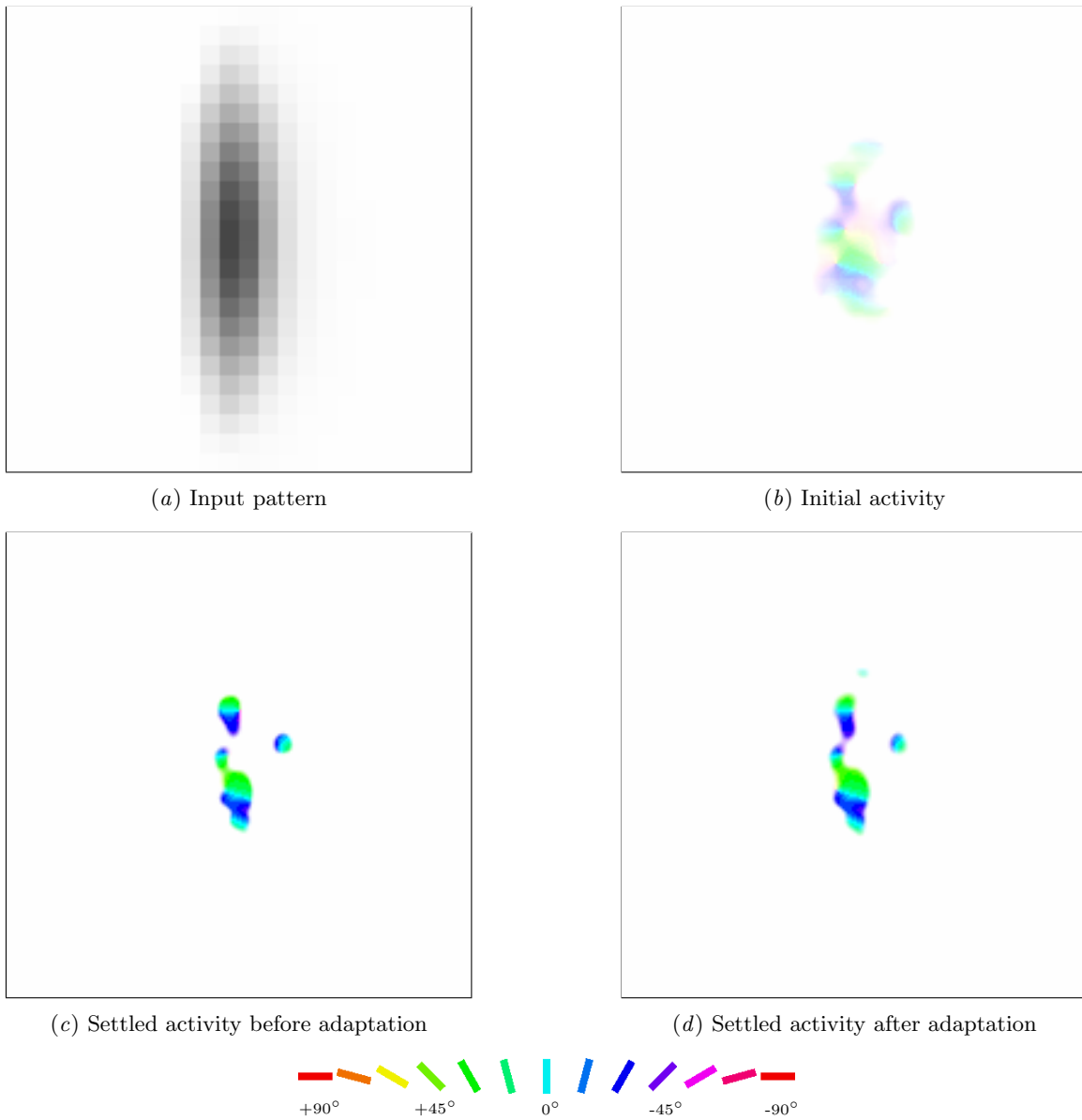
Figure 5.6: **Cortical response to the adaptation figure (*color figure*).** The initial (pre-settling) and settled activity are shown for the adaptation Gaussian (a). The initial activity (due to the afferent weights only) is wide and diffuse, like the input pattern. It does not change with adaptation, since the afferent weights were fixed in this experiment. The settled activity did change, becoming more concentrated near the center where the input was the strongest. Still, the average orientation remained relatively constant, with equal areas of colors on either side of 0° activated both before and after adaptation.

60

(*a*) Input pattern

(*b*) Initial activity

(*c*) Settled activity before adaptation

(*d*) Settled activity after adaptation

+90°   +45°   0°   -45°   -90°

Figure 5.7: **Cortical response at the peak direct effect (*color figure*).** For the input pattern shown (10° counterclockwise from the adaptation inputs), the settled activity changes significantly with adaptation. Fewer of the units encoding orientations greater than +10° (i.e., the cyan and blue areas) are activated. Slightly more of the regions encoding orientations less than +10° (the green and yellow-green areas) are activated. The net effect is an angle expansion: the average orientation changes from +14.5° to +16.9°. The change is perceived as the direct TAE.

61

in the direct TAE.

The changes are once again due to the strengthening of the inhibitory connections. As shown in figure 5.5a, the inhibitory connections between the green, cyan, and blue neurons were strengthened. This test line initially activated the green orientations most strongly, so lateral interactions allowed those neurons to silence the weakly-activated cyan and blue areas. The increased inhibition between neurons active during adaptation results in competition between them for activation — only a subset will remain active at a given time. Because these changes occurred only for neurons preferring orientations closer to 0° than 10°, the perceived orientation shifts away from 0°. Thus direct tilt aftereffects are occurring in the RF-LISSOM model exactly as predicted by the lateral inhibition theory.

5.4.3   Contraction of large angles (indirect effect)

For the peak of the indirect effect, the perceived orientation of a 60° test line shifted towards 0° by nearly a degree (figure 5.3). This change is again subtle, yet still evident in the activity patterns for the +60° test line in figure 5.8. After adaptation, more of the units encoding orientations smaller than +60° are activated (the green areas). The same units encoding orientations greater than +60° are activated both before and after adaptation; these units show no change because they were never activated during adaptation. The net effect is a shift of the perceived orientation *towards* 0°, resulting in the indirect TAE.

The indirect effect, true to its name, results only indirectly from the strengthening of the inhibitory connections. The inhibition between the green, cyan, and blue neurons was strengthened during adaptation. Of those, only the green neurons were activated by this test line, since it is quite distant from the adaptation line. Thus the stronger inhibitory connections themselves did not change the response pattern significantly, since the targets were already inactive.

However, since the total inhibitory connection strength to each neuron is limited (equation 3.4), the increases in strength among the green, cyan, and blue neurons come at the expense of all the other inhibitory connections to the green neurons. As shown in figure 5.5b, the weakened connections are spread out over neurons with a wide range of orientation preferences. The weakened connections include those to nearly all of the units activated for this test line. Thus those units inhibit the green areas much less than they did before adaptation. Consequently, more of the green neurons remain active throughout the settling process. Before adaptation, many of them had been silenced by the other neurons closer to the orientation of the test line. The net effect of these changes is that the perceived orientation shifts *towards* 0°.

The magnitude of the indirect effect is smaller than the direct effect because fewer adapted neurons are involved. Most of the neurons showed little change; only those few closest to the adaptation orientation were affected. For the direct effect, many neurons

(a) Input pattern

(b) Initial activity

(c) Settled activity before adaptation

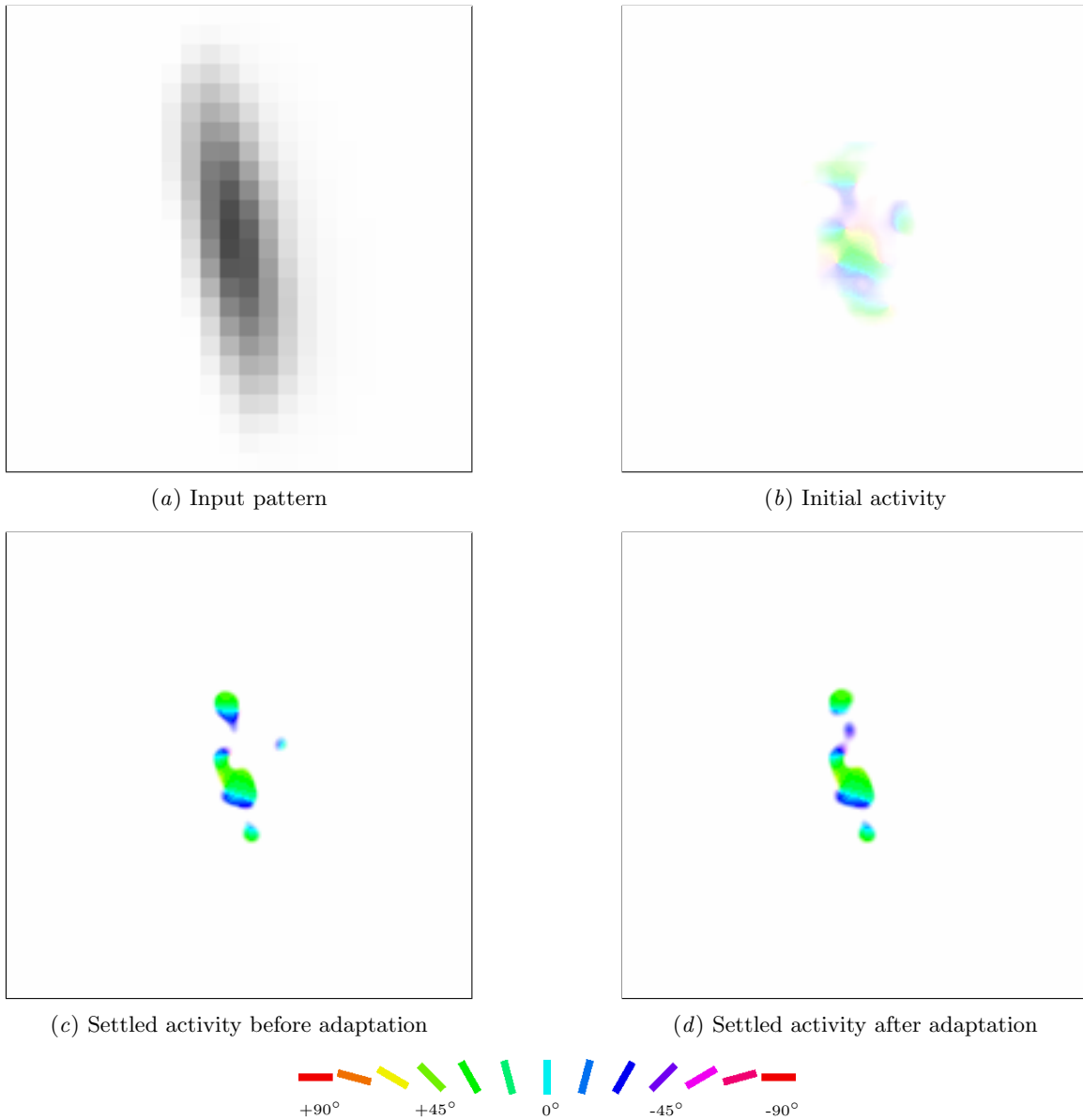(d) Settled activity after adaptation

+90°   +45°   0°   -45°   -90°
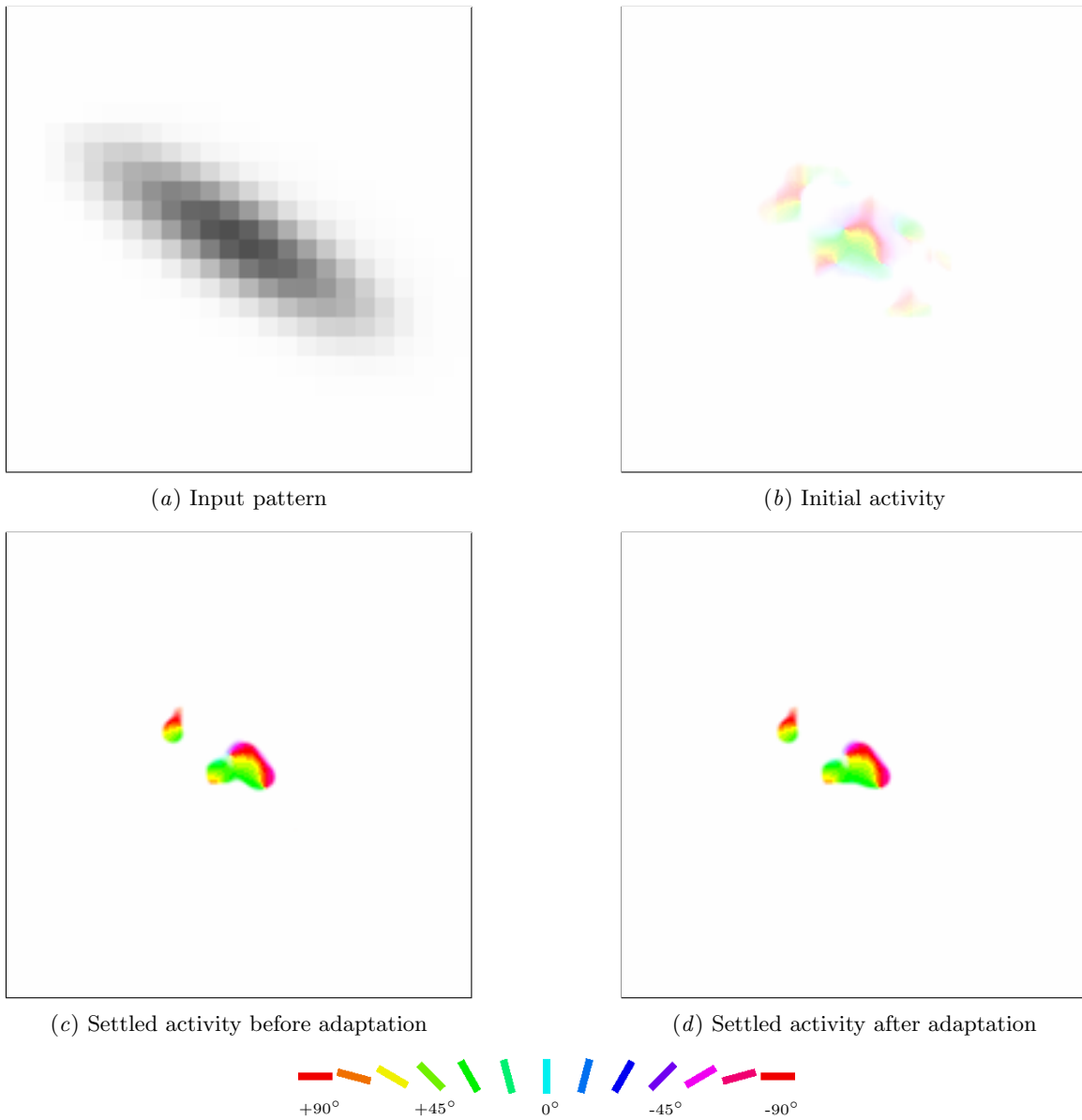
Figure 5.8: **Cortical response at the peak indirect effect (*color figure*).** For the input pattern shown (60° counterclockwise of the adaptation pattern), the settled activity again changes significantly with adaptation, but in the opposite direction from that of figure 5.7. More of the units encoding orientations greater +60° (i.e., the green areas) are activated, since they are no longer inhibited as strongly as before adaptation. The regions encoding orientations less than the testing orientation (the yellow and red areas) show little change. The net effect is an angle attraction: the average orientation changes from +58.3° to +57.4°. The change is perceived as the indirect TAE.

were affected — a number of those preferring orientations greater than that of the test line became less active, and some of the ones preferring smaller orientations became more active. These changes add up to a greater difference in the perceived orientation for the direct effect than for the indirect effect.

The indirect effect in RF-LISSOM is a simple consequence of the normalization of inhibitory weights after they have been adjusted by Hebbian learning. This explanation for the indirect effect was not discovered until the TAE was modeled here, thus underscoring the importance of building computational models. Chapter 6 will discuss the biological evidence for this theory of the indirect effect, and will compare it to the evidence for other theories that have been proposed.

The model has allowed us to look into the cortex as it is processing and see the changes as they occur. No other computational model has allowed the tilt aftereffect to be studied so concretely. As also discussed in chapter 6, it may be possible to obtain similar plots for the cortex to compare with these results.

5.5 Time course of the tilt aftereffect

In addition to the angular changes in the TAE described in the previous sections, the magnitude of the TAE in humans increases regularly with adaptation time (Gibson and Radner 1937). The equivalent of "time" in the RF-LISSOM model is an iteration, i.e. a single cycle of input presentation, activity propagation, settling, and weight modification. Figure 5.9 shows how the TAE varies for each angle as the number of adaptation iterations is increased. The same basic S-shaped curve is always evident, but its magnitude increases monotonically with adaptation. Since obtaining human data for even a single curve is extremely time consuming, equivalently comprehensive data for human subjects is unavailable. The plots are easily obtained for the model, however, which illustrates one reason why computational models are important — they make it possible to obtain quite detailed data that can constitute predictions for later experimental work.

The experimental work that has been done so far on the time course of the TAE in humans (Gibson and Radner 1937; Greenlee and Magnussen 1987; Magnussen and Johnsen 1986) corresponds to a single vertical slice through the direct TAE region in figure 5.9. It is measured by presenting an oriented adaptation figure for an extended period interrupted at intervals by the presentation of test and comparison lines. At each test presentation, the amount of the TAE present is measured using the standard techniques illustrated in figure 2.3. The duration of the test presentation is minimized so that it will not affect the magnitude of the TAE significantly.

When the time course of the direct TAE is measured in this way for human subjects, the increase is approximately logarithmic with time (Gibson and Radner 1937), as is evident
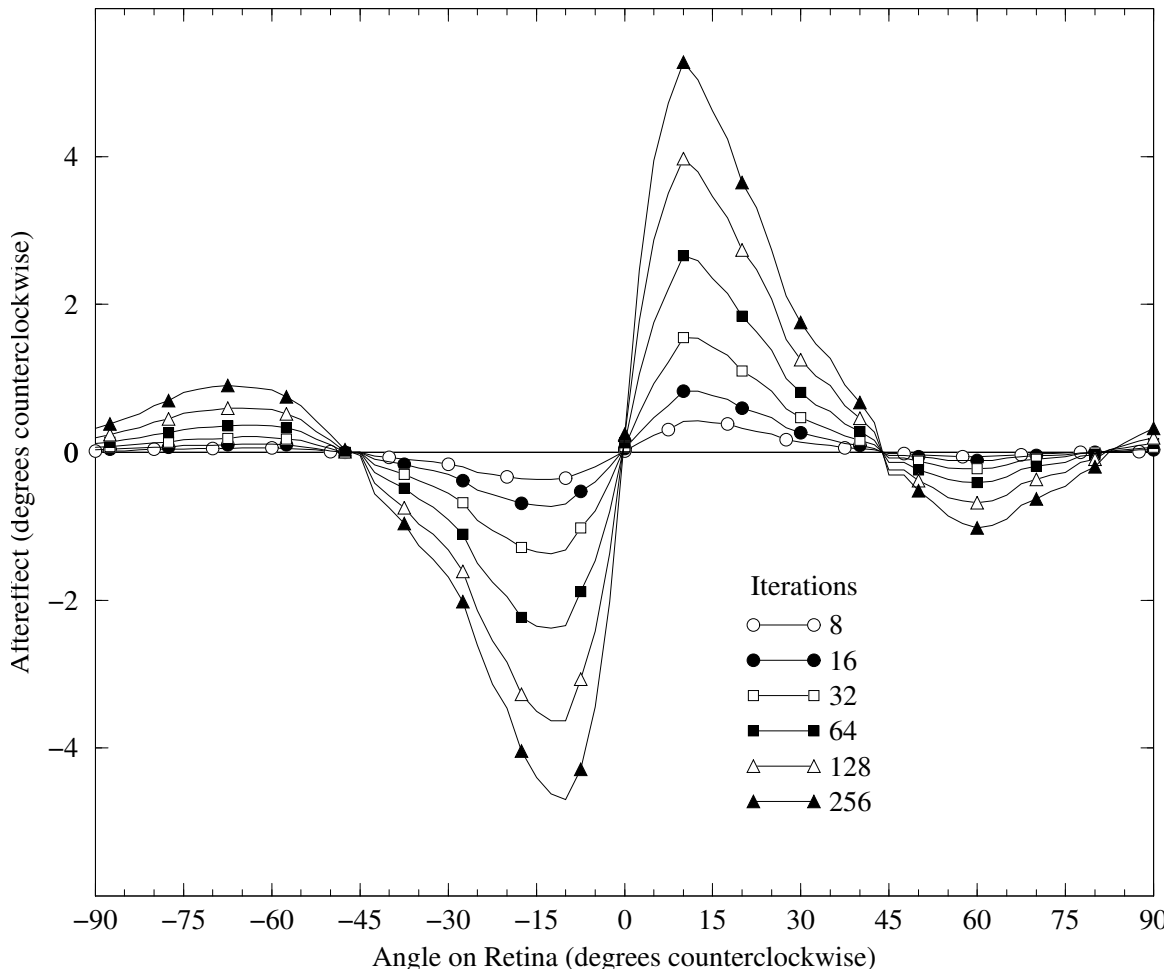
Figure 5.9: **Tilt aftereffect versus angle over the course of adaptation.** Each curve shows the average TAE in the RF-LISSOM model from figure 5.2 with a different amount of adaptation. The same basic S-shaped curve is seen regardless of the duration of adaptation, but the magnitude increases monotonically with adaptation.

for the model in figure 5.9. The magnitude of the TAE eventually reaches saturation at a level that depends upon the experimental protocol used (Greenlee and Magnussen 1987; Magnussen and Johnsen 1986). Figure 5.10 compares the shape of the TAE versus time curve for human subjects and for the RF-LISSOM model. The x axis for the RF-LISSOM and human data has different units, but the correspondence between the two curves might provide a rough way of quantifying the equivalent real time for an "iteration" of the model. The time course of the TAE in the RF-LISSOM model is similar to the human data, but the model does not show saturation effects over the adaptation amounts tested so far. This difference indicates that the biological implementation has additional constraints on the
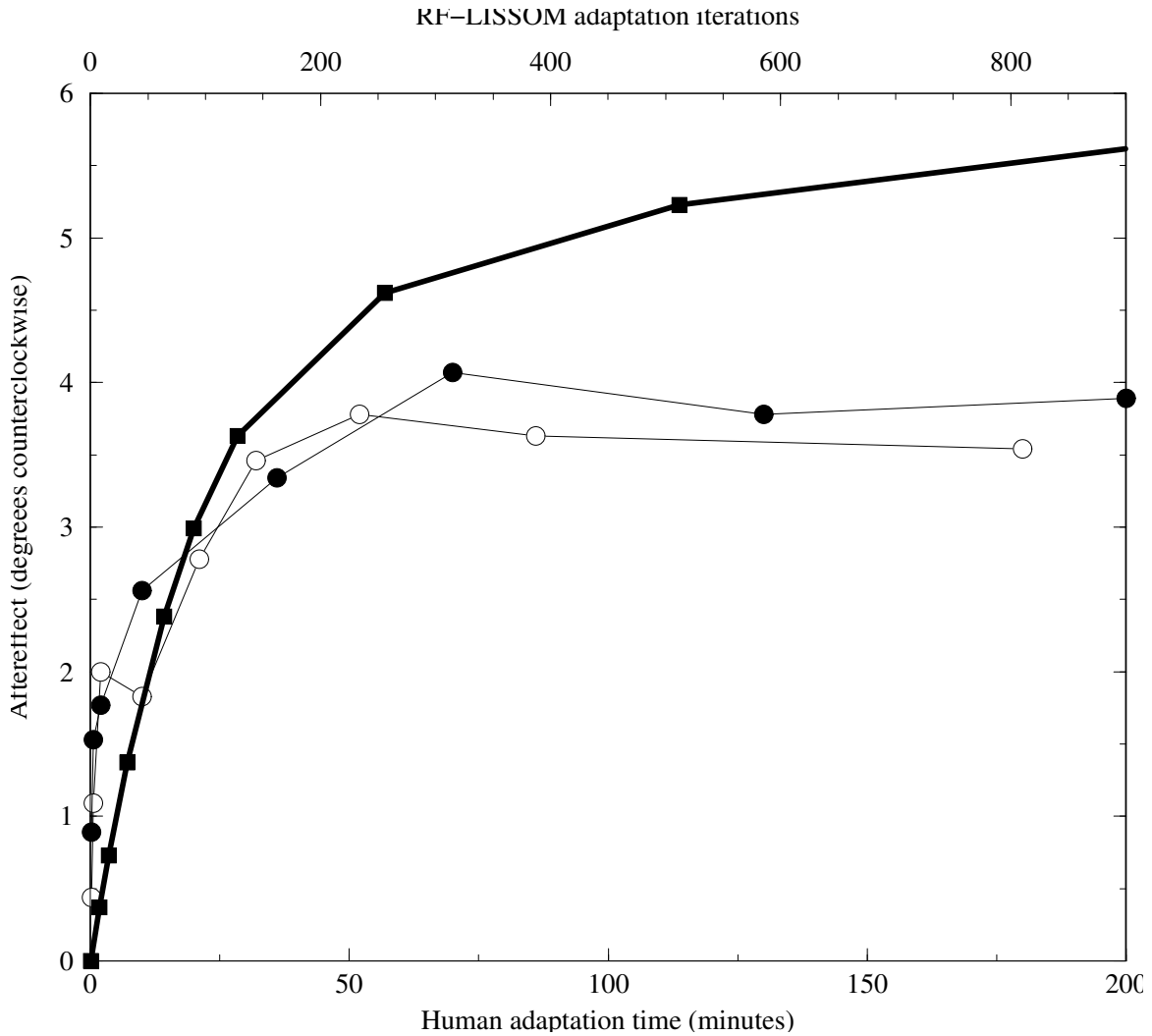
Figure 5.10: **Direct tilt aftereffect versus time.** The circles show the magnitude of the TAE as a function of adaptation time for human subjects MWG (unfilled circles) and SM (filled circles) from Greenlee and Magnussen (1987); they were the only subjects tested in the study. Each subject adapted to a single $+12°$ line for the time period indicated on the horizontal axis (bottom). To estimate the magnitude of the aftereffect at each point, a vertical test line was presented at the same location and the subject was requested to set it a comparison line at another location to match it. The vertical location of each point represents the average of five settings; the data for 0 – 10 minutes were collected separately from the rest. For comparison, the heavy line shows average TAE in the LISSOM model for a $+12°$ test line over 9 trials (with parameters as in figure 5.2). The horizontal axis (top) represents the number of iterations of adaptation, and the vertical axis represents the magnitude of the TAE at this time step. The RF-LISSOM results show a similar logarithmic increase in TAE magnitude with time, but do not show the saturation that is seen for the human subjects.

amount of learning that can be achieved over the time scale over which the tilt aftereffect is seen. This issue is explored further in the discussion in the next chapter, including possible ways that saturation might be incorporated into the RF-LISSOM model.

5.6   Conclusion

The RF-LISSOM model was shown to exhibit tilt aftereffects that are very similar to those measured in humans. This result links the detailed low-level behavior of the orientation map from chapter 4 with measurements of human performance at the level of conscious perception. It also demonstrates that the same self-organizing principles driving the development of the map can account for behavior in the adult. The model suggests a new explanation for the indirect effect, namely that limited synaptic resources require balancing of synaptic strengths. Having a large-scale computational model also permits quite detailed analysis that shows exactly how the TAE occurs. For instance, weight changes and the corresponding changes in activity patterns can be visualized directly in the model, and data for the time course of the TAE can be obtained at a level of detail impractical for human subjects. This analysis represents a comprehensive set of predictions of what may be going on in human perception, and thus provides opportunities for later experimental work. Chapter 6 discusses how the RF-LISSOM results relate to biological evidence, and describes specific predictions for human and animal experiments as well as suggesting areas for future research with the model.

Chapter 6

Discussion and Future Work

The results presented in the previous chapter demonstrate rigorously that activity-dependent plastic lateral interactions could be responsible for the tilt aftereffect. The same self-organizing principles that result in sparse coding and reduce redundant activation may also be operating over short time intervals in the adult, with quantifiable psychological consequences. This finding demonstrates a potentially important computational link between development, structure, and function.

To clarify exactly which aspects of human behavior have been replicated in the model, the following section will review the available psychophysical evidence and compare it to the results found in the model. The model accounts for the great majority of such evidence to date, but it does not yet demonstrate the saturation effect or decay of the TAE in darkness. Later sections will discuss how the cortex may implement the functions performed by the model, using the available biophysical evidence as constraints on the types of models considered to be plausible. The adult cortex might implement quite different mechanisms for plasticity than are present in the developing animal. Based on the results in chapter 5, a number of predictions for verification of the model by human and animal experiments are also discussed. For instance, one could measuring cortical activity analogously to the activity plots in the previous chapter. This thesis also opens up a variety of new directions for modeling function with RF-LISSOM, some of which are proposed at the end of this chapter.

6.1   Psychophysical evidence relating to the TAE

Even though the RF-LISSOM model was not developed as an explanation for the tilt aftereffect, it exhibits tilt aftereffects that have nearly all of the features of those measured in humans. With the appropriate extensions, it is expected to account for all of the known data on the TAE. The features of the TAE which have already been demonstrated in the

model include those in the following list. Except where noted, all of these features have been replicated by a number of researchers and are quite consistent across different studies, so any viable model of the tilt aftereffect would be expected to account for them.

Null at training angle: Adaptation does not change the perceived orientation of the stimulus used during adaptation (figure 5.2; (Gibson and Radner 1937).

Direct effect: Similar orientations are misperceived by human subjects as having a larger difference than they actually do (Gibson and Radner 1937), peaking between 5° and 20°, typically 10°– 13° (Howard and Templeton 1966,p.216; Campbell and Maffei 1971; Mitchell and Muir 1976). In RF-LISSOM, peaks at locations between 5° and 15° have been observed; 10° is typical (figure 5.2).

Null between direct and indirect effects: Human aftereffects return to zero somewhere between 25° and 50° (Campbell and Maffei 1971; Mitchell and Muir 1976; Muir and Over 1970). Zero-crossings between 30° and 60° have been observed in RF-LISSOM; 45° is typical (figure 5.2).

Indirect effect: Distant orientations are misperceived as having a smaller difference than they actually do (Gibson and Radner 1937), peaking somewhere between 60° and 85° (Muir and Over 1970; Campbell and Maffei 1971; Mitchell and Muir 1976). In RF-LISSOM, indirect effect peaks between 45° and 75° have been observed; 60° is typical. These effects vary significantly between studies and different individuals, as described in section 6.2.4.

Time course: The TAE magnitude increases at a diminishing rate with further adaptation (figure 5.10; Gibson and Radner 1937; Greenlee and Magnussen 1987; Magnussen and Johnsen 1986).

Orientation-independence: The TAE versus angle curve is similar for all pairs of test and adaptation angles which differ by the same amount, regardless of the absolute orientation on the retina (Mitchell and Muir 1976). Conflicting results were found by previous researchers, who could not demonstrate effects on oblique lines of adaptation to vertical lines. This discrepancy has been satisfactorily explained as a methodological problem of the earlier studies, and the assumption of orientation independence now seems well established.

Spatial localization: The TAE is localized to the area of the retina which was trained (Gibson and Radner 1937). Adaptation for a figure in one location has no measurable effect on test figures in other locations sufficiently distant. In the model, this occurs because weights are only adapted between active neurons, and a small stimulus activates only neurons in a small cortical area.

Several other aspects of the TAE will likely be exhibited by an RF-LISSOM model trained on more realistic input distributions. These simulations do not necessarily require any significant extensions to the model, but most will require larger cortex sizes and longer training times. When sufficient computing power is available and the model is extended as described, RF-LISSOM is expected to account for these experimental observations as well:

Higher variance at oblique orientations: The TAE versus angle curve shows greater variance for oblique testing orientations (Mitchell and Muir 1976), often sufficiently high to entirely mask the effect (Campbell and Maffei 1971). The variance may result from the smaller number of detectors in the fovea subserving angles that are neither horizontal nor vertical (Bauer et al. 1991; Mansfield 1974). Assuming the orientation is perceived with a mechanism similar to that in section 4.5, when fewer detectors are activated the response will vary more because the average is being computed from fewer neurons. An RF-LISSOM model trained on a non-uniform training distribution with a preponderance of horizontal and vertical lines develops more detectors for horizontal and vertical orientations. This type of adaptation has also been observed in kittens raised in deprived visual environments (Blakemore and van Sluyters 1975). Such an anisotropic distribution is probably typical of the early visual experience of humans (Mansfield 1974).

Frequency localization: The TAE is selective for spatial frequency (Ware and Mitchell 1974); adapting to a figure with narrow bars has no measurable effect upon a figure with wide bars, and vice versa. Spatial frequency selectivity has previously been demonstrated in the RF-LISSOM model (Miikkulainen et al. 1997; Sirosh et al. 1996). A unified orientation/spatial frequency RF-LISSOM model, obtained by training on oriented Gaussians of different sizes, would exhibit frequency localization for the same reason as for spatial localization. That is, only a small range of frequency detectors will be activated by a given stimulus, so only that range will show adaptation effects.

Movement direction specificity: The TAE is selective for the direction of movement (Carney 1982). Adapting on a pattern moving in one direction past a fixation point does not affect the orientation judgment of a pattern moving in the opposite direction. A unified orientation/movement direction RF-LISSOM model, obtained by training on oriented Gaussians moving in different directions, would exhibit this property as well. However, simulations have not yet been performed with moving stimuli for RF-LISSOM, and further work will be needed to determine how motion should be represented in the model.

Ocular transfer: The TAE transfers completely from one eye to the other (Campbell and Maffei 1971; Gibson and Radner 1937); adapting one eye causes equal effects

upon test lines in the same location in the visual field for either eye. A unified RF-LISSOM orientation/ocular dominance model could be obtained by training on oriented Gaussians at slightly offset positions in the two eyes. Such a model would exhibit ocular transfer if the neurons most selective for orientation were also binocular. If this does not turn out to be true for the model, it might be that the most plastic neurons in the cortex are also more likely to be binocular. The latter appears to be true for human cortex, with the input layer (layer IV) showing strong ocular dominance but lower adult plasticity, and the output layers (II, III, V, and VI) showing strong binocularity and high plasticity (Daw 1995,pp.139–140; Shatz and Stryker 1978). Modeling such layer-dependent effects would require a significant extension to RF-LISSOM because currently all the neurons in each column are grouped together for simulation.

Large-scale combined simulations allowing the study of some of these features were proposed by Sirosh (1995), but they generally require greater cortex sizes and training times to represent multiple dimensions on the same map. Current computer resources are not sufficient, but with forthcoming advances in technology they should soon be more practical.

Besides the above, there are a small number of features of the TAE not yet demonstrated by the current model, and not expected to be found unless extensions are made:

Saturation: The TAE saturates at approximately 4° (Campbell and Maffei 1971; Greenlee and Magnussen 1987; Magnussen and Johnsen 1986; Mitchell and Muir 1976) but in the RF-LISSOM model it will steadily increase up to at least 20° with sufficiently long adaptation times. Possible mechanisms for this limit to plasticity will be discussed further in section 6.2. It is a simple matter to artificially limit the plasticity allowed in the model, once it is known what the limiting factor should be.

Dark recovery: After the adaptation stimulus is removed, TAE magnitude gradually reduces in strength, even in the absence of visual input. In contrast, the RF-LISSOM model will ordinarily remain static until further input is received. Explanations for dark recovery in humans will be discussed further in section 6.2.3; adults may have connections which change strength rapidly but temporarily.

There are also some small effects due to the absolute orientation of the head with respect to gravity; these are presumably due to the vestibular system, not to processing in V1 (Howard and Templeton 1966; Wolfe and Held 1982). Apart from such data not expected to apply to V1, there is no known psychophysical evidence against the RF-LISSOM explanation of tilt aftereffects. The following section will examine the types of cellular processes that may underlie the psychological effects above, if the TAE is occurring in humans in the same way it does in the model.

6.2 Biological mechanisms underlying the TAE

Since RF-LISSOM models neurons at an abstract level, it allows cortical function to be studied without specifying a particular biochemical implementation. The same processes could be performed by a number of different mechanisms in different species, in different brain areas, at different stages of development, or at different time scales. Once the behavior is clear from computational simulations, candidates for its implementation can be proposed and tested. This section will relate the results from chapter 5 to experimental evidence, in an attempt to constrain the possible candidates for biological mechanisms underlying the TAE.

6.2.1 Is the TAE due to synaptic plasticity or to accumulation of inhibition?

Given a network of interconnected neurons, adaptation may conceivably occur either in the connections or nonspecifically across the neurons themselves. It is sometimes hard to make a distinction between these alternatives, but there are cases when they can clearly be differentiated. For instance, if only the connections are changed, being able to stimulate a neuron through a different unadapted pathway can demonstrate that the neuron itself is unchanged. If the entire neuron itself changes, then no matter what mechanism is used to stimulate it, it will show the effects of adaptation.

The TAE seen in RF-LISSOM results from changes in the connection weights, while the discredited theory of neural fatigue (chapter 2) postulated changes occurring within the neuron itself. However, echoes of the fatigue theory persist, even from researchers who have accepted the lateral inhibition theory of the TAE (e.g. Gelbtuch et al. 1986; Kurtenbach and Magnussen 1981; Masini et al. 1990; Tolhurst and Thompson 1975). These researchers describe the TAE as arising from "the prolonged effects of inhibition" on a neuron. They appear to interpret these effects as changes that occur within the neuron, perhaps a buildup of some intracellular inhibitory messenger, or a change in the properties of ion channels in the cell membrane. These changes would be activated by lateral inhibition, but would remain in force for a short time following it. If this were true, the TAE would be resulting from exactly the same biochemical processes that presumably mediate the tilt illusion (see section 6.4.1). The only difference should be that the aftereffect eventually decreases with time, while the illusion would persist.

However, several of these same researchers then go on to present evidence that a number of psychoactive substances modify one effect but not the other (Gelbtuch et al. 1986; Masini et al. 1990). This prompts them to propose explanations of this apparent paradox that tend to be rather complex and speculative. This thesis suggests a much simpler interpretation which is also supported by clear physiological evidence. When an

aftereffect occurs in the model, it is always a result of changes in the connection strength *between* neurons, not a nonspecific change in the neuron itself (cf. Barlow 1990). The model does not contain any representation of the state of the neuron, whose response changes only when its inputs or connection weights change.

Under these assumptions, there is no difficulty in explaining the differences between the effects of various drugs on the TAE and TI. Both effects would arise from lateral inhibitory interactions, accounting for the substantial similarities seen between them (see section 6.4.1). But there would be distinct differences in the biochemical mechanisms for the effects — the TI requires only that the lateral inhibitory connections be functional, while the TAE requires that their strength be modifiable. Many substances have been shown to affect plasticity without otherwise altering function (Daw 1995,ch.12). Similarly, since the inhibitory connections are polysynaptic, other substances could interfere with the effect of lateral inhibition on the target cell without disrupting adaptation occurring earlier in the pathway.

Furthermore, Vidyasagar (1990) has demonstrated that at least some types of orientation adaptation do not appear to be a result of changes within a single cell. Vidyasagar activated and inhibited single cells directly (using electrical and/or chemical stimulation), and was unable to find any adaptation effects in their responses. However, there was clear adaptation evident in the contrast-sensitivity threshold for the same cells when they were activated by a visual pattern that also activated other nearby cells. These results suggest that the effects occur somewhere along the connections between neurons (as in RF-LISSOM), rather than as a non-specific change within the individual neurons themselves.

6.2.2 Contribution of afferent and excitatory plasticity

In the RF-LISSOM model, modification of the lateral inhibitory connection weights was found to be sufficient for the model to exhibit realistic tilt aftereffects. However, the results do not provide a way to determine if afferent or lateral excitatory plasticity is also occurring, because similar results were found whether or not these connection types were modifiable. As long as the lateral inhibitory connections had plasticity at least comparable to that of the excitatory weight types, they dominated the adaptation because there are so many lateral inhibitory connections in the model.

If the afferent weights are modifiable in RF-LISSOM, the layout of the orientation map itself can undergo substantial reorganization. Areas representing the orientation used during adaptation increase in size, while those representing other orientations decrease. Such changes are known to occur in response to ordinary visual experience during development (Movshon and van Sluyters 1981). In the adult, however, they are thought to require cortical lesions or very long-term changes in visual input (Gilbert et al. 1996; Kapadia et al. 1994; Sugita 1996; see also Miikkulainen et al. 1997; Sirosh et al. 1996). On the other

hand, substantial plasticity has been demonstrated for neurons in the adult visual cortex *in vitro* (Artola and Singer 1987; Hirsch and Gilbert 1993; Kirkwood and Bear 1994). It has been proposed that such plasticity is controlled by a *plasticity gate* ordinarily held closed by the inhibitory connections from other active neurons (Kirkwood and Bear 1994; von der Malsburg 1987). It is not yet firmly established what circumstances cause such a gate to open.

If afferent plasticity is possible over periods less than one hour, it may be responsible for the saturation of the tilt aftereffect found by Greenlee and Magnussen (1987). The results computed in chapter 5 for the RF-LISSOM model assume that the neurons in the model encode a fixed orientation. After a marathon adaptation session with plastic afferent connections, this assumption would become invalid: the neurons in the map would develop substantially different orientation preferences. It is possible that higher levels in the cortex may begin to rearrange in the same way, resulting in a perceived orientation based upon the new map instead. Such higher-level adaptation would counteract the effects of adaptation in V1, because it would change the interpretation of activity on the map. The result could be that the perceived TAE levels off at a maximum, i.e. that it saturates.

A psychophysical test might be conducted to determine if hierarchical reorganization is occurring in this way. Adaptation to a grating of a particular orientation, spatial frequency, and contrast increases the ability to detect that similar stimuli actually differ very slightly from it along any of those dimensions (Albrecht et al. 1984; Greenlee and Heitger 1987). The decrease in the incremental threshold for detection of difference is ordinarily considered to be just another way of expressing what occurs in the TAE, and is presumed to result from the same mechanism. If it does, then it may provide a way to detect the phenomena causing the TAE even when the changes in perceived orientation have reached saturation. If reorganization is occurring, the lower level would continue becoming more sensitive to neighboring orientations, but the higher level would begin to cancel out the changes in the orientation map in order to restore veridical perception. It is not clear how the higher-level system would gauge the veridicality of the orientations in the map; it might compare responses in different parts of the map (likely receiving different levels of adaptation) and make appropriate corrections at local areas to ensure the map is consistent. If a process of this nature is occurring, it could be detected by testing if improved incremental sensitivity continues increasing past where the perceived orientation changes have saturated. If it does, then the saturation effect may be due to this reorganization at higher levels.

Alternatively, saturation may result from limits on the plasticity of the inhibitory connections, as proposed in the next section. Whether or not afferent and lateral excitatory connections are plastic in ordinary circumstances in the adult, the results from this thesis suggest that at least the inhibitory connections must retain some degree of plasticity. The

next section will examine what type of inhibitory plasticity would be required to account for the evidence.

6.2.3 Biophysical mechanisms of inhibitory plasticity in development and in adulthood

The mechanisms underlying adaptation of the lateral connections are not yet known in detail, although it is clear that long-term inhibitory adaptation does occur during development. The lateral connections are initially quite widespread and as modeled in RF-LISSOM only become selective between columns as a result of visual experience (Callaway and Katz 1990; Luhmann et al. 1986). This thesis shows that the same principles that can account for such development in infants could also be causing perceptual artifacts in the adult. However, since RF-LISSOM models neural processes at an abstract level, this result with the TAE does not depend upon an assumption that the biophysical processes are identical in each case. Although recent work suggests that there are a number of synaptic plasticity mechanisms which operate the same way in adults and infants (Artola and Singer 1987; Kandel and O'Dell 1992; Kirkwood and Bear 1994; Kirkwood et al. 1993), in general cortical plasticity appears much more limited in the adult (Daw 1995).

The steady decay of the tilt aftereffect in complete darkness may be a manifestation of differences between adult and developmental plasticity. For humans, it has been found that the TAE decays in darkness with approximately the same curve as when it increases in figure 5.10 (Greenlee and Magnussen 1987; Magnussen and Johnsen 1986). Other experiments, however, have found that small residual tilt aftereffects can be detected as long as two weeks after a four-minute adaptation (Wolfe and O'Connell 1986). In any case, the decay does not appear to rely on visual input, since it has the same time course whether the subject is in complete darkness, or if test patterns are presented at intervals during decay (Magnussen and Johnsen 1986).

In the RF-LISSOM model, decay will occur in the same way as adaptation occurs — subsequent inputs will cause the organization to return to equilibrium as long as they include orientations different from the fixation stimulus (cf. Wolfe 1984). However, if no inputs are presented (or, equivalently, blank inputs are presented), no weight changes will occur, and the tilt aftereffects will remain indefinitely. Random spontaneous retinal activity present in darkness is not expected to be sufficient to cause a return to equilibrium, since it is unlikely to contain oriented components at the spatial scale of the adaptation stimulus.

One quite speculative explanation of decay consistent with the present RF-LISSOM model is that it results from the same process of higher-level reorganization proposed in section 6.2.2. Reorganization that occurs while the adaptation stimulus is being presented would result in saturation, while reorganization that occurs after the map in V1 has stopped being modified would result in a steady decay in the perceived orientation difference. Thus

neither saturation nor decay would have been found in the experiments in chapter 5, as observed. From this perspective, there is no reason to suppose that the mechanisms of plasticity differ in the adult and the infant. However, this is unlikely to be the full explanation, since even single cells in V1 demonstrate decay of adaptation effects (Albrecht et al. 1984); though of course the single-cellular decay could be caused by feedback from higher levels as they reorganize.

Another less extravagant explanation may be that changes in synaptic effectiveness are merely a side-effect of limits on synaptic transmission at particular excitatory connections. This would act as a form of connection-specific fatigue (Geisler 1997). During presentation of an adaptation pattern, the ability of the connection to carry action potentials would decrease, resulting in a direct TAE. As resources are replenished, the effects of adaptation would decrease, thus explaining why decay occurs. This mechanism does not readily account for the saturation effect, since it is not obvious why there should be a limit to the amount of depletion possible.

Finally, a more compelling explanation for decay and saturation might be that adult and infant plasticity are two different (but possibly related) cellular mechanisms. This explanation would not depend upon reorganization of the orientation map in V1 or at higher levels. In this view, the mechanism underlying the TAE in the adult may be a separate fast, limited, and temporary version of the self-organizing process that captures longer-term correlations. The connection weights in this process would act as a small additive or multiplicative term on top of a larger long-term weight. For example, each inhibitory weight w could be represented as $w_o + \Delta_w$. The w_o portion would be comparatively static, keeping its value indefinitely in darkness and changing only with a long time constant, or perhaps not at all in the absence of cortical or retinal trauma. The Δ_w term, on the other hand, would adapt and decay very quickly, perhaps representing the short-term correlations between image elements.

Having this set of temporary, highly-plastic, strength-limited weights might be a quite deliberate feature of the cortex. Such a mechanism has been proposed by von der Malsburg (1987) as an explanation of visual object segmentation. He proposed that temporary plasticity allows the cortex to group elements of a visual scene into coherent objects, each composed of neurons firing in synchrony. The synchronization would be achieved by rapidly modulating lateral connection strength to temporarily strengthen connections between active units and weaken other connections. The amount of strengthening possible over this time scale would be quite limited, which would result in eventual saturation of the effects. Similarly, changes would be expected to decay after the visual input is removed, so that subsequent inputs are not affected if they are sufficiently far removed in time. But over short time scales, the interactions between subsequent inputs (e.g., the tilt aftereffect) could actually be beneficial: they would facilitate the segmentation of other similar collec-

tions of features. Analogous synchronization and segmentation effects have been modeled in RF-LISSOM already (Choe and Miikkulainen 1996; Miikkulainen et al. 1997).

Permanent changes in synaptic strength would require repeated presentation of inputs grouping on the short time scale, and the opening of some form of plasticity gate (Kirkwood and Bear 1994; von der Malsburg 1987). The long-term changes would thus represent long-term persistence of short-term correlations. The short-term and long-term processes might share some of the same biochemical pathways, or they might be entirely different processes that both implement the same function over different time scales. The TAE itself may be merely a minor consequence of this complex architecture for representing a wide range of correlations. The details of these mechanisms are abstracted in the current model, where only one type of reorganization mechanism is present. Future versions of RF-LISSOM may be extended to explore these issues in more detail.

6.2.4 Biophysical mechanisms of the indirect effect

The lateral inhibition theory for direct tilt aftereffects is widely accepted in one form or another. However, as described in section 2.3.4, no consensus has emerged about the cause of the indirect TAE. Over the years, it has been proposed to result from a "linkage" between the vertical and horizontal axes (Gibson and Radner 1937), adaptation of cells with multiple preferred orientations (Coltheart 1971), a three-lobed lateral interaction profile (O'Toole and Wenderoth 1977), and direct effects arising from an invisible "virtual axis" of symmetry (Wenderoth et al. 1989).

The linkage explanation was disproved by the finding that indirect and direct effects are similar at all absolute orientations (Mitchell and Muir 1976). The O'Toole and Wenderoth (1977) theory (i.e., inhibition at intermediate distances, but excitation at near and far distances) was abandoned by at least some of its original proponents when they found that the indirect effect arises significantly later than the direct effect (Wenderoth and Johnstone 1988; Wenderoth et al. 1989). If both effects result from very similar processes of lateral interactions, differing only in sign, one would have expected them to have a fairly similar time course as well.

Spivey-Knowlton (1993) has proposed that Coltheart's explanation in terms of cross-neurons may represent a possible neural substrate for the virtual axis theory, as described in section 2.3.3. Spivey-Knowlton's formulation of the virtual axis theory helps make it less abstract and more testable, but it does not provide any functional justification for indirect effects. Nevertheless, the virtual axis theory appears to be the working hypothesis for the indirect effect at present, if only because of the lack of competing theories compatible with the lateral inhibition theory of direct effects.

This thesis demonstrates, however, that a possibly quite simple, local mechanism in V1 is sufficient to produce the indirect effect. If the total synaptic resources at each neuron

are limited, for whatever reason, the strengthening of the lateral inhibitory connections between active neurons must also decrease the effectiveness of inactive connections to those neurons. Such a limit appears biologically plausible on the face of it, since a neuron has only a fixed surface area to which connections can made unless the neuron expands in size, and the neuron cannot become indefinitely large because the volume of the brain is fixed. Miller and MacKay (1994) note that there is widespread evidence of competition for a limited number of synaptic sites (Bourgeois et al. 1989; Hayes and Meyer 1988a,b; Murray et al. 1982; Pallas and Finlay 1991; Purves 1988; Purves and Lichtman 1985).

There is also extensive computational justification for synaptic resource conservation. One of the first computational models of Hebbian adaptation (Rochester et al. 1956) discovered that without such normalization, connection weights governed by a Hebbian rule will increase to infinity. Each time an input is presented, some connections will be strengthened, and the others will remain as they were. If each weight has a maximum strength, then each will eventually increase to its maximum value. Thus all connections will end up identical and will be unable to perform any useful function (Miller and MacKay 1994).

To prevent such unwanted behavior in a model, von der Malsburg (1973) proposed keeping the total synaptic strength constant. Thus when a connection strengthens, the other connections must be weakened. Such normalization could be accomplished by subtracting the weight change equally from each of the other connections. Miller and MacKay (1994) showed that doing so would result in a set of strictly binary connections, each either at full strength or zero strength. Such subtractive normalization seems implausible given that a variety of lateral connection strengths are found in the cortex (Hirsch and Gilbert 1991, 1993).

The multiplicative normalization used in RF-LISSOM (equation 3.4) results in each of the other connections being scaled down in proportion to its current strength. Multiplicative normalization preserves a variety of connection strengths. Other forms of normalization are possible, however, and the demonstration of indirect tilt effects with multiplicative normalization does not in itself rule out other possibilities for keeping the sum of connection strengths constant.

One must bear in mind that short-term strengthening might temporarily violate the above longer-term constraints. For instance, the total synaptic resources might not all be in use initially, so there could conceivably be a small delay before synaptic resource conservation would cause indirect effects. This delay could account for the findings that the indirect effect has a later onset than the direct effect (Wenderoth and Johnstone 1988; Wenderoth et al. 1989). In contrast, the direct and indirect effects in RF-LISSOM always occur simultaneously, since normalization is enforced as soon as any weights change. (Note the small error bars for the model over the range 45° to 90° in figure 5.2, indicating that most runs showed similar indirect effects.)

78

Variations in the utilization of synaptic resources in different individuals and under different circumstances may help explain why the indirect TAE varies greatly in magnitude between different studies, between different individuals in the same study, and for the same individual over different trials. (Note the large error bars for the human subject over the range 45° to 90° in figure 5.2.) Since the synaptic modifications are apparently temporary in adults (i.e., since the TAE decays in darkness), at any point in time some synaptic resources may be unused. The equilibrium state of some individuals may represent nearly full utilization of resources, and they would show relatively large indirect effects in general. For any individual, adaptation resulting from recent visual experience might change the amount of undedicated resources, causing wide variations in the indirect TAE depending upon the circumstances. But regardless of the state of resource utilization, the direct effect should show a consistent magnitude due to simple Hebbian adaptation, as is seen in the model and in humans.

The RF-LISSOM account of indirect effects is also consistent with many other possible explanations of the late onset and variation of the indirect effect. If the TAE is occurring in the cortex as in RF-LISSOM, the process of synaptic resource conservation could be mediated by entirely different cellular mechanisms from that of the direct effect. These mechanisms could easily have different time courses, be influenced differently by various drugs, etc. The dissociation between the mechanisms is strongly supported by the experiments of Wenderoth and Johnstone (1988).

In the absence of evidence to the contrary, one would assume that these mechanisms are operating within each neuron in V1 or in the neuron's lateral connections. No higher-level process is needed to explain the differences found in the model. This contrasts strongly with the explanation from Wenderoth and Johnstone (1988), who argued that their evidence indicated that the indirect effect arises beyond V1. In summary, the indirect TAE explanation supported by RF-LISSOM is biologically plausible, computationally justified, and requires only local mechanisms.

6.3   Specific predictions for experimental verification

If tilt aftereffects occur in the human visual system based on the same mechanisms as in the LISSOM orientation map model, then there are certain features we would expect to see in the human data. Many of these have already been discussed here and in chapter 5. However, several other consequences of the model have not yet been investigated specifically in humans, and RF-LISSOM has provided much more detailed information than that so far available for humans or animals.

The RF-LISSOM model predicts that for the indirect effect, the number of V1 neurons responding to a test pattern should increase as adaptation progresses. The net response

level should also increase. Other models such as satiation, fatigue, and the lateral axis theory (described in chapter 2 and section 6.2.4) would predict a decrease or no change in activity levels in V1 for that stimulus. In particular, the lateral axis theory would predict that changes would occur only at higher levels than V1. The RF-LISSOM prediction could be verified or refuted in humans by visualizing activity levels using optical imaging, visual-evoked-potential (VEP), or similar techniques. One would simply compare the response to a stimulus before and after adaptation. If cortical activity does indeed decrease for such a pattern, as it does in RF-LISSOM, it would be difficult to explain how this could arise from fatigued neurons or neurons with a buildup of some inhibitory substance.

If sufficient temporal and spatial resolution is available from the imaging process, plots like those presented for RF-LISSOM in section 5.4 could be computed in monkeys for comparison with the model. First, an orientation map would be computed using standard means (Blasdel and Salama 1986; Ts'o et al. 1990; Blasdel 1992a; Grinvald et al. 1994; Weliky et al. 1995). Next, the response of the cortex would be measured for test patterns at orientations typical of direct and indirect effects, (e.g. $10°$ and $60°$ from the adaptation angle). The imaging technique would need very high temporal resolution to make this measurement without prompting significant adaptation to the test pattern. Test pattern presentations vary in different psychophysical experiments, but they are typically on the order of a few seconds, so an image with the required spatial resolution (see below) would need to be obtainable within that time. Methods like PET (positron emission tomography) which require integration over long time periods would be unsuitable (Sejnowski and Churchland 1989).

Next, the cortex would be adapted to a stimulus of a particular orientation. Finally, the test patterns would be presented again, measuring the cortical response once more. The before-and-after activity plots could then be matched with the orientation map, showing exactly which areas change activity as a result of adaptation. In order to determine how the activity changes modify the overall orientation, the activity plots would need spatial resolution at least sufficient to resolve individual orientation columns, i.e. down to at least 0.1mm. When such a *before* plot is subtracted from an *after* plot, there should be a net decrease in activity for orientation detectors near the orientation used during adaptation, a net increase in activity of those with more distant orientations, and no change for very distant orientations.

If this type of spatial resolution is available, one would also be able to calculate perceived orientations as in section 4.5. The difference between the perceived orientations before and after should be within the range of TAE seen for human subjects. If this is found to be the case, it would represent strong support for the lateral inhibition theory of direct and indirect tilt aftereffects.

A further test of the model could be made by blocking intracortical inhibition me-

diated by $GABA_B$ receptors with bicuculline (Sillito 1979) and that mediated by $GABA_A$ receptors with phaclofen (Pfleger and Bonds 1995). Without inhibition, the sign of the TAE should be reversed and it should have a smaller magnitude (perhaps not even being measurable, if afferent plasticity is insignificant). Adaptation under those conditions should result in the afferent-only effects shown in figure 5.3. Note that blocking the inhibitory receptors does not necessarily prevent inhibitory adaptation (Vidyasagar 1990), since changes could still occur in the excitatory connections to inhibitory interneurons. However, the functional consequences of inhibitory adaptation would no longer be apparent. It is not possible to perform such tests in humans, since the chemicals disrupt brain function. It may be possible to devise suitable protocols for testing the TAE in animals, at which point such tests could be performed. If similar effects are found in these animals as in the model, it would also represent strong support for the lateral inhibition theory of tilt aftereffects.

6.4 Future work

Besides the specific topics for future study discussed so far (including ocular transfer, saturation, and dark decay), this research has opened up several areas that can now be studied computationally. These areas include running simulations that more closely parallel psychophysical experiments, testing low-level phenomena closely related to the tilt aftereffect, and examining illusions and aftereffects in other modalities. This section will sketch possible avenues for investigation of these topics.

6.4.1 Tilt illusions

In addition to the tilt aftereffects studied in this thesis, the RF-LISSOM model should exhibit direct tilt illusions between simultaneous spatially-separated stimuli (Calvert and Harris 1988; Carpenter and Blakemore 1973; Gilbert and Wiesel 1990; O'Toole 1979; Smith and Over 1977; Wenderoth and Johnstone 1988; Westheimer 1990). The stimuli would interact with each other as the lateral interactions settle, inhibiting the feature detectors tuned to orientations between those of the two patterns. This would drive the perceived orientation of each pattern away from that of the other.

However, this effect cannot yet be tested with the orientation map used for the other experiments in this thesis. The training inputs to the map were single oriented Gaussians, so correlations occurred only between local areas along the orientation of the Gaussian. Thus lateral connections develop along a single orientation: the orientation preference of that neuron (section 4.3). So tilt illusions would only occur for overlapping stimuli centered around the same location on the retina, because only those stimuli would have significant activation of feature-detectors linked by lateral connections.

Because the retinal representations of such lines overlap, the cortical responses will overlap as well, so it would not be possible to quantify the effect using the techniques in section 4.5. That is, the vector averaging procedure only allows a single perceived orientation to be calculated per local area of the retina. In contrast, the cortex appears to be able to perceive multiple overlapping orientations separately (as observed in humans, Blakemore et al. 1970; Carpenter and Blakemore 1973).

The segmentation mechanisms of the spiking neuron version of RF-LISSOM (Choe and Miikkulainen 1996; Miikkulainen et al. 1997) may provide a way to separate the representation of each line (Blakemore et al. 1970; Carpenter and Blakemore 1973). In this model, the set of activated units can divide itself into overlapping populations of neurons, each population firing out of phase with the others (von der Malsburg 1973). For two overlapping lines, the active units will form into two groups consisting of neurons firing in synchrony (as mentioned in section 6.2.3. The perceived orientation calculation from section 4.5 could then be performed on each group separately to yield a perceived orientation of each line. These perceived orientations could be compared to the orientation perceived when only one line is present, to determine how large a tilt illusion is present.

A simpler test procedure would be to self-organize an orientation map using inputs such as sinusoidal gratings that have longer-range correlations between similar orientations. Such patterns would represent objects with parallel lines or edges, which are very common in the visual environment. Long-range connections would then develop between widely separated orientation detectors in parallel directions, in addition to the relatively local connections now present. A version of the RF-LISSOM model trained on such patterns should demonstrate direct tilt illusions through lateral inhibition between two separated stimuli, as described above. Although such experiments require at least twice as large a cortex and retina as that used for this thesis, they should become practical in the near future.

Although indirect effects have been found for the TI much like those for the TAE, it is not yet known whether they will be found in the RF-LISSOM model. The simultaneous indirect effect may depend upon facilitation of weakly activated units by units at distant orientations. This facilitation would be mediated by lateral connections whose effective sign would depend upon local contrast, the extension to RF-LISSOM proposed in section 3.6. The indirect TI may also depend upon factors not yet considered, so it should be an interesting topic for research.

6.4.2 Effect of test grating

The RF-LISSOM model can be used to explore limitations of the psychophysical experiments, and thus it may help explain some results as artifacts of those experiments. As described in section 5.1, the protocol used in this thesis measured the effect of a fixed

adaptation stimulus on test lines of each different orientation. Psychophysical experiments generally measure the effect on a fixed test stimulus of adapting to test lines at each orientation. These two protocols will give identical results if two assumptions are met: (1) the angular function of the effect does not depend upon the absolute orientation on the retina, and (2) the change in perceived orientation can be measured without affecting later measurements. Assumption 1 appears to be valid for humans, who show similar effects at all angles (Mitchell and Muir 1976). It must be true in the RF-LISSOM as well, because the model has no preferred direction. That is, the choice of which orientation to call "vertical" in the model is arbitrary.

Assumption 2 is clearly true for the model, since learning can be turned off entirely so that presentation of test patterns does not have any lasting effect. However, it is clearly not true in general for humans, because adaptation cannot be measured psychophysically without presenting a test pattern. The cortex will adapt to all patterns presented, not just those the experimenter has labeled "adaptation" stimuli. Only in a computational model such as RF-LISSOM can the TAE be evaluated without any side effects, although various presentation techniques have been devised to limit the residual effects of the test pattern in psychophysical experiments.

The contribution of those side effects could be evaluated in RF-LISSOM by duplicating the psychophysical experiments in greater detail. The RF-LISSOM learning rate would remain at the same level throughout the experiment, thus causing adaptation to every stimulus presented. A particular experiment would be chosen for replication, e.g. Mitchell and Muir (1976). The reports of this experiment would need to indicate how long the test stimulus was presented, relative to how long the training stimulus was presented, so that these parameters could be duplicated in the model. The experiment would then be recreated as follows.

First, a fixed testing stimulus would be chosen and presented for the period of time used in the psychophysical experiment, in order to compute a baseline for the perceived orientation of the test line. Next, the model would adapt for a fixed amount of time to another line at a particular orientation. The testing stimulus would then be presented again, with learning still on, and the magnitude of the TAE would be measured. Finally, the cortex would be returned to equilibrium in order to account for the decay of the TAE, either by resetting all weights to their initial values, or by presenting randomly oriented stimuli long enough to erase the effects of adaptation. This procedure would be repeated for each point on the angular function of the TAE from figure 5.2.

The curve from this procedure could then be compared to one where learning is turned off for the testing pattern, in order to show the effect of the presentation of the test pattern. In humans, large aftereffects (greater than 5°) have been found for very short test presentations (Wolfe 1984). With the tests above it might be possible to decide whether

different mechanisms are operating over these short time scales (Harris and Calvert 1989; Wolfe 1984), or whether the effect of the test pattern is sufficient to explain the larger aftereffect found in those circumstances. Computing a single point on the TAE curve with the above procedure takes more computing time than it took to create the entire set of curves in figure 5.9, but limited tests of this nature should be computationally feasible.

6.4.3 Contrast effects

When both the adapting and test gratings have the same contrast, and test stimuli are presented for a relatively long period, the magnitude of the TAE does not depend upon the contrast (Parker 1972). However, in any other circumstance, the absolute and relative contrasts systematically affect the magnitude of the TAE (Harris and Calvert 1989; Parker 1972; Ross et al. 1993; Ross and Speed 1996). For instance, with long presentation times, a larger TAE will be seen for a low-contrast test grating than for one which has high contrast. Similarly, a low-contrast adaptation pattern will cause a smaller TAE for a high-contrast test-grating (Parker 1972). Such effects may be a straightforward consequence of the activity-dependent adaptation modeled in RF-LISSOM. Larger adaptation will occur for higher-contrast inputs, but adaptation will also occur if the test inputs have high contrast, so the effects may cancel out. These effects are expected to be seen in the current RF-LISSOM model, if learning is left on for testing using the protocol proposed in section 6.4.2.

However, some of the effects noted by Harris and Calvert (1989) appear to suggest influences of the contrast-dependent lateral connections proposed in section 3.6. For some combinations of inputs, the TAE appears much weaker at low adapting contrasts than at high contrasts. This could be a result of having long-range lateral connections that are excitatory at very low contrasts. For low-contrast inputs, the inhibitory lateral interactions would be negligible, which should result in a much smaller TAE as observed. Experiments with a version of RF-LISSOM extended with such connections may help determine which mechanisms can account for these effects.

6.4.4 Masking phenomena

Most studies of the tilt aftereffect are performed using stimuli at a fixed contrast, and require the subject to judge the orientation of the stimuli by some means. An alternative way to study the effects of adaptation is to measure the detection threshold for each orientation before and after adaptation to a particular stimulus. The results from such procedures are generally similar to the results for the TAE measurements: adapting to one orientation *masks* stimuli at nearby orientations, effectively raising the detection threshold for those orientations and lowering them for somewhat more distant orientations (Blakemore and

Nachmias 1971; Over et al. 1972; Ross et al. 1993; Smith and Over 1977; Vidyasagar 1990; Virsu and Taskinen 1974; Waugh et al. 1993).

However, there are numerous differences between masking and the TAE. For instance, there doesn't appear to be any facilitation observed for very distant orientations, yet the TAE shows an indirect effect for those orientations (Over et al. 1972). In addition, masking persists in the presence of inhibitory receptor blockade (Vidyasagar 1990) that would be expected to prevent the TAE (see section 6.3). The differences may provide clues about how the contrast-dependent lateral interactions are implemented. In the masking paradigm, the stimulus used during adaptation is high-contrast, while that used to test the detection threshold is at a very low contrast, by definition. Thus this test condition should show how the adaptation of the lateral inhibitory connections affects the long-range lateral excitatory connections. The evidence suggests that the excitatory connections decrease in strength with adaptation, but further study would be needed to explain how or why that occurs.

6.4.5 Aftereffects in visual hierarchies

If higher levels such as V2 have an organization similar to that of V1 as modeled by RF-LISSOM, they should also show tilt aftereffects in much the same way. This may account for tilt aftereffects that have been demonstrated for purely subjective contours. Such contours contain orientations perceived by an observer but not physically present in the image (Berkley et al. 1993; Paradiso et al. 1989; van der Zwan and Wenderoth 1994, 1995). Subjective contour perception is generally thought to arise at levels higher than V1 (van der Zwan and Wenderoth 1995), and is thus not addressed by the current RF-LISSOM model. If each level has the same structure, the patterns produced by a given level could interact in that level and later levels (Berkley et al. 1993), conceivably causing tilt illusions and tilt aftereffects between features not present in the original image. Hierarchical levels of RF-LISSOM models have been proposed by Sirosh (1995) as a possible implementation of the visual processing hierarchy described by Van Essen et al. (1992). Future simulations with such models may allow studies of the tilt aftereffect between illusory and real contours.

6.4.6 Hyperacuity

Besides experiments with tilt adaptation, certain other psychophysical tests also appear to give information about learning processes within the primary visual cortex. For instance, Fahle et al. (1995) and Weiss et al. (1993) found that performance in hyperacuity tasks, such as deciding whether two lines of same orientation are separated by a small perpendicular offset, improves with practice. The improvement occurs even without verbal or other feedback indicating whether each judgment is correct. The effect is specific to position and

orientation, but transfers between eyes to some degree. This is thought to indicate that at least some part of the effect arises in V1, since V1 is the first stage in the visual pathway where binocular inputs are combined.

Shiu and Pashler (1992) reported similar results for orientation discrimination tasks, although they found that the effect also depends on cognitive factors. The ability to distinguish the orientations of two similarly oriented lines only improved with practice if the subjects were directed to pay attention to the orientation. However, the effect was specific to the location on the retina on which examples had been presented, so it did not consist of some more effective deliberate strategy that the subject learns during the experiment. This suggests that attentional mechanisms may activate circuitry in V1 (or other early visual areas) that regulates plasticity.

The RF-LISSOM model should be able to account for such psychophysical learning phenomena. The active feature detectors and lateral connections between them would adapt during repeated presentations. Over time, this would expand the area of the cortical feature map responding to those features. This would result in representation and discrimination of smaller differences. Since the orientation discrimination testing paradigm is simple yet shows clear attentional effects, it might also form a good testbed for extensions of RF-LISSOM that include feedback from higher cortical levels. Such studies might help clarify how and when adaptation occurs in the early visual system.

6.4.7 Other visual aftereffects

Aftereffects appear to be a nearly universal feature of cortical sensory processing (Barlow 1990). Many visual aftereffects similar to the TAE have been documented in humans, including aftereffects of curvature, motion, spatial frequency, size, position, and color (Barlow 1990; Howard and Templeton 1966; Wolfe 1984). In all of these, the cortex appears to adapt to a long-lasting stimulus, changing the perceived value of subsequent stimuli. For instance, after prolonged viewing of a moving stimulus, stationary stimuli appear to be moving in the opposite direction (the movement aftereffect, also known as the waterfall illusion.)

Since topographically-organized detectors for most of these features have been found, RF-LISSOM is expected to be able to account for their aftereffects by the same process of decorrelation mediated by self-organizing lateral connections. The current RF-LISSOM model is clearly suitable for investigating some of these aftereffects, such as those of spatial frequency, size, and position. Cortical maps for these dimensions have already been demonstrated in RF-LISSOM, so these maps can be tested for aftereffects using techniques similar to those used in this thesis. Development of maps for the other visual dimensions has not yet been modeled in RF-LISSOM, so testing their aftereffects will have to wait until the maps have been studied.

Analogous aftereffects have also been found for other modalities such as hearing,

touch, muscle positioning, and posture (Howard and Templeton 1966,p.84–85,p.162–163). For instance, hearing a sound in one location can influence the perceived location of later sounds. After adaptation, sounds presented in nearby locations appear to be farther away than they actually are, and the effect appears to peak at a certain distance, much like the direct TAE. Aftereffects might also be present for taste and smell, though these parameters are relatively difficult to control in an experimental setting, and they have not been tested as extensively as vision, hearing, or touch.

If development in these areas can be modeled with RF-LISSOM, as is expected, aftereffects will be present in the behavior of the organized maps. Finding that the after-effects found in the model match those observed for each modality would indicate that the same decorrelating processes studied for vision in RF-LISSOM also apply to other types of perception, both during development and in the adult. It would thus strongly suggest that similar computations are being performed in areas of the cortex performing very different tasks. Combined with studies of the tilt aftereffect at different levels in the visual hierarchies, it would represent evidence that processes like those in RF-LISSOM are a ubiquitous feature of the cortex. Thus these simple decorrelating principles may account for a large part of the apparent complexity of the cortex. Achieving a unified explanation of such disparate phenomena would represent a significant advance in our understanding of the functioning of the brain.

6.5   Conclusion

In addition to providing a simple computational model of the development of many structures in the visual cortex, RF-LISSOM exhibits tilt aftereffects that are very similar to those measured in humans using psychophysical methods. The model is detailed enough to support specific, testable predictions about cortical function. The model is expected to account for a variety of other phenomena in its current formulation. Future computational studies may show how the model should be extended to account for other phenomena.

Chapter 7

Conclusion

7.1 Summary of the thesis

This thesis presents the first detailed explanation for the entire angular function of the tilt aftereffect.

Chapter 1 explained the basic effect, including a simple demonstration. After adapting to an oriented pattern, subsequent patterns of slightly different orientation appear to have a much larger orientation difference (the direct effect). Subsequent patterns nearly perpendicular to the adaptation pattern instead appear to be tilted towards the adaptation orientation (the indirect effect).

Chapter 2 surveyed previous research on the TAE, including other computational models of the TAE. It was argued that the lateral inhibitory theory is the only viable explanation proposed so far for the direct tilt aftereffect. It was also argued that previous models have not convincingly explained the presence of the indirect tilt aftereffect.

Chapter 3 explained the RF-LISSOM system of Sirosh and Miikkulainen (1994a, 1997) in detail, including the network architecture, activity calculation, and connection weight learning mechanisms. It also explored the biological plausibility of the model, suggesting that the current model is valid for the typical case of high-contrast inputs only. It also suggested possibilities for the role of a self-organizing model in the explanation of environmental and genetic factors observed in development.

Chapter 4 described how a realistic cortical orientation map was trained using parameters set by Sirosh (1995). It also examined possible algorithms for estimating the orientation perceived using such a map. It was shown that taking the vector average of the orientation preferences of activated neurons represents a sufficiently accurate estimate of the actual orientation.

Chapter 5 described the aftereffect experiments and results using the self-organized orientation map, and demonstrated that the results from the model are nearly identical to

the psychophysical data for the tilt aftereffect in humans. It presented a quite detailed analysis of the mechanism of the TAE in RF-LISSOM, and this analysis represents predictions of phenomena that may be seen in human or monkey cortex.

Chapter 6 related psychophysical and biophysical evidence with the results found for the TAE in the RF-LISSOM model. It explored possible cortical implementations of the self-organizing processes in the model, and presented specific experimental predictions that could be used to verify or refute these proposed mechanisms. Many directions for future work were suggested, including an examination of tilt illusions and studies of aftereffects in other modalities.

7.2   Conclusion

The experiments reported in this thesis lend strong computational support to the theory that tilt aftereffects result from Hebbian learning of the strengths of lateral connections between neurons. Furthermore, the aftereffects occur as a result of the same decorrelating process that is responsible for the initial development of the orientation map. This process tends to deemphasize constant features of the input, resulting in short-term perceptual anomalies such as aftereffects. The same model should also apply to other aftereffects and to simultaneous tilt illusions.

Because RF-LISSOM is a computational model, it can demonstrate many phenomena in high detail that are difficult to measure experimentally, thus presenting a view of the cortex that is otherwise not available. For instance, this thesis showed direct visualizations of aftereffects as they were occurring in the simulated cortex, making it clear exactly which processes contributed to the effect. This type of analysis can provide an essential complement to experimental work with humans and animals.

RF-LISSOM is the first model to provide a comprehensive and fundamental account of how both cortical structure and function emerge by Hebbian self-organization in the primary visual cortex. It is also the first to show how both indirect and direct tilt aftereffects could arise from simple, biologically plausible mechanisms in the primary visual cortex. Thus a single simple model may explain an unprecedented number and scope of cortical phenomena, which contributes substantially to our understanding of the cortex.

Bibliography

Albrecht, D. G., Farrar, S. B., and Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology*, 347:713–739.

Amari, S.-I. (1980). Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, 42:339–364.

Artola, A., and Singer, W. (1987). Long-term potentiation and NMDA receptors in rat visual cortex. *Nature*, 330:649–652.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Barlow, H. B. (1985). The twelfth Bartlett memorial lecture: The role of single neurons in the psychology of perception. *Quarterly Journal of Experimental Psychology*, 37A:121–145.

Barlow, H. B. (1990). A theory about the functional role and synaptic mechanism of visual after-effects. In Blakemore, C., editor, *Vision: Coding and Efficiency*, 363–375. New York: Cambridge University Press.

Bauer, R., Leferink, J., Eckhorn, R., and Jordan, W. (1991). Complementary global maps for orientation coding in upper and lower layers of the cat striate cortex and their possible functions. *Journal of Comparative Neurology*, 305:282–288.

Berkley, M. A., Debruyn, B., and Orban, G. (1993). Illusory, motion, and luminance-defined contours interact in the human visual system. *Vision Research*, 34(2):209–216.

Blakemore, C., and Carpenter, R. H. S. (1971). Lateral thinking about lateral inhibition. *Nature*, 234:418–419.

Blakemore, C., Carpenter, R. H. S., and Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228:37–39.

Blakemore, C., and Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, 228:477–478.

Blakemore, C., and Nachmias, J. (1971). The orientation specificity of two visual after-effects. *Journal of Physiology (London)*, 213:157–174.

Blakemore, C., and van Sluyters, R. C. (1975). Innate and environmental factors in the development of the kitten's visual cortex. *Journal of Physiology (London)*, 248:663–716.

Blasdel, G. G. (1992a). Differential imaging of ocular dominance columns and orientation selectivity in monkey striate cortex. *Journal of Neuroscience*, 12:3115–3138.

Blasdel, G. G. (1992b). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12:3139–3161.

Blasdel, G. G., and Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, 321:579–585.

Bourgeois, J. P., Jastreboff, P. J., and Rakic, P. (1989). Synaptogenesis in visual cortex of normal and preterm monkeys: Evidence for intrinsic regulation of synaptic overproduction. *Proceedings of the National Academy of Sciences, USA*, 86:4297–4301.

Bunt, S. M., Horder, T. J., and Martin, K. A. C. (1979). The nature of the nerve fibre guidance mechanism responsible for the formation of an orderly central visual projection. In Freeman, R. D., editor, *Developmental Neurobiology of Vision*, 331–343. New York: Plenum.

Burkhalter, A., Bernardo, K. L., and Charles, V. (1993). Development of local circuits in human visual cortex. *Journal of Neuroscience*, 13:1916–1931.

Callaway, E. M., and Katz, L. C. (1990). Emergence and refinement of clustered horizontal connections in cat striate cortex. *Journal of Neuroscience*, 10:1134–1153.

Calvert, J. E., and Harris, J. P. (1988). Spatial frequency and duration effects on the tilt illusion and orientation acuity. *Vision Research*, 28(6):1051–1059.

Campbell, F. W., and Maffei, L. (1971). The tilt aftereffect: A fresh look. *Vision Research*, 11:833–840.

Carney, T. (1982). Directional specificity in tilt aftereffect induced with moving contours: A reexamination. *Vision Research*, 22(10):1273–1275.

Carpenter, R. H. S., and Blakemore, C. (1973). Interactions between orientations in human vision. *Experimental Brain Research*, 18:287–303.

Casagrande, V. A., and Norton, T. T. (1989). Lateral geniculate nucleus: A review of its physiology and function. In Leventhal, A. G., editor, *The Neural Basis of Visual Function*, vol. 4 of *Vision and Visual Dysfunction*, 41–84. Boca Raton, Florida: CRC Press.

Catsicas, M., and Mobbs, P. (1995). Waves are swell. *Current Biology*, 5(9):977–979.

Choe, Y., and Miikkulainen, R. (1996). Self-organization and segmentation with laterally connected spiking neurons. Technical Report AI96-251, Department of Computer Sciences, The University of Texas at Austin.

Coltheart, M. (1971). Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, 78(2):114–121.

Constantine-Paton, M., Cline, H. T., and Debski, E. (1990). Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Annual Review of Neuroscience*, 13:129–154.

Dacey, D. M. (1994). Physiology, morphology, and spatial densities of identified ganglion cell types in primate retina. In Bock, G. R., and Goode, J. A., editors, *Higher-Order Processing in the Visual System (Ciba Foundation Symposium 184)*, 12–28. Chichester: Wiley.

Dalva, M. B., and Katz, L. C. (1994). Rearrangements of synaptic connections in visual cortex revealed by laser photostimulation. *Science*, 265:255–258.

Daw, N. (1995). *Visual Development.* New York: Plenum Press.

De Valois, R. L., Yund, E. W., and Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22:531–544.

Diamond, S. (1974). Four hundred years of instinct controversy. *Behavior Genetics*, 4:237–252.

Dong, D. W. (1995). Associative decorrelation dynamics: A theory of self-organization and optimization in feedback networks. In (Tesauro et al. 1995), 925–932.

Dong, D. W. (1996). Associative decorrelation dynamics in visual cortex. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu /users/nn/web-pubs/htmlbook96.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., and Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269:981–985.

Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural Computation*, 7(3):425–468.

Fahle, M., Edelman, S., and Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, 35:3003–3013.

Feller, M. B., Wellis, D. P., Stellwagen, D., Werblin, F. S., and Shatz, C. J. (1996). Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science*, 272:1182–1187.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 4:2379–2394.

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

Fisken, R. A., Garey, L. J., and Powell, T. P. S. (1975). The intrinsic, association and commissural connections of area 17 of the visual cortex. *Philosophical Transactions of the Royal Society of London Series B*, 272:487.

Fitzpatrick, D., Schofield, B. R., and Strote, J. (1994). Spatial organization and connections of iso-orientation domains in the tree shrew striate cortex. In *Society for Neuroscience Abstracts*, vol. 20, 837.

Geisler, W. S. (1997). Private communication.

Gelbtuch, M. H., Calvert, J. E., Harris, J. P., and Phillipson, O. T. (1986). Modification of visual orientation illusions by drugs which influence dopamine and GABA neurones: Differential effects on simultaneous and successive illusions. *Psychopharmacology*, 90:379–383.

Gibson, J. J., and Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology*, 20:453–467.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9:1–13.

Gilbert, C. D., Das, A., Ito, M., Kapadia, M., and Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences, USA*, 93:615–622.

Gilbert, C. D., Hirsch, J. A., and Wiesel, T. N. (1990). Lateral interactions in visual cortex. In *Cold Spring Harbor Symposia on Quantitative Biology, Volume LV*, 663–677. Cold Spring Harbor Laboratory Press.

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3:1116–1133.

Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9:2432–2442.

Gilbert, C. D., and Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30(11):1689–1701.

Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H. (1986). User's guide for NPSOL (version 4.0): A Fortran package for nonlinear programming. Technical Report SOL 86-2, Department of Operations Research, Stanford University, Stanford, CA 94305.

Goldman-Rakic, P. S. (1980). Morphological consequences of prenatal injury to the primate brain. In McConnell, P. S., Boer, G. J., Romijn, H. J., van de Poll, N. E., and Corner, M. A., editors, *Adaptive Capabilities of the Nervous System*, vol. 53 of *Progress in Brain Research*, 3–19. New York: Elsevier.

Goodhill, G. (1993). Topography and ocular dominance: A model exploring positive correlations. *Biological Cybernetics*, 69:109–118.

Greenlee, M. W., and Heitger, F. (1987). The functional role of contrast adaptation. *Vision Research*, 28(7):791–797.

Greenlee, M. W., and Magnussen, S. (1987). Saturation of the tilt aftereffect. *Vision Research*, 27(6):1041–1043.

Grinvald, A., Lieke, E. E., Frostig, R. D., and Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14:2545–2568.

Grossberg, S. (1976). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics*, 21:145–159.

Harris, J. P., and Calvert, J. E. (1989). Contrast, spatial frequency, and test duration effects on the tilt aftereffect: Implications for underlying mechanisms. *Vision Research*, 29(1):129–135.

Hata, Y., Tsumoto, T., Sato, H., Hagihara, K., and Tamura, H. (1993). Development of local horizontal interactions in cat visual cortex studied by cross-correlation analysis. *Journal of Neurophysiology*, 69:40–56.

Hayes, W. P., and Meyer, R. L. (1988a). Optic synapse number but not density is constrained during regeneration onto surgically halved tectum in goldfish: HRP-EM evidence that optic fibers compete for fixed numbers of postsynaptic sites on the tectum. *Journal of Computational Neurology*, 274:539–559.

Hayes, W. P., and Meyer, R. L. (1988b). Retinotopically inappropriate synapses of subnormal density formed by misdirected optic fibers in goldfish tectum. *Developmental Brain Research*, 38:304–312.

Henry, G. H. (1989). Afferent inputs, receptive field properties and morphological cell types in different laminae of the striate cortex. In Leventhal, A. G., editor, *The Neural Basis of Visual Function*, vol. 4 of *Vision and Visual Dysfunction*, 223–245. Boca Raton, Florida: CRC Press.

Hirsch, J. A., and Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *Journal of Neuroscience*, 11:1800–1809.

Hirsch, J. A., and Gilbert, C. D. (1993). Long–term changes in synaptic strength along specific intrinsic pathways in the cat visual cortex. *Journal of Physiology*, 461:247–262.

Howard, I. P., and Templeton, W. B. (1966). *Human Spatial Orientation*. London: Wiley.

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

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154.

Hubel, D. H., and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.

Hubel, D. H., and Wiesel, T. N. (1967). Cortical and callosal connections concerned with the vertical meridian of visual fields of the cat. *Journal of Neurophysiology*, 30:1561–1573.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243.

Jouvet, M. (1980). Paradoxical sleep and the nature-nurture controversy. In McConnell, P. S., Boer, G. J., Romijn, H. J., van de Poll, N. E., and Corner, M. A., editors, *Adaptive Capabilities of the Nervous System*, vol. 53 of *Progress in Brain Research*, 331–346. New York: Elsevier.

Kaas, J. H. (1991). Plasticity of sensory and motor maps in adult animals. *Annual Review of Neuroscience*, 14:137–167.

Kandel, E. R., and O'Dell, T. J. (1992). Are adult learning mechanisms also used for development? *Science*, 258:243–245.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1991). *Principles of Neural Science.* New York: Elsevier. Third edition.

Kapadia, M. K., Gilbert, C. D., and Westheimer, G. (1994). A quantitative measure for short-term cortical plasticity in human vision. *Journal of Neuroscience*, 14:451–457.

Kaplan, E. (1989). The receptive field structure of retinal ganglion cells in cat and monkey. In Leventhal, A. G., editor, *The Neural Basis of Visual Function*, vol. 4 of *Vision and Visual Dysfunction*, 10–40. Boca Raton, Florida: CRC Press.

Katz, L. C., and Callaway, E. M. (1992). Development of local circuits in mammalian visual cortex. *Annual Review of Neuroscience*, 15:31–56.

Kirkwood, A., and Bear, M. F. (1994). Hebbian synapses in visual cortex. *Journal of Neuroscience*, 14(4):1634–1645.

Kirkwood, A., Dudek, S. M., Gold, J. T., Aizenman, C. D., and Bear, M. F. (1993). Common forms of synaptic plasticity in vitro. *Science*, 260:1518–1521.

Köhler, W., and Wallach, H. (1944). Figural after-effects: An investigation of visual processes. *Proceedings of the American Philosophical Society*, 88:269–357.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kohonen, T. (1989). *Self-Organization and Associative Memory.* Berlin; New York: Springer. Third edition.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6:895–905.

Kurtenbach, W., and Magnussen, S. (1981). Inhibition and disinhibition and summation among orientation detectors in human vision. *Experimental Brain Research*, 43:193–198.

Lippe, W. R. (1994). Rhythmic spontaneous activity in the developing avian auditory system. *Journal of Neuroscience*, 14(3):1486–1495.

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255:209–212.

Luhmann, H. J., Martínez Millán, L., and Singer, W. (1986). Development of horizontal intrinsic connections in cat striate cortex. *Experimental Brain Research*, 63:443–448.

Maffei, L., and Galli-Resta, L. (1990). Correlation in the discharges of neighboring rat retinal ganglion cells during prenatal life. *Proceedings of the National Academy of Sciences, USA*, 87:2861–2864.

Magnussen, S., and Johnsen, T. (1986). Temporal aspects of spatial adaptation: A study of the tilt aftereffect. *Vision Research*, 26(4):661–672.

Magnussen, S., and Kurtenbach, W. (1980a). Adapting to two orientations: Disinhibition in a visual aftereffect. *Science*, 207:908–909.

Magnussen, S., and Kurtenbach, W. (1980b). Linear summation of tilt illusion and tilt aftereffect. *Vision Research*, 20:39–42.

Mansfield, R. J. W. (1974). Neural basis of orientation perception in primate vision. *Science*, 186:1133–1135.

Marder, E., and Calabrese, R. L. (1996). Principles of rhythmic motor pattern generation. *Physiological Reviews*, 76(3):687–717.

Marks, G. A., Shaffery, J. P., Oksenberg, A., Speciale, S. G., and Roffwarg, H. P. (1995). A functional role for REM sleep in brain maturation. *Behavioural Brain Research*, 69:1–11.

Masini, R., Antonietti, A., and Moja, E. A. (1990). An increase in strength of tilt aftereffect associated with tryptophan depletion. *Perceptual and Motor Skills*, 70:531–539.

McGuire, B. A., Gilbert, C. D., Rivlin, P. K., and Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *Journal of Comparative Neurology*, 305:370–392.

Meister, M., Wong, R. O. L., Baylor, D. A., and Shatz, C. J. (1991). Synchronous bursts of action-potentials in the ganglion cells of the developing mammalian retina. *Science*, 252:939–943.

Merigan, W. H., and Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, 16:369–402.

Merzenich, M. M., Recanzone, G. H., Jenkins, W. M., and Grajski, K. A. (1990). Adaptive mechanisms in cortical networks underlying cortical contributions to learning and non-declarative memory. In *Cold Spring Harbor Symposia on Quantitative Biology, Vol. LV*, 873–887. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (1997). Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In Goldstone, R. L., Schyns, P. G., and Medin, D. L., editors, *Perceptual Learning*, vol. 36 of *Psychology of Learning and Motivation*, 257–308. San Diego, CA: Academic Press.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *Journal of Neuroscience*, 14:409–441.

Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: Analysis and simulation. *Science*, 245:605–615.

Miller, K. D., and MacKay, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Computation*, 6:100–126.

Mitchell, D. E., and Muir, D. W. (1976). Does the tilt aftereffect occur in the oblique meridian? *Vision Research*, 16:609–613.

Mooney, R., Madison, D. V., and Shatz, C. J. (1993). Enhancement of transmission at the developing retinogeniculate synapse. *Neuron*, 10:815–825.

Mountcastle, V. B. (1968). Neural mechanisms in somesthesia. In Mountcastle, V. B., editor, *Medical Physiology*, vol. 2, 1372–1423. Saint Louis: C. V. Mosby. 12th edition.

Movshon, J. A., and van Sluyters, R. C. (1981). Visual neural development. *Annual Review of Psychology*, 32:477–522.

Muir, D., and Over, R. (1970). Tilt aftereffects in central and peripheral vision. *Journal of Experimental Psychology*, 85:165–170.

Murray, M., Sharma, S., and Edwards, M. A. (1982). Target regulation of synaptic number in the compressed retinotectal projection of goldfish. *Journal of Computational Neurology*, 209:374–385.

Obermayer, K., Blasdel, G. G., and Schulten, K. J. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Physical Review A*, 45:7568–7589.

Obermayer, K., Ritter, H. J., and Schulten, K. J. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences, USA*, 87:8345–8349.

O'Toole, B., and Wenderoth, P. (1977). The tilt illusion: Repulsion and attraction effects in the oblique meridian. *Vision Research*, 17:367–374.

O'Toole, B. I. (1979). Exposure-time and spatial-frequency effects in the tilt illusion. *Perception*, 8(5):557–564.

Over, R., Broerse, J., and Crassini, B. (1972). Orientation illusion and masking in central and peripheral vision. *Journal of Experimental Psychology*, 96:25–31.

Pallas, S. L., and Finlay, B. L. (1991). Compensation for population-size mismatches in the hamster retinotectal system: Alterations in the organization of retinal projections. *Visual Neuroscience*, 6:271–281.

Paradiso, M. A., Shimojo, S., and Nakayama, K. (1989). Subjective contours, tilt aftereffects, and visual cortical organization. *Vision Research*, 29(9):1205–1213.

Parker, D. M. (1972). Contrast and size variables and the tilt after-effect. *The Quarterly Journal of Experimental Psychology*, 24:1–7.

Pettet, M. W., and Gilbert, C. D. (1992). Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences, USA*, 89:8366–8370.

Pfleger, B., and Bonds, A. B. (1995). Dynamic differentiation of GABAA-sensitive influences on orientation selectivity of complex cells in the cat striate cortex. *Experimental Brain Research*, 104:81–88.

Purves, D. (1988). *Body and Brain: A Trophic Theory of Neural Connections*. Cambridge, MA: Harvard University Press.

Purves, D., and Lichtman, J. W. (1985). *Principles of Neural Development*. Sunderland, MA: Sinauer.

Reh, T. A., and Constantine-Paton, M. (1985). Eye-specific segregation requires neural activity in three-eyed *rana pipiens*. *Journal of Neuroscience*, 5(5):1132–1143.

Rochester, N., Holland, J. H., Haibt, L. H., and Duda, W. L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, 2:80–93. Reprinted in **?**.

Roffwarg, H. P., Muzio, J. N., and Dement, W. C. (1966). Ontogenetic development of the human sleep-dream cycle. *Science*, 152:604–619.

Ross, J., and Speed, H. D. (1996). Perceived contrast following adaptation to gratings of different orientations. *Vision Research*, 36(12):1811–1818.

Ross, J., Speed, H. D., and Morgan, M. J. (1993). The effects of adaptation and masking on incremental thresholds for contrast. *Vision Research*, 33(15):2051–2056.

Sejnowski, T. J., and Churchland, P. S. (1989). Brain and cognition. In Posner, M. I., editor, *Foundations of Cognitive Science*, chapter 8, 315–356. Cambridge, MA: MIT Press.

Shatz, C. J. (1990). Impulse activity and the patterning of connections during CNS development. *Neuron*, 5:745–756.

Shatz, C. J. (1996). Emergence of order in visual system development. *Proceedings of the National Academy of Sciences, USA*, 93:602–608.

Shatz, C. J., and Stryker, M. P. (1978). Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *Journal of Physiology (London)*, 281:267–283.

Shiu, L.-P., and Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, 52:582–588.

Sillito, A. M. (1979). Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge. *Journal of Physiology*, 289:33–53.

Sirosh, J. (1995). *A Self-Organizing Neural Network Model of the Primary Visual Cortex*. PhD thesis, Department of Computer Sciences, The University of Texas at Austin, Austin, TX. Technical Report AI95-237.

Sirosh, J., and Miikkulainen, R. (1994a). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71:66–78.

Sirosh, J., and Miikkulainen, R. (1994b). Modeling cortical plasticity based on adapting lateral interaction. In Bower, J. M., editor, *The Neurobiology of Computation: The Proceedings of the Third Annual Computation and Neural Systems Conference*, 305–310. Dordrecht; Boston: Kluwer.

Sirosh, J., and Miikkulainen, R. (1995). Ocular dominance and patterned lateral connections in a self-organizing model of the primary visual cortex. In (Tesauro et al. 1995), 109–116.

Sirosh, J., and Miikkulainen, R. (1996). Self-organization and functional role of lateral connections and multisize receptive fields in the primary visual cortex. *Neural Processing Letters*, 3:39–48.

Sirosh, J., and Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9:577–594.

Sirosh, J., Miikkulainen, R., and Bednar, J. A. (1996). Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

Smith, A. T., and Over, R. (1977). Orientation masking and the tilt illusion with subjective contours. *Perception*, 6(4):441–447.

Somers, D. C., Toth, L. J., Todorov, E., Rao, S. C., Kim, D.-S., Nelson, S. B., Siapas, A. G., and Sur, M. (1996). Variable gain control in local cortical circuitry supports context-dependent modulation by long-range connections. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

Spivey-Knowlton, M. J. (1993). Simulating tilt illusions with lateral inhibition and a virtual axis. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, 953–958. Hillsdale, NJ: Erlbaum.

Stemmler, M., Usher, M., and Niebur, E. (1995). Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science*, 269:1877–1880.

Stent, G. S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proceedings of the National Academy of Sciences, USA*, 70:997–1001.

Steriade, M., Paré, D., Bouhassira, D., Deschênes, M., and Oakson, G. (1989). Phasic activation of lateral geniculate and perigeniculate thalamic neurons during sleep with ponto-geniculo-occipital waves. *Journal of Neuroscience*, 9(7):2215–2229.

Sugita, Y. (1996). Global plasticity in adult visual cortex following reversal of visual input. *Nature*, 380:523–526.

Sulston, J. E., and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *caenorhabditis elegans*. *Developmental Biology*, 56:110–156.

Sutherland, N. S. (1961). Figural after-effects and apparent size. *Quarterly Journal of Psychology*, 13:222–228.

Swindale, N. V. (1980). A model for the formation of ocular dominance stripes. *Proceedings of the Royal Society of London Series B*, 215:243–264.

Tesauro, G., Touretzky, D. S., and Leen, T. K., editors (1995). *Advances in Neural Information Processing Systems 7*. Cambridge, MA: MIT Press.

Thomson, A. M., and Deuchars, J. (1994). Temporal and spatial properties of local circuits in neocortex. *Trends in Neurosciences*, 17(3):119–126.

Tolhurst, D. J., and Thompson, P. G. (1975). Orientation illusions and aftereffects: Inhibition between channels. *Vision Research*, 15:967–972.

Ts'o, D. Y., Frostig, R. D., Lieke, E. E., and Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science*, 249:417–420.

van der Zwan, R., and Wenderoth, P. (1994). Psychophysical evidence for area V2 involvement in the reduction of subjective contour tilt aftereffects by binocular rivalry. *Visual Neuroscience*, 11:823–830.

van der Zwan, R., and Wenderoth, P. (1995). Mechanisms of purely subjective contour tilt aftereffects. *Vision Research*, 35(18):2547–2557.

Van Essen, D. C., Anderson, C. H., and Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255:419–423.

Vidyasagar, T. R. (1990). Pattern adaptation in cat visual cortex is a co-operative phenomenon. *Neuroscience*, 36(1):175–179.

Virsu, V., and Taskinen, H. (1974). Central inhibitory interactions in human vision. *Experimental Brain Research*, 23:65–74.

von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, 15:85–100. Reprinted in **?**.

von der Malsburg, C. (1987). Synaptic plasticity as basis of brain organization. In Changeux, J.-P., and Konishi, M., editors, *The Neural and Molecular Bases of Learning*, 411–432. New York: Wiley.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Ware, C., and Mitchell, D. E. (1974). The spatial selectivity of the tilt aftereffect. *Vision Research*, 14:735–737.

Waugh, S. J., Levi, D. M., and Carney, T. (1993). Orientation, masking, and vernier acuity for line targets. *Vision Research*, 33(12):1619–1638.

Weiss, Y., Edelman, S., and Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718.

Weliky, M., Kandler, K., Fitzpatrick, D., and Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron*, 15:541–552.

Wenderoth, P., and Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. *Vision Research*, 28(2):301–312.

Wenderoth, P., van der Zwan, R., and Johnstone, S. (1989). Mechanisms of orientation illusions. In Vickers, D., and Smith, P., editors, *Human Information Processing: Measures, Mechanisms, and Models*, 83–106. Bellingham, Washington: SPIE.

Westheimer, G. (1990). Simultaneous orientation contrast for lines in the human fovea. *Vision Research*, 30:1913–1921.

Wilson, H. R., and Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision Research*, 33(8):1133–1149.

Wolfe, J. M. (1984). Short test flashes produce large tilt aftereffects. *Vision Research*, 24(12):1959–1964.

Wolfe, J. M., and Held, R. (1982). Gravity and the tilt aftereffect. *Vision Research*, 22(8):1075–1078.

Wolfe, J. M., and O'Connell, K. M. (1986). Fatigue and structural change: Two consequences of visual pattern adaptation. *Investigative Ophthalmology and Visual Science*, 27(4):538–543.

Wong, R. O. L., Meister, M., and Shatz, C. J. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, 11(5):923–938.

Vita

James Bednar was born in Houston, Texas on February 28, 1971. He has received two Bachelor's degrees from the University of Texas at Austin. The first was a Bachelor of Science in Electrical Engineering in the Computer Engineering block, received December 1993. The second was a Bachelor of Arts in Philosophy with a minor of Mathematics, received May 1994. He entered the University of Texas at Austin as a graduate student in the Fall of 1994, and is currently in the PhD. program in the Department of Computer Sciences.

Permanent Address: 9134 Troulon
Houston, Texas 77036
USA
jbednar@cs.utexas.edu
http://www.cs.utexas.edu/users/jbednar/

This thesis was typeset with $\text{\LaTeX}\,2_\varepsilon$[1] by the author.

---

[1]$\text{\LaTeX}\,2_\varepsilon$ is an extension of $\text{\LaTeX}$. $\text{\LaTeX}$ is a collection of macros for $\text{\TeX}$. $\text{\TeX}$ is a trademark of the American Mathematical Society. The macros used in formatting this thesis were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin.