Color version of *Neural Computation* 12(7):1721-1740, 2000.

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

James A. Bednar (jbednar@cs.utexas.edu)
Risto Miikkulainen (risto@cs.utexas.edu)
Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712 USA

September 22, 2000

Abstract

RF-LISSOM, a self-organizing model of laterally connected orientation maps in the primary visual cortex, was used to study the psychological phenomenon known as the tilt aftereffect. The same self-organizing processes that are responsible for the long-term development of the map are shown to result in tilt aftereffects over short time scales in the adult. The model permits simultaneous observation of large numbers of neurons and connections, making it possible to relate high-level phenomena to low-level events, which is difficult to do experimentally. The results give detailed computational support for the long-standing conjecture that the direct tilt aftereffect arises from adaptive lateral interactions between feature detectors. They also make a new prediction that the indirect effect results from the normalization of synaptic efficacies during this process. The model thus provides a unified computational explanation of self-organization and both the direct and indirect tilt aftereffect in the primary visual cortex.

1 Introduction

The tilt aftereffect (TAE, Gibson & 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). The effect resembles an afterimage from staring at a bright light, but it represents changes in orientation perception rather than in color or brightness.

Most modern explanations of the TAE are loosely based on the feature-detector model of the primary visual cortex (V1), which characterizes this area as a set of orientation-detecting neurons. Experiments showed that these neurons became more difficult to excite during repeated presentation of oriented visual stimuli, and the desensitization persisted for some time (Hubel & Wiesel, 1968). This observation led to the *fatigue* theory of the TAE: perhaps active neurons become fatigued due to repeated firing, causing the response to a test figure to change during adaptation.

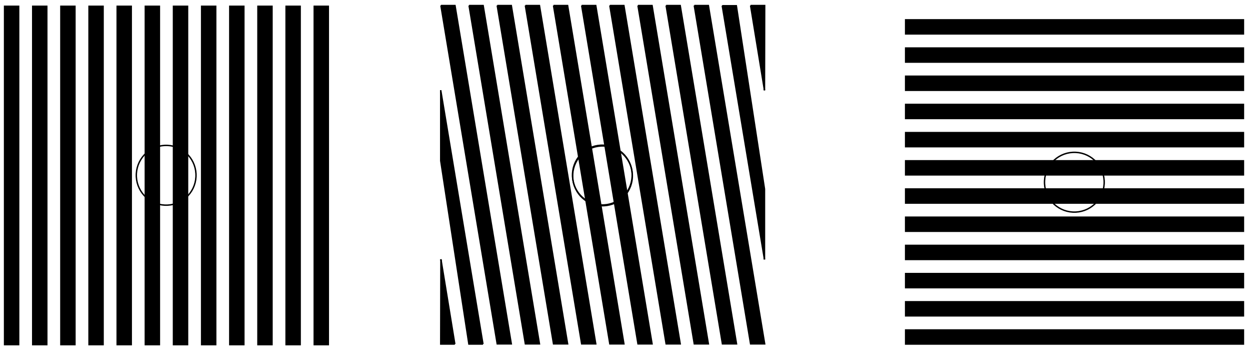


Figure 1: **Tilt aftereffect patterns.** Fixate your gaze on the circle inside the central diagram for at least thirty seconds, moving your eye slightly inside the circle to avoid developing strong afterimages. Now fixate on the diagram at the left. The vertical lines should appear slightly tilted clockwise; this phenomenon is called the direct tilt aftereffect. If you instead fixate upon the horizontal lines at the right, they should appear barely tilted counterclockwise, due to the indirect tilt aftereffect. (Adapted from Campbell & Maffei, 1971.)

Assuming the perceived orientation is some sort of average over the orientation preferences of the activated neurons, the final perceived orientation would thus show the direct TAE (Coltheart, 1971). The fatigue theory has been discredited for a number of reasons, chief among which is that individual V1 neurons actually do not appear to fatigue (Finlayson & Cynader, 1995; McLean & Palmer, 1996). In fact, their response to direct stimulation is essentially unchanged by adaptation to a visual stimulus (Vidyasagar, 1990). The now-popular *inhibition* theory postulates that tilt aftereffects instead result from changing inhibition between orientation-detecting neurons (Tolhurst & Thompson, 1975).

The inhibition hypothesis has recently been incorporated into theories of the larger purpose and function of the cortex. Barlow (1990) and Földiák (1990) have proposed that the early cortical regions are acting to reduce the amount of redundant information present in the visual input. They suggest that 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 (see also Field, 1994; Miikkulainen et al., 1997; Sirosh et al., 1996). Based on these theories, Dong (1995) has shown analytically that perfect decorrelation can result in direct tilt aftereffects which are similar to those found in humans. Only very recently, however, has it become computationally feasible to test the inhibition/decorrelation theory of the TAE in a detailed model of cortical function, with limitations similar to those known to be present in the cortex.

A Hebbian self-organizing process (the Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map, or RF-LISSOM; Miikkulainen, Bednar, Choe, & Sirosh, 1997; Sirosh, 1995; Sirosh & Miikkulainen, 1994a, 1997) has been shown to develop feature detectors and specific lateral connections that could produce such 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 the 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 first model that also shows how the long-range lateral connections self-organize as an integral part of the process. The lateral-

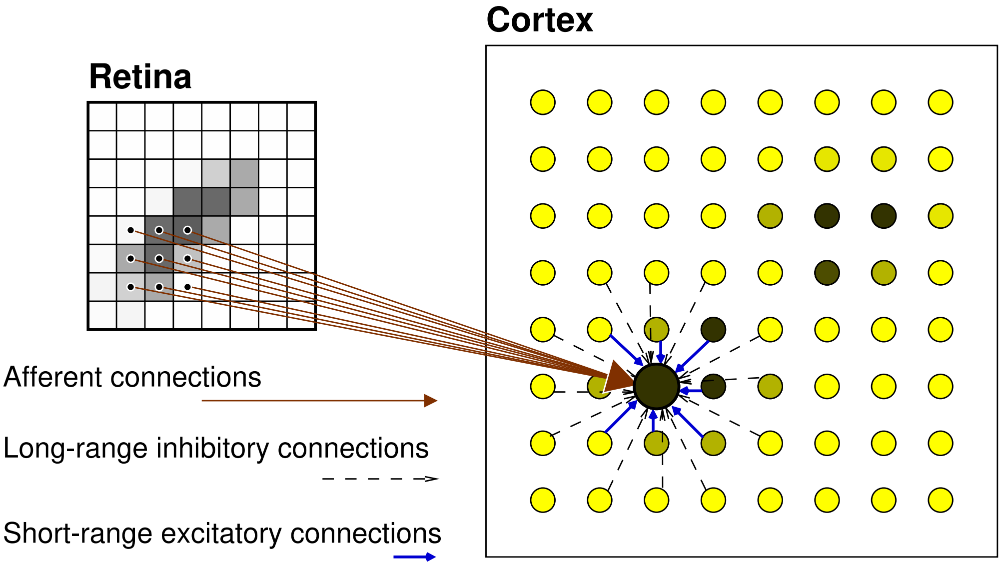


Figure 2: **Architecture of the RF-LISSOM network.** A small RF-LISSOM network and retina are shown, along with connections to a single neuron (shown as a large circle). The input is an oriented Gaussian activity pattern on the retinal ganglion cells (shown by grayscale coding); the LGN is bypassed for simplicity. The afferent connections form a local anatomical receptive field (RF) on the simulated retina. Neighboring neurons have different but highly overlapping RFs. Each neuron computes an initial response as a scalar product of its receptive field and its afferent weight vector. The responses then repeatedly propagate within the cortex through the lateral connections and evolve into activity "bubbles". After the activity stabilizes, weights of the active neurons are adapted.

ly connected model has also been shown to account for many of the dynamic aspects of the adult visual cortex, such as reorganization following retinal and cortical lesions (Miikkulainen et al., 1997; Sirosh & Miikkulainen, 1994b; Sirosh, Miikkulainen, & Bednar, 1996). These findings suggest that the same self-organization processes that drive development may be acting in the adult (see Fregnac, 1996; Gilbert, 1998 for review). We explore this hypothesis here with respect to the TAE.

The current work is a first study of the *functional* behavior of the RF-LISSOM model, specifically the response to stimuli similar to those known to cause the TAE in humans. The model permits simultaneous observation of activation and connection patterns between large numbers of neurons. This makes it possible to relate higher-level phenomena to low-level events, which is difficult to do experimentally. The results provide detailed computational support for the idea that tilt aftereffects result from a general self-organizing process that aims at producing an efficient, sparse encoding of the input through decorrelation.

2 Architecture

The cortical architecture for the model 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 cortex corresponds to a vertical column of cells through the six layers of the primate cortex. The cortical network is modeled with a sheet of interconnected neurons and the retina with a sheet of retinal ganglion cells (figure 2). Neurons receive afferent connections from broad overlapping circular patches on the retina. The $N \times N$ network is projected on to a central region of the $R \times R$ retinal ganglion cells, and each neuron is connected to ganglion cells in an area of radius r around the projections. Thus, neurons at a particular cortical location receive afferents from a corresponding location on the retina. Since the lateral geniculate nucleus (LGN) accurately reproduces the receptive fields of the retina, it has been bypassed for simplicity.

Each neuron also has reciprocal excitatory and inhibitory lateral connections with itself and other neurons. Lateral excitatory connections are short-range, connecting each neuron with itself and its close neighbors. Lateral inhibitory connections run for comparatively long distances, but also include connections to the neuron itself and to its neighbors.[1]

The input to the model consists of 2-D ellipsoidal Gaussian patterns representing retinal ganglion cell activations (as in figures 2 and 3a). For training, the orientations of the Gaussians are chosen randomly from the uniform distribution in the full range $[0, \pi)$, and the positions are chosen randomly within the retina. The elongated spots approximate natural visual stimuli after the edge detection and enhancement mechanisms in the retina. They can also be seen as a model of the intrinsic retinal activity waves that occur in late pre-natal development in mammals (Bednar & Miikkulainen, 1998; Shatz, 1990). The RF-LISSOM network models the self-organization of the visual cortex based on these natural sources of elongated features.

The afferent weights are initially set to random values, and the lateral weights are preset to a smooth Gaussian profile. The connections 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) is calculated as a weighted sum of the retinal activations:

$$\eta_{ij} = \sigma\left(\sum_{a,b} \xi_{ab}\mu_{ij,ab}\right), \tag{1}$$

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

$$\eta_{ij}(t) = \sigma\left(\sum \xi\mu + \gamma_e \sum_{k,l} E_{ij,kl}\eta_{kl}(t-1) - \gamma_i \sum_{k,l} I_{ij,kl}\eta_{kl}(t-1)\right), \tag{2}$$

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)

[1]For high-contrast inputs, long-range interactions must be inhibitory for proper self-organization to occur (Sirosh, 1995). Recent optical imaging and electrophysiological studies have indeed shown that long-range interactions in the cortex are inhibitory at high contrasts, even though individual lateral connections are primarily excitatory (Grinvald et al., 1994; Hata et al., 1993; Hirsch & Gilbert, 1991; Weliky et al., 1995). The model uses explicit inhibitory connections for simplicity since all inputs used are high-contrast, and since it is the high-contrast inputs that primarily drive adaptation in the Hebbian model (Bednar, 1997).

during the previous time step. 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 cortical activity pattern starts out diffuse and spread over a substantial part of the map (as in figure 3*c*), but within a few iterations of equation 2, converges into a small number of stable focused patches of activity, or activity bubbles (figure 3*d*). 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:

$$w_{ij,mn}(t+\delta t) = \frac{w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}}{\sum_{mn}\left[w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}\right]}, \tag{3}$$

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). 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.

At long distances, very few neurons have correlated activity and therefore most long-range connections eventually become weak. The weak connections are eliminated periodically, resulting in patchy lateral connectivity similar to that observed in the visual cortex. The radius of the lateral excitatory interactions starts out large, but as self-organization progresses, it is decreased until it covers only the nearest neighbors. Such a decrease is one way of varying the balance between excitation and inhibition to ensure that global topographic order develops while the receptive fields become well-tuned at the same time (Miikkulainen et al., 1997; Sirosh, 1995). Similar effects can be achieved by changing the scaling factors γ_e and γ_i over time, or by using connections whose sign depends upon the activation level of the neuron (as in Stemmler et al., 1995).

3 Experiments

The model consisted of an array of 192×192 neurons, and a retina of 36×36 ganglion cells. The center of the anatomical receptive field of each neuron was placed at the location in the central 24×24 portion of the retina corresponding to the location of the neuron in the cortex, so that every neuron would have a complete set of afferent connections. The RF had a circular shape, consisting of random-strength connections to all ganglion cells within 6 units from the RF center. The cortex was self-organized for $20,000$ iterations on oriented Gaussian inputs with major and minor axes of half-width $\sigma = 7.5$ and 1.5, respectively.[2] The training took 2.5 hours on 16 processors of a Cray

[2] 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 0.00005 were pruned away at $20,000$ iterations. The lateral inhibitory connections were initialized to a Gaussian profile with $\sigma = 100$, and the lateral excitatory connections to a Gaussian with $\sigma = 15$, with no connections outside the nominal circular radius. The lateral excitation

T3E supercomputer at the Texas Advanced Computing Center. The model requires 1.5 gigabytes of physical memory to represent the 400 million connections in this small section of the cortex.

3.1 Orientation map organization

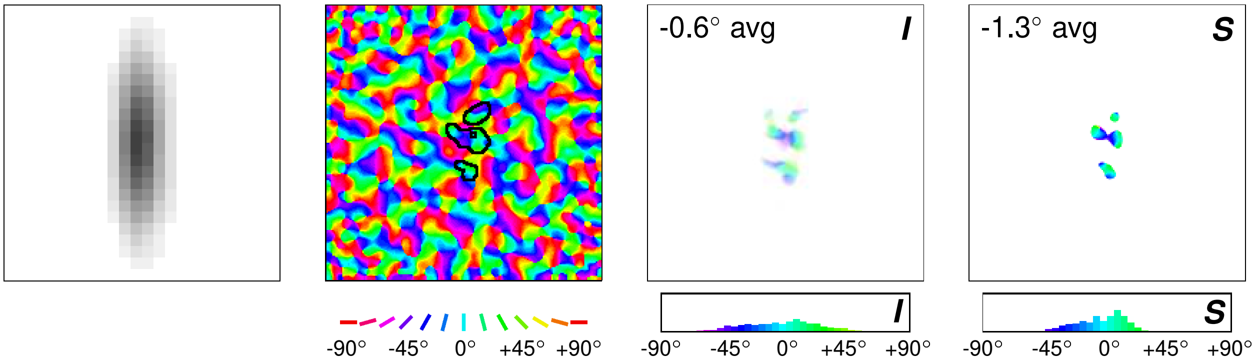
In the self-organization process, the neurons developed oriented receptive fields organized into orientation columns very similar to those observed in the primary visual cortex (figure 3b). The strongest lateral connections of highly-tuned cells link areas of similar orientation preference, and avoid neurons with the orthogonal orientation preference. 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 subsequently found experimentally by Bosking et al., 1997.) 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 (Sirosh et al., 1996). Since the long-range lateral connections are inhibitory, the net result is *decorrelation*: redundant activation is removed, resulting in a sparse representation of the novel features of each input (Barlow, 1990; Field, 1994; Sirosh et al., 1996). As a side effect, illusions and aftereffects may sometimes occur, as will be shown below.

3.2 Aftereffect simulations

In psychophysical measurements of the TAE, a fixed stimulus is presented at a particular location on the retina. To simulate these conditions in the RF-LISSOM model, the position and angle of the inputs were fixed to a single value for a number of iterations. To permit more detailed analysis of behavior at short time scales, the learning rates were reduced from those used during self-organization to $\alpha_A = \alpha_E = \alpha_I =$0.000005. All other parameters remained as in self-organization.

To compare with the psychophysical experiments, it is necessary to determine what orientation the model "perceives" for any given input. Precisely how neural responses are interpreted for perception remains quite controversial (see Parker & Newsome, 1998 for review), but results in primates suggest that behavioral performance approaches the statistical optimum given the measured properties of cortical neurons (Geisler & Albrecht, 1997). Accordingly, we extracted the perceived orientation using a vector sum procedure, which has been shown to be optimal under conditions present in the model (Snippe, 1996). For this procedure, each active neuron was represented by a vector whose magnitude corresponded to the activation level and whose direction corresponded to the orientation preference of the neuron. Perceived orientation was then measured as a vector sum over all neurons that responded to the input. In the model, the perceived orientation was found to match the absolute orientation of the input pattern to within a few degrees (Bednar, 1997).

γ_e and inhibition strength γ_i were both 0.9. The learning rate α_{A} was gradually 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 were 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 increased to 13 over the course of training. These parameter settings were used by Sirosh et al. to model development of the orientation map, and were not tuned or tweaked for the tilt aftereffect simulations. Small variations produce roughly equivalent results (Sirosh, 1995).



(*a*) Input pattern (*b*) Orientation map (*c*) Initial activity (*d*) Settled activity

Figure 3: **Orientation map activation.** The orientation color key underneath (*b*) applies to all of the graphs in (*b-d*). After being trained on inputs like the one in (*a*) with random positions and orientations, the RF-LISSOM network developed the orientation map shown in (*b*). Each neuron is colored according to the orientation it prefers. The black outline shows the extent of the patchy self-organized lateral inhibitory connections of one neuron (marked with a black square) which has a vertical orientation preference. The strongest connections of each neuron are extended along its preferred orientation and link columns with similar orientation preferences, avoiding those with orthogonal preferences. The brightness of the colors in (*c,d*) shows the strength of activation for each neuron to pattern (*a*). The initial response of the organized map is spatially broad and diffuse (*c, top*), like the input, and its cortical location at, above, and below the center of the cortex indicates that the input is vertically extended around the center of the retina. The response is patchy because the network is also encoding orientation, and the neurons that encode orientations far from the vertical do not respond (compare *c* to *b*). The histogram (*c, bottom*) sums up the orientation coding of the response. Each bin represents a range of 5°, which is the precision to which the orientation map was measured. A wide range of neurons preferring orientations around 0° are activated, but the average orientation is approximately 0° (-0.6° for this particular run). After the network settles through lateral interactions, the activation is much more focused, both spatially (*d, top*) and in representing orientation (*d, bottom*), but the spatial and orientation averages continue to match the position and orientation of the input, respectively. The average orientation of the settled response (-1.3° here) is taken to be the perceived orientation for the TAE experiments. Animated demos of these figures can be seen at `http://www.cs.utexas.edu/users/nn/pages/research/selforg.html`.

To determine the tilt aftereffect in the model, the perceived orientation was first computed for inputs of all orientations. The model then adapted to a fixed input for 90 iterations, and the perceived orientation was again computed for all inputs. The difference between the initial perceived angle and the one perceived after adaptation was taken as the magnitude of the TAE. Figure 4 plots these differences for the different angles. For comparison, figure 4 also shows the most detailed data available for the TAE in human foveal vision (Mitchell & 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 exhibited angle repulsion effects nearly identical to those found in the RF-LISSOM model; the data was most complete for the subject shown. The magnitude of this *direct* TAE increases very rapidly to a maximum angle repulsion at 8-10°, falling off somewhat more gradually to zero as the angular separation increases. The simulations with larger angular separations (from 40° to 85°) show a smaller angle attraction, i.e. the *indirect* effect. Although there is a greater inter-subject variability in the psychophysical

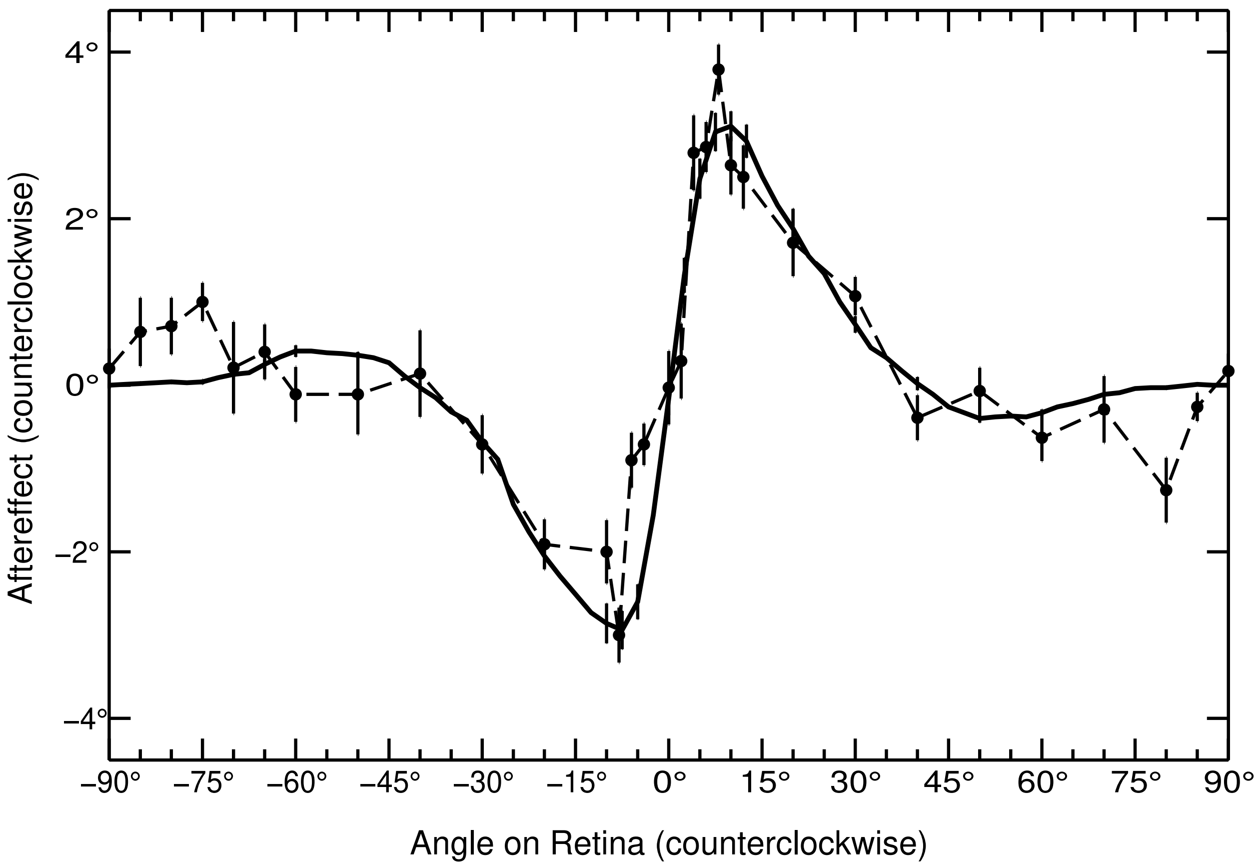


Figure 4: **Tilt aftereffect at different angles.** The open circles represent the tilt aftereffect for a single human subject (DEM) from Mitchell and Muir (1976) averaged over ten trials. For each angle in each trial, the subject adapted for three minutes on a sinusoidal grating of a given angle, then 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 *x*-axis range $\pm 40^\circ$; the indirect TAE for the larger angles varied widely between $\pm 2.5^\circ$. 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 ten trials with different adaptation angles. Error bars indicate ± 1 standard error of measurement; in most cases they are too small to be visible since the results were very consistent between different runs. The network adapted to an oriented line at the center of the retina for 90 iterations, then the TAE was measured for test lines oriented at each angle. The duration of adaptation was chosen so that the magnitude of the human data and the model match; this was the only parameter fit to the data. The result from the model closely resembles the curve for humans, showing both direct and indirect tilt aftereffects.

literature for the indirect effect than the direct effect, those found for the RF-LISSOM model are well within the range seen for human subjects.

In parallel with the angular changes in the TAE, its peak magnitude in humans increases logarithmically with adaptation time (Gibson & Radner, 1937), eventually saturating at a level that depends upon the experimental protocol used (Greenlee & Magnussen, 1987; Magnussen & Johnsen, 1986). As the number of adaptation iterations is increased in the model, the magnitude of the TAE also increases logarithmically (figure 5). (The single curve that best matched the magnitude of the human data was shown in figure 4, but the ones for different amounts of adaptation all had the same basic shape.) The time course of the TAE in the RF-LISSOM model is qualitatively similar to the human data, but it does not completely saturate over the adaptation amounts tested so far.

3.3 How does the TAE arise in the model?

The TAE seen in figure 4 must result from changes in the connection strengths between neurons, since no other component of the model changes as adaptation progresses. Simulations performed with only one type of weight adapting (either afferent, lateral excitatory, or lateral inhibitory) show that the inhibitory weight changes determine the shape of the curve for all angles (Bednar, 1997).

In what way do the changing inhibitory connections cause these effects? We will demonstrate this process in a simulation where only inhibitory weights adapted, the adaptation period was longer, and a higher learning rate was used, all in order to exaggerate the effect and make its causes more clearly visible. First, the connections change in a way that leaves the perceived orientation of the adaptation line unchanged. Figure 6*a* shows the initial response (***I***), the settled response (***S***), and the settled response after adaptation (***A***) for a vertical input (0°, marked with a vertical line.) The response after adaptation is more diffuse because the most active neurons have become inhibited (equation 3). Throughout adaptation, the distribution of active orientation detectors is centered around approximately the same angle, and a constant angle is perceived.

The initial and settled responses to a test line with a slightly different orientation (e.g. 10°, marked with a dotted line in figure 6*b*) are again centered around that orientation (6*b****I*** and ***S***). However, comparing the histograms of settled activity before and after adaptation (6*b****S*** and ***A***), it is clear that fewer neurons close to 0° respond after adaptation, but an increased number of those representing distant angles (over 10°) do. This is because during adaptation, inhibition was strengthened primarily between neurons close to the 0° adaptation angle, and not between those that prefer larger orientations.

Such adaptation increases the ability of the map to detect small differences from the adaptation line. Before adaptation, the settled histograms for 0° and 10° are fairly similar, with averages differing by only 9.2° in this simulation (6*a****S*** and 6*b****S***). After adaptation, the histograms are very different, resulting in a 23° difference in perceived orientation (compare 6*a****A*** with 6*b****A***). This adaptation is manifested at the psychological level as a shift of the perceived orientation *away* from the adaptation line, that is, the direct TAE.

Meanwhile, the response to a very different test line (e.g. 50°, the dotted line in figure 6*c*) becomes broader and stronger after adaptation (compare 6*c****S*** and ***A***). Adaptation occurred only in activated neurons, so neurons with orientation preferences greater than 50° are unchanged.

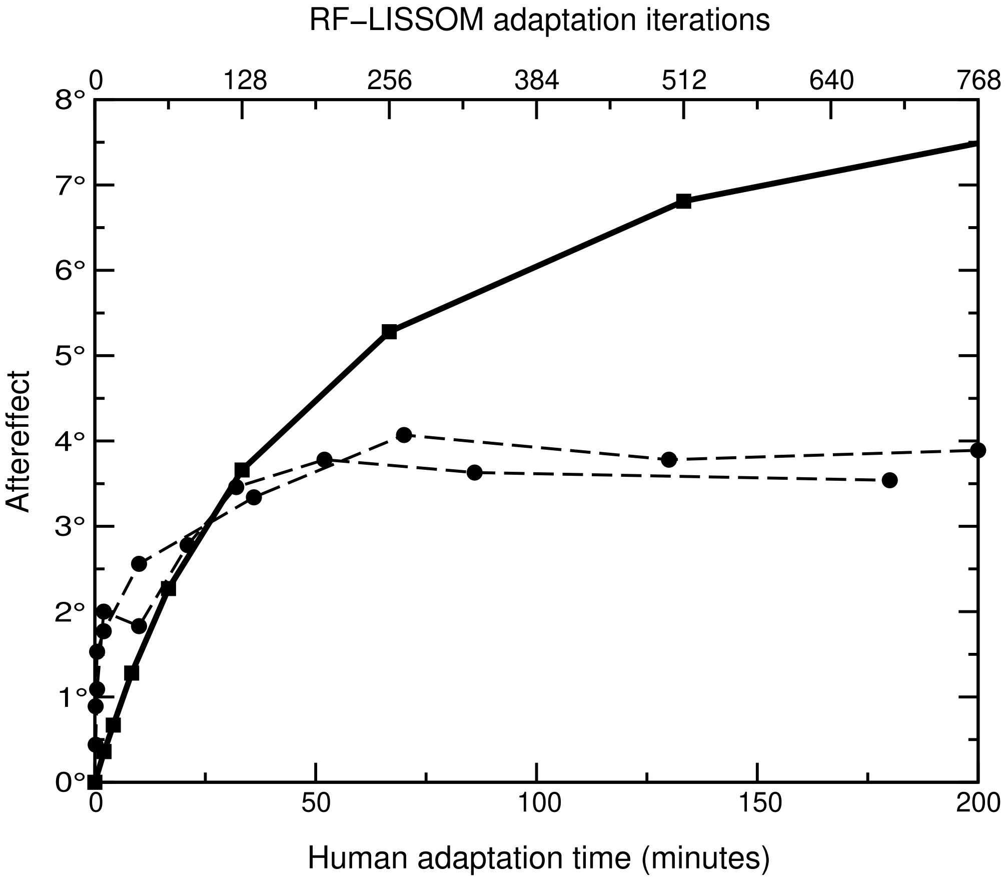
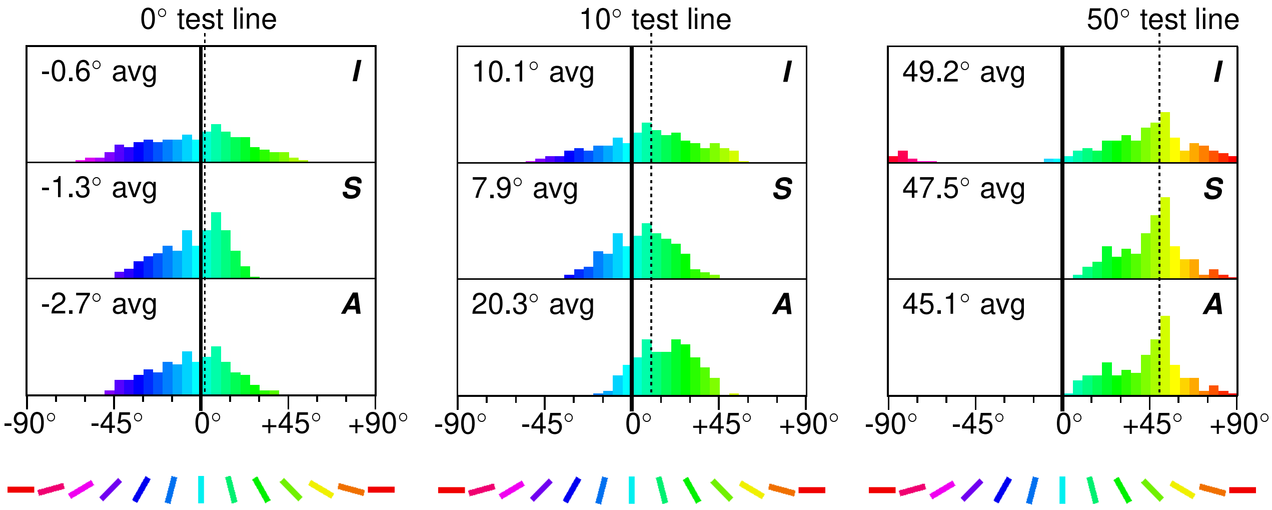


Figure 5: **Direct TAE over time.** The circles show the magnitude of the TAE as a function of adaptation time for human subjects MWG (unfilled circles) and SM (filled circles) from Greenlee and Magnussen (1987). They were the only subjects tested in the study. Each subject adapted to a single $+12^{\circ}$ 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 a comparison line at another location to match it. The plots represent averages of five runs; 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 $+12^{\circ}$ test line over 9 trials (with parameters as in figure 4). 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.

However, those with preferences less than 50° actually now respond more strongly (6c**S** and **A**). The reason is that during adaptation, the inhibitory connections of these neurons with each other became stronger. Because of normalization (equation 3), their connections to other neurons, i.e. those representing distant angles such as 50°, became weaker. As a result, the 50° line now inhibits them less than before adaptation. Thus they are more active, and the perceived orientation shifts towards 0°, causing the indirect tilt aftereffect.

The indirect effect is therefore true to its name, caused indirectly by the strengthening of inhibitory connections during adaptation. This explanation of the indirect effect is novel, and emerges automatically from the RF-LISSOM model. The model thus shows computationally how



(*a*) Adaptation (*b*) Direct effect (*c*) Indirect effect

Figure 6: **Explanation of the TAE.** This figure shows activation histograms similar to those in figure 3*c* and 3*d* for several test lines. The histograms focus on the orientation-specific aspects of the cortical response by abstracting out the spatial content, and demonstrate how changes in the response cause the TAE. The top row of histograms (marked ***I***, for Initial) shows the initial response to a vertical line (0°), a 10° line, and a 50° line, each marked by dotted lines. In each case, the initial response is roughly centered around the orientation of the input line. The next row (marked ***S***, Settled) shows the settled response, which has been focused by the lateral connections but is still centered around the input orientation. The bottom row (marked ***A***, Adapted) shows the settled response to the same input *after* adapting to a vertical (0°) line, marked by a vertical line on the plots. To magnify and clarify the effect for explanatory purposes, only the inhibitory weights were modifiable in this simulation, their learning rate was increased to 0.00005, and the adaptation lasted for 256 iterations. In (*a*), the settled response to the 0° line broadens with adaptation, as the inhibition between the active units increases (*a****S***→***A***). Since the response remains centered around 0°, there is little change in the perceived orientation and the TAE is close to 0° (as can also be seen in figure 4). In contrast in (*b*), a dramatic orientation shift is evident: while the settled histogram before adaptation was centered around 7.9°, after adaptation it is centered around 20.3° (*b****S***→***A***), inducing a direct effect of 12.4°. The direct TAE is caused by the same changes that caused the broadening in (*a*): the activity around 0° has decreased, while the activity at larger angles has increased. The changes are more subtle for the indirect effect (*c*). For the 50° stimulus, only the neurons around 0° in (*c****I***) fall in the range of orientations initially activated by the adaptation line (*a****I***), and thus those are the only ones that change their behavior between (*c****S***) and (*c****A***). During adaptation, their inhibition to and from neurons around 0° was increased, and the weight normalization caused a corresponding decrease to other neurons, including those around 50°. As a result, they are now less inhibited than before adaptation, and the average response shifts *towards* the adaptation angle (the indirect TAE). Animated demos of these examples can be seen at `http://www.cs.utexas.edu/users/nn/pages/research/selforg.html`.

both the direct and indirect effects can be caused by the same activity-dependent adaptation process, and that it is the same process that drives the development of the map.

4 Discussion and Future Work

Our results suggest that the same self-organizing principles that result in sparse coding and reduce redundant activation in the visual cortex may also be operating over short time intervals in the adult, with quantifiable psychological consequences such as the TAE. This finding demonstrates a potentially important computational link between development, structure, and function.

Although the RF-LISSOM model was not originally developed as an explanation for the tilt aftereffect, it exhibits tilt aftereffects that have nearly all of the features of those measured in humans. The effect of varying angular separation between the test and adaptation lines is similar to human data, the time course is approximately logarithmic, and the TAE is localized to the retinal location that experienced the stimulus. With minor extensions, the model should account for other features of the TAE as well, such as higher variance at oblique orientations, frequency localization, movement direction specificity, and ocular transfer (Bednar, 1997).

One difference, however, shows up in prolonged adaptation: the TAE in humans eventually saturates near 4-5° (Greenlee & Magnussen, 1987; Magnussen & Johnsen, 1986; Wolfe & O'Connell, 1986). The TAE in the RF-LISSOM model does not saturate quite as quickly (figure 5), and it also appears to saturate for different reasons. In the model, saturation occurs only after the inhibitory weights have been strengthened so much that the cortical response to the adaptation line is entirely suppressed. However, in humans, the adaptation line is still easily detectable even after saturation (Magnussen & Johnsen, 1986). Apparently, the human visual system has additional constraints on how much adaptation can occur in the short term, and those constraints are currently not part of the model.

Another feature of the human TAE that does not directly follow from the model is the gradual recovery of accurate perception even in complete darkness (Greenlee & Magnussen, 1987; Magnussen & Johnsen, 1986). With the purely Hebbian learning used in the model, only active units are adapted, and thus no weight changes will occur for blank inputs. One possible explanation for both saturation and dark recovery is that the inhibitory weights modified during tilt adaptation are a set of small, temporary weights adding to or multiplying more permanent connections. Changes to these weights would be limited in magnitude, and the changes would gradually decay in the the absence of visual stimulation. Such a mechanism was proposed by von der Malsburg (1987) as an explanation of visual object segmentation, and was implemented in the RF-LISSOM model of segmentation by Choe and Miikkulainen (1998) and Miikkulainen et al. (1997). A variety of physical substrates for temporary modifications in efficacy have been found (reviewed in Zucker, 1989). Such effects could also be included in the TAE model, which would allow it to replicate the saturation and dark recovery aspects of the human TAE.

The main contribution of the RF-LISSOM model of the TAE is its novel explanation of the indirect aftereffect. Proponents of the lateral inhibitory theory of the TAE and tilt illusion have generally ignored indirect effects, or postulated that they occur only at higher cortical levels (Wenderoth & Johnstone, 1988; van der Zwan & Wenderoth, 1995), partly because it has not been clear how they could arise through inhibition in V1. However, recent theoretical models have suggested that indirect effects could occur as early as V1 for simultaneous stimuli (Mundel et al., 1997), and RF-LISSOM demonstrates that a quite simple, local mechanism in V1 is also sufficient to produce indirect aftereffects: If the total synaptic strength at each neuron is limited, bolstering the lateral inhibitory connections between active neurons eventually weakens their inactive inhibitory connections, causing the indirect aftereffect.

Such weight normalization is computationally necessary: weights governed by a pure Hebbian rule would otherwise increase indefinitely, or would have to reach a fixed maximum strength (Rochester et al., 1956; von der Malsburg, 1973; Miller & MacKay, 1994), and neither of these outcomes is biologically plausible. Furthermore, very recent experimental results demonstrate that several processes of normalization are actually occurring in visual cortex neurons near the time scales at which the indirect effect is seen (Turrigiano et al., 1994, 1998; Turrigiano, 1999). Such regulation changes the efficacy of each synapse while keeping their relative weights constant (Turrigiano et al., 1998), as in the RF-LISSOM model. The current RF-LISSOM results predict that similar whole-cell regulation of total synaptic strength underlies the indirect tilt aftereffect in V1.

The RF-LISSOM results may also help explain why there is relatively large inter-subject variability in the shape and magnitude of the angular function of the indirect effect (Mitchell & Muir, 1976). Although the overall fit in figure 4 is quite good, there is a small discrepancy at very large angles. This may be an interesting artifact of the visual environment of modern humans. As humans orient themselves with respect to manmade objects such as books, pictures, etc., they will often see angles near 90°, whereas the model was trained with only single lines. If the model were trained on more realistic visual inputs that included corners, crosses, and approximately orthogonal angles, it would develop more connections between neurons with orthogonal preferences. These connections would have an inhibitory effect, and would decrease through normalization during adaptation. As a result, the model would show an increased indirect effect near $\pm 90°$, matching the human data. The prediction is, therefore, that the indirect effect differences between subjects arise from differences in their long-term visual experience, due to factors such as attention and different environments.

Through mechanisms similar to those causing the TAE, the RF-LISSOM model should also be able to explain tilt illusions between two stimuli presented simultaneously. Such an explanation was originally proposed by Carpenter and Blakemore (1973), and the principles have been demonstrated recently in an abstract model of orientation (Mundel et al., 1997). Testing this hypothesis in RF-LISSOM for spatially-separated stimuli would require a much larger inhibitory connection radius than used for these experiments; such simulations are not yet practical because of computational constraints. Bayesian analysis may offer a way to extract the response to two simultaneously-presented overlapping lines (Zemel et al., 1998), which could allow the tilt illusion to be measured in the current model. The tilt illusion can also be tested with existing computing hardware by first combining the current orientation map simulation with an ocular dominance simulation (such as that of Sirosh & Miikkulainen, 1997). Then perceived orientations can be computed for inputs in different eyes; if the tilt illusion is occurring then the perceived orientation for an input in one eye will change when a different orientation is presented to the other eye (as found in humans, Carpenter & Blakemore, 1973).

In addition, many similar phenomena such as aftereffects of curvature, motion, spatial frequency, size, position, and color have been documented in humans (Barlow, 1990). Since specific detectors for most of these features have been found in the cortex, RF-LISSOM should be able to account for them by the same process of decorrelation mediated by self-organizing lateral connections.

5 Conclusion

The experiments reported here lend strong computational support to the theory that tilt aftereffects result from Hebbian adaptation of the 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. The same model should also apply to other aftereffects and to simultaneous tilt illusions.

Computational models such as RF-LISSOM can demonstrate many visual phenomena in high detail that are difficult to measure experimentally, thus presenting a view of the cortex that is otherwise not available. This type of analysis can provide an essential complement to both high-level theories and to experimental work with humans and animals, significantly contributing to our understanding of the cortex.

A Acknowledgments

Thanks to Joseph Sirosh for supplying the original RF-LISSOM code. This research was supported in part by the National Science Foundation under grants IRI-9309273 and IIS-9811478. Computer time for the simulations was provided by the Texas Advanced Computing Center at the University of Texas at Austin, and by the Pittsburgh Supercomputing Center under grant IRI-940004P.

Software for the RF-LISSOM model and demos of the data presented in the paper are available at `http://www.cs.utexas.edu/users/nn/`.

References

Anderson, J. A., & Rosenfeld, E. (Eds.) (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

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

Bednar, J. A. (1997). *Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex*. Master's thesis, Department of Computer Sciences, The University of Texas at Austin. Technical Report AI97-259.

Bednar, J. A., & Miikkulainen, R. (1998). Pattern-generator-driven development in self-organizing models. In Bower, J. M. (Ed.), *Computational Neuroscience: Trends in Research, 1998* (pp. 317–323). New York: Plenum.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, *17* (6), 2112–2127.

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

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

Choe, Y., & Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, *21*, 139–157.

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

Dong, D. W. (1995). Associative decorrelation dynamics: A theory of self-organization and optimization in feedback networks. In Tesauro, G., Touretzky, D. S., & Leen, T. K. (Eds.), *Advances in Neural Information Processing Systems 7* (pp. 925–932). Cambridge, MA: MIT Press.

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

Finlayson, P. G., & Cynader, M. S. (1995). Synaptic depression in visual cortex tissue slices: An in vitro model for cortical neuron adaptation. *Experimental Brain Research*, *106* (1), 145–155.

Földiák, P. (1990). Forming sparse representations by local anti–Hebbian learning. *Biological Cybernetics*, *64*, 165–170.

Fregnac, Y. (1996). Dynamics of functional connectivity in visual cortical networks: An overview. *Journal of Physiology (Paris)*, *90*, 113–139.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience*, *14* (5), 897–919.

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

Gilbert, C. D. (1998). Adult cortical dynamics. *Physiological Reviews*, *78* (2), 467–485.

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

Grinvald, A., Lieke, E. E., Frostig, R. D., & 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.

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

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

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

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

McLean, J., & Palmer, L. A. (1996). Contrast adaptation and excitatory amino acid receptors in cat striate cortex. *Visual Neuroscience*, *13* (6), 1069–1087.

Miikkulainen, R., Bednar, J. A., Choe, Y., & 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., & Medin, D. L. (Eds.), *Perceptual Learning* (Vol. 36 of Psychology of Learning and Motivation, pp. 257–308). San Diego, CA: Academic Press.

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

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

Mundel, T., Dimitrov, A., & Cowan, J. D. (1997). Visual cortex circuitry and orientation tuning. In Mozer, M. C., Jordan, M. I., & Petsche, T. (Eds.), *Advances in Neural Information Processing Systems 9* (pp. 887–893). Cambridge, MA: MIT Press.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience*, *21*, 227–277.

Rochester, N., Holland, J. H., Haibt, L. H., & 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 Anderson & Rosenfeld, 1988.

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

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

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

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

Sirosh, J., & 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., & 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., & Choe, Y. (Eds.), *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.

Snippe, H. P. (1996). Parameter extraction from population codes: A critical assessment. *Neural Computation*, *8* (3), 511–529.

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

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

Turrigiano, G., Abbott, L. F., & Marder, E. (1994). Activity-dependent changes in the intrinsic properties of cultured neurons. *Science*, *264*, 974–977.

Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends in Neurosciences*, *22* (5), 221–227.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, *391*, 845–846.

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

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

von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, *15*, 85–100. Reprinted in Anderson & Rosenfeld, 1988.

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

Weliky, M., Kandler, K., Fitzpatrick, D., & 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., & Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. *Vision Research*, *28*, 301–312.

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

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10* (2), 403–430.

Zucker, R. S. (1989). Short-term synaptic plasticity. *Annual Review of Neuroscience*, *12*, 13–31.