LETTER Communicated by Friedrich Sommer

A Sparse Generative Model of V1 Simple Cells with Intrinsic Plasticity

Cornelius Weber
c.weber@fias.uni-frankfurt.de
Jochen Triesch
j.triesch@fias.uni-frankfurt.de
Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Frankfurt am Main 60438, Germany

Current models for learning feature detectors work on two timescales: on a fast timescale, the internal neurons' activations adapt to the current stimulus; on a slow timescale, the weights adapt to the statistics of the set of stimuli. Here we explore the adaptation of a neuron's intrinsic excitability, termed intrinsic plasticity, which occurs on a separate timescale. Here, a neuron maintains homeostasis of an exponentially distributed firing rate in a dynamic environment. We exploit this in the context of a generative model to impose sparse coding. With natural image input, localized edge detectors emerge as models of V1 simple cells. An intermediate timescale for the intrinsic plasticity parameters allows modeling aftereffects. In the tilt aftereffect, after a viewer adapts to a grid of a certain orientation, grids of a nearby orientation will be perceived as tilted away from the adapted orientation. Our results show that adapting the neurons' gain-parameter but not the threshold-parameter accounts for this effect. It occurs because neurons coding for the adapting stimulus attenuate their gain, while others increase it. Despite its simplicity and low maintenance, the intrinsic plasticity model accounts for more experimental details than previous models without this mechanism.

1 Introduction

Primary visual cortex V1 is one of the best-studied brain areas because it is easily accessible at the brain surface in many mammalian species, and it represents information that can be easily visualized. The most prominent response property of V1 simple cells is that they respond to localized edges in the visual field, their receptive fields characterizable as Gabor filters (for a review of V1 cell properties, see, e.g., Hirsch & Martinez, 2006). During early development, the filter characteristics are strongly susceptible to learning (Sengpiel, Stawinski, & Bonhoeffer, 1999). On a timescale of several seconds, however, filter characteristics can vary by the neurons adapting to the incoming stimuli (Blakemore & Campbell, 1969). This can improve

discriminability of visual patterns by reducing correlations (Müller, Metha, Krauskopf, & Lennie, 1999; Wainwright, 1999), and leads to visual aftereffects.

In this letter, we present a new generative model describing adaptation and learning at these different timescales. Its principal innovation is the utilization of an intrinsic plasticity mechanism rooted in information theory (Triesch, 2007). Here we show how the fast-adapting neuronal variables of this mechanism allow modeling of aftereffects and help learning of V1 simple-cell-like receptive fields.

1.1 Models for the Development of Edge Detectors. Learning edge detectors from natural images became popular with sparse coding generative models (Olshausen & Field, 1997) and independent component analysis (ICA) models (Bell & Sejnowski, 1996), both of which are equivalent under certain assumptions (Olshausen & Field, 1997). ICA is mostly performed with PCA-based preprocessing for dimensionality reduction based on principal component analysis (PCA) and using a fast algorithm version (Hyvärinen, Karhunen, & Oja, 2001). These models assume that V1 neurons are sparsely active and hence must code an image in terms of features that are sparsely distributed in images. Learning optimal image reconstructions under this constraint leads to neurons becoming localized edge detectors in accordance with V1 simple cells. A recent variant, the sparse set coding network (Rehn & Sommer, 2007), eliminates small activations and thereby obtains a minority of circular receptive fields, as found in macaque.

In a generative model, the input data will be reconstructed from the hidden code (latent variables) using top-down feedback connections. The error between this reconstruction and the data is the main learning signal for these feedback connections. Recent evidence indicates that V1 feedback to lateral geniculate nucleus (LGN) is inhibitory in those LGN regions from which the corresponding V1 cell receives its input, and excitatory to the opposite (ON versus OFF) LGN layer (Wang, Jones, Andolina, Salt, & Sillito, 2006), supporting the negative feedback used for learning by the generative models.

The most commonly used models (Olshausen & Field, 1997; Rao & Ballard, 1997) use the reconstruction error in the input to reestimate and refine the V1 code. They learn the feedback projections and obtain the forward, LGN-V1, connections by matrix transposition or, in the case of ICA, by a nonlocal learning rule (Bell & Sejnowski, 1996). The wake-sleep algorithm (Hinton, Dayan, Frey, & Neal, 1995), which we have already used to model V1 simple cells (Weber, 2001), involves no iterative computations of the V1 code. Feedback and forward connections are both learned by local learning rules in separate phases, as we describe below.

A generative model has the advantage that as one neuron generates a specific feature of the data stimulus, other neurons need only to consider other, nonexplained, features. This happens because only the difference between the data and its reconstruction contributes to learning. Hereby the

neurons' firing becomes decorrelated, assuming that the latent variables are uncorrelated in their prior. Hence, no inhibitory lateral interaction between the hidden neurons is necessary for feature learning.

Another class of models enforces sparse coding via lateral weights (Földiák, 1990; Bednar & Miikkulainen, 2000), which decorrelate the neurons' firing. These models do not use feedback connections to the input layer. They produce localized edge detectors with larger receptive fields than those produced by sparse generative models (Falconbridge, Stamps, & Badcock, 2006).

1.2 Models for the Tilt Aftereffect. In the tilt aftereffect (TAE), the perceived orientation of a test grating is altered. For instance, after adapting to a vertical grating, slightly off-vertical test gratings are perceived as being more strongly tilted. This effect arises after several seconds; its magnitude rises logarithmically in time, until it saturates after approximately 1 hour of adaptation (Greenlee & Magnussen, 1987). In these experiments, the maximally perceived orientation error was approximately 4 degrees at a test orientation 12 degrees away from the adapting orientation. The effect decays with the same rate constant as it arises. Orientations near 90 degrees away from the adapting orientation are perceived tilted toward the adapting stimulus's orientation. This weaker effect is termed the indirect TAE.

The simplest explanation of the tilt aftereffect is that the V1 simple cells that perceive the adapting stimulus selectively reduce their excitability. This involves reducing their mean membrane potential (Carandini, 2000). In addition, there have been found response increases away from the adapting orientation (Dragoi, Sharma, & Sur, 2000). A test stimulus that is tilted slightly away from the adapting orientation will excite cells at the adapting orientation as well as cells farther away. With cells at the adapting orientation being less excitable and cells farther away having stronger responses, the perception of the orientation of the test stimulus will be biased away from the adapting orientation.

In order to find out how adaptation is implemented in the brain, it has been studied how varying the neuronal model parameters influences the shape and size of the TAE curve. For example, Wilson and Humanski (1993) adapt one parameter per neuron, which increases divisive feedback to the input units. This is done to match an experiment that shows increased thresholds for judging contrast increments, and to explain the TAE experiment for a given adaptation time. Clifford, Wenderoth, and Spehar (2000) allow only two operations that resemble additive and multiplicative changes of the tuning curves and model successfully the TAE, tilt illusion (TI), and motion-related direction aftereffect (DAE) as well as a color aftereffect. Jin, Dragoi, Sur, and Seung (2005) mold neuronal tuning curves by scaling, changing the widths, and shifting. They conclude that repulsive shifts of preferred orientations (an effect of lateral interactions) are important for achieving accurate predictions of the TAE.

Models that aim at maximizing information transmission show that aftereffects can be modeled by adjusting a local parameter such as the variance of the hidden units' responses (Dayan, Sahani, & Deback, 2002) or a response envelope function over the hidden units (Wainwright, 1999). Schwabe and Obermayer (2006) investigate the adaptation of different parameters and find that optimizing the values of the neuronal gain accounts best for experimental data on attentional modulation, while optimizing the strength of recurrent connections explains changes observed in perceptual learning experiments.

In the LISSOM model (Bednar & Miikkulainen, 2000), lateral inhibitory connections gain strength by fast Hebbian learning during adaptation. Hence, neurons that code for the adapting orientation increase their mutual inhibition, thereby attenuating their firing and causing the TAE. Furthermore, the indirect TAE arises because as a consequence of inhibitory weight normalization, neurons coding for the adapting stimulus are less inhibited from the far orientations that are tested in the indirect TAE. This induces a bias toward the adapting orientation. In another study (Ciroux, 2006), the LISSOM model also explains the McCollough effect, in which a color is negatively associated with an orientation (McCollough, 1965). However, this aftereffect naturally decays only after hours or days, much slower than it emerges, unlike the TAE. Therefore, a different mechanism is likely to be involved.

1.3 Motivation for New Model. Recently we have suggested that a neuron adjusts its transfer function parameters, slope and threshold, to maintain its firing rate in a sparse exponential regime (Triesch, 2007). It is important that a neuron can respond to perturbations in a homeostatic manner (Davis & Bezprozvanny, 2001), and we propose this mechanism for firing rate homeostasis. Since our neuron was shown to discover heavy-tailed directions in the input (Triesch, 2007), we wish to extend the model to multiple neurons. After having solved the bars test in initial explorations (Butko & Triesch, 2007), we show here that intrinsic plasticity can enforce sparse coding in the hidden neurons of a generative model, enabling the learning of realistic feature detectors. Since the intrinsic plasticity parameters adapt faster than the weights but slower than the activations, we investigate short-term adaptation. We show that the model can explain the TAE with local neuronal adaptation parameters, without the use of lateral interactions. Hence it successfully integrates the fields of unsupervised learning and explanation of physiological data in terms of optimal coding on several timescales.

2 Methods

Our model for the development of V1 simple cells is a sparse coding Helmholtz machine that was previously shown to produce localized edge

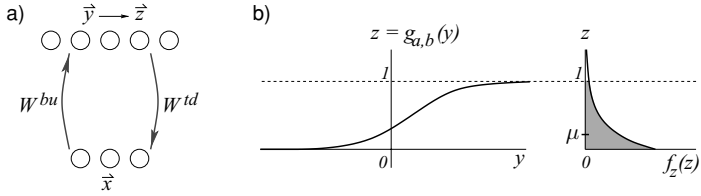

Figure 1: Model architecture and variables. (a) The input x is conveyed by the bottom-up weights W^{bu} to the "net input" y of the hidden neurons. The input can be reconstructed from the hidden neurons' output z by the top-down weights W^{td}. (b) Left: The transfer function $g_{a,b}(y)$ computes z given y and depends on the parameters a and b. Right: The sparseness of the exponential prior density function $f_z(z)$ on z is parameterized by its mean μ.

detectors (Weber, 2001). The first essential property is that it is a generative model; hence, during learning, the model is tested on whether it can reconstruct the incoming data from its internal representation via feedback weights. The wake-sleep algorithm (Hinton et al., 1995) supplies local learning rules for both the feedback and feedforward weights and is an approximation to the generative Helmholtz machine model. The second essential model ingredient is sparseness of the internal representation. Following Triesch (2007) we enforce sparseness by modifying the simple cells' transfer function parameters so that the firing rate distribution during presentation of the data approximates a decaying exponential function.

Hence, there are three groups of variables and associated timescales being adjusted: (1) neuronal activations adapt completely to a current stimulus and have no memory; (2) transfer function parameters of the hidden neurons adapt fairly fast but need to sample several occurrences of data points in order to estimate and adjust the firing rate statistics to be sparse; and (3) neuronal weights analyze the statistics of the entire data set and learn slowly.

2.1 Architecture and the Neuron Model. The model architecture and its variables are shown in Figure 1. An input data point $\vec{x}$ is conveyed to the net activation $\vec{y}$ of the hidden units by the bottom-up weight matrix W^{bu}. For hidden unit i and inputs j, this gives

$$y_i = \sum_j w_{ij}^{bu} x_j. \tag{2.1}$$

The logistic transfer function then conveys y_i to a neuronal output z_i:

$$z_i = g_{a,b}(y_i) = \frac{1}{1 + \exp(-(a_i y_i + b_i))}. \tag{2.2}$$

The transfer function has two modifiable parameters: a scales the input and resembles a gain, or slope. b shifts the curve and resembles a threshold, because it determines the point at which the output starts to rise sharply. The weights are normalized so that $\sum_j (w_{ij}^{bu})^2 = 1$ for each hidden neuron i.

2.2 Sparseness. Given the input distribution over $\vec{x}$ and relatively slow changing weights W^{bu}, the transfer function parameters are adjusted to keep the neuronal output z_i of each hidden neuron individually in an approximately exponential regime, as indicated on the right of Figure 1. This exponential function

$$f_{\exp}(z_i) = \frac{1}{\mu} e^{-\frac{z_i}{\mu}} \tag{2.3}$$

is parameterized by μ, which determines the mean for an exponential defined over the positive half-axis. Exponential spike count distributions were found in V1 responses to video in anesthetized cats (Baddeley, 1997). They argue that an exponential rate distribution minimizes metabolic consumption because it maximizes the entropy for a given mean firing rate. As long as μ is set well below 0.5, we can assume that the cutoff at 1 by the transfer function does not change the shape of the exponential relevantly, and sparse firing is guaranteed by the shape of the exponential, which has a Fisher kurtosis of 6. It was argued that distributions with kurtosis larger than zero are suited for ICA (Bell & Sejnowski, 1996) and that distributions with large kurtosis produce a sparse code (Olshausen & Field, 1997).

We adjust the parameters a and b to minimize the Kullback-Leibler divergence $d(f_z \| f_{\exp})$ between the hidden neuron's firing rate distribution $f_z(z_i)$, which depends on a and b, and the desired distribution $f_{\exp}(z_i)$. As we have shown previously (Triesch, 2007), gradient descent yields incremental update rules:

$$\Delta a_i = -\eta_a \frac{\partial}{\partial a_i} d(f_z(z_i) \| f_{\exp}(z_i)) = \eta_a \left(\frac{1}{a_i} + y_i - 2 y_i z_i - \frac{1}{\mu} y_i z_i + \frac{1}{\mu} y_i z_i^2 \right) \tag{2.4}$$

$$\Delta b_i = -\eta_b \frac{\partial}{\partial b_i} d(f_z(z_i) \| f_{\exp}(z_i)) = \eta_b \left(1 - 2 z_i - \frac{1}{\mu} z_i + \frac{1}{\mu} z_i^2 \right). \tag{2.5}$$

2.3 Wake-Sleep Algorithm. The wake-sleep algorithm learns the generative and recognition weights in two separate, alternating phases. First, we consider only learning of the generative weights W^{td} from the data in a so-called wake phase. After we obtain the hidden code $\vec{z}$ using equations 2.1 and 2.2, the network reconstructs the data, and one obtains the reconstruction error,

$$\hat{x} = \vec{x} - W^{td} \vec{z} \, . \tag{2.6}$$

An error-based Hebbian-like learning rule,

$$\Delta w_{ji}^{td} = \eta_w^{td} \hat{x}_j z_i, \tag{2.7}$$

adjusts W^{td} to minimize the reconstruction error, equation 2.6. η_w^{td} is the learning step size. Unlike the recognition weights, the generative weights W^{td} are not normalized. They scale automatically to minimize the error, equation 2.6, and there are no transfer function parameters on the linear input units that would adapt.

The recognition weights W^{bu} are trained to functionally invert the generative weights in the sleep phase in which no data are present. For this purpose, "fantasy" vectors $\tilde{z}$ are presented on the hidden units. These are projected as $\tilde{x} = W^{td}\tilde{z}$ to the input units, and the recognition weights perform the functional inversion $W^{bu}\tilde{x}$ by attempting to reconstruct the fantasy vectors. The reconstruction error is

$$\hat{z} = \tilde{z} - g_{a,b}(W^{bu}\tilde{x}), \tag{2.8}$$

using the transfer function, equation 2.2. Again, an error-based Hebbian-like learning rule,

$$\Delta w_{ij}^{bu} = \eta_w^{bu} \hat{z}_i \tilde{x}_j, \tag{2.9}$$

adjusts W^{bu} to minimize the reconstruction error, equation 2.8.

The fantasy vectors used in the sleep phase are sparse random activation patterns with activation values drawn from the exponential distribution, equation 2.3, independently for each unit.

The generative model $W^{td}\vec{z}$ is a linear ICA model that generates the data from sparse, overcomplete latent variables. In contrast, the recognition model $g_{a,b}(W^{bu}\vec{x})$ involves a nonlinear transfer function. Hence, both models cannot invert each other exactly, so the recognition model only approximates inference of the latent variables from the data. The nonlinearity is required as part of the intrinsic plasticity mechanism to create the sparse latent variables from which the linear generative model is learned. Without the sparsifying nonlinearity of the elsewise linear recognition model, the linear generative model would not learn sparse features. So if one incorporated on each hidden neuron the inverse of the nonlinearity into the generative model by using $W^{td}g_{a,b}^{-1}(\vec{z})$, then the generative model could invert the recognition model exactly, but that would also cancel the sparsifying effect of the nonlinearity.

Since the IP parameters a and b take part in transforming the $\vec{y}$ into the $\vec{z}$ latent variables, they may be regarded as part of the recognition model. On the other hand, we adjust them in the wake phase when the generative weights are learned. We do this because a and b must adapt to

the data distribution that is available only in the wake phase. They assist in maximizing the likelihood of the generative model in the weight update, equation 2.7, by keeping the latent variable distribution stationary over the course of learning.

2.4 Images. Patches were taken from 62 gray-scale images of natural scenes of which 14 included some artifacts such as buildings. Preprocessing consisted of filtering the images with a low-pass filter that is described in frequency space as $f \cdot \exp(-(\frac{f}{f_0})^4)$, with $f_0 = 200$ cycles per picture, as described in Olshausen and Field (1997). By the factor f, which attenuates small frequencies, this filter eliminates large luminance changes and counterbalances the $1/f^2$ frequency spectrum of natural images ("whitening"). By its exponential term, which attenuates very high frequencies, it eliminates noise and sampling artifacts. Olshausen and Field refer to comparable response characteristics of retinal ganglion cells. For each filtered image patch $\vec{x}$ shown on the retina, the mean was subtracted.

2.5 Parameters. The mean activation was set to $\mu = 0.01$. The weight learning rates were $\eta_w^{td} = 0.02$ and $\eta_w^{bu} = 0.1$. The reason to make η_w^{bu} larger than η_w^{td} is to compensate for smaller average absolute values of the learning terms $|\hat{z}_i|$ and $|\tilde{x}_j|$ in equation 2.9 compared to $|\hat{x}_j|$ and $|z_i|$ in equation 2.7.

The intrinsic plasticity parameters used in the simulations were $\eta_a = \eta_b = 0.001$. Learning is robust against their value. This is because the adaptation of a and b to one image patch is done after its processing and thus has an effect only on the next, randomly selected, image patch.

The network had 16×16 input units, matching the size of the gray-scale image patches, and the hidden layer consisted of 24×24 units. The resulting weights were obtained after 15 million learning steps. Each learning step consisted of a wake phase at which an image patch was presented, and computation of equations 2.1, 2.2, 2.6, and 2.7. Then equations 2.4 and 2.5 were computed in the wake phase. Then followed at each learning step a sleep phase for learning the recognition model via creation of a fantasy vector and computation of equations 2.8 and 2.9.

2.6 Receptive Field Analysis. In order to quantify the receptive fields, they were fit with Gabor functions on an elliptical gaussian support. These Gabors had eight continuous parameters, as described in Ringach (2002): the orientation, frequency f, and phase of the cosine wave function; the position of the gaussian envelope along dimensions 1 and 2 of the input area, its height, and its standard deviations σ_x, σ_y. The half-axis along σ_x was assumed parallel to the cosine wave; hence, σ_y was along the wave ridges, or parallel to the edge that the edge detector would respond to. We minimized the squared difference between a receptive field (i.e., every weight value) and its Gabor reconstruction using the simplex method (Press, 1988). In

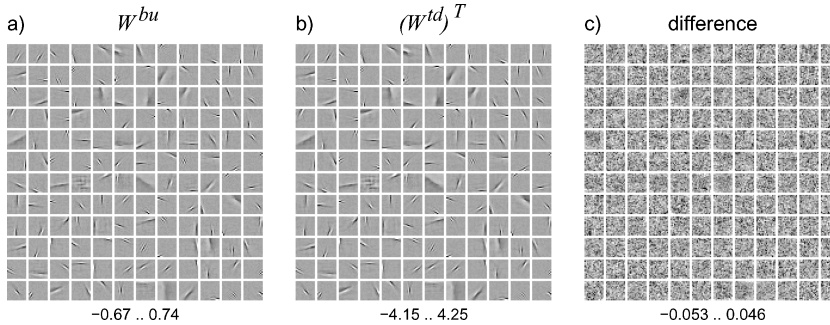

Figure 2: Trained weights. (a) Receptive fields. Each small square represents a row of W^{bu} with positive weights shown bright and negative weights dark. (b) Projective fields: columns of W^{td}, or rows of $(W^{td})^T$. (c) The difference between the weights shown in a minus those shown in b after normalizing the columns of W^{td}, just as done to the rows of W^{bu} during learning. The numbers below show the minimum and maximum occurring values for each weight matrix. A quarter of all weights are shown.

order to prevent receptive field center fits to be outside the input area, the fitting was performed on a larger area with zero padding around the original input area. All eight parameters were optimized simultaneously by the simplex method. Since this method easily gets stuck in a local minimum, 16 automated attempts from different initial positions for each receptive field were made, and the best fit was taken. Neurons that had a bad fit (55 of 576) due to their irregular shape or to the minimization getting stuck in a local minimum were excluded from the parts of the analysis that depended on the Gabor fits.

Gabor fitting may result in a tendency to return spatial frequencies that are too low. First, the discarded "noisy" receptive fields are mainly of high frequencies. Second, as a general effect of Gabor fitting, the wavelength of the fitted cosine can be much larger than the receptive field, because the gaussian envelope will attenuate the surround, allowing the cosine to have large values outside the receptive field. The wavelength obtained by a Gabor fit can therefore be much larger than, for example, the dominant wavelength that would be obtained from a Fourier transform.

3 Results

3.1 Emergence of Feature Detectors. A quarter of all hidden neurons' receptive fields (RF), randomly selected, are shown in Figure 2a. Most of them are localized edge detectors. Figure 2b shows the top-down feedback weights that correspond to the hidden neurons' projective fields toward the input neurons. These weights have the same shape as the RF weights with

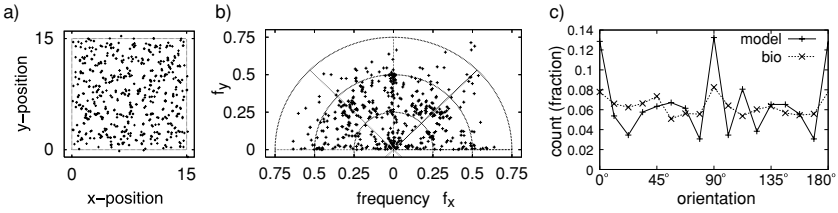

Figure 3: (a) Receptive field centers (each cross denotes one RF center) are distributed evenly over the input area. (b) Distribution of the filters in orientation-frequency space (orientation denoted by angle and frequency by radius as 1/(pixel unit)). (c) Distribution of the orientation preference of model neurons (solid line) and cat V1 simple cells (dotted line). Data from cat are from Figure 2 of B. Li, Peterson, and Freeman (2003).

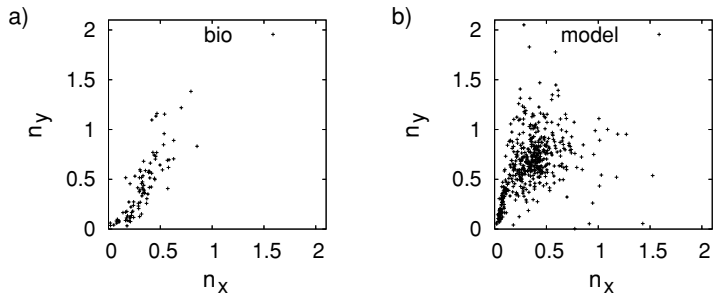

Figure 4: RF shape distributions. (a) Macaque data from Ringach (2002). (b) Our model. The axes express the RF length along the wave (n_x) and width (n_y) per wavelength (see text for details). The model RFs display a large variety of shapes.

a small difference of random structure remaining, as shown in Figure 2c. Two principal differences remain. First, the feedback weights are larger here than the feedforward weights. This is because, unlike the feedforward weights, which are normalized, the feedback weights have to account for the variance of the data (in our case in the order of 1), which they reconstruct (see equation 2.6). Smaller input data would have yielded smaller feedback weights (while increasing in the forward pathway the gain parameters $\{a\}$). Second, the feedback weights should be regarded as sign-reversed, constituting the negative part of the reconstruction error, equation 2.6. The experimental results of Wang et al. (2006) support the notion of a V1 to LGN feedback structure that matches the geometry of the LGN to V1 connections with a reversed sign.

The receptive fields (RFs) are analyzed in Figures 3 and 4. Their centers are approximately evenly distributed over the input plane, as seen in Figure 3a. Receptive field orientations and spatial frequencies vary

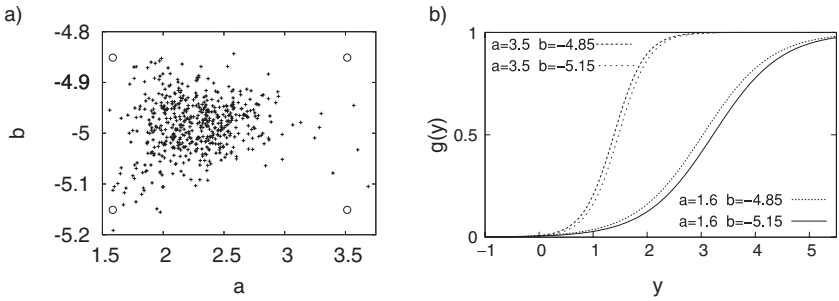

Figure 5: (a) Scatter plots of the occurring a and b values. (b) The transfer function for two different a and b parameter values and their possible combinations. The selected a and b values are near the occurring extremes and are indicated in (a) by circles.

substantially, as can be seen in Figure 3b, while a stronger presence of the cardinal orientations at 0 and 90 degrees is evident in Figure 3c (solid line). Data from 2598 cat simple cells (B. Li, Peterson, & Freeman, 2003) (dotted line) show that there are also more cells coding for the cardinal orientations. Observations that cardinal orientations are physiologically more finely resolved than oblique orientations have led to the term *oblique effect*. The model, however, displays an exaggerated preference to cardinal orientations together with a reduction at nearby orientations (± 11.25 degrees away in Figure 3c). We conjecture that an alignment of high-frequency edge detectors with the image pixel rows is a reason for this reduction of slightly slanted orientations. Figure 3b supports this view in that we can see cell clusters at the cardinal orientations near a high frequency of 0.5 cycles per pixel.

The scatter plots in Figure 4 compare the shapes of model RFs with experimental data from 93 macaque monkey neurons (Ringach, 2002). Given the wavelength of the cosine $T = 1/f$, he defines $n_x = \sigma_x/T$ and $n_y = \sigma_y/T$, which give the RF elongations in multiples of the wavelength along the wave progression and along the wave ridges, respectively. In the model data in Figure 4b, there is a tendency toward larger n_y than n_x values, compared with the experimental data in Figure 4a. Hence, the model RFs tend to be more elongated along the edges that the neurons respond to. The largest discrepancy to macaque data is the near lack of round, nonoriented RFs, as they appear near the origin of Figure 4a. Currently, only the model of Rehn and Sommer (2007) also predicts these shapes, in which sparse-set coding limits the fraction of active units. Our exponential prior distribution defines sparse activations but not a sparse set.

We find in Figure 5a that the a values extend more than twofold, while the b values are within a small interval less than $\pm 4\%$ around their average. Figure 5b visualizes this, showing that slope variations of the transfer function are larger than variations of its offset. We believe that the reason for the

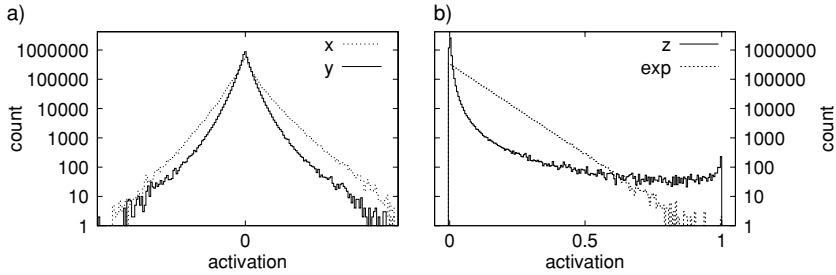

Figure 6: Histograms of log counts of model variables x_j, y_i, and z_i for randomly chosen units j and i. (a) Both x_j and y_i are approximately distributed like a double exponential (Laplacian), as evident from the approximate linear histogram decrease away from zero. The y distribution (kurtosis = 12.7) has a stronger upward bend than the x distribution (kurtosis = 6.8). (b) The latent unit output z_i in comparison with data sampled from an exponential distribution. Given that y_i is approximately exponentially distributed, the logistic transfer function makes z_i deviate from an exponential.

b parameter to be so tightly constrained is that varying b but not a changes $g_{a,b}(y = 0)$. The peak of the incoming data is at $y = 0$, and approximately half of the inputs y to a neuron are negative (cf. Figure 6a). Hence, varying $g_{a,b}(y = 0)$ by varying b changes the mean of the data significantly, while varying a does much less so. Furthermore, increasing b would involve the risk that a previously quiet neuron can start to fire.

3.2 Variable Distributions. We found that the hidden variables $\{z\}$ do not perfectly approximate an exponential distribution. Figure 6a shows that a pixel value x (after filtering) approximates a double-exponential distribution. So does the linear net input y of a hidden unit, but it has a heavier tail than the pixel distribution, as we expect from the edge-filtering operation by the weights W^{bu}. Feeding the y through the logistic transfer function causes the z distribution in Figure 6b to deviate largely from an exponential, in being distributed even more sparsely. Negative y are collapsed to near-zero values smaller than $g_{a,b}(y = 0)$. Positive y are stretched by the upward bend of the logistic transfer function for $z < 0.5$, decreasing smaller inputs more than larger inputs. In effect, z tends to consist of a small number of large values and a large number of near-zero values. Hence, the transfer function renders the latent variables $\{z\}$ sparse in that IP manages to keep them in a range where the transfer function has an upward bend.

If the z variables were to be exponentially distributed like the x variables, then a trivial one-to-one connection, together with a transfer function in a linear regime, would optimally result from learning to produce this exponential z-distribution. Instead, our model reconstructs the image from the

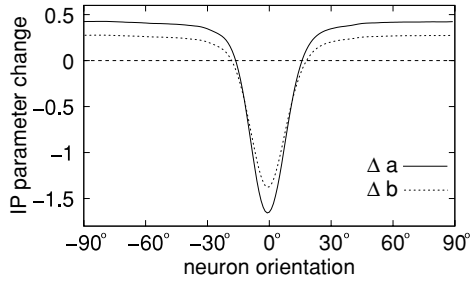

Figure 7: The changes of a (solid line) and b (dashed line) during adaptation, dependent on the orientation preference of the neurons. Zero on the x-axis means that the neurons respond most to the orientation of the adapting stimulus.

components of a much sparser z. Thereby it can extract components from the data that are more sparse than exponential.

A hierarchical visual system that extracts features across many levels might also face the problem that a higher level must code more sparsely than the next lower level, while sparseness cannot increase arbitrarily toward the highest levels. A speculative solution can be the layered structure of the cortex in which a feedback layer reconstructs the input using a sparsified code, while a feedforward layer projects a denser code to the next higher level. For example, in V1, the backprojecting layer 6 would code sparsely, in that its input from the LGN (Blasdel & Lund, 1983) could be weaker than the principal LGN input into layer 4, while the forward-projecting layers 2 and 3 would code more densely, for example, by pooling simple cell responses into less selective complex cell responses.

3.3 Pattern Adaptation. Previously we have shown that in response to sudden sensory deprivation, a neuron with intrinsic plasticity would adapt its parameters in order to maintain firing rate homeostasis (Triesch, 2007). Here we present a sinusoidal grating stimulus to the trained network. This continuously activates edge detectors that are responsive to the given orientation, while those perceiving other orientations will be deprived. This is done for several learning steps during which the phase of the adaptation grating is varied, so that all neurons coding for that specific orientation receive stimulation. The intrinsic plasticity parameters a and b are allowed to adapt as during training. The weight learning rule was not invoked during this short adaptation. Note that in the learning rule, only the feedback weights are learned during presentation of data anyway.

As a response, shown in Figure 7, the parameters a and b decrease strongly for neurons responsive to the adapting orientation and increase slightly away from the adapting orientation. The rate of adaptation can be scaled by η_a and η_b and is open to interpretation, since we have not made a

quantitative match to real time. One possibility is to account for homeostatic mechanisms on timescales of approximately 24 hours (Desai, Rutherford, & Turrigiano, 1999). Here we choose to interpret the neurons' adaptation to occur on a timescale of several seconds. On this timescale, Carandini, Movshon, and Ferster (1998) found an elevation of contrast threshold when a test grating matched an adapting orientation; the size of the effect dropped to half this maximal value for test gratings at orientations ± 8 degrees away from the adapting orientation. For test gratings beyond ± 20 degrees from the adapting location, contrast thresholds could decrease. These data match our observed decrease of gain- and negative-threshold parameters a and b, respectively, near the adapting orientation and the a and b increases farther away. This timescale is also relevant for the TAE.

3.4 Tilt Aftereffect. The tilt aftereffect experiment measures how short-term adaptation to an oriented grating stimulus changes the perceived orientation of test stimuli of differring orientation. To measure the orientation perceived by the network, we sum over the activation z_n multiplied by the orientation θ_n of each unit n that has been obtained from the Gabor fits (i.e., population vector decoding).[1]

After adaptation, which is done as above, test gratings are shown to the network in order to measure whether the perceived orientations have changed. Perceived orientations are measured for stimuli of all orientations varied in 1 degree steps from 0 to 180 degrees, and the measured orientations are compared to those measured before adaptation. The difference between these perceived orientations is plotted in Figure 8.

The results show that in the vicinity of the adapting stimulus (shown at 0 degrees), the perceived orientation of the test stimulus is tilted away from the orientation of the adaptation stimulus. All curves in Figure 8 result from performing 180 adaptation trials varying the orientation of the adaptation stimulus in 1 degree steps and averaging between these trials. This is done in order to decrease noise resulting from irregularities in the trained map and the scatter of a and b that results from their adaptation to the last few training patterns.

Figure 8a shows that the shape of the TAE as a function of orientation difference depends on whether a only, b only, or a and b are adapted. If b is involved in adaptation, then orientations far away from the adapting orientation are affected, yielding an unnaturally wide-reaching effect. Hence, the results are biologically plausible only if the b parameter does not change on a short timescale.

[1] In order to account for the 180 degree periodicity in the orientation, we multiply the units' orientations by two and perform the sum as a vector sum in two dimensions. This yields $z_{net}\, e^{2i\,\theta_{net}} = \sum_n z_n\, e^{2i\,\theta_n}$, where θ_{net} is the network orientation response. z_{net} is a response magnitude that does not interest us further.

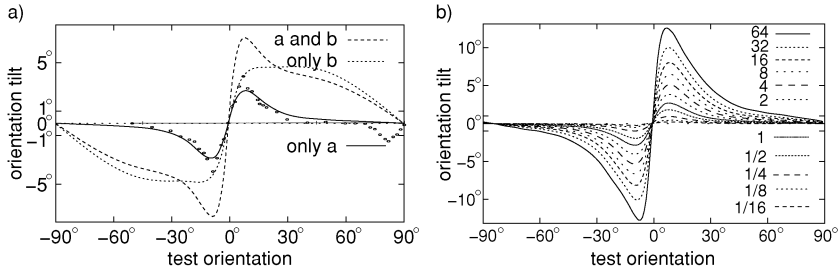

Figure 8: Tilt aftereffect. The x-axis denotes the difference between the orientation of test and adaptation pattern. The y-axis plots the tilt of the perceived orientation of the test pattern compared to preadaptation perception. It is perceived as tilted away from the orientation of the adaptation pattern. Every curve is an average over 180 experiments, each with a different adaptation orientation. (a) If both intrinsic plasticity parameters, a and b, are allowed to adapt (highest curve), then the model tilt aftereffect is much wider than experimental data (small circles from Campbell & Maffei, 1971). Similarly, if only b is varied (wide curve). However, if b is fixed during adaptation ($\eta_b = 0$), then the TAE curve (lower curve) matches the experimental data well; $32 \cdot 180$ adaptation steps were performed. (b) $\eta_b = 0$, and the adaptation duration is varied, from 180 to $256 \cdot 180$ adaptation steps, doubling from each curve to the next.

If we compare the TAE curve (when only a is adapted) to the decrease of a in Figure 7, we find that the maximum of the TAE is at the steepest change of a (judged by visual inspection). The values of both a and b adapt similarly (see Figure 7); despite this, adaptation of b differs in that it produces a much wider-ranging TAE (see Figure 8a). This difference in the TAE must therefore be explained in that first the transfer function reacts differently to changes of a and b, and then the pooling of the activations that is performed to compute the perceived orientation leads to different TAE curves.

In Figure 8b it can be seen that the model TAE (where only a is allowed to adapt) increases approximately logarithmically with longer adaptation time, as found experimentally (Greenlee & Magnussen, 1987): curve levels increase linearly while adaptation durations double from curve to curve. This increase does not saturate, unlike the case in humans (Greenlee & Magnussen, 1987).

There has been discussion about whether the TAE can be explained by neuronal fatigue (Carandini, 2000; Crowder et al., 2006; Dragoi et al., 2000). We tested a fatigue condition by allowing a only to decrease, not increase, in Figure 9 (solid curve). Conversely, we allowed a to increase, but not decrease, in Figure 9 (dashed curve). The decreases of a are near the adapting orientation (see Figure 7), and the maximum of the TAE is close to the adapting orientation—approximately 8 degrees away from it. The increases of a are away from the adapting orientation, and the maximum

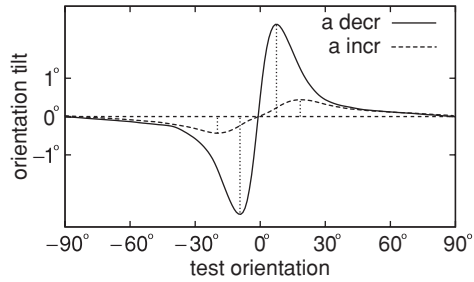

Figure 9: Tilt aftereffect when a is allowed only to decrease (solid curve) and when a is allowed only to increase (dashed curve). In both cases, b was held constant. The vertical dotted lines indicate the extrema of the curves, where the maximum of the TAE would occur.

of the TAE is also shifted farther away, approximately 18 degrees from the adapting orientation. Hence, quantifying the TAE distinguishes fatigue (gain decrease at the adapting orientation) from increased excitability (gain increase away from the adapting orientation). Experimental accounts of the adaptation and test orientation difference which yields the maximum TAE are at 8 degrees (Campbell & Maffei, 1971), 10 to 15 degrees (Wenderoth & Johnstone, 1988), 15 degrees measured with texture edges (Hawley & Keeble, 2006), 10 to 20 degrees (Paradiso, Shimojo, & Nakayama, 1989), or 20 degrees (Noudoost, Adibi, & Sanayei, 2003). We might explain the observed differences in that experimental conditions (e.g., contrast of adaptation- and test stimuli) might have favored a decrease of excitability leading to a TAE maximum at smaller orientation differences or an increase of excitability leading to a TAE maximum at larger orientation differences.

4 Discussion

4.1 Robustness of Model. Our model presents an attractive alternative to previous generative sparse coding and ICA models. It has local learning rules and accounts for the development of both LGN-V1 feedforward and feedback weights without the need of a matrix transposition or inversion. Allowing the neurons' transfer function parameters a and b to be modifiable relieves the modeler from exploring their optimal setting and allows the modeling of more experimental details than models without this mechanism can account for. This introduces three other parameters: the mean firing rate μ and the learning rates η_a, η_b for these two parameters. We find that setting a and b by hand is difficult, because the effect of a is dependent on the value of b, and vice versa. In contrast, setting μ is intuitive and directly related to sparseness. The learning rates can be chosen from within a few orders of magnitude to be effective.

4.2 Other Forms of Adaptation. Adaptation happens on different timescales. On timescales of approximately 24 hours, firing rates stabilize by homeostatic mechanisms (Desai et al., 1999). They exposed cultured cortical pyramidal neurons to activity blockade and found the initial slope of the *f-I* curve increased and the threshold reduced after 24 hours. These effects were not discernible after just 2.5 hours. Moulder, Jiang, Taylor, Olney, and Mennerick (2006) electrically stimulated hippocampal pyramidal neurons for several hours. This inactivated individual presynaptic glutamate terminals, revealing a slow, synapse specific form of homeostatic plasticity.

There is also adaptation on a very short timescale. Following an initial pulse in the white matter, Ramoa and Sur (1996) find suppression of the response of rat visual cortical neurons to a second pulse. Suppression was dominant less than 100 ms after the first pulse and was absent in young rats (<P10), which have not yet developed their inhibitory intracortical circuitry. This suggests inhibitory mechanisms for this fast timescale.

Fast effects of opposite sign also exist. Repetitive correlated spiking induces rapid long-term potentiation in two connected hippocampal culture neurons (Zhang & Linden, 2003). In another experiment, cortical cells' receptive fields could be expanded into previously unresponsive regions within minutes after associative costimulation (Eysel, Eyding, & Schweigart, 1998). Since these effects lasted for hours, synaptic plasticity might be the underlying mechanism for this effect. Sharpee et al. (2006) found that V1 neurons adapt their filters to stimulus statistics beyond the mean and variance. Since this happens over minutes and has a longer timescale than, for example, the TAE, they suggest that the underlying mechanisms may be distinct from those of other aftereffects. Future work should address whether other forms of adaptation can also be incorporated and explained by a generative model with information maximization.

4.3 Sites of Adaptation. There is early light adaptation based on local mean intensity in the retina and fast contrast gain control based on the root-mean-square contrast in the LGN (Bonin, Mante, & Carandini, 2006). Accordingly, LGN responses are well explained by dividing the receptive field input by a suppressive local surround field. Evidence that there is further cortical adaptation comes from the orientation and frequency specificity of adaptation (Blakemore & Campbell, 1969), whereas LGN cells are poorly tuned to orientation. LGN neurons, which respond only transiently to a stimulus, adapt less than cortical neurons do when reexposed to a stimulus (Ohzawa, Sclar, & Freeman, 1985). Furthermore, adaptation to oriented gratings and the TAE display interocular transfer (Campbell & Maffei, 1971), and binocular cells first come into play in V1. Pattern specific adaptation is uniquely cortical (Carandini et al., 1998).

In higher areas, response reduction following adaptation is reported and termed a familiarity effect (L. Li, Miller, & Desimone, 1993) or repetition suppression (James & Gauthier, 2006). In an event-related fMRI experiment,

after 20 seconds or longer, adaptation to an oriented pattern, the BOLD signal in V1, V2, V3/VP, V3A, and V4 to a test stimulus decreased at small angular difference between the adapting and test stimuli and increased at large angular differences (Fang, Murray, Kersten, & He, 2005). Together these findings hint at the existence of a mechanism, acting within seconds and minutes, which aims at keeping a neuron's firing rate at a constant level.

4.4 Mechanisms of Adaptation. A host of experiments investigate those mechanisms of adaptation that we think underlie the TAE. Prolonged viewing of high-contrast gratings raises the detection threshold of a low-contrast test grating that has similar orientation and frequency (Blakemore & Campbell, 1969). A substantial reduction in firing rate and a hyperpolarization is observed in V1 neurons after a few seconds of stimulation (Carandini et al., 1998). These changes are localized to the orientation and spatial frequency of the adapting stimulus. These adaptation effects appear gradually within a minute, and the decay time course is similar.

Adaptation to gratings of varying contrast is termed *contrast gain control*, since contrast response functions shift laterally along a log-contrast axis, implying gain adaptation (Ohzawa et al., 1985). In vivo (Carandini et al., 1998; Sanchez-Vives, Nowak, & McCormick, 2000) studies find a hyperpolarization of the membrane potential in V1 cortical cells, which may result from the activation of Ca^{2+} and Na^{+}-dependent K^{+} conductances. This hyperpolarization is large (5–10 mV) and long lasting (10–20 seconds) (Carandini, 2000). Hence, an intrinsic cell property contributes to the postadaptation suppression of activity. This hyperpolarization and a decrease in the spike response are elicited by either current injections or visual stimuli. Application of GABA, which prevents firing of the cell, does not reduce the degree of adaptation (Vidyasagar, 1990). Congruently, adaptation is not related to the spiking activity of the cells (Crowder et al., 2006). Consequently, one may prefer to interpret our model neuron variable z, which is used in the adaptation terms 2.4 and 2.5, not directly as a firing rate but rather as a signal, computed at the soma, that is proportional to the rate if firing is not suppressed by GABA.

As another possible mechanism, shunting inhibition may reduce neuronal gain, as shown in cerebellar cultured granule cells (Mitchell & Silver, 2003) and in neurons of cat primary visual cortex (Borg-Graham, Monier, & Frégnac, 1998). In a Hodgkin-Huxley-type model neuron, the gain can be selectively controlled by concerted changes of ionic currents, namely, the transient outward current I_A and the hyperpolarization-activated inward current I_H (Burdakov, 2005). Furthermore, active dendritic spike backpropagation may multiplicatively increase gain and be downregulated by dendritic inhibition to decrease gain (Mehaffey, Doiron, Maler, & Turner, 2005).

4.5 Network Effects. Our model neurons' gain changes locally in a single neuron. However, V1 neurons are interconnected by horizontal

connections, including inhibitory interneurons. This network may be an additional source of a neuron's adaptability.

Pharmacological studies involving neurotransmitters address neuronal interactions and thereby assess the network's contribution. For example, neuronal excitability was increased by depletion of serotonin, and the size of the TAE was found to be increased (Masini, Antonietti, & Moja, 1990). Yet it is unclear whether predominantly excitatory or inhibitory channels were facilitated. An influence of inhibition is implied in a study involving nomifensine, which potentiates dopaminergic transmission and enhances the TAE. Haloperidol, which blocks dopaminergic receptors, reduced the TAE (Gelbtuch, Calvert, Harris, & Phillipson, 1986).

By demonstrating the dependence of the TAE on neuronal interactions, these studies hint at a network influence on neuronal adaptability.

A network effect is particularly implicated in the tilt illusion (TI), in which two lines that cross at, say, 15 degrees seem to cross at a larger angle such as 17 degrees, but the mechanisms seem to be different; for example, lorazepam (which potentiates the activity of GABA) produced a dose-related increment in the size of the TI but had no effect on the TAE. By contrast, nomifensine and haloperidol did not affect the TI while affecting the TAE. In this study, Gelbtuch et al. (1986) suggest that the differential effects may reflect their differing actions on two processes: lateral inhibition (involved in the TI) and adaptation in visual channels (involved in the TAE). Our model neurons' local gain changes represent the latter process.

More direct evidence for a network influence on the TAE is that the orientation tuning curves of single neurons shift during adaptation (Dragoi, Sharma, Miller, & Sur, 2002). After adaptation, the initial response (56 ms after stimulus onset) to a test stimulus resembles the unadapted response, while over the following 24 ms, the tuning curve gradually shifts away from the adapting stimulus orientation.

Model studies also demonstrate network effects: balanced increases in background excitation and inhibition decrease gain multiplicatively by increasing current noise and conductance (Chance, Abbott, & Reyes, 2002). Murphy and Miller (2003) show that multiplicative gain changes arise robustly from the simple addition of excitation or inhibition alone, provided the modulating excitation or inhibition is small relative to the peak of the tuning curve of the driving excitation. Similar to divisive normalization that is based on neighboring neurons' responses (Schwartz & Simoncelli, 2001), such activation-based gain adaptation schemes do not take into account the stimulus history over tens of seconds, as does the adaptation of intrinsic plasticity parameters.

4.6 How Local Is Adaptation? Adaptation was found to be a response to the main region of a cell's receptive field from which excitatory discharges are elicited, but not to a stimulus at the surround (Ohzawa et al., 1985). Adaptation is furthermore similar in all cortical layers, V1 simple cells,

complex cells, and V2 cells (Crowder et al., 2006), despite the fact that there are more complex cells as well as stronger long-range lateral connectivity in supragranular layers. On the other hand, orientation-selective adaptation was most common in areas where orientation preference changes little over the cortical surface (Crowder et al., 2006). Together this indicates some influence from the vicinity on adaptation, but makes the influence of long-range lateral connections unlikely. In an architectural view, local inhibitory interneurons may be involved in adaptation, but not the long-range horizontal connections in layers 2/3 of V1.

Recently it emerged that cross-orientation suppression, an effect commonly attributed to horizontal V1 connections, can be accounted for by the behavior of LGN cells, together with cortical amplification by threshold. These conclusions have been made based on the fast response (Smith, Bair, & Movshon, 2006), because cross-oriented stimuli suppress synaptic inhibition and excitation (Priebe & Ferster, 2006), and because LGN neurons show contrast saturation (B. Li, Thompson, Duong, Peterson, & Freeman, 2006). While these recent studies do not address aftereffects, they point out that the feedforward visual pathway yields a richer physiology than previously thought. This view is complemented by our model in that it demonstates the capability of the forward pathway to produce aftereffects without lateral influences.

5 Conclusion

We have presented a generative, sparse coding model for the development of V1 simple cells that differs from standard models in two independent design principles: the generative framework is implemented by the wake-sleep algorithm, which provides local learning rules for the bottom-up as well as the top-down weights, and sparseness is enforced by the intrinsic plasticity of two transfer function parameters of the hidden units, which maintain homeostasis of their firing rate. Training results are characterized by a large variety of receptive field shapes as well as of the gain parameters a of individual neurons. Since the sparse hidden representation is exploited for the learning of feature detectors, the intrinsic plasticity mechanism is supposed to adapt on a faster timescale than the weights. For the a parameter adapting on a timescale of tens of seconds, we find that the tilt aftereffect can be accounted for. The threshold parameters b are less variable, and their fast adaptation would lead to a tilt aftereffect that is broader than experimentally observed. Varying b changes the output of neurons with zero input and hence causes a data-independent change of behavior with profound potential influence in a network. A possibility remains that adaptation of b maintains firing rate homeostasis on a timescale of around 24 hours, accounting for the experiments of Desai et al. (1999).

Since gain modulation serves not only adaptation but also other perceptual effects such as attention or routing of information, one will expect to

find several mechanisms for gain modulation acting on distinct time scales. The simple proposed model does not reproduce all details of these effects, such as the shift of a neuron's orientation tuning curve (Dragoi et al., 2002) or the indirect TAE. The question remains whether lateral effects can be explained by individual neurons' adaptation and the spread through fixed lateral connections, or whether modifiable connections are needed to explain such effects. Our next step for increasing model complexity will be to include learned lateral connections, as in Weber (2001), in order to test whether more visual aftereffects can be accounted for.

Acknowledgments

We acknowledge financial support by the European Union through projects FP6-2005-015803 ("Daisy") and MEXT-CT-2006-042484 ("PLICON") and by the Hertie Foundation. We gratefully acknowledge support by the Frankfurt Center for Scientific Computing; Philipp Wolfrum, Cristina Savin, Arthur Franz, and Urs Bergmann for discussions; and the anonymous reviewers for substantial feedback on the model.

References

Baddeley, R. A. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B*, *264*, 1775–1783.

Bednar, J., & Miikkulainen, R. (2000). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neur. Comp.*, *12*(7), 1721–1740.

Bell, A., & Sejnowski, T. (1996). Edges are the "independent components" of natural scenes. In M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems*, *9* (pp. 831–837). Cambridge, MA: MIT Press.

Blakemore, C., & Campbell, F. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.*, *203*, 237–260.

Blasdel, G. G., & Lund, J. S. (1983). Termination of afferent axons in macaque striate cortex. *J. Neurosci.*, *3*(7), 1389–1413.

Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *J. Neurosci.*, *26*(23), 6346–6353.

Borg-Graham, L., Monier, C., & Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, *393*, 369–373.

Burdakov, D. (2005). Gain control by concerted changes in I_A and I_H conductances. *Neur. Comp.*, *17*, 991–995.

Butko, N., & Triesch, J. (2007). Learning sensory representations with intrinsic plasticity. *Neurocomputing*, *70*(7–9), 1130–1138.

Campbell, F., & Maffei, L. (1971). The tilt after-effect: A fresh look. *Vis. Res.*, *11*(8), 833–840.

Carandini, M. (2000). Visual cortex: Fatigue and adaptation. *Curr. Biol.*, *10*(16), R605–607.

Carandini, M., Movshon, J., & Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology*, *37*, 501–511.

Chance, F., Abbott, L., & Reyes, A. (2002). Gain modulation from background synaptic input. *Neuron*, *35*, 773–782.

Ciroux, J. (2006). *Simulating the McCollough effect in a selforganizing model of the primary visual cortex.* Unpublished master's thesis, University of Edinburgh.

Clifford, C., Wenderoth, P., & Spehar, B. (2000). A functional angle on some aftereffects in cortical vision. *Proc. R. Soc. Lond. B*, *267*, 1705–1710.

Crowder, N., Price, N., Hietanen, M., Dreher, B., Clifford, C., & Ibbotson, M. R. (2006). Relationship between contrast adaptation and orientation tuning in V1 and V2 of cat visual cortex. *J. Neurophysiol.*, *95*, 271–283.

Davis, G., & Bezprozvanny, I. (2001). Maintaining the stability of neural function: A homeostatic hypothesis. *Annu. Rev. Physiol.*, *63*, 847–869.

Dayan, P., Sahani, M., & Deback, G. (2002). Adaptation and unsupervised learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems*, *15* (pp. 221–228). Cambridge, MA: MIT Press.

Desai, N. S., Rutherford, L. C., & Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neurosci.*, *2*(6), 515–520.

Dragoi, V., Sharma, J., Miller, E., & Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature Neurosci.*, *5*(9), 883–891.

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, *28*, 287–298.

Eysel, U. T., Eyding, D., & Schweigart, G. (1998). Repetitive optical stimulation elicits fast receptive field changes in mature visual cortex. *Neuroreport*, *9*, 949–954.

Falconbridge, M., Stamps, R., & Badcock, D. (2006). A simple Hebbian/anti-Hebbian network learns the sparse, independent components of natural images. *Neur. Comp.*, *18*, 415–429.

Fang, F., Murray, S., Kersten, D., & He, S. (2005). Orientation-tuned fMRI adaptation in human visual cortex. *J. Neurophysiol.*, *94*, 4188–4195.

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

Gelbtuch, M., Calvert, J., Harris, J., & Phillipson, O. (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.

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

Hawley, S., & Keeble, D. (2006). Tilt aftereffect for texture edges is larger than in matched subjective edges, but both are strong adaptors of luminance edges. *J. Vision*, *6*, 37–52.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, *268*, 1158–1161.

Hirsch, J., & Martinez, L. (2006). Circuits that build visual cortical receptive fields. *Trends in Neurosciences*, *29*(1), 30–39.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

James, T., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human Brain Mapping*, *27*, 37–46.

Jin, D., Dragoi, V., Sur, M., & Seung, H. (2005). Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex. *J. Neurophysiol.*, *94*, 4038–4050.

Li, B., Peterson, M., & Freeman, R. (2003). Oblique effect: A neural basis in the visual cortex. *J. Neurophysiol.*, *90*, 204–217.

Li, B., Thompson, J., Duong, T., Peterson, M., & Freeman, R. (2006). Origins of cross-orientation suppression in the visual cortex. *J. Neurophysiol.*, *96*, 1755–1764.

Li, L., Miller, E., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.*, *69*(6), 1918–1929.

Masini, R., Antonietti, A., & Moja, E. (1990). An increase in strength of tilt aftereffect associated with tryptophan depletion. *Percept. Mot. Skills*, *70*(2), 531–539.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science*, *149*(3688), 1115–1116.

Mehaffey, W., Doiron, B., Maler, L., & Turner, R. (2005). Deterministic multiplicative gain control with active dendrites. *J. Neurosci.*, *25*(43), 9968–9977.

Mitchell, S., & Silver, R. (2003). Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron*, *38*, 433–445.

Moulder, K., Jiang, X., Taylor, A., Olney, J., & Mennerick, S. (2006). Physiological activity depresses synaptic function through an effect on vesicle priming. *J. Neurosci.*, *26*(24), 6618–6626.

Müller, J., Metha, A., Krauskopf, J., & Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. *Science*, *285*(5432), 1405–1408.

Murphy, B., & Miller, K. (2003). Multiplicative gain changes are induced by excitation or inhibition alone. *J. Neurosci.*, *23*(31), 10040–10051.

Noudoost, B., Adibi, M., & Sanayei, M. (2005). Attention sharpens selectivity. In *European Conference on Visual Perception Abstracts* (p. 36).

Ohzawa, I., Sclar, G., & Freeman, R. (1985). Contrast gain control in the cat's visual system. *J. Neurophysiol.*, *54*(3), 651–667.

Olshausen, B. A., & Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.*, *37*, 3311–3325.

Paradiso, M., Shimojo, S., & Nakayama, K. (1989). Subjective contours, tilt after-effects, and visual cortical organization. *Vis. Res.*, *29*(9), 1205–1213.

Press, W. H. (1988). *Numerical recipes in C*. Cambridge: Cambridge University Press.

Priebe, N., & Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat. Neurosci.*, *9*(4), 552–561.

Ramoa, A., & Sur, M. (1996). Short-term synaptic plasticity in the visual cortex during development. *Cerebral Cortex*, *6*, 640–646.

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties of the visual cortex. *Neur. Comp.*, *9*(4), 721–763.

Rehn, M., & Sommer, F. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.*, *22*(2), 135–146.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.*, *88*, 455–463.

Sanchez-Vives, M., Nowak, D., & McCormick, D. A. (2000). Membrane mechanisms underlying contrast adaptation in cat area 17 in vivo. *J. Neurosci.*, *20*(11), 4267–4285.

Schwabe, L., & Obermayer, K. (2006). Adaptivity of tuning functions in a generic recurrent network model of a cortical hypercolumn. *J. Neurosci.*, *25*(13), 3323–3332.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neurosci.*, *4*(8), 819–825.

Sengpiel, F., Stawinski, P., & Bonhoeffer, T. (1999). Influence of experience on orientation maps in cat visual cortex. *Nature Neurosci.*, *2*, 727–732.

Sharpee, T., Sugihara, H., Kurgansky, A., Rebrik, S., Stryker, M., & Miller, K. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*, 936–942.

Smith, M., Bair, W., & Movshon, J. (2006). Dynamics of suppression in macaque primary visual cortex. *J. Neurosci.*, *26*(18), 4826–4834.

Triesch, J. (2007). Synergies between intrinsic and synaptic plasticity mechanisms. *Neur. Comp.*, *19*, 885–909.

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

Wainwright, M. (1999). Visual adaptation as optimal information transmission. *Vis. Res.*, *39*, 3960–3974.

Wang, W., Jones, H., Andolina, I., Salt, T., & Sillito, A. (2006). Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neurosci.*, *9*(10), 1330–1336.

Weber, C. (2001). Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In G. Dorffner, H. Bischof, & K. Hornik (Eds.), *Proc. ICANN* (pp. 1147–1152). Berlin: Springer-Verlag.

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

Wilson, H., & Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vis. Res.*, *33*(8), 1133–1149.

Zhang, W., & Linden, D. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nat. Rev. Neurosci*, *4*(11), 885–900.

Received February 19, 2007; accepted July 18, 2007.