How are complex cell properties adapted to the statistics of natural stimuli?

Konrad P. Körding*[12] , Christoph Kayser[2], Wolfgang Einhäuser[2] & Peter König[2]

(1)   Institute of Neurology, UCL London, Queen Square, London, WC1N 3BG,

United Kingdom.

(2)   Institute of Neuroinformatics, University of Zürich and ETH Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

email: {koerding, kayser, weinhaeu, peterk}@ini.phys.ethz.ch

Correspondence should be addressed to K.P.K (koerding@ini.phys.ethz.ch, Tel +44 (0)20 7837 3611 x4404, Fax +44 (0)20 7813 3107)

Running head: How are Complex Cells adapted to natural scenes?

Word count : 26.409 characters

Abstract:

Sensory areas should be adapted to the properties of their natural stimuli. What are the underlying rules that match the properties of complex cells in primary visual cortex to their natural stimuli? To address this issue we sampled movies from a camera carried by a freely moving cat, capturing the dynamics of image motion as the animal explores an outdoor environment. We use these movie sequences as input to simulated neurons. Following the intuition that many meaningful high-level variables, e.g. identities of visible objects, do not change rapidly in natural visual stimuli, we adapt the neurons to exhibit firing rates that are stable over time. We find that simulated neurons, which have optimally stable activity, display many properties that are observed for cortical complex cells. Their response is invariant with respect to stimulus translation and reversal of contrast polarity. Furthermore, spatial frequency selectivity and the aspect ratio of the receptive field quantitatively match the experimentally observed characteristics of complex cells. Hence, the population of complex cells in the primary visual cortex can be described as forming an optimally stable representation of natural stimuli.

Introduction

Most neurons in the primary visual cortex can be classified into one of two generic cell types. The simple cells respond selectively to bars and gratings presented at a specific position, orientation, spatial frequency and contrast polarity (Hubel and Wiesel. 1962; Schiller et al. 1976b). The neurons of the other type, complex cells, also respond to bars or gratings of adequate orientation and spatial frequency. They, however, respond equally well regardless of the contrast polarity of the stimulus and its precise location within the region of the receptive field (Hubel and Wiesel. 1962; Kjaer et al. 1997).

The properties of sensory neurons, including the complex cells, can be expected to be well adapted to the statistics of the stimuli they are exposed to under natural conditions.

The most prominent hypothesis of how neural properties should be adapted to the statistics of natural scenes is called "sparse coding". It states that sensory neurons should be selective to specific features, only responding strongly to a small subset of stimuli, but otherwise showing low activities (Barlow 1961; Fyfe and Baddeley 1995; Olshausen and Field 1996). This theory could well explain the properties of simple cells in primary visual cortex (Bell and Sejnowski 1997; Olshausen and Field 1996; Van Hateren and van der Schaaf 1998).

Under what assumption about the objective of adaptation do simulated neurons develop the same properties as complex cells? To derive such an objective we start with the insight that it is one of the tasks of the brain to extract relevant sensory features (Barlow 1961). Relevant variables, such as the description of a visual scene in terms of objects, change on a slower time scale than low level features, such as luminance in a small spatial region. If we, for example, see an animal such as a tiger, it usually stays around for some time. However, the position of the image of its stripes on the retina changes on a shorter timescale. Such insight has lead to the development of criteria that measure the stability or temporal coherence of the responses of simulated neurons (Becker 1999; Einhäuser et al. 2002; Földiak 1991; Kayser et al. 2001; Klopf 1982; Stone and Harper 1999; Sutton and Barto 1981; Wallis and Rolls 1997; Wiskott and Sejnowski 2002). These studies have successfully applied this criterion to the representations of artificial stimuli such as moving bars to establish that such a mechanism could lead to complex-type neurons (Földiak 1991; Wiskott and Sejnowski 2002). However, by using such simple stimuli the population of neurons does not obtain a rich enough distribution to be thoroughly compared to physiology.

Here we apply a similar stability criterion to the representations of natural stimuli. We then compare the resulting neuronal response properties, i.e. their selectivity to orientation and spatial frequency as well as their response modulation and aspect ratio, to those of complex cells in primary visual cortex.

Methods

Stimuli

We study the response properties of simulated neurons after adaptation to image sequences of natural scenes. A freely moving cat explores the forest located next to the campus in Zürich while carrying a miniature CCD camera (for details see Einhäuser et al. 2002) on its head that samples the natural visual input. This procedure is carried out in accordance with institutional and national guidelines of animal care. A video of 3000 frames, recorded at 25 frames per second, digitized at a resolution of 4.5 pixel/deg and converted to grayscale using the MATLAB rgb2gray function, is used for this study. Ideally we would like to take a single long sequence from the central region of the video. Such a sequence however would need to be prohibitively long to uniformly sample the stimulus material. That is why we instead take pairs of patches measuring 30 by 30 pixels from randomly selected, but matching locations within two subsequent frames in the movie. Temporal coherence is evaluated between the patches of the same pair, approximating the optimal sampling process. The patches are first multiplied pointwise with a Gaussian kernel centered over the patch whose standard deviation (width) was 10 pixels. This procedure has a limited effect on the amount of information available in the input stream, but avoids edge effects and the anisotropy inherent in square patches. Repeating the simulations below without this windowing leads to qualitatively similar results (data not shown). The receptive field obtained in such simulations are localized, do not cover the full patch and are approximately round too. The resulting patches are decomposed into their principal components. The first component, representing the mean patch brightness, is removed. Components 2 through 100 carry more than 95% of the variance and define a vector I, which defines the input to the optimization algorithm. As the activity of each subunit linearly depends on the input, the preprocessing of the input by a principal component analysis, which is also linear transformation, has no influence on the optimization process. Discarding the higher order components, however, does have an effect. As these components carry only a small part of the total variance we do not expect an influence of this step on the results obtained. Indeed, this assumption is supported by the results of a recent study (Kayser et al. 2001). On the positive side, as the number of

dimensions of the optimization problem is reduced by a factor of 9 a significant increase in computational efficiency is achieved.

Simulated Neurons

Complex cells, in contrast to simple cells, display several strong nonlinear properties (Chance et al. 1999; Movshon et al. 1978; Ohzawa et al. 1997; Spitzer and Hochstein 1988). Hence, it is not possible to describe them adequately by linear models, and we have to consider nonlinear model neurons. Identical to the choice in a number of other studies (e.g. Hyvärinen and Hoyer 2000) we chose the the two subunit energy model (Adelson and Bergen 1985; Hyvärinen and Hoyer 2000).

Each such model neuron consists of two subunits (Figure 1A). Each of the subunits computes the scalar product of the same input patch (I) with a weight vector ($W_{1,i}$, $W_{2,i}$ respectively). Hence each neuron is characterized by two linear receptive fields. Both outputs are subsequently squared and summed to define the neurons activity: $A_i = \sqrt{\left(W_{1,i} I\right)^2 + \left(W_{2,i} I\right)^2}$.

These simulated neurons can, given appropriate weights, exhibit a large variety of response properties. Most of these properties are never observed for real neurons. The simulated neurons can however also act like a complex cell if both subunits have Gabor-wavelet like receptive fields with identical orientation and spatial frequency, and the two wavelets have a relative phaseshift of 90º (Figure 1B). If such a neuron is excited by a visual stimulus in form of a bar that is moved over it's receptive field, each subunit has an activity that depends on the bar's position. As the bar is shifted the subunits alternate in having large squared activity. Thus, the neurons activity, the sum of the squared subunits activities, changes only little as the bar is moved within the receptive field. Given the large number of parameters (twice the length of the weight vector) involved in determining the response properties of these model neurons, such complex cell like properties are only one among many other conceivable outcomes.

Optimization

The input consists of image patches that are extracted from successive frames of the movies. To simulate the adaptation process we optimize the parameters of a population of 100 neurons so that their responses are maximally coherent over time while being decorrelated from one another. This is done by maximizing the following objective function:

$$\Psi \equiv \underbrace{-\sum_i \frac{1}{\langle \overline{A}_i^2 \rangle} \left\langle \left(\overline{A}_{i,t} - \overline{A}_{i,t-40ms} \right)^2 \right\rangle}_{\Psi_{stable}} \underbrace{-\sum_{i \neq j} \frac{\langle \overline{A}_i \overline{A}_j \rangle^2}{\langle \overline{A}_i^2 \rangle \langle \overline{A}_j^2 \rangle}}_{\Psi_{decorr}} \quad ;$$

Here, $<>$ denotes the average over all stimuli, and thus over time; $\overline{A}_{i,t}$ is the activity of neuron i at time t, minus its mean over all times. Ψ_{stable} takes on large negative values if the output activities change fast. It thus punishes fast temporal variations. The 40 ms lag between two successive time points used in that objective function is well within the range of strong correlations of orientations in natural stimuli (Einhäuser et al. 2002). Ψ_{Decorr} on the other hand takes on large negative values in the case of correlated activities of different neurons and thus punishes such correlations. The average squared value of each subunit's activity is multiplicatively normalized to be one each iteration of the algorithm.

The parameters of the model neurons are optimized by scaled gradient descent. For Ψ_{stable} this leads to a local Hebb-type learning rule. The weight change is local to the synapse and depends only on pre- and postsynaptic activities at two subsequent points in time.

We furthermore compare our results to the work of Hyvärinen and Hoyer (2000). In this work they simulate a set of optimally sparse neurons that are modeled as 4-subunit energy models. All subunits are constrained to have uncorrelated output thus effectively enforcing a phase shift of 90 degrees. We repeat their simulations, using their code with our data as input. In this simulation 24 energy detector neurons with four subunits are used. We also perform a number of control simulations where we substitute Ψ_{stable} with one of a number of alternative definitions of sparseness.

Data analysis

In analogy to physiological experiments we characterize the response properties of the model neurons by several indices. The orientation tuning width is calculated as the range of orientations for which the response to a bar of optimal position is above $1/\sqrt{2}$ of the maximal activity. The best orientation φ is defined as the stimulus orientation that leads to maximal responses. The selectivity for spatial frequency is defined via the range of spatial frequencies to which the response exceeds $1/\sqrt{2}$ of the maximal level (Schiller et al. 1976b). The difference between the lower and upper bound of this range is then multiplied by 100. We measure the responses of neurons to drifting sinusoidal gratings of optimal orientation and spatial frequency. The neurons AC/DC-ratio is the maximum minus the minimum divided by the mean of the resulting activity.

The models that are used for the modeling of complex cells, such as the two subunit energy model used here always respond to moving gratings with twice the temporal frequency of the moving grating as they respond equally well to bright and dark edges. This implies that the simulated neurons have a vanishing first harmonic (F1) while the second harmonic (F2) does not vanish. Real complex cells however show such frequency doubling only to a limited degree, and both components are small (Spitzer and Hochstein 1985; Heeger 1992). How should the AC/DC ratios of such simulated neurons be compared to the relative modulation of real neurons? Either we could compare the AC/DC ratio to the F2/F0 ratio of real neurons, assuming that the frequency doubling is just an artifact of the simulation method. Alternatively we could compare the AC/DC ratio of the simulated neurons to the F1 of the real neurons, which is the preferable method to distinguish complex cells from simple cells. In this scenario followed in this paper the simulated neurons should have small AC/DC ratio, compared to the relative modulation of real neurons.

The envelope of the receptive field is defined as: $E_i(x,y) \equiv W_{i,1}(x,y)^2 + W_{i,2}(x,y)^2$. The length L_i and width V_i (defined via the standard deviations) of the receptive field is calculated (using the abbreviation $[.]_+ \equiv \max(.,0)$):

$$L_i \equiv \sqrt{\sum_{x,y}\left(x\sin(\varphi)+y\cos(\varphi)\right)^2\left[E_i(x,y)-0.5\,\mathrm{std}(E_i)\right]_+}$$

$$V_i \equiv \sqrt{\sum_{x,y}\left(x\cos(\varphi)-y\sin(\varphi)\right)^2\left[E_i(x,y)-0.5\,\mathrm{std}(E_i)\right]_+}$$

Where x and y are the positions relative to the center of gravity of the receptive field. The aspect ratio is defined as L_i/V_i. The subtraction and rectification prevents points with low values, lying far from the receptive field, from strongly influencing the aspect ratio. This is comparable to removing values below the noise level in physiological experiments. Histograms are compared using a one-sided Kolmogorov-Smirnov (KS) test yielding the probability of both histograms being drawn from the same distribution.

Parametric studies

In parametric studies we characterize the dependence of Ψ_{stable} on the receptive field properties. To elucidate why sparse coding alone is not expected to result in complex cell type responses we also measure the dependence of a specific definitions of sparseness on the receptive field properties:

8

$$\Psi_{Kurtosis} \equiv \sum_i \frac{\left\langle \overline{A}_i^4 \right\rangle}{\left\langle \overline{A}_i^2 \right\rangle^2}$$

We repeat this simulation with the objective function derived from the Cauchy prior and the standard deviation obtaining essentially the same results. We use the same two-subunit model as in the optimization procedures above albeit with simplified receptive fields. Since the optimization methods result in Gabor type receptive fields and neuronal receptive fields are well approximated by these, we choose the subunits to be Gabor wavelets of fixed orientation and spatial frequency. The phase and aspect ratio of each subunit, however, remain free parameters:

$$G(a,s,s_x,s_y,x,y) = \sin(180x/a+s)*\exp(-x^2/(as_x)^2 - y^2/(as_y)^2),$$

where a, which is fixed to a value of 5 pixels, is the size of the Gabor, s is the relative shift between the subunits, s_x and s_y are the relative length and width and x and y the relative positions of the pixels. For Figure 4B and C we choose identical shapes: W_1=G(5,0,1,1), W_2=G(5,s,1,1) and vary the shift, s, between the subunits. For Figure 4C we choose a fixed shift of 90°: W_1= G(5,0,λ,w), W_2=G(5,90°,λ,w) and vary length, λ, and width, w, between 0.5 and 4 in steps of 0.1. Aspect ratios are binned in steps of 0.2 between 0.2 and 5.

Results

We simulate neurons and adapt them to display optimally stable activity over time. The resulting response properties are characterized by the receptive fields of their two subunits (Figure 2A). Most of the subunits exhibit a receptive field that is well described by a Gabor wavelet. They thus have receptive fields that are localized in the visual space and that are selective to orientation and spatial frequency. Most neurons exhibit a phase shift between the Gabor wavelets representing the receptive fields of each of its subunits that is close to a quarter cycle (90 degrees). This suggests that the response properties of the simulated neurons exhibit some translation invariance (sketched in Figure 1B), a key property of complex cells. The neurons are furthermore tuned to orientation and spatial frequency (Figure 2B,C see also Webster and De Valois 1985).

In the following we quantitatively compare the simulated neurons' responses to bars and gratings to those of real neurons. First we investigate the orientation specificity. In response to a bar of optimal width the population of optimized neurons displays a narrow orientation tuning (38° width, Figure 2A). This specificity is somewhat tighter than the tuning width of real complex cells

(56°, p<0.001 KS-test, Schiller et al. 1976a). The simulated neurons also exhibit a tight tuning (index of 51.9) to spatial frequency comparable to the tuning index of cortical neurons (average index of 46.9, Schiller et al. 1976b), although the small difference is significant (p<0.01 KS test).

Next we compare real and simulated neurons on the basis of their response to moving gratings. In primary visual cortex, a bimodal distribution of relative modulation strengths is observed (Skottun et al. 1991) (Figure 3C). Complex cells are defined as having a relative modulation below 1.0, while simple cells are defined by larger values of the modulation ratio. In our simulations a wide bimodal distribution of AC/DC values is also observed. The AC/DC ratios of the optimally adapted complex cells have a mean (0.41) that is not significantly larger than the experimentally observed relative modulations (0.40, p>0.3 KS-test).

Last we compare the aspect ratios of the receptive fields, defined as the ratio of its width relative to its length. Real complex cells have an aspect ratio of 1.02 +- 0.2 (Ohzawa and Freeman 1997) (Figure 3D). The optimally adapted neurons have an aspect ratio of 1.09±0.3, closely matching the experimental values (p>0.3, t-test).

AC/DC ratio and aspect ratio define the invariant processing performed by complex cells. Thus, the simulated neurons with optimally stable activity result in good fits to the measured properties of complex cells in the primary visual cortex.

It has been proposed that combining sparse coding with appropriate boundary conditions also leads to complex cells (Hyvarinen and Hoyer 2000). We repeat that simulation using our stimulus database. This simulation yields neurons with an orientation selectivity of 37° and a spatial frequency selectivity of 40.5, both well in the range of the physiological values (56°, 46.9 respectively) and comparable to optimizing a stability objective (38°, 51.9 respectively). For the AC/DC ratio this simulation, however, results in a value of 0.65 that is far larger than the physiological value (0.40) and the result of optimizing a stability objective (0.41) (p<0.001 KS-test). Thus, combining a sparseness objective with additional boundary conditions does not result in sufficiently translation invariant neurons. Furthermore, the aspect ratio of 1.73 is far larger than the one observed for real complex cells (1.02, p<0.001 t-test). Similar results and equally significant deviations are found if we exchange Ψ_{stable} in our simulations by the objective function derived from a Cauchy prior as used by Olshausen and Field (1996) or the Kurtosis. This suggests that only the objective of stability adequately explains the properties of complex cells.

The head mounted camera does not register changes in gaze associated with movements of the eyes. However, recent results indicate that under the conditions the stimuli were recorded eye movements contribute little to stabilizing

the retinal image (Möller et al. 2003). To control for possible residual stabilizing effects of eye-movements we perform two experiments: (1) We simulate eye movements that randomly stabilize 50% of the patches. (2) We randomly shuffle 10% of the patches. The resulting receptive field properties are essentially unchanged in both cases. In particular in both cases they are translation invariant and have AC/DC ratios close to the relative modulation of physiological data (p>0.3 for both controls, KS-test). Therefore we do not expect major changes of the reported results if eye-movements of the cats under free viewing conditions were taken into account.

To investigate if the results generalize to a more general nonlinear model or if the results are due to the way we constructed our model neurons we perform an additional simulation (Figure 4A). Simulated neurons consisting of 8 half-squaring subunits are modeled. The neural properties resulting from optimizing Ψ_{stable} are similar to those found for the two-subunit energy model described above. Importantly, the AC/DC ratio distribution is not significantly larger than the relative modulations of real complex cells (p>0.3, KS-test). Thus, the results do not critically depend on the constraints on the model neurons' nonlinear properties defined by the two-subunit energy model. The type of the nonlinearity is set in our simulations. For the neurons to exhibit complex cell properties however the subunits need to obtain identical orientation and spatial frequency as well as the right phase shift. This simulation thus shows that these properties can be obtained from natural scenes even for varied neuron models.

To better understand the above results we proceed to characterize some important nonlinear statistical properties of videos natural scenes. To do so we measure the objective values of simulated neurons in response to the videos of natural scenes. We choose the subunits of the same model as above to be Gabor wavelets of fixed orientation and spatial frequency, leaving the aspect ratio and the relative phase as free parameters. With this more restricted set of subunit receptive fields, we can systematically analyze the influence of the receptive field properties on various objective functions. Varying the relative phase of the subunits reveals that Ψ_{stable} is maximal if the simulated neuron is translation invariant and the wavelets have a relative phase of 90 deg (Figure 4B). Neurons then represent localized oriented energy detectors and are translation invariant, as are real complex cells. We furthermore analyze the influence of the aspect ratio on the objective functions (Figure 4C). Ψ_{stable} reaches its highest value for spherical receptive fields with an aspect ratio of about 1 similar to the value of real complex cells (Ohzawa and Freeman 1997). For comparison with other studies we also plot sparseness as a function of phase and aspect ratio which peaks at values that are far from those found in physiology. It thus seems that stability is a good candidate

for an adaptation criterion that links complex cells with the statistics of natural scenes.

Discussion

We have show that adaptation to a stability objective leads to simulated neurons sharing important spatial properties of complex cells in the primary visual cortex.

Sparseness can be derived from several ideas such as minimizing energy consumption, optimal channel coding or searching for a meaningful representation of data. Stability can also be derived from various ideas: High level variables such as object identities are stable, stable variables can be transmitted through channels with lower bandwidth and learning is easier in a system where variables change slowly.

Recently Hurri and Hyvarinen (2003) have proposed that optimizing stability of linear neurons in response to natural stimuli leads to receptive fields like those of simple cells. The stability of linear neurons however is always considerably lower than the stability of the nonlinear complex cells in our study. The authors furthermore use a slightly different objective that biases the neurons to be both stable and sparse. These results might still indicate that both simple and complex cell responses could be understood in a coherent framework derived from the idea of stability.

In our simulations each neuron only saw the input stimulus windowed by a Gaussian. Parts of the properties of the neurons, in particular the aspect ratio could thus be affected by this preprocessing. Some of the simulated neurons however do have receptive fields that are smaller than the size of the Gaussian. There is a tendency for neurons to obtain localized receptive fields. It would be interesting for future studies to analyze if the distribution of receptive field sizes can be obtained exclusively from optimizing stability. Such studies would however need very large numbers of simulated neurons as they would need to jointly encode the retinal space in addition to the orientation and spatial frequency space.

Do neurons found in primary visual cortex exhibit sparse or stable or maybe both types of response properties? Both objectives seem useful for processing in the nervous system. The question of which objective links the properties of natural scenes to the properties of complex cells is experimentally accessible. On one hand, for these analyses recordings from neurons in response to natural scenes would need to be compared to response to artificial stimuli such as bars or gratings. With respect to sparseness some experiments started to

address this issue (Baddeley et al. 1997; Vinje and Gallant 2000). If a large set of natural visual patterns is presented in sequence most of these are not effectively stimulating the recorded neuron. A small subset of stimuli, however, can activate the neuron strongly and elicit very high firing rates. Similar experiments could address how stable neural responses are.

The fact that complex cells of adult animals are well described as an adaptation to a stability objective raises the question, whether this adaptation occurs on onto- or phylogenetic timescales. If there is an ontogenetic component to the development of complex cells, it allows the following experimental test of the stability hypothesis. Changing the environment during an animal's critical period (e.g. by strobe rearing) would impair the development of complex cell type receptive fields. In particular there should be a range of strobe rates in which complex cells are severely affected, while simple cells are not. From measurements of correlation times in natural videos (Kayser et al. 2003) this rate is expected to be of the order of 10Hz.

If simple cells optimize a sparseness criterion and complex cells optimize a stability criterion, it is tempting to speculate, whether such a division of labor is repeated in higher areas. Indeed in a widely used architecture for invariant object recognition, the Neocognitron (Fukushima 1980), a hierarchical network with an alternation of simple and complex type cells is used. Hence it is interesting to build larger systems consisting of several layers, each optimizing an adequate objective. This could result in a hierarchical system allowing to predict the response properties of neurons in higher cortical areas and to relate the response properties of such neurons to the statistics of the real world.

Acknowledgements

Figure captions

Fig. 1.

(A) Patches of size 30 by 30 are extracted from each frame, windowed by a circular kernel and fed into a network of neurons modeled as two subunit energy detectors. Each subunit sums up the common input weighted by its weightvector. The outputs are squared and added to obtain the output of the neuron. (B) The way such a model neuron can become translation invariant is sketched. Both subunits need to have receptive fields that are Gabor wavelets with a relative phaseshift of 90 deg. A bar of optimal orientation and spatial frequency is moved through the neurons receptive field. Whenever one subunit has a very positive or very negative activity the other one has an activity of about zero. The outputs of each subunit thus vary a lot while the sum of the outputs of the two subunits does not change much as the bar is moved.

Fig. 2.

Qualitative properties of the simulated neurons. (A) Pairs of receptive fields are shown of neurons with optimally stable activity. (B) The responses of two representative neurons to bars of changing orientations and widths (displayed at the optimal positions) are shown. (C) Responses of the same neurons to gratings of optimal phase and orientation but varying spatial frequency are shown. (D) The optimal orientation and optimal spatial frequency are plotted for all the simulated neurons.

Fig. 3.

Density distribution of properties of complex cells in primary visual cortex and of neurons with optimally stable activity. (A) The orientation tuning widths are shown for cortical complex cells in monkey cortex (Schiller et al. 1976a) and for the simulated neurons. (B) The selectivities to spatial frequency are shown for cortical complex cells in monkey cortex (Schiller et al. 1976b) and for the simulated neurons. (C) The relative modulation strengths are shown for a collection of 1061 cat complex cells replotted from (Skottun et al. 1991) along with the AC/DC ratio of the simulated neurons. (D) The distribution of aspect ratios is sketched for cat cortical neurons (Ohzawa and Freeman 1997). This is compared with the aspect ratios of the simulated neurons.

Fig. 4.

The influence of the parameters. (A) Neurons consisting of 8 halfsquaring ($f(x)=x^2$ for $x>0$, $f(x)=0$ otherwise) subunits are modeled. The histogram of

AC/DC ratios is shown for cat complex cells (Skottun et al. 1991) and for the optimized neurons with 8 halfsquaring subunits each. (B) Back in the simple two-subunit energy model, the objective functions Ψ_{stable} (thick lines) ,and $\Psi_{kurtosis}$ (thin lines) are plotted as a function of the relative phase between subunits with Gabor shaped receptive fields. (C) Objective functions are plotted as a function of the aspect ratio. Error bars denote the s.e.m.

References

Adelson EH and Bergen JR. Spatiotemporal Energy Models for the Perception of Motion. *JOpticSocAm A* 2: 284-299, 1985.

Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, Wakeman EA, and Rolls ET. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc R Soc Lond B Biol Sci* 264: 1775-1783, 1997.

Barlow HB. Possible principles underlying the transformation of sensory messages. In: *Sensory Communication*, edited by W. Rosenblith: M.I.T. Press, Cambridge MA, 1961, p. 217.

Becker S. Implicit learning in 3D object recognition: the importance of temporal context. *Neural Comput* 11: 347-374, 1999.

Bell AJ and Sejnowski TJ. The "independent components" of natural scenes are edge filters. *Vision Res* 37: 3327-3338, 1997.

Chance FS, Nelson SB, and Abbott LF. Complex cells as cortically amplified simple cells. *Nat Neurosci* 2: 277-282, 1999.

Einhäuser W, Kayser C, König P, and Körding KP. Learning the invariance properties of complex cells from natural stimuli. *EurJNeurosci* 15: 475-486, 2002.

Földiak P. Learning invariance from transformation sequences. *Neural Comput* 3: 194-200, 1991.

Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *BiolCybern* 36: 193-202, 1980.

Fyfe C and Baddeley R. Finding compact and sparse- distributed representations of visual images. *Network: Computation in Neural Systems* 6: 333-344, 1995.

Heeger DJ Half-squaring in responses of cat striate cells. *Vis Neurosci* 9: 427-43, 1992.

Hubel DH and Wiesel. TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160: 106-154, 1962.

Hyvärinen A and Hoyer P. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput* 12: 1705-1720, 2000.

Kayser C, Einhäuser W, Dümmer O, König P, and Körding K. Extracting slow subspaces from natural videos leads to complex cells. In: *ICANN*, edited by Dorffner G, Bischoff H and Kornik K. Berlin, Heidelberg: Springer, 2001, p. 1075-1080.

Kayser C, Einhäuser W, and König P. Temporal Correlations of Orientations in Natural Scenes. *Neurocomputing* (in press), 2003.

Kjaer TW, Gawne TJ, J.A. H, and Richmond BJ. Insensitivity of V1 complex cell responses to small shifts in the retinal image of complex patterns. *JNeurophysiol* 78: 3187-3197, 1997.

Klopf AH. *The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence.* Washington, DC: Hemisphere., 1982.

Möller G, Einhäuser W, and König P. Cats' eye movements under natural conditions. *Soc. Neurosci. Abstr.*, 2003.

Movshon JA, Thompson ID, and Tolhurst DJ. Receptive field organization of complex cells in the cat's striate cortex. *J Physiol* 283: 79-99, 1978.

Ohzawa I, DeAngelis GC, and Freeman RD. Encoding of binocular disparity by complex cells in the cat's visual cortex. *J Neurophysiol* 77: 2879-2909, 1997.

Ohzawa I and Freeman W. Spatial pooling of subunits in complex cell receptive fields. *Soc Neurosci Abstr* 23: 1669, 1997.

Olshausen B and Field D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607-609, 1996.

Schiller PH, Finlay BL, and Volman SF. Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *JNeurophysiol* 39: 1320-1333, 1976a.

Schiller PH, Finlay BL, and Volman SF. Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *JNeurophysiol* 39: 1334-1351, 1976b.

Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, and Bonds AB. Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31: 1079-1086, 1991.

Spitzer H and Hochstein S. Complex-cell receptive field models. *Prog Neurobiol* 31: 285-309, 1988.

Spitzer H and Hochstein S. Simple and complex-cell response dependences on stimulation parameters. *J Neurophysiol* 53: 1244-1265, 1985.

Stone JV and Harper N. Temporal constraints on visual learning: a computational model. *Perception* 28: 1089-1104, 1999.

Sutton RS and Barto AG. An adaptive network that constructs and uses an internal model of its world. *Cognition and Brain Theory* 3: 217-246, 1981.

Van Hateren JH and van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. *ProcRSocLond B Biol Sci* 265, 1998.

Vinje WE and Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273-1276, 2000.

Wallis G and Rolls ET. Invariant face and object recognition in the visual system. *ProgNeurobiol* 51: 167-194, 1997.

Webster M and De Valois RL. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J Opt Soc Am A* 2: 1124-1132., 1985.

Wiskott L and Sejnowski TJ. Slow Feature Analysis: Unsupervised Learning of Invariances. *Neural Comput* 14: 715-770, 2002.

16

Figure 1

A

100 neurons

B

squared
subunit activities S

neuron activity A

Figure 2

A

B

activity

orientation [deg]

0   180

C

activity

spatial frequency [cyc/30 pixels]

0.5   2   4   16

D

orientation [deg]

90

180   0

spatial frequency [cyc/30 pixels]

0.5   2   8

Figure 3

A

Physiology

Simulation

number of neurons

0   50   100   120

specificity for orientation [deg]

B

Physiology

Simulation

number of neurons

0   50   100

specificity to spatial frequency

C

Physiology
Complex   Simple

number of neurons

relative Modulation
Simulation

0   1   >2

AC/DC ratio

D

Physiology

mean+-2 std

Simulation

number of neurons

0   1   4

aspect ratio

Figure 4