European Journal of Neuroscience, Vol. 15, pp. 475–486, 2002

Learning the invariance properties of complex cells from their responses to natural stimuli

Wolfgang Einhäuser, Christoph Kayser, Peter König and Konrad P. Körding
Institute of Neuroinformatics, University of Zürich and ETH Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Keywords: development, primary visual cortex, temporal continuity, unsupervised learning

Abstract

Neurons in primary visual cortex are typically classified as either simple or complex. Whereas simple cells respond strongly to grating and bar stimuli displayed at a certain phase and visual field location, complex cell responses are insensitive to small translations of the stimulus within the receptive field [Hubel & Wiesel (1962) *J. Physiol. (Lond.)*, **160**, 106–154; Kjaer *et al.* (1997) *J. Neurophysiol.*, **78**, 3187–3197]. This constancy in the response to variations of the stimuli is commonly called invariance. Hubel and Wiesel's classical model of the primary visual cortex proposes a connectivity scheme which successfully describes simple and complex cell response properties. However, the question as to how this connectivity arises during normal development is left open. Based on their work and inspired by recent physiological findings we suggest a network model capable of learning from natural stimuli and developing receptive field properties which match those of cortical simple and complex cells. Stimuli are drawn from videos obtained by a camera mounted to a cat's head, so they should approximate the natural input to the cat's visual system. The network uses a competitive scheme to learn simple and complex cell response properties. Employing delayed signals to learn connections between simple and complex cells enables the model to utilize temporal properties of the input. We show that the temporal structure of the input gives rise to the emergence and refinement of complex cell receptive fields, whereas removing temporal continuity prevents this processes. This model lends a physiologically based explanation of the development of complex cell invariance response properties.

Introduction

Hubel & Wiesel (1962) proposed a feedforward model to explain the response properties of simple as well as complex cells. Simple cells obtain their selectivity by receiving input from appropriate neurons in the lateral geniculate nucleus. Complex cells pool input from simple cells, which are specific to different positions and polarities but share similar orientation tuning. Several extensions to this model have been proposed and the quantitative contribution of different afferents to complex cells is still debated (Hoffmann & Stone, 1971; Movshon, 1975; Wilson & Sherman, 1976; Toyama *et al*., 1981; Ferster & Lindstrom, 1983; Malpeli *et al*., 1986; Douglas & Martin, 1991; Reid & Alonso, 1996; Sompolinsky & Shapley, 1997; Alonso & Martinez, 1998; Mel *et al*., 1998; Chance *et al*., 1999). Nevertheless, the Hubel and Wiesel model still adequately serves as a foundation for understanding the invariance of complex cell responses.

Despite its popularity, the classical model fails to address the basic question as to how this precise synaptic connectivity is achieved during development. A number of modelling studies have examined how invariance of complex cells evolves from learning rules applied to artificial visual stimuli. For example, Schraudolph & Sejnowski (1992) perform anti-Hebbian learning and Becker (1996) uses spatial smoothness to gain translation invariance. The principle of temporal continuity used in this study was first applied to the learning of complex cell invariance in the trace rule formulation proposed by Földiak (1991). Stimuli in this study were artificially created bars similar to those often used in physiological experiments. Progressing towards more natural stimuli, static images of faces have been used; these were manually rotated or translated within the visual field to impose a well-defined 'temporal' structure (Becker, 1999; Wallis & Rolls, 1997). Applying variants of the trace rule to the stimuli these studies achieved translation or viewpoint invariance similar to that of cells in higher cortical areas. A recent study uses an analytical approach – related to independent component analysis – with natural images to gain insight into translation and phase-invariant representations (Hyvärinen & Hoyer, 2000), leaving the question of the physiological basis of that model open.

Here we propose a network model which is inspired by physiological findings and trained with natural stimuli. Stimuli are obtained from videos recorded with a camera mounted to a cat's head. These videos provide a continuous stream of stimuli, approximating the input which the visual system is naturally exposed to, and preserving its temporal structure. This input projects to a set of neurons which employs a competitive Hebbian scheme to learn simple cell type response properties. These neurons in turn project to another set of neurons, which additionally utilize the temporal structure of their input for learning. This temporal learning rule relates to recent experimental results reporting that the induction of LTP vs. LTD depends on the relative timing of pre- and postsynaptic spikes (Markram *et al*., 1997; Bi & Poo, 1998). We show that this temporal learning rule, in combination with the temporal structure of natural stimuli, leads to the emergence of complex cell response properties.

Correspondence: Dr Wolfgang Einhäuser, as above.
E-mail: weinhaeu@ini.phys.ethz.ch

Received 31 August 2001, revised 11 December 2001, accepted 18 December 2001

Materials and methods

Network model

Based on the Hubel & Wiesel (1962) model, we implement a three-layer feed-forward network whose layers are referred to as input layer, middle layer and top layer, respectively. The neurons of the input layer represent the thalamic afferents projecting to primary visual cortex. Each input neuron is connected to each neuron in the middle layer, these in turn project to all top layer neurons. Learning in the middle and top layer is competitive between neurons within the same layer. Furthermore, learning of top layer cells depends not only on the current activity but also on its past trace, and thus allows for utilization of the temporal structure of the input.

Notation

Throughout this paper $\mathbf{A}^{(\mathrm{I})}$, $\mathbf{A}^{(\mathrm{M})}$, $\mathbf{A}^{(\mathrm{T})}$ denote the activities of input, middle and top layer neurons, respectively; $\mathbf{W}^{(\mathrm{IM})}$ and $\mathbf{W}^{(\mathrm{MT})}$ refer to the weight matrices from input to middle and from middle to top layers. Bold letters indicate vectors or matrices. Subscript indices denote a single element of an activity vector or weight matrix, arguments in parentheses express time dependence. Explicit notation of time is omitted where unnecessary. Brackets $\langle x\rangle_\nu$ denote a smooth temporal average of x over ν computations of value x, i.e.

$$\langle x\rangle_\nu(t) = \frac{x(t)}{\nu} + \left(1 - \frac{1}{\nu}\right)\langle x\rangle_\nu(t-\tau),$$

where τ is the simulated time since the last calculation of $\langle x\rangle_\nu$. $[x]_+$ denotes rectification, i.e. $[x]_+ = \max(x,0)$. The temporal distance between two consecutive stimuli is denoted by Δt, and δ_{ij} denotes the Kronecker delta (1 for $i = j$, 0 otherwise)

Input layer

Stimuli consist of 10 by 10 pixel patches drawn from natural videos as described below and determine the activity in the input layer.

$\mathbf{A}^{(\mathrm{I})}(t)$ contains the pixel values of the current stimulus.

Middle layer

The activity of the middle layer neurons is calculated as:

$$\mathbf{A}^{(\mathrm{M})}(t) = \left[\frac{\mathbf{A}^{(\mathrm{I})}(t)\mathbf{W}^{(\mathrm{IM})}(t-\Delta t)}{\langle \mathbf{A}^{(\mathrm{M})}(t)\rangle_\nu} - I\left(\frac{\mathbf{A}^{(\mathrm{I})}(t)\mathbf{W}^{(\mathrm{IM})}(t-\Delta t)}{\langle \mathbf{A}^{(\mathrm{M})}(t)\rangle_\nu}\right)\right]_+ \qquad (1a)$$

where divisions by vectors are pointwise. The scalar I models the effect of a fast inhibitory circuit. The inhibition is proportional to the sum of all these activities. Inhibition changes the neurons' activities, which in turn influence inhibition strength. Assuming the inhibition to be fast in comparison to the change in the input, the activity of the middle layer neurons quickly reaches a stable state. For computational efficiency this is calculated as a fixpoint of:

$$I = \langle [\mathbf{A}^{(\mathrm{M},0)} - I]_+\rangle_{neurons} \qquad (1b)$$

where the brackets $\langle . . .\rangle$ here denote the mean over all middle layer neurons and $\mathbf{A}^{(\mathrm{M},0)}$ is the middle layer cells' activity without inhibition. The exact form of this inhibition is not a crucial issue, as it can be replaced by half rectification without a qualitative change in the results (data not shown). In equation 1a the activity of each neuron is normalized by the temporal average of its activity to prevent explosion of activities and weights.

Learning in the middle layer employs a 'winner-takes-all' scheme, which allows only the neuron of highest activity to learn. This neuron of highest activity will be called the 'learner' L.

$$L(t) = \arg\max_i(A_i^{(\mathrm{M})}(t)) \qquad (2)$$

We implement a threshold mechanism, which only allows a small subset of stimuli to effectively trigger learning. If and only if the learner in the middle layer exceeds its threshold (T_i), i.e.

$$A_L^{(\mathrm{M})}(t) > T_L(t-\Delta t), \qquad (3)$$

the weight matrix $\mathbf{W}_{i\mathrm{L}}^{(\mathrm{IM})}$ is changed by the Hebbian rule

$$W_{i\mathrm{L}}^{(\mathrm{IM})}(t) = (1-\alpha^{(\mathrm{M})})W_{i\mathrm{L}}^{(\mathrm{IM})}(t-\Delta t) + \alpha^{(\mathrm{M})}A_i^{(\mathrm{I})} \qquad (4a)$$

where $\alpha^{(\mathrm{M})}$ is the learning-rate for the middle layer. Otherwise the weights remain unchanged:

$$W_{i\mathrm{L}}^{(\mathrm{IM})}(t) = W_{i\mathrm{L}}^{(\mathrm{IM})}(t-\Delta t). \qquad (4b)$$

Also if and only if condition (3) is fulfilled, the threshold T_L is updated by:

$$T'_\mathrm{L}(t) = A_\mathrm{L}^{(\mathrm{M})}(t) \qquad (5a)$$

All thresholds decay as:

$$\mathbf{T}(t) = (1-\eta)\mathbf{T}'(t) \qquad (5b)$$

Stimuli which fulfill condition (3) will be referred to as 'effective' stimuli throughout this paper.

The combination of the winner-takes all mechanism of equation 4a with the threshold criterion (3) and the temporal normalization of equation 1a favours neurons which are strongly active for a small number of stimuli (to be the winner L and to exceed threshold), but weakly active for the rest of the stimuli (avoiding down-regulation by temporal averaging), i.e. whose activity is sparsely distributed. The threshold stabilizes learning, as only types of stimulus patterns which consistently reappear in the stimulus set can repeatedly exceed threshold. The threshold decay ensures that the network remains plastic and an equal fraction of stimuli is effective throughout the learning process.

Top layer

The top layer activity $\mathbf{A}^{(\mathrm{T})}(t)$ is calculated as,

$$A_j^{(\mathrm{T})}(t) = \frac{\max_i(A_i^{(\mathrm{M})}(t)W_{ij}^{(\mathrm{MT})}(t-\Delta t))}{\langle A_j^{(\mathrm{T})}(t)\rangle_\nu}, \qquad (6)$$

where the smooth average is taken over the effective stimuli. Equation 6 is equivalent to equation 1 apart from two obvious exceptions: the sum over i (which is implicit in the matrix multiplication) is replaced by a 'max' operation and there is no inhibition.

Analogous to the 'learner' L in the middle layer case, a learner M is defined for the top layer:

$$M(t) = \arg\max_i A_i^{(\mathrm{T})}(t) \qquad (7)$$

For the temporal structure to influence learning in the top layer, its weight update depends not only on the present activity of the cells but also on the past activity of their afferents. Therefore the weight between a middle layer cell i and a top layer cell j is increased if i was the most active middle layer cell in the previous time step, i.e. $i = \mathrm{L}(t - \Delta t)$, and j is currently the most active cell in the top layer, i.e $j = \mathrm{M}(t)$. The weight is thus increased if strong presynaptic activity preceeds strong postsynaptic activity as found in recent physiological studies (Markram *et al.*, 1997; Bi & Poo, 1998):

$$W_{i\mathrm{M}(t)}^{(\mathrm{MT})}(t) = (1 - \alpha^{(\mathrm{T})})W_{i\mathrm{M}(t)}^{(\mathrm{MT})}(t - \Delta t) + \alpha^{(\mathrm{T})}\delta_{i\mathrm{L}(t-\Delta t)}, \qquad (8a)$$

where $\alpha^{(\mathrm{T})}$ is the top layer learning rate and δ denotes the Kronecker delta. This weight update is only performed if condition (3) was fulfilled at time-step $t - \Delta t$, otherwise no changes are applied to the weights:

$$\mathbf{W}^{(\mathrm{MT})}(t) = \mathbf{W}^{(\mathrm{MT})}(t - \Delta t). \qquad (8b)$$

The fact that in equation 8a M is taken from the current time-step, whereas L is taken from the previous one, enables the temporal properties of the stimuli to affect learning. The top layer cells thereby extract information which tends to be correlated over time, while discarding uncorrelated information. They thus gain invariance to rapidly varying variables, while remaining specific to slowly varying variables.

Simulations are performed using MATLAB (Mathworks 2000, Natick MA) – the source code is available from the authors upon request. Unless otherwise stated, $n_\mathrm{M} = 60$ middle and $n_\mathrm{T} = 4$ top neurons are simulated with learning rates of $\alpha^{(\mathrm{M})} = 0.025$ and $\alpha^{(\mathrm{T})} = 2 \times 10^{-5}$. The threshold decay parameter is set to $\eta = 10^{-4}$ and smooth averages are taken over $\nu = 100$ iterations. When starting the simulation the weights between the input and middle layers are initialized uniformly with ± 0.01% uniformly distributed noise. Thresholds are initialized to the mean weight and smooth temporal averages to $1/\nu$. Weights between the middle and top layers are initialized uniformly to $1/n_\mathrm{M}$. Please note that we will show that the network properties do not depend critically on either precise tuning of the parameters or the initial conditions.

Natural stimuli

We obtained video sequences from a camera mounted to a cat's head. The cat wore chronic skull implants (for procedures see Siegel *et al.*, 1999) for the purpose of physiological experiments. The implant features two nuts, to which we attached a removable micro ccd camera (Conrad Electronics, Hirschau, Germany). Being lightweight (34 g) the camera did not affect the cat's head movements. The camera's output was transferred via a cable attached to the leash to a VCR (Lucky Goldstar, Seoul, Korea) carried by the experimentor, while the cats were taken for walks in various environments. All procedures were in compliance with Institutional and National guidelines for experimental animal care. Videos were digitized using a miroVIDEO DC 30 graphics card (Pinnacle Systems, Mountain View, CA) and Adobe Premiere software (San Jose, CA) at a sampling rate of 25 frames per second and a resolution of 320 by 240 pixels. As the camera spans a visual angle of 71 by 53° each pixel corresponded to about 13 min of arc. Data were converted to grayscale by the standard MATLAB rgb2gray function. Each image was low-pass filtered with a 3 × 3 binomial kernel:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

and convolved with a 3 × 3 laplacian kernel:

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

negative values were set to 0. This procedure mimics part of the spatial filtering in the lateral geniculate nucleus. From the image sequences thus obtained, 10 × 10-pixel wide patches were extracted and used as input to the network.

A total of 36 min of video were used. Patches were drawn from 25 different locations centred at gridpoints at vertical or horizontal distances of 0, ± 20 and ± 40 pixels from the image centre. This yielded 900 min (25 locations, each 36 min) of stimuli, which we repeatedly presented to the network. We wish to emphasize that this enlargement of the effective stimulus set should not be mistaken as a form of batch learning. In the system presented here, weights are updated using stimuli of not more than one frame temporal distance – the system learns online.

Properties of the stimulus patches are analysed using standard methods (Jähne, 1997). Orientation (θ_i) within an image patch is defined as the direction of the axis with the smallest moment of inertia. The ratio of the difference to the sum of smallest and longest axes of inertia will be referred to as 'bar-ness' throughout this paper. It is 0 for an isotropic structure and 1 for a perfectly orientated structure. The bar is defined as the line parallel to the orientation containing the centre of gravity of the gradient perpendicular to the patch's orientation. We define the 'position' (r_i) of a bar as its distance from the patch's centre, with a positive sign if above the horizon and negative otherwise.

Network analysis

Each column of the weight matrix $\mathbf{W}^{(\mathrm{IM})}$ corresponds to the receptive field of one middle layer cell. The topographic representation of each cell's receptive field can be visualized as an '*x*–*y*-diagram'. The same bar-ness measure used for the stimuli can also be applied to these *x*–*y*-representations.

In order to compare the obtained receptive fields with physiology, we used typical physiological stimuli for testing network's responses. We created a set of bars of varying orientations (θ) and positions (r), both defined analogue to θ_i and r_i of the inertia method. The bars had a Gaussian profile with a width of 1 pixel perpendicular to their orientation. They were presented as test stimuli as input to a converged network, whose weight updates were switched off. Each neuron's activity was recorded on presentation of these stimuli. Plotting the activities colour-coded as a function of θ and r yields a diagram that allows comparison of the responses of middle and top layer neurons with those of physiological simple or complex cells. A schematic diagram of this representation, which will be called a 'θ–r-

diagram' throughout this paper, is shown in Fig. 2b. Sine-wave gratings were used as an additional measure to characterize the neurons' responses. With the weight updates switched off, moving gratings of different orientations were presented to the network and the responses of the neurons were recorded. A common measure for categorizing the responses of a cell is the AC/DC (F1/F0) ratio. Like De Valois *et al*. (1982) we define the AC component as the peak to peak amplitude, and consider a cell to be complex if its AC/DC ratio is < 1 and simple otherwise. Note that these artificial stimuli were used only for testing a converged network; in no part of this study were any artificial stimuli used for training.

In order to compare this study with other proposed objective function approaches, we investigated the time course of sparseness and temporal slowness during training. For the definition of sparseness we followed Vinje & Gallant (2000):

$$sparseness = \frac{1 - \frac{1}{N}\left(\frac{\left(\sum\limits_{t=1}^{N} \mathbf{a}(t)\right)^2}{\sum\limits_{t=1}^{N} \mathbf{a}^2(t)}\right)}{1 - \frac{1}{N}} \tag{9}$$

Regarding temporal smoothness, we used the formulation proposed in Kayser *et al*. (2001):

$$slowness = -\frac{\left\langle \left(\frac{\partial \mathbf{a}(t)}{\partial t}\right)^2 \right\rangle_N}{\mathrm{var}_N(\mathbf{a}(t))} \tag{10}$$

The objective functions were evaluated by showing N sequential natural stimuli to the network. In equations 9 and 10 t is to be understood as a discrete stimulus index and N denotes the number of stimuli shown to evaluate the objective function; the brackets $\langle \ldots \rangle$ denote averaging over the stimulus set. $\mathbf{a}(t) = (\mathbf{A}(t)/\langle \mathbf{A}(t) \rangle_N)$ is the normalized activity of middle and top layer neurons, respectively, when a set of N natural stimuli is shown to the network. The derivative in equation (10) is implemented as a finite difference, i.e. $\mathbf{a}(t) - \mathbf{a}(t - \Delta t)$, and var_N denotes each neuron's temporal variance over the stimulus set.

Results

Natural stimuli

In order to understand how the network's response properties are a consequence of natural input, we first investigated the relevant statistical properties of this input.

Figure 1a shows four example frames of the videos taken from the cat's perspective. Because bars or gratings are commonly used as stimuli in physiological experiments, we investigated to what degree our natural videos contained such orientated structures. Therefore we used the bar-ness measure defined by the inertia method, which

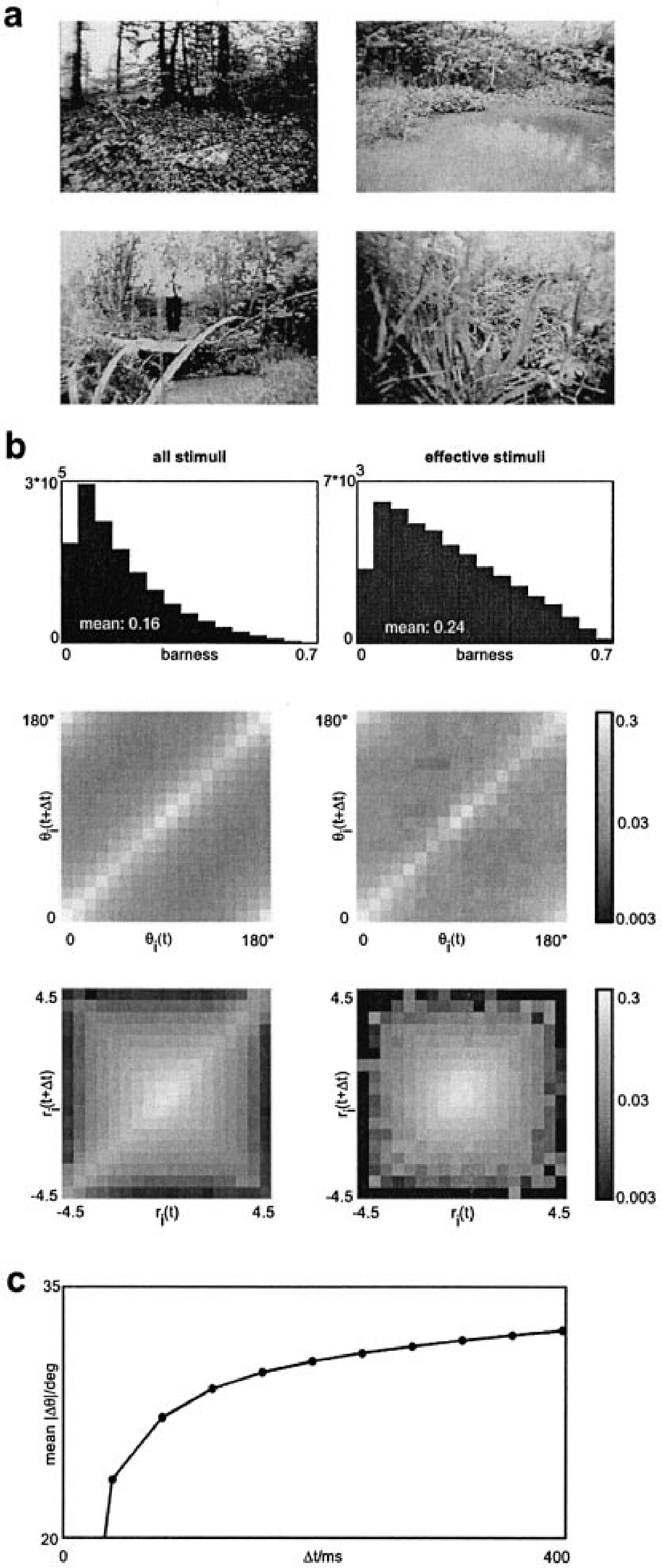

FIG. 1. Input properties. (a) Four examples out of the total 53884 frames are shown. (b) Left column, statistics of the complete set of stimuli; right column, statistics of effective stimuli (definition see methods). From top to bottom, distribution of bar-ness, correlation of orientation in subsequent stimuli, correlation of position in subsequent stimuli. Orientation is given in degrees, position in pixels, the index i indicates that data are obtained by the inertia method described in the methods. In all correlation plots incidences are colour-coded, normalized by the total histogram of orientation, or position and share the same logarithmic colour bar (r is defined to be negative if the bar's centre is below the horizon). (c) Mean change in orientation between subsequent frames over inter-frame distance for natural stimuli. (Note that the curve does saturate at a difference < 45°, as orientations are not uniformly distributed in natural images.)

approaches 1.0 for perfectly orientated structures. We observed a mean bar-ness of 0.16, thus only a small percentage of the naturally obtained stimuli were dominated by orientated structures (Fig. 1b, top left). This suggests that the use of bars or gratings as training stimuli may be inappropriate for mimicking natural conditions.

Although all stimuli from the videos were presented to the network, only a subset had an impact on learning. Using the threshold criterion (equation 3), the learning rule itself selects these ‘effective’ stimuli, in baseline conditions about 3.5% of the total. Investigating the bar-ness of the effective stimuli revealed a tendency towards orientated structures compared with the complete stimulus set (Fig. 1b, top, mean bar-ness, 0.24). This can be understood, as the threshold criterion favours types of patterns which tend to reappear consistently throughout the stimulus set. Although orientated structures appeared rarely, they were still more consistent than random patterns. Despite this relative increase in bar-ness, the bar-ness of the majority of the effective stimuli was still substantially lower than for ideal bars or gratings.

The inertia method also provides the locally dominant orientation and position for each stimulus patch. As learning in the top layer depends on pairs of sequential stimuli (equation 8a), we analysed the correlation of orientation and the correlation of position between sequential stimuli. Figure 1b reveals a strong temporal correlation in orientation (middle left) whereas position is much less correlated (bottom left). We therefore conclude that local orientation changes on a slower time scale than local position.

In order to analyse the impact of the inter-stimulus interval Δt on this correlation we measured the absolute change of orientation between subsequent stimuli depending on Δt. We found that the change in orientation increased with increasing Δt, i.e. orientation of more distant frames was less correlated (Fig. 1c).

The middle layer learns simple cell response properties

After training the network with natural stimuli most neurons in the middle layer have acquired receptive fields similar to simple cells in primary visual cortex. Figure 2a shows all these receptive fields in *x–y* representation after 450 simulated hours of training. The majority of neurons resemble detectors for bars of a certain orientation and position. In order to further quantify the localization in orientation-position space we used the *θ–r*-diagrams schematically represented in Fig. 2b as described in the methods. Figure 2c shows the same receptive fields as Fig. 2a, but in *θ–r*-representation. Most of them are well localized with respect to orientation and position.

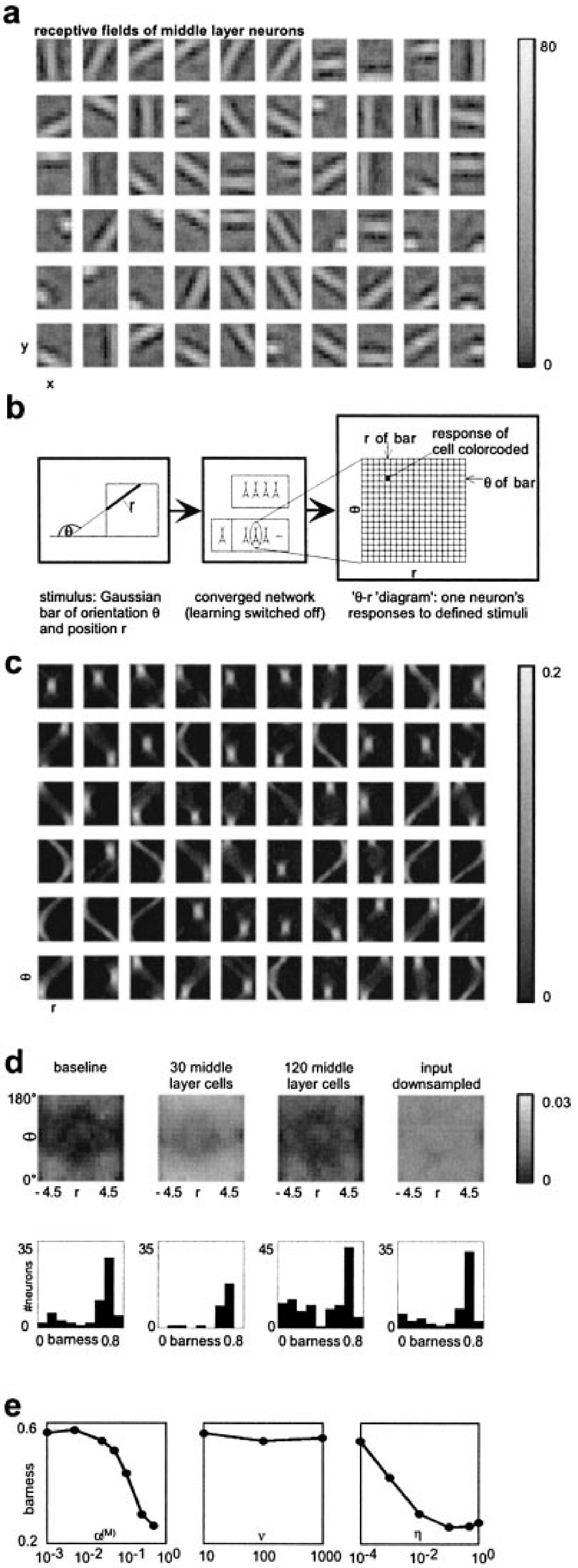

FIG. 2. Middle layer properties. (a) Receptive fields of all middle layer cells after 450 simulated hours are shown in *x–y*-representation. (b) Schematic for calculation of *θ–r*-diagrams is shown. Left, bars with a Gaussian profile perpendicular to their orientation are created for different orientations (*θ*) and positions (*r*). These bars are presented to a converged network, that has been trained with natural stimuli and whose learning mechanisms are switched off (middle). Each neuron’s response is recorded and represented (gray-scale-coded) in a neuron specific diagram whose axes are given by the orientation and position of the stimulus, the so-called *θ–r*-diagrams. (c) *θ–r*-diagrams are shown for the cells of part (a). Note that, by definition of *r* and *θ*, the upper left corner is connected to the lower right of the diagram. (d) Means of *θ–r*-diagrams (top) and distribution of middle layer bar-ness (bottom) are shown for different network properties. The left panels represent the baseline condition, the second and third networks with 30 and 120 middle layer neurons, the right panel a 60 middle layer neuron network trained with 2 times down-sampled input. All plots are taken after 450 simulated hours. (e) Dependence of mean middle layer cell bar-ness on network parameters (from left to right, middle layer learnrate, temporal average window, threshold parameter) after 450 simulated hours.

In order to control for the influence of network size and the chosen input, we simulated networks with 30 or 120 instead of 60 neurons in the middle layer. Furthermore we trained the network with input drawn from the same videos, but down-sampled by a factor of two before processing. The percentage of effective stimuli increases with the number of neurons (2.0% for 30, 3.5% for 60 and 6.2% for 120 neurons) as the threshold is individual to each middle layer cell. Downsampling of the input does not affect the number of effective stimuli (3.5%). Calculating the mean of all θ–r-diagrams we obtained a measure of how well the middle layer cells cover the stimulus space (Fig. 2d, top). This can further be quantified by 1 minus the standard deviation over all data-points divided by its mean. For a perfectly homogenous distribution this value is 1. For 30, 60 (baseline), 120 middle layer cells and down-sampled input stimulus space, coverage is 0.75, 0.87, 0.91 and 0.91, respectively. We found that more middle layer cells better cover the stimulus space, which is a direct consequence of the competition in the learning rule in the middle layer. A further way to quantify the simple cell-like properties of the middle layer cells was to apply the bar-ness measure to the receptive field representation in Fig. 2a. This revealed that most of the middle layer cells show a high bar-ness (mean, 0.63, 0.54, 0.46 and 0.54 for 30, 60, 120 middle layer neurons and down-sampled input, respectively, Fig. 2d, bottom). The decrease of average bar-ness with increasing network size is a saturation effect. Competition prevents any two middle layer cells from acquiring identical receptive fields and occupying the same location in the stimulus space.

In order to investigate other parameters possibly affecting learning in the middle layer, we measure the middle layer bar-ness in a converged network in terms of $\alpha^{(M)}$, ν and η. For all 'learnrates' $\alpha^{(M)} < 0.05$, the bar-ness of the receptive fields exceeds 0.5; i.e. middle layer cells acquire simple cell properties for these $\alpha^{(M)}$ values (Fig. 2e, left). The speed of convergence decreases with decreasing learnrate, but even for $\alpha^{(M)} = 10^{-3}$, middle layer cells converge after 300 h of simulated time to the level of baseline simulation ($\alpha^{(M)} = 0.025$). The number of iterations ν, over which the smooth average is taken, also only slightly affects the network in converged state (Fig. 2e, middle). Increasing the threshold parameter η, i.e. using more and more stimuli for learning ($\eta = 1$ implies that all stimuli are effective), reduces middle layer cells' receptive field's bar-ness. This is because noise patterns, even if not consistent over the input, are easily learnt and unlearnt again, preventing middle layer cells from exhibiting stable receptive fields. Decreasing η demands more stimuli, as only a smaller percentage will be effective, preventing the network from learning new structures. However, η can be varied over a wide range around baseline without qualitatively impairing the simple cell-like properties of the middle layer cells. (Fig. 2e, right). In conclusion, although the values of $\alpha^{(M)}$, ν and η influence the speed of convergence, their precise tuning is not a critical issue for the middle layer to learn simple cell properties in a converged network.

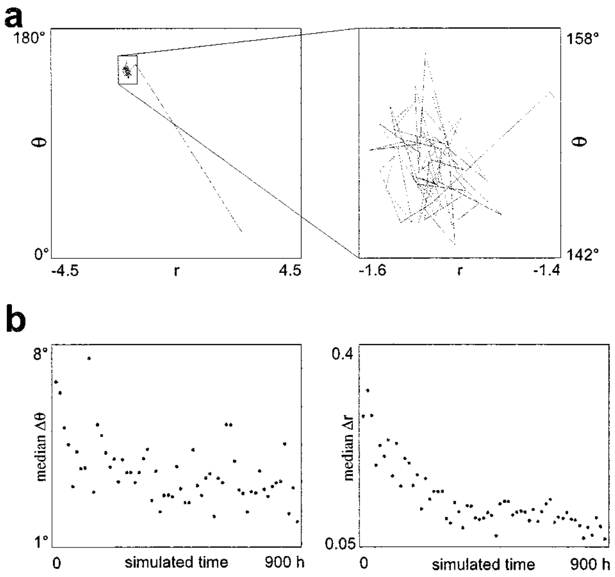

FIG. 3. Stability of middle layer properties. (a) The development of localization of the centre of gravity of the θ–r-diagram for a typical middle layer cell in baseline simulation. Each data-point corresponds to 15 h steps in simulated time. The left panel is taken over the complete possible range for orientation and position, the right panel shows a magnification of the indicated area. (b) The development of change in the centre of gravity of orientation (left) and position (right) for all middle layer cells is shown.

In the baseline simulation the weights between input and middle layer were initialized uniformly and a small percentage (0.01%) of uniformly distributed noise added. In order to control for the influence of initial conditions, we performed three additional simulations: with no noise, uniform random initialization, and presetting the middle layer receptive fields to white circles with radii between 2 and 6 pixels on a grey background. After 15 simulated hours, all these simulations reach mean bar-ness values between 0.51 and 0.56, which match the range observed for different random initializations in the 0.01% noise baseline case after the same time. Differences between the different initial conditions exceed this intrabaseline variability only within the first 10 simulated hours. Therefore we conclude that the exact form of the initial conditions has little effect on convergence speed and no effect on the properties in the converged network.

All these controls show that learning of simple cell response properties by the middle layer cells is robust with respect to changes in network size as well as input, and does not depend strongly on either the parameters chosen or the initial conditions.

In hierarchical networks it is vital for faithful higher level representation that the receptive fields of the lower levels remain stable over time. Figure 3a shows the trajectory of the centre of gravity of the θ–r-diagram of a typical middle layer cell. The datapoints are taken every 15 h of simulated time over a total of 900 h of simulated time. The receptive field of the neuron converges rapidly and then stays within a range of 0.2 pixels and 16°, as shown in the detail in the right panel of Fig. 3a. Movements in this space are restricted mainly by cells occupying neighbouring positions in θ–r space, due to the competition in learning. Figure 3b shows that for all cells in the middle layer the change in orientation and position (absolute change after each 15 h of simulated time) declines and also stays in the range observed for the example chosen in Fig. 3a. We conclude that the model converges well, revealing simple cells whose receptive fields remain stable over prolonged periods of online learning.

The top layer learns complex cell properties

The top layer cells utilize the temporal structure of the input to acquire their response properties. Here we show that these cells learn complex cell-like response properties, when presented with natural stimuli if and only if the natural temporal structure of the natural stimuli is preserved.

As a first qualitative characterization of each top layer neuron we plotted the receptive fields of the 10 middle layer cells with highest connection strengths (Fig. 4a). Middle layer cells with strongest

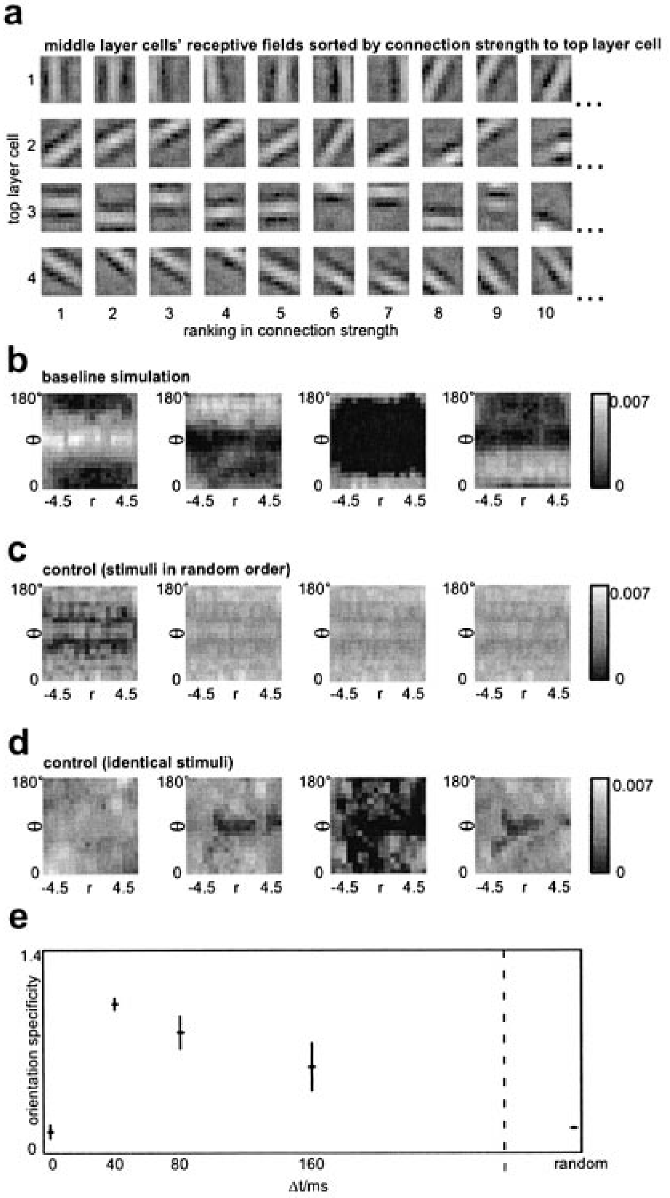

FIG. 4. Properties of top layer. (a) Receptive fields of the 10 middle layer cells connected strongest to each top layer cell. Each top layer cell corresponds to one row; middle layer cells are sorted descending in connection strength from left to right. Data are taken from the simulation of Fig. 2a ('baseline') after 450 simulated hours. (b) θ–r-diagrams for all top layer cells of Fig. 4a. (c) θ–r-diagrams for top layer cells, when stimuli are presented in random order. The top layer cells no longer exhibit translation-invariant orientation tuning as temporal correlation is lost. (d) θ–r-diagrams for top layer cells, when learning only takes place on identical stimuli. The top layer cells do no longer exhibit translation-invariant orientation tuning as temporal correlation is lost either. (e) Dependence of orientation specificity on the inter-stimulus interval Δt of the network. The baseline condition of panels (a) and (b) is found at 40 ms, the control condition of (d) at 0 ms. The control of (c) is represented to the far right, as stimuli presented in random order mimic a situation of large temporal distances between subsequent stimuli.

connections to a top cell all share the same preferred orientation but differ in position. This indicates that each top layer cell codes for one specific orientation regardless of position and thus exhibits the property of translation invariance observed in cortical complex cells.

In order to investigate this more quantitatively we plotted the θ–r-diagrams of the top layer cells (Fig. 4b). They show selective responses to stimuli of different orientations with a tuning width of ≈ 20° (half width at half maximum), but are insensitive to translation. In the θ–r-diagram this is represented by strong modulation along the orientation axis and small modulation along the position axis. To compare orientation and position tuning quantitatively, a dimensionless measure was introduced. For every orientation (or position) the mean over all positions (or orientations) is taken from the θ–r-diagrams. The standard deviation of the resulting vector, normalized by the mean of all responses $\times\sqrt{2}$, defines the orientation (or position) 'specificity' of a given neuron. Specificity is 0 if a neuron's response is independent of the respective dimension, whereas sinewave-modulated tuning (one cycle within the θ–r-diagram) yields a specificity of 1. Calculating orientation specificity for the neurons of Fig. 4b yields 0.955, 1.053, 1.055 and 1.084, whereas they are much less specific to position (0.227, 0.096, 0.161 and 0.148). A dark bar on a grey background, instead of a bright bar, yields similar results. Therefore the top layer cells in our model are invariant to translation and contrast polarity, as are cortical complex cells.

In order to control for the influence of input and network size, top layer cells' response properties were analysed for the same control conditions as in the middle layer case. After 450 simulated hours for 30 middle layer neurons the mean orientation specificity of the top layer cells was 0.849 and their mean position specificity 0.209; 120 middle layer neurons yielded 1.209 for mean orientation specificity and 0.160 for mean position specificity for top layer cells. Thus, top layer properties are robust with respect to the exact number of middle layer cells projecting to each top layer cell. Orientation specificity (mean, 0.836) of the top layer cells for the down-sampled input falls slightly below the baseline condition (mean, 1.037), which has the same number of middle layer neurons (60)., However, their position specificity (mean, 0.322) is almost twice as large as that in the baseline condition (mean, 0.171). This is explained by the fact that effective movement in the down-sampled input is only half as fast as in the baseline input, yielding a stronger correlation between subsequent positions. In the baseline simulation, weights from middle to top layer are initialized at the same value without any noise. In order to control for the influence of this initial condition we added 100% random, uniformly distributed, noise to the initial values. Under these extreme initial conditions it takes 360 simulated hours for the network to converge (i.e. orientation specificity stays within 5% around its final value), whereas it takes 165 simulated hours in baseline conditions. However, the relative differences (after 450 simulated hours) in the final values of orientation and position specificity compared with the baseline simulation are $< 0.5\%$. As with the middle layer case, one can find an upper bound for the learnrate $\alpha^{(T)}$, which is sufficient to yield the observed complex cell properties. We find orientation as well as position specificity reaches the value observed in baseline ($\alpha^{(T)} = 2 \times 10^{-5}$) for all learnrates $< 10^{-4}$. Obviously convergence speed decreases with decreasing $\alpha^{(T)}$, but $\alpha^{(T)} = 10^{-5}$ is still sufficient for the top layer to converge within 450 simulated hours. In conclusion, the learning of complex cell properties is robust with respect to network size, parameters and initial conditions.

We have shown so far that in our model the temporal structure of natural scenes is sufficient to gate the learning of complex cell properties. The following controls will address the question, whether temporal continuity is also necessary for the learning of complex cell properties in the investigated framework. As a first control we used the same stimulus set as that in the baseline condition but presented the patches in random order. Middle layer cells were not impaired, as their learning did not require temporal continuity, while top layer cells no longer exhibited complex cell type properties. In the θ–r-diagrams of the top layer cells after 450 simulated hours shown in Fig. 4c, specificity to orientation is lost (mean orientation specificity, 0.180). As a second control we showed each stimulus twice, with weight updates turned off between different stimuli and turned on between identical stimuli. Temporal information is thereby lost, although the stimuli are still presented in the same order as in the baseline condition. Middle layer cells were again not impaired but the top layer cells again showed largely reduced orientation specificity (mean 0.214, Fig. 4d).

In order to further quantify the relation between temporal correlation in the input and learning in the top layer we changed the temporal distance Δt between subsequent frames. Figure 4e shows that, with increasing Δt, the orientation specificity of top layer cells decreases. This is explained by the fact, that the correlation in orientation between frames of larger temporal distance is reduced in comparison to shorter inter-frame intervals (cf. Figure 1c). Thus we can conclude that, in the proposed scheme, temporal continuity of natural scenes is sufficient and necessary to the learning of complex cell properties**.**

A further important criterion is the stability of the acquired network properties over time. We investigated the orientation and position specificity as a function of simulated time. Figure. 5a and b show the results for all the conditions discussed above. All simulations reach a steady state after a maximum of 300 simulated hours and remain stable from then on. At first sight, the transients of the various conditions seem different in Fig. 5a and b. However, Fig. 5c and d show that the transients are nearly identical if one plots the specificity measures vs. the number of effective stimuli instead of simulation time. This shows that the different transients in Fig. 5a and b are explained by the fact that with an increase of the number of middle layer cells the number of effective stimuli is also increased. We can therefore conclude that the network converges well and learning of complex cell response properties is a stable process. Furthermore the convergence properties show that cells which have already gained some complex properties, can further refine them. In combination with the fact that the network learns online, our model thus not only applies to learning of complex cells from random initial conditions but also to experience dependent refinement of complex cell receptive fields.

Relation to objective function approaches

It has been proposed that optimizing the neurons' sparseness with respect to natural stimuli leads to the emergence of simple and complex type receptive fields (Olshausen & Field, 1996; Hyvärinen & Hoyer, 2000). Therefore we measure the sparseness (equation 9) of the middle and top layer cells of our network during training. We find that sparseness increases during learning for both the middle and the top layer neurons (Fig. 6a), but the middle layer cells reach higher sparseness values (23% vs. 2.7%). This supports the idea that

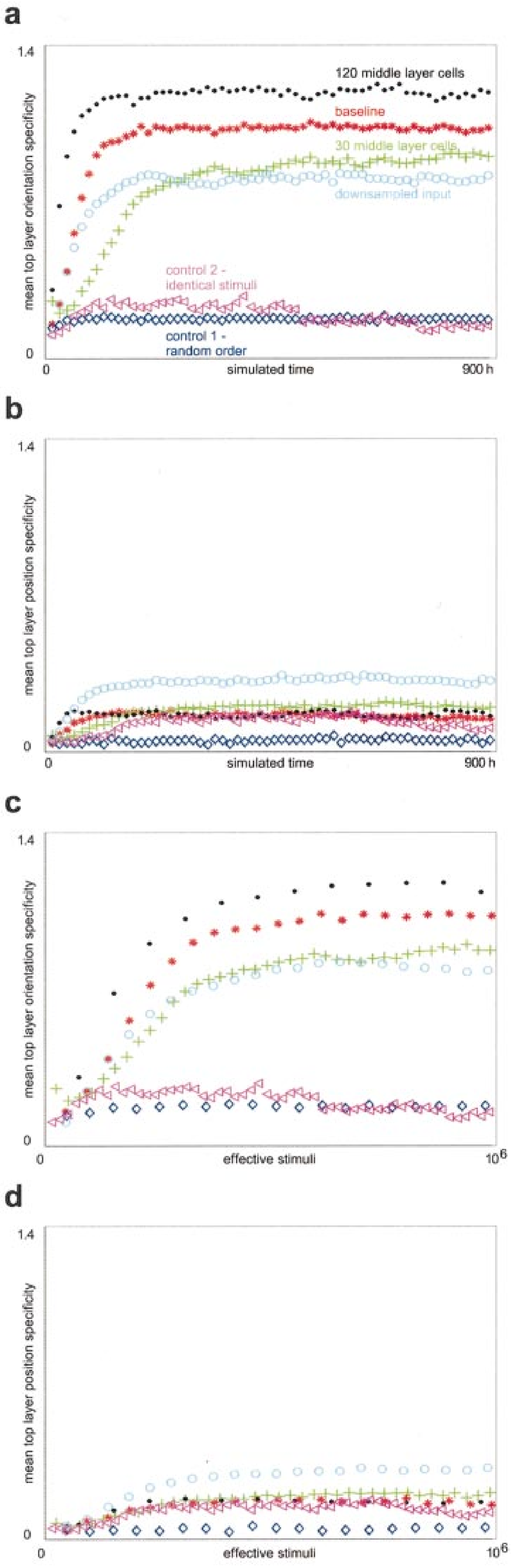

FIG. 5. Stability of top layer properties. (a) Orientation specificity as a function of time averaged over all top layer cells. Colours indicate the simulation; all simulations shown in Figs 2d and 4b–d are investigated. Each data point corresponds to 15 h simulated time. (b) Position specificity as a function of time for all simulations and controls. Colours and point-markers as indicated in (a). (c) Transients of (a), plotted vs. effective stimuli instead of simulated time. Each data point corresponds to 15 h simulated time. (d) Same as (c) but for position specificity instead of orientation specificity.

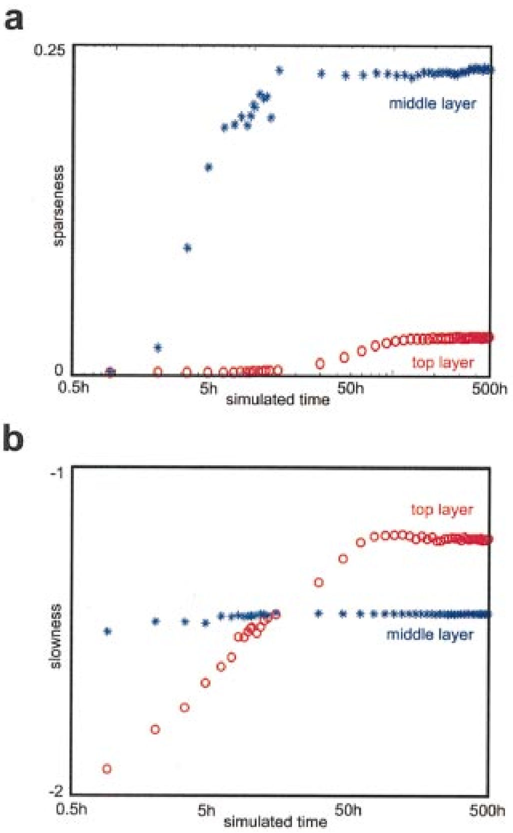

FIG. 6. Objective functions. (a) Development of mean sparseness of middle layer (blue stars) and top layer (red circles) cells during training. (b) Development of mean slowness of middle layer (blue stars) and top layer (red circles) cells during training.

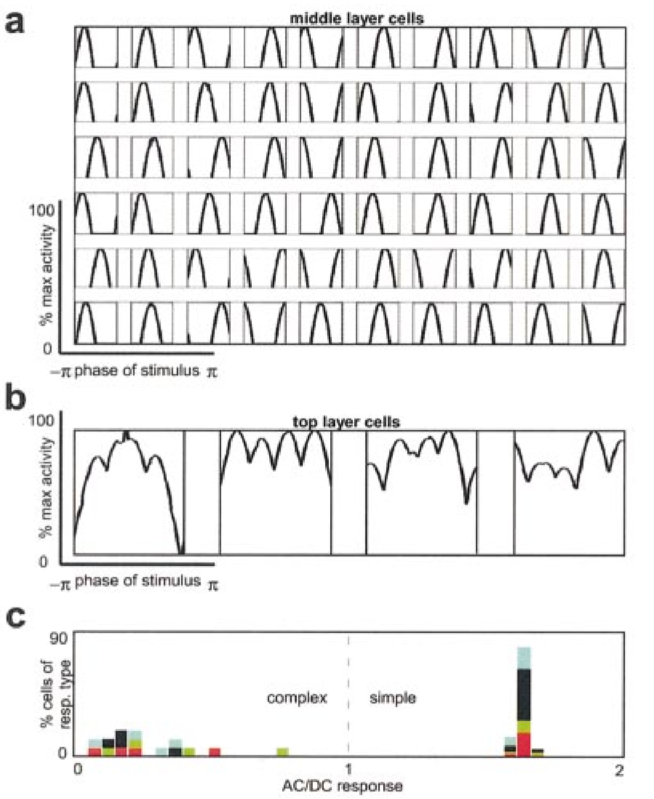

FIG. 7. Middle and top layer Cell responses to sine wave gratings (a) Responses of all middle layer cells to a drifting grating of spatial frequency 0.25 per pixel and optimal orientation of the respective cell are shown. The scale indicated on the left refers to all panels individually. (b) Same analysis as in (a) for the top layer cells from baseline simulation is shown. (c) Distribution of AC/DC response ratio (AC component is defined as peak to peak amplitude) for middle and top layer cells from 4 simulation conditions (baseline, 30 and 120 middle layer cells, down-sampled input – colours correspond to those of Fig. 5). Note that the vertical axis is normalized for each cell type (middle layer/top layer) individually.

especially the middle layer cells favour a sparse code, although it remains to be shown, that their sparseness is indeed optimized by the given learning rule**.**

Besides sparseness, temporal slowness (equation 10) has also been proposed as an objective function to explain complex cell invariance properties in response to natural stimuli (Kayser *et al.,* 2001). Temporal slowness increases during learning for top layer cells, while remaining nearly constant for middle layer cells (Fig. 6b). This suggests that the top layer cells favour slowly varying output**.**

In order to compare the responses of the simulated neurons directly with the responses of cortical neurons, we used an additional measure. A measure commonly used in physiology to define simple from complex cells, is the AC/DC ratio (see methods) when the visual system is exposed to moving sinewave gratings. Figure 7a shows the responses of middle layer cells in a baseline simulation after being trained with natural stimuli for 450 simulated hours. The responses are plotted for the preferred orientation (the sinewave stimulus the cell responds maximally to) at a fixed spatial frequency of 0.25 per pixel. All middle layer cells show responses similar to those of physiologically characterized simple cells. Figure 7b shows the same measure for the top layer cells, which exhibit responses typical for complex cells. Calculating the AC/DC ratio as described in the methods section for both middle and top layer cells yields the histogram in Fig. 7c. Whereas the AC/DC ratio in all top layer cells is <0.5, in all middle layer cells it ranges between 1.5 and 1.7. Therefore, according to the definition of De Valois *et al*. (1982), all top layer cells are complex (AC/DC $<$ 1) and all middle layer cells simple (AC/DC $>$ 1).

Discussion

Relating the statistics of the real world to neuronal properties is an emerging field within neuroscience. Recently there has been much progress leading not only to powerful computational algorithms (Hyvärinen, 1999; Schwartz & Simoncelli 2001) but also to compact models of neuronal properties (Olshausen & Field, 1997) Within this approach this study shows that simple and complex cells can be learnt from real world stimuli in a biologically realistic scheme. The proposed model is robust with respect to its parameters and exhibits stable receptive fields, whose responses match those observed in physiology. Learning complex cell invariance properties depends on the temporal structure of the visual input and it is suppressed if temporal continuity is removed.

In the present study we assume a purely feedforward architecture, in which thalamic inputs drive middle layer cells which, after learning, exhibit simple cell type properties. These in turn drive top layer cells, which exhibit the complex cell properties. Chance *et al.* (1999) propose that complex cell properties could also be exhibited by cells with simple cell-like bottom-up connections and recurrent connections from other simple cells. However, their study does not address how the necessary recurrent connectivity could be acquired. It remains an interesting problem for further research how such a type of recurrent connectivity can be learnt by the mechanisms proposed in this study.

In the present study the input to a cat's retina is approximated by a camera mounted to the animal's head. Unlike primates, cats have an oculomotor range (± 28°), which only covers a small fraction of their field of view (primates: ± 70°). Furthermore, unrestrained cats seldom perform eye-saccades to rapidly change the direction of gaze, but rather move their head – using saccades only for minor smoothing of the movement (Guitton *et al*., 1984). Thus we can conclude that, by neglecting possible eye movements, the camera only receives input which is present on the cat's retina, and it does not significantly alter the temporal structure of the stimuli.

The learning of simple cell response properties has previously been investigated in a number of studies (Olshausen & Field, 1996; Bell & Sejnowski, 1997; van Hateren & van der Schaaf, 1998). They show that searching for sparse representations, de-correlated or independent components in real world images results in receptive fields similar to those of cortical simple cells. The winner-takes-all learning rule used in the present study to learn the connectivity from input to middle layer, in combination with the proposed threshold mechanism, may be viewed as a sparse prior guiding synaptic plasticity. Indeed we have shown that sparseness increases for middle and top layer cells of our network. We here used the formulation of sparseness introduced by Vinje & Gallant (2000) which itself is a rescaling of the definition by Rolls & Tovee (1995). This definition measures how sparsely a single neuron's activity is distributed over the stimulus set. Other definitions of sparseness measure instead how sparsely the activity of a population is distributed over a single stimulus. The former formulation seems more appropriate for our purpose, as the stimulus set is large compared with the number of neurons. However, if one adds additional constraints, e.g. like decorrelating the neurons' activities, as is often done in optimizing objective functions, both formulations in practice lead to similar results. Although the equivalence of our model to the maximization of sparseness remains to be shown, our results demonstrate how optimization of sparseness might be implemented in a physiologically plausible framework.

The trace rule originally proposed by Földiak (1991) revealed that temporal continuity can be exploited to produce complex cell-like receptive fields. Stimuli for this study were bars of four different orientations as often used in physiological experiments. This concept was applied by Wallis & Rolls (1997) and Becker (1999) to the problem of face recognition. These studies used static images of faces, which were subjected to well-controlled transformations such as rotations or translations. Stimuli used in the present study are not only taken from natural images, but also preserve the natural temporal structure of the real-world input to the visual system. Therefore our model has – as do biological visual systems – to cope with additional difficulties. Firstly, the movements of objects on the retina are not constant, but contain rapidly changing movements as well as nearly immobile fixation periods. Secondly, the used sequences contain a continuous variety of objects, unlike the usually small number of instances used for training and testing in previous studies. Thirdly, a lot of the encountered stimuli might even be unsuitable for learning. Thus the system has to select valuable stimuli for itself. Therefore we consider our stimuli, obtained from a camera mounted on freely behaving cat, the critical test for the temporal continuity hypothesis.

Other recent studies on learning of complex cell properties from natural images use analytic approaches, searching for example for sparse subspaces (Hyvärinen & Hoyer, 2000), or – more closely related to this study – slowly varying features (Kayser *et al., 20*01; Wiskott & Sejnowski, 2001). We show that sparseness as well as temporal slowness increase for the top layer cells during training. This is evidence that our network may provide a physiological substrate for these types of objective function approaches.

Objective function studies provide theoretical insight into simple and complex cell response properties, but neither depend on, nor provide, a definite physiological basis for their models. Therefore they allow comparisons of the resulting receptive fields with physiological findings, but not the underlying mechanisms. Here we implement a network model which itself is based on mechanisms observed in cortical physiology. The model's cells acquire simple and complex cell response properties and also increase commonly used objective functions. Although a direct analytic link between the objective function formulations and the present study remains an issue for further research, the present model suggests a link between objective functions and cortical mechanisms**.**

To our knowledge, the present study is the first to combine natural stimuli, which preserve the temporal structure of real world scenes, with a physiologically plausible model to learn complex cell properties. However, several issues regarding the physiological realism remain to be discussed. The ratio between the number of simple and complex cells in our model does not match the ratio typically observed in physiological experiments (Skottun *et al*., 1991). This restriction is closely linked to the fact that we use full connectivity in and between layers. Thus the considered complex cells can be regarded as a fully connected subset of the total number of complex cells. Building on the property that the number of middle layer cells can change over a wide range in our model without substantially affecting the network's behaviour, matching simple/complex ratio and connectivity more closely to physiological data remains an interesting issue for further research.

A major characteristic of the proposed learning rule is the implementation of competition between different neurons of the same layer. This results in a learning rule in which only the neuron with the highest activity can learn at each stimulus presentation. In a Hebbian scheme such competition can for example be generated by strong lateral inhibition (Ellias & Grossberg, 1975; Hertz *et al*., 1991). Körding & König (2000a) propose an alternative model, in which timing of the cell's firing is crucial. The network exhibits global oscillations, in which cells receiving stronger input fire earlier within a common cycle; only the cells that fire first can learn. The mechanism thus implements a winner take all circuit on learning, while implementing a linear circuit for representation. The study presented here is inspired by a different mechanism for strong competition; it seems that bursts of action potentials are necessary for the induction of LTP (Pike *et al*., 1999). *In vivo* studies on layer V pyramidal cells show that such bursts are typically associated with calcium spikes in the apical dendrites (Larkum *et al*., 1999a). Further data show that even very low levels of inhibition are sufficient to block the generation of calcium spikes (Larkum *et al*., 1999b). Combining the two findings about LTP and calcium spikes suggests a model whose learning is highly competitive but inputs are still faithfully represented (Körding & König, 2000b). Furthermore several physiological experiments (Lisman, 1989, 1994; Artola *et al*., 1990; Bear & Malenka, 1994) argue in favour of a scheme in which synapses are only changed if the neurons exhibit high levels of activity. Note therefore that various different cortical mechanisms could provide the foundation for the thresholded competitive learning used in our model.

We use a 'max norm' to integrate the input from the middle layer cells onto the top layer cells. Whereas some studies propose a simple linear sum (Fukushima, 1980; Mel, 1997) the max norm has been suggested to fit physiological findings better when a hierarchy of levels is considered (Riesenhuber & Poggio, 1999). This is also supported by recent physiological evidence (Lampl *et al*., 2001). In order to exploit the temporal structure of the stimuli, the learning in the top layer depends on past presynaptic activity as well as current postsynaptic activity. Electrophysiological evidence suggests that plasticity depends not only on the present activity but also on its temporal evolution (Yang & Faber, 1991; Ngezahayo *et al*., 2000). Markram *et al*. (1997) show that the change of synaptic efficacy is increased only if presynaptic input precedes the postsynaptic spike. The time scale of this process is similar to our inter-frame delay of 40 ms. This is our rationale for considering adjacent video frames instead of, for example, using a weighted average over several frames. Further experiments show that synaptic efficacy in the hippocampus (Debanne *et al*., 1996) and in cultured neurons (Bi & Poo, 1998) also depends on the relative timing between pre- and postsynaptic activity. This dependence is thus likely to be a common cortical feature also present in primary visual cortex.

Our model predicts that temporal continuity of natural scenes is exploited by the visual system to gain or refine the receptive fields of complex cells. In order to check this prediction we propose the following experimental test. Temporal continuity of the real world can be impaired by the use of stroboscopic light. As correlation between visual scenes declines with increasing temporal distance, receptive field properties of complex cells in strobe reared animals should depend on the inter-strobe interval. Betsch *et al*. (2001) have shown, using natural stimuli of sizes similar to cats' complex cell receptive fields, that orientation is uncorrelated after about 300 ms. Therefore the environment of an animal strobe-reared at strobe rates <3 Hz should closely resemble the control condition of randomly shuffled stimuli. As in the control condition, complex cells in our model no longer exhibit their typical properties, we predict a significant impairment of complex cells in such strobe-reared animals.

Acknowledgements

We would like to thank B.Y. Betsch for help in collecting the video data. We are also grateful to T. Freeman and the anonymous referees for their helpful comments on previous versions of the manuscript. The work was financially supported by the 'Studienstiftung des deutschen Volkes', Honda R & D Europe (Germany) (W.E.), the ''Neuroscience Center Zurich'' (C.K.), the 'Swiss National Fonds' (P.K. – grant no. 3100–51059.97), the 'Boehringer Ingelheim Fond' and the Collegium Helveticum (K.P.K.).

References

Alonso, J.M. & Martinez, L.M. (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature Neurosci*., **1**, 395–403.

Artola, A., Brocher, S. & Singer, W. (1990) Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, **347**, 69–72.

Bear, M.F. & Malenka, R.C. (1994) Synaptic plasticity: LTP and LTD. *Curr. Opin. Neurobiol*., **4**, 389–399.

Becker, S. (1996) Mutual information maximization: models of cortical self-organization. *Network*, **7**, 7–31.

Becker, S. (1999) Implicit learning in 3D object recognition: The importance of temporal context. *Neural Computation*, **11**, 347–374.

Bell, A.J. & Sejnowski, T.J. (1997) The 'independent components' of natural scenes are edge filters. *Vision Res*., **37**, 3327–3338.

Betsch, B.Y., Körding, K.P. & König, P. (2001) The visual environment of cats. *Annual Meeting of the Assoc. for Research in Vision and Ophthalmology, Fort Lauderdale*, **3308**, 615.

Bi, G.Q. & Poo, M.M. (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci*., **18**, 10464–10472.

Chance, F.S., Nelson, S.B. & Abott, L.F. (1999) Complex cells as cortically amplified simple cells. *Nature Neurosci*., **2**, 277–282.

De Valois, R.L., Albrecht, D.G. & Thorell, L.G. (1982) Spatial frequency selectivity of cells in Macaque visual cortex. *Vision. Res*., **22**, 545–559.

Debanne, D., Gähwiler, B.H. & Thompson, S.M. (1996) Cooperative interactions in the induction of long-term potentiation and depression of synaptic excitation between hippocampal CA3-CA1 cell pairs in vitro. *Proc. Natl. Acad. Sci. USA*, **93**, 11225–11230.

Douglas, R.J. & Martin, K.A.C. (1991) A functional microcircuit for cat visual cortex. *J. Physiol. (Lond.)*, **440**, 735–769.

Ellias, S.A. & Grossberg, S. (1975) Pattern formation, contrast control and oscillations in the short term memory of shunting on-center off-surround networks. *Biol. Cybern*., **20**, 69–98.

Ferster, D. & Lindstrom, S. (1983) An intracellular analysis of geniculocortical connectivity in area 17 of the cat. *J. Physiol. (Lond.)*, **342**, 181–215.

Földiak, P. (1991) Learning invariance from transformation sequences. *Neural Comput*., **3**, 194–200.

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

Guitton, D., Douglas, R.M. & Volle, M. (1984) Eye-head coordination in cats. *J. Neurophysiol*., **52**, 1030–1050.

van Hateren, J.H. & van der Schaaf, A. (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Lond. B*., **265**, 359–366.

Hertz, J., Krogh, A. & Palmer, R.G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley, New York.

Hoffmann, K. & Stone, J. (1971) Conduction velocity of afferents to cat visual cortex: a correlation with cortical receptive field properties. *Brain Res*., **32**, 460–466.

Hubel, D.H. & Wiesel, T.N. (1962) Receptive fields, binoculular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)*, **160**, 106–154.

Hyvärinen, A. (1999) Survey on independent component analysis. *Neural Computing Surveys*, **2**, 94–128.

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

Jähne, B. (1997) *Digital Image Processing – Concepts, Algorithms and Scientific Applications*, 4th compl. rev. edn. Springer, Berlin.

Kayser, C., Einhäuser, W., Dümmer. O., König, P. & Körding, K.P. (2001) Extracting slow subspaces from natural videos leads to complex cells. In Dorffner, G., Bischoff, H. & Hornik, K. (eds), *Artificial Neural Networks – ICANN 2001 Proceedings*. Springer, Berlin, Heidelberg, pp. 1075–1080.

Kjaer, T.W., Gawne, T.J., Hertz, J.A. & Richmond, B.J. (1997) Insensitivity of V1 complex cell responses to small shifts in the retinal image of complex patterns. *J. Neurophysiol*., **78**, 3187–3197.

Körding, K.P. & König, P. (2000a) A learning rule for dynamic recruitment and decorrelation. *Neural Netw*., **13**, 1–9.

Körding, K.P. & König, P. (2000b) Learning with two sites of synaptic integration. *Network*, **11**, 25–39.

Lampl, I., Riesenhuber, M., Poggio, T. & Ferster, D. (2001) The max operation in cells in the cat visual cortex. *Soc. Neurosci. Abstr*., **27**, 619.30.

Larkum, M.E., Kaiser, K.M. & Sakmann, B. (1999a) Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. *Proc. Natl. Acad. Sci. USA*, **96**, 14600–14604.

Larkum, M.E., Zhu, J. & Sakmann, B. (1999b) A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, **398**, 338–341.

Lisman, J. (1989) A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. USA*, **86**, 9574–9578.

Lisman, J. (1994) The CaM kinase II hypothesis for the storage of synaptic memory. *Trends Neurosci*., **17**, 406–412.

Malpeli, J., Lee, C., Schwark, H. & Weyand, T. (1986) Cat area 17. I. Pattern of thalamic control of cortical layers. *J. Neurophysiol*., **56**, 1062–1073.

Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, **275**, 213–215.

Mel, B.W. (1997) SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput*, **9**, 777–804.

Mel, B.W., Ruderman, D.L. & Archie, K.A. (1998) Translation-invariant orientation tuning in visual 'complex' cells could derive from intradendritic computations. *J. Neurosci.*, **18**, 4325–4334.

Movshon, J. (1975) The velocity tuning of single units in cat striate cortex. *J. Physiol. (Lond.)*, **249**, 445–468.

Ngezahayo, A., Schachner, M. & Artola, A. (2000) Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *J. Neurosci.*, **20**, 2451–2458.

Olshausen, B.A. & Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, **381**, 607–609.

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

Pike, F.G., Meredith, R.M., Olding, A.W.A. & Paulsen, O. (1999) Postsynaptic bursting is essential for 'Hebbian' induction of associative long-term potentiation at excitatory synapses in rat hippocampus. *J. Physiol. (Lond.)*, **518**, 571–576.

Reid, R.C. & Alonso, J.M. (1996) The processing and encoding of information in the visual cortex. *Curr. Opin. Neurobiol.*, **6**, 475–480.

Riesenhuber, M. & Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nature Neurosci.*, **2**, 1019–1025.

Rolls, E.T. & Tovee, M.J. (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.*, **73**, 713–726.

Schraudolph, N.N. & Sejnowski, T.J. (1992) Competitive anti-Hebbian learning of invariants. In Moody, J.E., Hanson, S.J. & Lippmann, R.P. (eds), *Advances in Neural Information Processing Systems*, **Vol. 4**. Morgan Kaufmann, San Mateo, pp. 1017–1024.

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

Siegel, M., Sarnthein, J. & König, P. (1999) Laminar distribution of synchronization and orientation tuning in area 18 of awake behaving cats. *Soc. Neurosci. Abstr.*, **25**, 678.

Skottun, B.C., De Valois, R.L., Grosof, G.H., Movshon, J.A., Albrecht, D.G. & Bonds, A.B. (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Res.*, **31**, 1079–1086.

Sompolinsky, H. & Shapley, R. (1997) New perspectives on the mechanisms for orientation selectivity. *Curr. Opin. Neurobiol.*, **7**, 514–522.

Toyama, K., Kimura, M. & Tanaka, K. (1981) Organization of cat visual cortex as investigated by cross-correlation technique. *J. Neurophysiol.*, **46**, 202–214.

Vinje, W.E. & Gallant, J.L. (2000) Sparse coding and decorrealtion in primary visual cortex during natural vision. *Science*, **287**, 1273–1276.

Wallis, G. & Rolls, E.T. (1997) Invariant face and object recognition in the visual systems. *Prog. Neurobiol.*, **51**, 167–194.

Wilson, J. & Sherman, S. (1976) Receptive-field characteristics of neurons in cat striate cortex: changes with visual field eccentricity. *J. Neurophysiol.*, **39**, 512–533.

Wiskott, L. & Sejnowski, T. (2001) Slow feature analysis: unsupervised learning of invariances. *Neural Comp.*, in press.

Yang, X.D. & Faber, D.S. (1991) Initial synaptic efficacy influences induction and expression of long-term changes in transmission. *Proc. Natl. Acad. Sci. USA*, **88**, 4299–4303.