Learning distinct and complementary feature-selectivities from Natural Colour Videos

Wolfgang Einhäuser[1], Christoph Kayser, Konrad P. Körding & Peter König

Institute of Neuroinformatics (UNI/ETH Zürich) Winterthurerstrasse 190, 8057 Zürich,

Switzerland; Phone: +41-1-6353044; Fax.: +41-1-6353053

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

[1]Corresponding author

Abstract

Many biological and artificial neural networks require the parallel extraction of multiple

features, and meet this requirement with distinct populations of neurons that are selective to

one property of the stimulus while being non-selective to another property. In this way,

several populations can resolve a set of features independently of each other, and thus achieve

a parallel mode of processing. This raises the question how an initially homogeneous

population of neurons segregates into groups with distinct and complementary response

properties. Using a colour image sequence recorded from a camera mounted to the head of a

freely behaving cat, we train a network of neurons to achieve optimally stable responses, that

is, responses that change minimally over time. This objective leads to the development of

colour selective neurons. Adding a second objective, de-correlating activity within the

network, a subpopulation of neurons develops with achromatic response properties. Colour

selective neurons tend to be non-oriented while achromatic neurons are orientation-tuned. The

proposed objective thus successfully leads to the segregation of neurons into complementary

populations that are either selective for colour or orientation.

Introduction

A striking feature of the visual system is the existence of cells that are selective to one property of the visual input while being non-selective to another. An important example of this coexistence of selectivity and invariance are complex neurons in primary visual cortex, which are selective for spatial frequency and orientation, but are insensitive to small translations of the stimulus (Hubel & Wiesel 1962). Another example of this feature can been seen in inferotemporal cortex, where neurons respond selectively to complex objects like faces, but are invariant to translation, rotation, scaling and to changes in contrast and illumination (Rolls 1992, Hietanen et al. 1992, Ito et al 1995). Experimental evidence suggests that neurons with comparable response properties are organized into distinct functional pathways (Livingstone & Hubel 1984, DeYoe & VanEssen 1985, Zeki & Shipp 1988, Livingstone & Hubel 1988). The notion of a distinct neural pathway implies that a group of neurons will be selective to one stimulus dimension, but invariant to other dimensions. It is generally agreed that within a given sensory processing level, neurons can be segregated according to their selectivity to a particular stimulus dimension. However, this segregation may not be strictly maintained across multiple processing levels in the brain (Merigan & Maunsell 1993). At higher processing stages, selectivity to a given stimulus property may be increased following exposure to that stimulus (Logothetis et al. 1995, Sigala & Logothetis 2002), suggesting that selectivity may be established through experience, and depend on properties of the sensory input.

Recently, several learning schemes have been proposed that link the response properties of visual neurons to statistical properties of the visual input. The principle of sparse coding, for example, can predict how visual experience gives rise to the receptive field properties of simple (Olshausen & Field 1996, Bell & Sejnowski 1997, van Hateren & van der Schaaf 1998) and of complex (Hyvärinen & Hoyer 2000) neurons in primary visual cortex. At

the highest level of the visual system, neurons in inferotemporal cortex also appear to form sparse representations (Rolls & Tovee 1995). Thus, sparse representations may be ubiquitous in the visual system.

Another learning scheme able to predict response properties of neurons in the visual system is the principle of temporal coherence. This learning scheme favours responses that change minimally (i.e., are stable) over time. Originally based on the trace rule proposed by Földiak (1991), neuron simulations using temporal coherence as an objective function display response properties characteristic of simple (Hurri & Hyvärinen 2002) and complex (Kayser et al. 2001, Einhäuser et al. 2002) cells, as well as response-invariant behaviour characteristic of higher levels of the primate visual system (Stone 1996, Wallis & Rolls 1997, Wiskott & Senjowski 2002). Thus, the principle of temporal coherence explains response-invariant behaviour at different levels of the visual hierarchy.

However, the question as to how an initially homogeneous population of neurons could segregate into subpopulations that are selective to one particular stimulus property while being non-selective to another, remains unresolved. Here we address this question by training a network of simulated neurons with natural, colour image sequences using a learning scheme derived from the principle of temporal coherence. This scheme leads to the emergence of neurons that are selective to the colour of the stimulus while being non-selective to stimulus orientation, and to a complementary group of neurons displaying the opposite stimulus preferences.

Methods

Stimuli

We recorded natural image sequences using a removable lightweight CCD-camera attached to the head of a freely behaving cat, as described in detail elsewhere (Einhäuser et al. 2002). The cat is accompanied by an animal handler, but is otherwise free to explore an outdoor

environment as it chooses. All procedures are in compliance with Institutional and National guidelines for experimental animal care.

A total of 4900 consecutive frames from this sequence is then digitised into the RGB format at a colour depth of 24 bits. From each frame, we extract 20 patches (30 x 30 pixels) at random locations. Patches are then extracted at the same locations from the following frame in the sequence, yielding 20 separate stimulus pairs per frame pair (Figure 1). As using the square patch as such would introduce an orientation bias, each colour channel from the images is smoothed with a Gaussian kernel (12 pixels in width). This ensures a smooth isotropic aperture. For computational efficiency, a principal component analysis (PCA) is performed on the patch to reduce its dimensionality. As the absolute luminance of a visual stimulus is filtered out at the retina and cortical responses are mostly independent of the global illumination level, we discard the mean intensity by excluding the first principal component. Unless otherwise stated, principal components 2 to 200 are used in the subsequent analysis.

Neuron models

The activity of each neuron is computed as

$$A_i = \varphi\left(\sum_j W_{ij} I_j\right)$$

where I is the input vector and W the weight matrix; $\varphi(x)$ defines the neuron model. In this study we adopted two different neuron models that are in common use: the linear-threshold model ($\varphi(x) = \max(x,0)$) and the full-wave-rectifying model ($\varphi(x) = |x|$). In each case the output of a neuron cannot be less than zero, reflecting the fact that real neurons cannot spike at negative rates.

Objective functions

The objective function used in this study is known as temporal coherence, which we implement by minimising the squared temporal derivative of the neurons' activity. To avoid

the trivial solution of a neuron whose weights are all equal to zero, this squared derivative is normalized by the variance of a neuron's output over time, yielding:

$$\Psi^{stable} = \sum_i \psi_i^{stable} = \sum_i - \frac{\left\langle \left(\frac{d}{dt} A_i(t) \right)^2 \right\rangle_t}{\text{var}_t(A_i(t))}$$

where $\langle . \rangle_t$ and var_t denote the mean and variance over time, respectively. The derivative is implemented as a finite difference $A_i(t+\Delta t) - A_i(t)$, where Δt is the inter-frame interval, which in this case was 40ms. As this objective favours neurons whose output varies slowly over time, we will hereafter refer to it as the "stability" objective. The neuron specific value ψ_i^{stable} will be referred to as the individual stability of neuron i.

The stability objective depends exclusively on the properties of the stimulus input (akin to a feed-forward mechanism in biological systems) and does not include interaction between neurons in the network. Thus, optimising Ψ^{stable} alone would lead to a population of neurons with identical receptive fields. We introduce interactions between neurons in the network by adding a de-correlation term. This forces neurons to acquire dissimilar receptive fields:

$$\Psi^{decorr} = -\frac{1}{(N-1)^2} \left\langle \sum_i \sum_{i \neq j} \left(\sigma_{ij}^2(t) \right) \right\rangle_t$$

$\sigma_{ij} \equiv \frac{\left(A_i - \langle A_i \rangle_t \right) \left(A_j - \langle A_j \rangle_t \right)}{\sqrt{\text{var}_t(A_i) \text{var}_t(A_j)}}$ denotes the coefficient of correlation and N, the number of neurons

(N=200).

The total objective to be optimised is then defined as

$$\Psi^{total} = \Psi^{stable} + \beta \Psi^{decorr}$$

where β scales the contribution of the de-correlation term.

Optimization

At the beginning of the optimisation process, the weight matrix W is initialised to random values drawn from a normal distribution of mean 0 and unit variance. W is then normalized such that the variance corresponding to one input dimension over all neurons equals 1/N. Starting from these conditions Ψ^{total} is maximised using the gradient ascent method: For each iteration of the optimisation process the function Ψ^{total} and its analytically determined gradient $\frac{d\Psi^{total}}{dW}$ are computed over the entire set of natural stimuli. The weight matrix W is then updated in the direction of this gradient. The magnitude of this change is defined by the adaptive step-size procedure, as used in the implementation of Hyvärinen & Hoyer (2000). We regard the optimisation process to have converged when the relative change in the value of the objective function between successive iterations falls below 0.1%. We additionally perform a control analysis in which we disrupt the original temporal ordering of the image sequence used to train the network. The two frames making up a stimulus pair in this case are selected at random from the original set of frames.

All computations are performed using MATLAB (Mathworks, Natick, MA).

Analysis

We analyse the properties of the receptive fields in the network after convergence. First, the neuron's receptive field representation in input space is obtained by inverting the PCA on the weight matrix, W. Then, each receptive field is scaled individually such that the values of each pixel fall between 0 and 1. The scaling is the same for each colour channel and thus does not bias either the chromatic or the spatial properties of the optimised receptive field. To analyse colour and spatial content independently, the receptive fields are transformed from a colour channel representation (RGB format) into a representation separating hue, saturation and brightness channels (HSV format), achieved using a standard function in MATLAB ('rgb2hsv.m').

We characterise the chromatic selectivity of a receptive field by calculating the mean saturation across pixels. Neurons with a mean saturation greater than 0.2 are classed as chromatic neurons, while the remaining neurons are classed as achromatic. Spatial properties of the receptive field are assessed by measuring the anisotropy of the receptive field using standard methods (Jähne 1997). Briefly, the tensor of inertia is computed on the values from the brightness channel. Anisotropy is defined as the ratio of the difference between the tensor's long and short principal axis, divided by their sum. This measure is 0 for an isotropic (non-oriented) receptive field and approaches 1 for a perfectly oriented receptive field. Neurons with an anisotropy value less than 0.2 were considered to be non-oriented.

Results

Under all conditions analysed here, the optimisation process converges rapidly and reaches steady state after less than 60 iterations (Figure 2a). The optimised receptive fields are analysed in input space (by inverting the PCA on the weight-matrix, see Methods). As a starting point, we chose the following parameters values: 200 neurons, using PCA components 2-200, $\beta=5$ and full-wave rectifying neurons. Using this as a baseline, approximately 80% of the neurons exhibit colour selectivity. Figure 2b shows a complete set of receptive fields, sorted by the individual stability value (ψ_i^{stable}).

The stability objective (Ψ^{stable}) does not include interactions between neurons in the network. The de-correlation term (Ψ^{decorr}) force neurons to acquire different receptive fields. As stability and de-correlation are competing mechanisms being added into a single objective function, neurons with sub-optimal stability emerge as well as neurons with suboptimal de-correlation. As one expects, activities of neurons with high stability values ψ_i^{stable}, tend to be more correlated than those with low stability values (Figure 2c).

Visual inspection of the neurons in figure 2b already suggests that chromatic neurons tend to have higher values of ψ_i^{stable} than achromatic neurons. Quantitative analysis reveals that chromatic neurons indeed have higher stability (mean ψ_i^{stable}, -0.0041) than achromatic neurons (mean ψ_i^{stable}, -0.0053). There is a pronounced relation between a neuron's chromaticity and its individual stability (correlation coefficient: r=0.70, Figure 3a). This demonstrates that chromatic neurons have optimally stable responses to natural stimuli, while the achromatic stimuli are a consequence of sub-optimal stability due the de-correlation objective.

We also investigate the spatial properties of the neurons in the optimised network. Orientation selectivity was estimated from the degree of anisotropy in the receptive field (see Methods). In the simulation using the baseline parameter values described above, 66% of the neurons are non-oriented. In the simulations using the linear-threshold model, 81% are non-oriented. These results suggest that there is a relationship between spatial and chromatic properties of our model neurons. We find a strong correlation between chromaticity (mean saturation) and isotropy (defined as 1 minus anisotropy), both for the linear-threshold model (correlation coefficient: 0.72, Figure 3b) and full-wave rectifying model (correlation coefficient: 0.79, Figure 3c). Thus, colour selective neurons tend to be non-oriented, while achromatic neurons tend to be tuned for orientation.

Our results show that achromatic neurons emerge as a consequence of adding a de-correlation term (Ψ^{decorr}) to our objective function. The de-correlation objective permits interactions *between* neurons in the network, and results in receptive fields that are sub-optimal with respect to the stability objective alone. We went on to test whether increases in the relative contribution of the de-correlation objective (using the term β, see Methods) would further increase the fraction of achromatic neurons in the network. The limit case of $\beta \rightarrow \infty$ is simulated by omitting stability from the objective function altogether. For the simulations

using the linear-threshold model, the population average of mean saturation (chromaticity) is reduced from 0.50 where β=0 (no contribution from Ψ^{decorr}) to 0.19 where $\beta \to \infty$ (no contribution from Ψ^{stable}). The results from the simulations using the full-wave rectifying neurons are qualitatively similar, with a reduction in mean saturation from 0.5 (β=0) to 0.25 ($\beta \to \infty$). For both neuron types a 50% drop in mean saturation is achieved for values of β=20 (Figure 4a).

The reduction in mean saturation is associated with a reduction in the proportion of chromatic neurons. In the simulations using the linear-threshold model, increasing β from 1 to 20 reduces the proportion of chromatic neurons from 100% to 61%. In the full-wave rectifying model simulation, the same increment in β reduces the proportion of chromatic neurons from 93% to 63%. These results show that the proportion of chromatic versus achromatic neurons as well as the average chromaticity across the network is determined by the relative contribution of the de-correlation versus the stability objective. Therefore a single parameter, β, is sufficient to determine the proportion of chromatic versus achromatic neurons in the network.

Next, we investigate the robustness of our model to reductions in the dimensionality of the training input. We determine the change in mean saturation when as few as 25 principle components were used, roughly 1/8th of the principle components used in the main simulations (see Methods). The change in mean saturation across both linear and full-wave rectifying models, and for values of β between 5 and 10, was not larger than 53% (Figure 4b). This minimal dependence on the number of input dimensions contrasts with a study by Hoyer & Hyvärinen (2000) using independent component analysis, which reported a strong correlation between input dimensionality and the proportion of colour selective neurons. Increasing input dimension by a factor of 2.5 yielded an increase in the number of chromatic neurons of approximately 290% (estimated from Figure 10 in Hoyer & Hyvärinen 2000).

Doubling the number of independent components (from 100 to 200 dimensions) yielded a 77% increase in chromatic neurons, while in our simulations, which use PCA, the same increase in dimensionality produces an increase of just 11% (Figure 4c). This indicates that in simulations using ICA, the emergence of colour selective neurons depends on dimensionality of the input, while in simulations using the stability objective, colour selectivity is determined by the relative strength of the de-correlation objective.

We sought to verify that the results of our simulations are indeed a consequence of the temporal structure in the natural stimuli used as training input. Additional simulations were performed in which we destroy this temporal structure by randomly shuffling the frames comprising a stimulus pair. We compare the value of the objective function between initialisation and steady state conditions. In the simulations of full-wave rectified neurons, $\beta=10$, steady state values are 8% higher than at initialisation (from -2.23 to -2.04). In the simulations using the linear-threshold model, an increase of 4% is found (from -2.05 to -1.96). These values are modest compared to those found in the simulations with natural temporal ordering in the training input (from -1.96 to -1.32, or 33% in the full-wave rectified model simulations and from -1.72 to -1.25, or 27% in the linear-threshold model simulations).

The impact of shuffling the stimulus pairs was even more dramatic when the objective function was defined by Ψ^{stable} alone ($\beta=0$). In this case, the increase in the value of the objective function when using the natural temporal stimuli is more than 10 times larger (31% increase for the full-wave rectified model, 43% increase for the linear-threshold model) than for the simulations using the temporally shuffled stimuli (3% increase for the full-wave rectified model, 4% increase for the linear-threshold model).

We also determined the effect of shuffling the stimulus pairs on the emergence of colour selectivity in the network. Mean pixel saturation was 0.19 in the network trained on the shuffled stimuli, half the value that is obtained when the natural temporal order is intact (0.38). This resulted in the proportion of chromatic neurons in the network falling from 65%

in the naturally ordered condition to just 19% in the shuffled condition. Furthermore, most neurons in the simulation using the linear-threshold model are close to the chromatic/achromatic threshold of 0.2 (96% fall between 0.1 and 0.3 mean pixel saturation), whereas for naturally ordered condition, mean saturation shows a wider distribution (43% between 0.1 and 0.3). While the change in mean saturation is less pronounced for the full-wave rectifying model (0.24 compared to 0.29), a similar narrowing in chromaticity is observed (63% compared to 30% of neurons fall between 0.1 and 0.3 mean saturation, Figure 4d). This indicates that in the shuffled condition, most neurons cannot be clearly identified as either chromatic or achromatic.

Taken together, these findings indicate that natural temporal structure is critical to the attainment of an optimal solution for the stability objective function at the level of the network, and for the emergence of distinct chromatic and achromatic neuron populations.

Discussion

In this study we address how neurons selective for different stimulus dimensions can emerge from an initially homogeneous population. We have shown that optimising the stability objective alone yields non-oriented chromatic neurons. Forcing neurons, to acquire dissimilar receptive fields (and thus sub-optimal stability) leads to the emergence of a second subpopulation of oriented achromatic neurons. The stability of each neuron serves as system inherent measure to separate the two groups of neurons.

Furthermore, our simulations show that the relative size of each subpopulation is determined by a single parameter. Thereby, we have shown that the proposed objective function successfully segregates neurons into distinct populations that are selective to one property of the stimulus while being relatively non-selective to another property. By adopting complementary selectivities, a small number of neuronal populations can encode a complete

set of features in the stimulus independently of each other, and thus achieve a parallel mode of processing.

We considered two distinct neuronal models in our simulations, the linear-threshold model and the full-wave rectifying model. The results are similar regardless of which neuronal model is used, indicating that our objective function can succeed irrespective of the type of model. This result highlights the potential utility of our approach. Our objective function may provide insights into the mechanisms underlying learning and development, not just at early stages of the visual system, but at higher levels of the visual hierarchy as well. The generalisability of our approach also holds promise for its application in the development of artificial vision systems.

Simulations of the development of colour selective responses using natural stimuli as input has been addressed in several recent studies. These adopt a version of the sparseness principle, using independent component analysis (ICA), and use standard colour images (Hoyer & Hyvärinen 2000; Tailor et al., 2000) or hyperspectral images as training input (Wachtler et al. 2001). All studies find colour selective neurons, similar to those described here. However, neither study quantifies the relation of the neurons' spatial receptive fields to their chromatic properties. Here we find a strong correlation between chromatic and spatial properties. Another remarkable difference between the stability objective and the sparseness objective, as modelled using ICA, is the dependence of the latter approach on the dimensionality of the training input. This indicates that when using the ICA approach, the segregation of chromatic and achromatic subpopulations will depend on properties of the dimensionality of the external input. In the present approach this segregation is regulated by an internal parameter specifying the strength of the interactions between neurons within the network.

Earlier studies using the temporal coherence objective to learn invariant responses delivered a clear proof of concept using artificial (Földiak 1991) as well as artificially

transformed natural stimuli (Stone 1996; Wallis & Rolls 1997). This studies were motivated by the idea that, if different transformations of the visual input naturally happen on different timescales, selectivity and invariance towards these transformations can be extracted using solely this fact. For example, the invariance to position and selectivity to orientation results from local orientations being correlated over longer time scales than positions (Kayser et al. 2003). As the temporal structure of the stimulus is therefore the decisive property exploited by this type of objective function, it is crucial that we demonstrate that these results are confirmed when using stimuli that preserve temporal structure in the natural input. Our unique stimuli, derived from recordings made from a camera mounted on a freely behaving cat, provide image sequences that preserve the natural temporal structure. Note however, that we found it technically infeasible to use more realistic colour representations, as the current sampling rates of devices capable of recording hyperspectral images are several orders of magnitude below the correlation time constants observed in our natural videos (Kayser et al. 2003). Thus, we were not able to capture this aspect of the natural input. Note that the results of the ICA studies using hyperspectral images (Wachtler et al. 2001) are not qualitatively different to those in which the standard RGB-representation was used (Tailor et al. 2000, Hoyer & Hyvärinen 2000). Thus, we can be confident that the RGB format is adequate for our purposes. Furthermore, the present 'Catcam' videos were not recorded at sufficient temporal resolution to model the temporal properties of receptive fields. Modelling chromatic spatio-temporal receptive fields with the stability objective thus remains an interesting issue for future research.

There is physiological evidence that colour-sensitive neurons in primate V1 tend to be non-oriented, whereas achromatic neurons tend to exhibit the precise orientation tuning seen to be characteristic of V1 simple and complex neurons (Gouras, 1974; Lennie et al. 1990). This is further supported by psychophysical experiments that show that humans can resolve higher spatial frequencies for isochromatic patterns than for isoluminant patterns (Webster et

al. 1990, Sekiguchi et al. 1993). This raises the issue as to how our particular objective function might be implemented physiologically. A possible mechanism for the de-correlation objective is provided by inhibitory lateral connections in primary visual cortex. Physiologically, this process does not necessarily require synaptic changes at the tangential connections, but can also exploit modifications to afferent synapses that are in turn influenced by lateral interactions (Körding & König 2000). This mechanism utilizes action potentials propagating retrogradely through the dendritic tree (Stuart & Sakmann 1994). Building on the same physiological mechanism, a temporally asymmetric learning rule (Makram et al. 1997, Larkum et al. 1999) could plausibly subserve the function of the stability objective (Einhäuser et al. 2002).

The biological plausibility of our approach may go beyond the similarity between our simulated receptive fields and those found in visual cortex. We speculate that important properties of the visual system, including the establishment of distinct functional pathways, may develop using learning rules resembling the stability objective.

Acknowledgements

References

Bell, A.J. & Sejnowski, T.J. (1997) The "independent components" of natural scenes are edge filters. *Vision Res.* **37** 3327-3338.

DeYoe, E.A. & VanEssen, D.C. (1985) Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature* **317** 58-61.

Einhäuser, W., Kayser, C., König, P. & Körding, K.P. (2002) Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur. J. Neurosci.* **15** 475-486.

Földiak, P. (1991) Learning Invariance from Transformation Sequences. *Neural Comput.* **3** 194-200.

Gouras, P. (1974) Opponent-colour cells in different layers of foveal striate cortex. *J. Physiol* **199** 533-547.

Hietanen, J.K., Perrett, D.I., Oram, M.W., Benson, P.J. & Dittrich, W.H. (1992) The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp. Brain Res.* **89** 157-171.

Hoyer, P.O., Hyvärinen, A. (2000) Independent Component Analysis Applied to Feature Extraction From Colour and Stereo Images. *Network* **11** 191-210.

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.

Hurri, J.& Hyvärinen, A. (2002) Simple-Cell-Like Receptive Fields Maximize Temporal Coherence in Natural Video. *submitted*

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

Ito, M., Tamura, H., Fujita, I. & Tanaka, K. (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73** 218-26.

Jähne, B. (1997) Digital Image Processing - Concepts, Algortihms and Scientific Applications, 4th compl. rev. edn. Springer-Verlag, Berlin Heidelberg New York

Kayser, C., Einhäuser, W., Dümmer O., König P. & Körding K.P (2001) Extracting slow subspaces from natural videos leads to complex cells. In G. Dorffner, H. Bischoff, K. Hornik (eds.) *Artificial Neural Networks - (ICANN) LNCS 2130,* Springer-Verlag, Berlin Heidelberg New York pp. 1075-1080.

Kayser, C., Einhäuser, W. & König, P. (2003) Temporal correlations of orientations in natural scenes. *Neurocomputing, in press.*

Körding, K.P. & König, P. (2000) A learning rule for dynamic recruitment and de-correlation. *Neural Netw.* **13** 1-9.

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

Lennie, P., Krauskopf, J. & Sclar, G. (1990) Chromatic Mechanisms in Striate Cortex of Macaque. *J. Neurosci.* **10** 649-669.

Livingstone, M.S. & Hubel, D.H. (1984) Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* **4** 309-356.

Livingstone, M.S. & Hubel, D.H. (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* **240** 740-749.

Logothetis, N.K., Pauls, J., Poggio , T. (1995) Shape representation in the inferior temporal cortex of monkeys. *Current Biol.* 5 552-563.

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

Merigan, W.H. & Maunsell, J.H. (1993) How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16** 369-402.

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.

Rolls, E.T. (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos. Trans. R. Soc. Lond. B*. **335** 11-20.

Rolls, E.T. & Tovee, M.J. (1995) Sparseness of the Neuronal Representation of Stimuli in the Primate Temporal Visual Cortex. *J. Neurophysiol.* **73** 713-725.

Sekiguchi, N., Williams, D.R. & Brainard D.H. (1993) Efficiency in detection of isoluminant and isochromatic interference fringes. *J. Opt. Soc Am. A* **10** 2118-2133.

Sigala, N. & Logothetis, N.K. (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415** 318-320.

Stone, J.V. (1996) Learning Perceptually Salient Visual Parameters using spatiotemporal smoothness constraints. *Neural Comput.,* **8**:1463-1492.

Stuart, G. J. & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, **367** 69–72.

Tailor D.R., Finkel L.H. & Buchsbaum G. (2000) Color-opponent receptive fields derived from independent component analysis of natural images. *Vis. Res.* **40**:2671-2676.

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.

Wachtler, T., Lee, T.W. & Sejnowski, T. (2001) Chromatic Structure of natural scenes. *J. Opt. Soc. Am. A* **18** 65-77.

Wallis, G. & Rolls, E.T. (1997) Invariant Face and Object Recognition in the Visual Systems. *Prog. Neurobiol.* **51** 167-194.

Webster, M.A., DeValois, K.K. & Switkes, E. (1990) Orientation and spatial-frequency discrimination for luminance and chromatic gratings. *J. Opt. Soc. Am. A*. **7** 1034-1039.

Wiskott, L. & Sejnowski, T. (2002) Slow Feature Analysis: Unsupervised Learning of Invariances. *Neural Comp*. **14** 715-770.

Zeki, S. & Shipp, S. (1988) The functional logic of cortical connections. *Nature* **335** 311-317.

Captions

Figure 1

Three examples of the two consecutive images that are selected from the natural image sequence. Each pair of images forms the basis of the input to our model. The white squares indicate three of the twenty 30x30 pixel patches that are sampled from the same location in each frame pair.

Figure 2

(a) Development of objective function Ψ^{total} in the course of the optimisation process for a simulation of full-wave-rectifying neurons with $\beta=5$.

(b) Receptive fields produced following 60 iterations of the simulation shown in (a). The receptive fields are sorted according their individual stability value, with low stability values at the top of the figure and high stability values at the bottom.

(c) Coefficient of correlation between all the activities over one stimulus presentation of the neurons of panel b.

Figure 3

(a) Dependence of neuron chromaticity on individual objective value ψ_i^{stable}. Chromaticity was assessed by calculating mean saturation for each receptive field, averaged across pixels.

(b) Chromaticity versus anisotropy for $\beta=5$ and full-wave rectifying neurons.

(c) Chromaticity versus anisotropy for $\beta=5$, linear-threshold model.
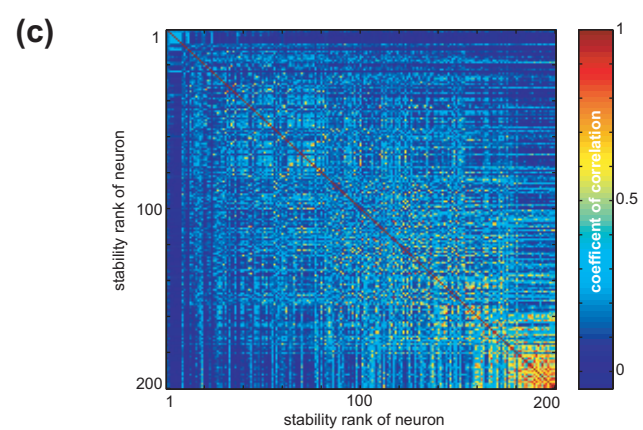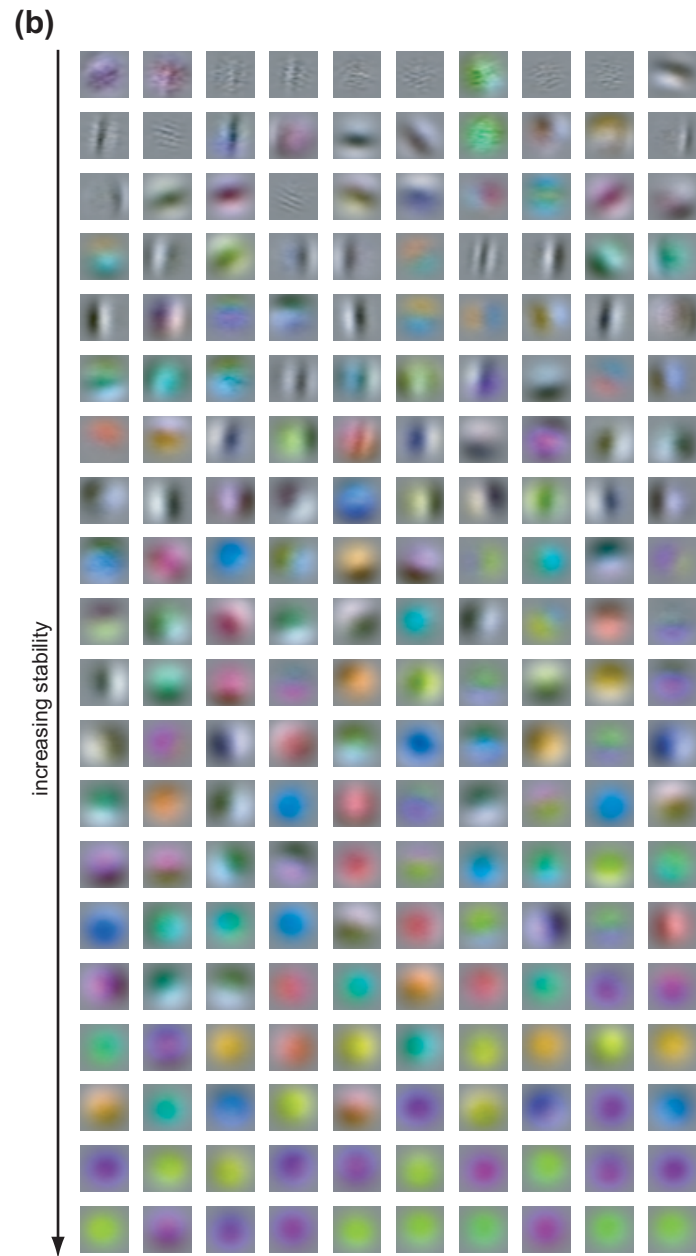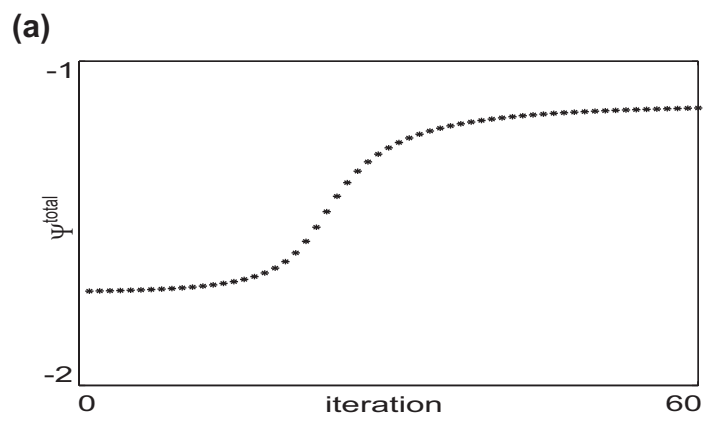
Figure 4
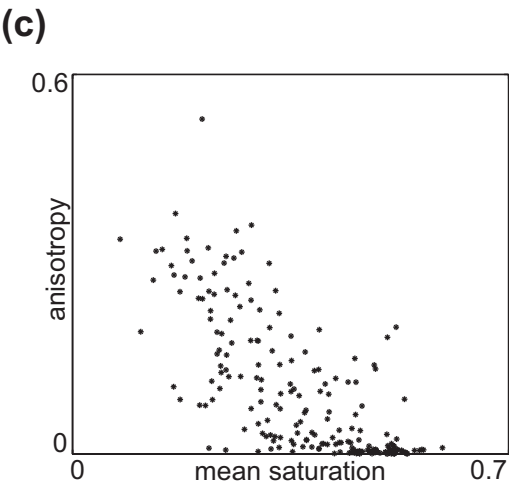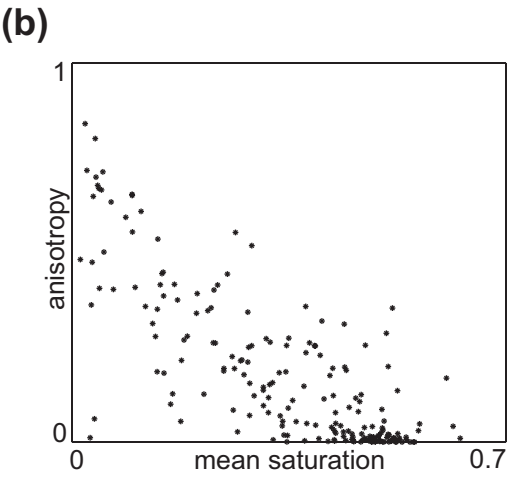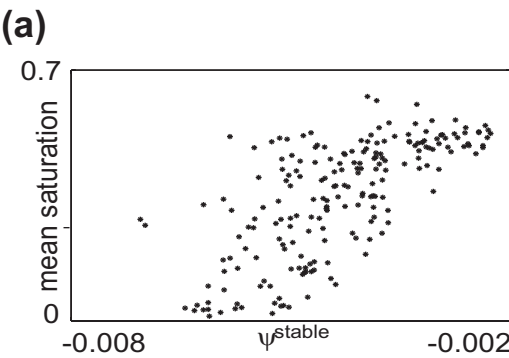
(a) Mean chromaticity versus β for the full-wave rectifying model (stars) and linear-threshold model (circles).

(b) Dependence of mean neuron chromaticity on PCA dimension for different values of β and the two neuron models.

(c) Percentage of chromatic neurons (mean saturation > 0.2) of (a) compared to results from ICA, redrawn from Hoyer & Hyvärinen (2000).

(d) Distribution of mean pixel saturation for the control condition of shuffled stimuli (left) compared to stimuli that` preserve the natural temporal structure (right). Top row shows results from the linear-threshold model, bottom row, the full-wave rectifying model.

Einhäuser et al. Figure 1

(a)

(b)

increasing stability

(c)

Einhäuser et al. Figure 2

(a)

(b)

(c)

Einhäuser et al. Figure 3

Einhäuser et al. Figure 4