AVAILABLE AT
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Neural Networks 16 (2003) 1353–1371

Neural Networks

www.elsevier.com/locate/neunet

2003 Special Issue

Modeling the adaptive visual system: a survey of principled approaches

Lars Schwabe*, Klaus Obermayer

Department of Computer Science and Electrical Engineering, Berlin University of Technology, FR2-1, Franklinstrasse 28/29, Berlin 10587, Germany

Received 2 December 2002; revised 28 July 2003; accepted 28 July 2003

Abstract

Modeling the visual system can be done at multiple levels of description ranging from computer simulations of detailed biophysical models to firing rate and so-called 'black-box' models. Re-introducing David Marr's analysis levels for the visual system, we motivate the use of more abstract models in order to answer the question of what the visual system is computing. The approaches we selected to review in this article concentrate on modeling the changes of sensory representations. The considered time-scales, range from the developmental time-scale of receptive field formation to fast transient neuronal dynamics during a single stimulus presentation. Common to all approaches is their focus on providing functional interpretations, instead of 'only' explanations in terms of mechanisms. Although the concrete approaches can be distinguished along different lines, a common theme is emerging which may qualify as a paradigm for providing functional interpretations for changes of receptive field properties, i.e. the dynamic adjustment of sensory representations to varying external or internal conditions.

Keywords: Visual cortex; Models; Infomax; Bayesian inference; MDL principle

1. Introduction

When should an organism exploit its environment in order to benefit from its previously acquired knowledge, and when should an organism continue to explore its environment to acquire even more knowledge for later exploitation? This famous and well-known *exploration–exploitation tradeoff* (see, e.g., Kaelbling, Littman, & Moore, 1996) is a rather high-level problem relevant for understanding the behavior of biological organisms and for building intelligent artificial agents. This problem, however, has a low-level counter part which in turn is highly relevant for understanding biological sensory systems and, of course, for building the sensing devices of intelligent artificial agents. The so-called *plasticity–stability tradeoff* (Grossberg, 1976) refers to the problem of deciding when and how to adapt the internal sensory representation of the outside world.

On the one hand, subsequent processing stages rely on stable and reproducible sensory representations when performing their computations like, e.g., the decision between exploration and exploitation. On the other hand, sensory representations need to be plastic in order to ensure a high fidelity representation of sensory stimuli whose behavioral relevance may not be known beforehand. In other words, sensory representations need to be adaptive. Indeed, adaptation is a widespread phenomenon in nervous systems, and it happens on multiple time-scales. In the context of the early visual system, the activity-dependent refinement of cortical maps happens during a critical phase which lasts weeks, perceptual learning occurs during hours and days, contrast-adaptation in the primary visual cortex works on the order of a few seconds, and the fast dynamics of receptive field properties on the order of a few tens or hundred milliseconds may also be viewed as an adaptive process.

Together with the apparent general applicability of the idea of adaptation comes a certain vagueness: for example, what exactly do contrast-adaptation and the development of cortical maps have in common? Is it indeed justified to think of dynamic receptive field properties or even attentional top–down modulations in the visual cortex as some sort of adaptation? Clearly, understanding adaptation only in the narrow sense of a neuronal fatigue after prolonged stimulation falls too short. However, with this review we want to demonstrate that adaptation, when understood in

* Corresponding author. Tel.: +49-30-314-24753; fax: +49-30-314-73121.
E-mail address: schwabe@cs.tu-berlin.de (L. Schwabe).

0893-6080/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2003.07.009

a broader sense, may serve as a paradigm for providing functional interpretations of experimentally characterizable changes of receptive field properties on multiple time-scales. We suggest to conceptualize these changes as adjustments of sensory representations to changing external or internal conditions to ensure their optimal usage. In order to apply this view to a concrete phenomenon, however, one needs to make explicit (i) which conditions are changing, and (ii) what is the measure according to which the optimality of a representation is to be judged. Here we review models of the visual system at multiple time-scales and focus on so-called principled approaches, because they make explicit both aspects.

This article is structured as follows: In Section 1.1 we give a very short introduction to the experimentally well characterized early part of the primate visual pathway. In Section 1.2 we justify our focus on principled approaches by re-introducing the three analysis levels for the visual system proposed by David Marr more than 20 years ago (Marr, 1982), and in Section 1.3 we suggest some taxonomies for grouping organization principles for the visual system. In Sections 2–4 we review models for adaptation processes in the visual system, and in Section 5 we finish with a short summary and a sketch of what we think are promising directions for future research.

1.1. The primate early visual pathway

Most (but not all) approaches we review in this article are intended to be simple models of primary visual cortex. In these models the inputs are pixel images, and the neuronal responses are modeled as functions of these images. Thus, these models abstract from the early visual pathway. Often this can be justified, because explicitly modeling the signal transmission from the retina to the primary visual cortex would make the model too complicated and hide the main idea behind it. For completeness, however, let us briefly summarize the anatomy and physiology of the early visual pathway. A more detailed description can be found in (Kandel, Schwarz, & Jessel, 1996).

Fig. 1(a) sketches the overall anatomical structure of the early visual pathway. When a cat or monkey fixates its environment, light that falls into the eye is focused by the cornea and the lens to form two images on the left and right retinae. The part of the world that contributes to the image formed on the two retinae is called the visual field of the animal. Each retina transforms the incoming light intensity distribution into spike patterns, which are transmitted by the two optic nerves into the central nervous system. Each optic nerve consists of roughly 10^6 fast conducting axons, almost all of which target two structures within the thalamus, which are called Lateral Geniculate Nuclei or LGN. At the optic chiasm, each optic nerve branches such that one half of the fibers targets the 'contralateral' LGN at the side opposite to its origin, whereas the rest contacts the 'ipsilateral' LGN at the same side as the eye of origin. The cross-over of the nerve fibers occurs in a highly ordered fashion, which ensures that each LGN receives the fibers from the ipsilateral parts of both retinae. Thus, each LGN processes the signals from the contralateral hemisphere of the visual field.

Proceeding from the LGN, another less concentrated bundle of fibers, which is referred to as the optic radiation, contacts the primary visual cortex (also referred to as area V1). In contrast to the optic nerve, the optic radiation does not cross hemispheres. Hence, analogously to the LGN, each hemisphere of the primary visual cortex processes visual information from the contralateral hemisphere of the visual field. From V1, two major output streams can be divided. The first stream projects from V1 to higher visual areas, whereas the second stream projects back to the LGN and other deep structures. So far we have summarized anatomical basics of the pathway from the retina to V1. Let us now move on to the prototypical response properties of cells in the retina, LGN and V1.

Neurons in the retina and the LGN are sensitive to stimuli only within a small, roughly circular part of the visual field, which is called their receptive field. If this receptive field is illuminated with a small bright spot, the responses of most cells are not uniform but depend on the spot's position within the receptive field. Some neurons are excited, when the spot hits the center of the receptive field, whereas their response is suppressed by illumination of the surround. These cells are called ON-center cells. The so-called OFF-center cells are inhibited by illumination in the center and excited by surround illumination. Diffuse illumination evokes almost no response. However, compared to neurons in the retina and the LGN which have nearly circular receptive fields, cortical response characteristics show a qualitatively new feature: many cortical neurons respond selectively to contrast lines, bars, or gratings of a certain orientation within the receptive field.

Fig. 1(b) shows the spike train of a cortical neuron in macaque V1 in response to an oriented bar moved across the classical receptive field (after (Hubel & Wiesel, 1968)). The neuron responds most strongly for a given orientation of the bar, and its response declines strongly with increasing difference between the stimulus orientation and the preferred orientation of the cell. In Fig. 1(c) the average spike rate of an orientation selective neuron is plotted as a function of the orientation of a moving grating used for visual stimulation (after (Levitt & Lund, 1997)). This curve is called an 'orientation tuning curve'. The response of this neuron is tuned to an orientation of approximately 30°. The neuron responds similarly to an orientation of 210°, which corresponds to a grating of the same orientation but moving into the opposite direction. Fig. 1(d) shows a hypothetical wiring scheme between the neurons in the LGN and an orientation-tuned neuron in V1: a projection from ON- and OFF-center cells in the LGN with their receptive field centers being located along an axis whose orientation matches the preferred orientation of the cortical neuron

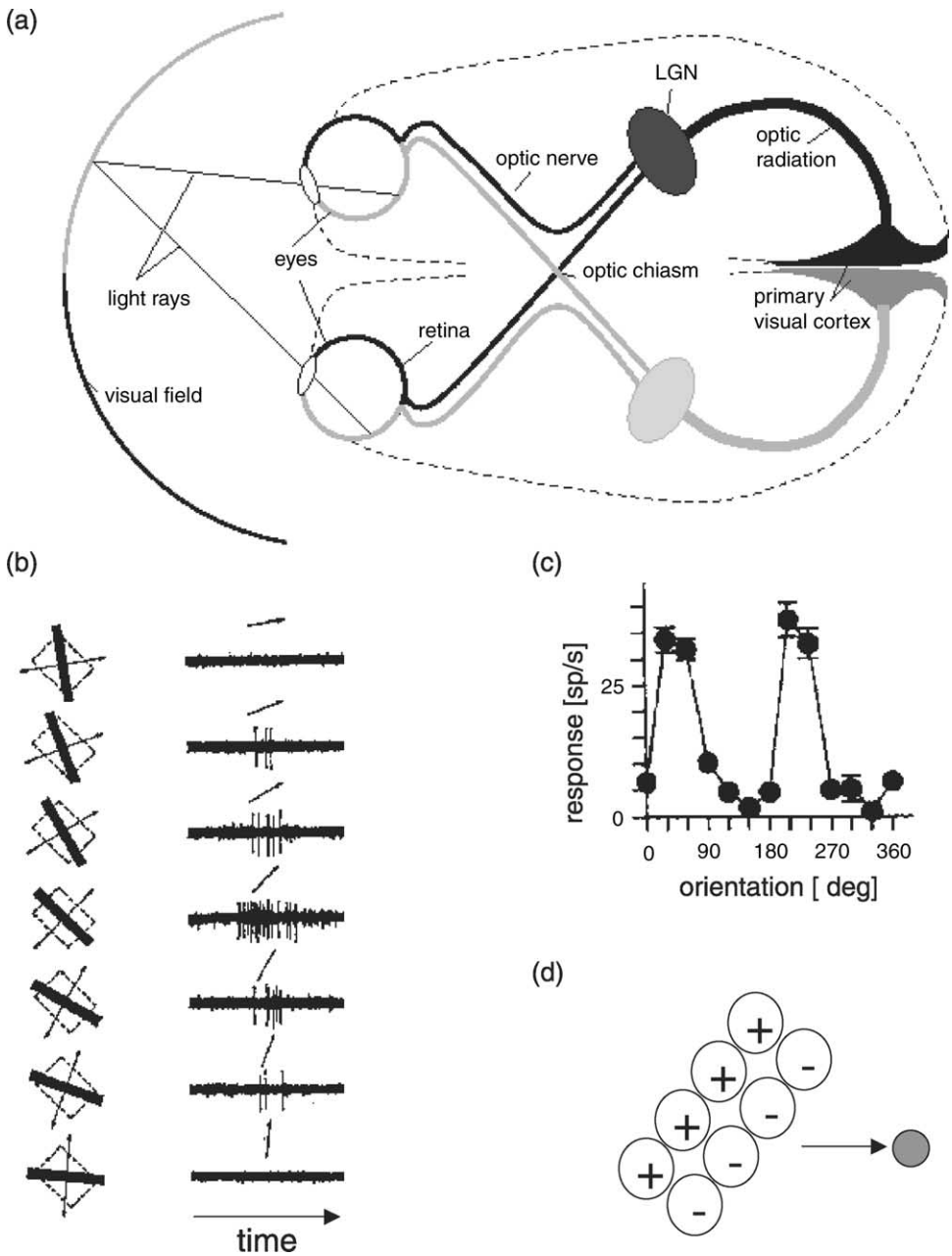

Fig. 1. (a) The primate early visual pathway. (b) Responses of an orientation-selective neuron in the primary visual cortex when an oriented bar is moved across the classical receptive field (rectangular box) (c) Typical orientation tuning curve with the mean firing rate plotted against the orientation of the moving visual stimulus. (d) Spatial layout of how to connect ON-center and OFF-center neurons in the LGN to cells in primary visual cortex for producing orientation selectivity predicted by a feedforward model.

would induce orientation selectivity. In other words, the cortical neuron becomes a detector of oriented edges with the strongest response to bright illumination in the ON-center regions and absence of illumination in the OFF-center regions. The response vanishes for diffuse stimulation.

Of course, response properties in the primary visual cortex have been characterized in much more detail than sketched here, but this so-called orientation selectivity is the paradigmatic example for feature selectivity in the visual system. It may come as a surprise, but until now the details of the possibly species-specific mechanisms underlying orientation selectivity are not known, and the wiring scheme in Fig. 1(d) is only one possible explanation among a continuum ranging from the feed-forward scheme suggested by Fig. 1(d) which involves the LGN and V1 (Hubel & Wiesel, 1977) to pure recurrent models relying only on the intra-cortical circuitry within V1 (Adorjan, Levitt, Lund, & Obermayer, 1999a). Furthermore, it also applies only to simple cells, which in addition to a preferred orientation have a preferred phase, i.e. reversing the light and dark regions of a stimulus would silence such neurons.

The phase-invariance of complex cells can't be explained with such a wiring scheme. Recent and in-depth reviews of this topic can be found in (Ferster & Miller, 2000; Sompolinsky & Shapley, 1997). However, for our considerations the concrete wiring scheme itself is only of secondary interest. The approaches reviewed here ask for why the receptive fields of cells in primary visual cortex are oriented at all (see (Olshausen & Field, 1996) in Section 2.2) or why complex cells are phase invariant (see (Wiskott & Sejnowski, 2002 in Section 2.2). Let us now re-introduce David Marr's analysis levels for the visual system in order to justify the principled approaches.

1.2. David Marr's analysis levels

David Marr (together with Tomasio Poggio) developed a general account of information-processing systems in general and of visual systems in particular in terms of three levels of analysis (Marr, 1982): (i) the level of the computational theory of the system, (ii) the level of algorithms and representations, which are the mathematical descriptions of the computations, and (iii) the level of the neuronal implementation of these algorithms and representations. In computer science a similar distinction between the (formal) specification of a problem, its algorithmic solution and the concrete implementation has emerged as the ideal case for developing algorithms/software. The challenge, however, of actually applying Marr's three levels of analysis is twofold: first, in the neurosciences one is not designing systems as in computer science, but analyzing them. Second, a clear separation between the three levels may not always be possible (or even desirable), because in neuronal systems the actually realized algorithms are certainly not independent of the 'neuronal hardware' they are running on. In other words, although a recurrent neuronal network can in principle simulate an arbitrary Turing machine (see (Minsky, 1967) and (Siegelmann & Sontag, 1991)), i.e. every conceivable algorithm can be implemented, this does not mean that the brain's microcircuits should be viewed in this particular way. Let us now illustrate Marr's basic idea with a metaphor taken from computer science.

Although in computer science algorithms are often developed in an ad hoc manner, the ideal case is a step-by-step development starting from an abstract specification of the problem to be solved and ending with executable code. Consider the problem of sorting a sequence of numbers $S = (i_0, i_1, \ldots, i_N)$ according to a relation $<$. The first step for developing a sorting algorithm is to formalize the requirements via, e.g., using the predicate *sorted* defined as

$$\text{sorted}\,(S): \; i_0 < i_1 < \ldots < i_N$$

Every sorting algorithm *sort* has to transform an input sequence S_{in} into an output sequence $S_{\text{out}} = sort\,(S_{\text{in}})$ so that after the sorting algorithm has terminated *sorted* (S_{out}) holds, i.e. the output sequence S_{out} is sorted. We like to point out three important aspects of this metaphor. First, only algorithms for which the predicate sorted holds for every input sequence S_{in} should be called a sorting algorithm. Second, multiple sorting algorithms could solve the problem of sorting a sequence of numbers, e.g. quick-sort, bubble-sort or selection-sort. They may differ, however, in terms of their computational efficiency. Third, each sorting algorithm could be formulated, compiled and executed on a variety of different hardware platforms, e.g. a personal computer, a mechanical computing device, a Nintendo Gamecube or a mobile phone.

Applied to David Marr's levels of analysis, the formulation of the predicate sorted corresponds to level (i). An algorithm can only be developed if its requirements have been formulated. Correspondingly, the coarsest description of a sensory system is a (verbal) description of what the system is computing, even if the details of this computation are not known yet. The formulation of a particular sorting algorithm like quick-sort or bubble-sort corresponds to level (ii). In the same way as the problem of sorting a sequence of numbers could be solved by a multitude of different algorithms in different ways, a sensory system could compute representations of the outside world in a more or less efficient manner. Another parallel is the type of the used representations. Sorting algorithms are usually formulated for particular data-types like linked lists, arrays or heaps. For a particular sensory system one can ask, whether the representation is discrete or continuous, noisy or reliably. Finally, the hardware used to execute the software implementation of a sorting algorithm corresponds to level (iii). For example, once assumed that the representations of a particular sensory system are continuous, one has to make explicit the neuronal substrate of these representations. Continuous representations could be realized at the neuronal level in a multitude of different ways. It is conceivable that the time-averaged firing rate of individual neurons corresponds to the represented continuous quantity, but the average activity of a population of neurons as well as time-delays between action potentials are also candidate substrates. On the other hand, discrete representations could be realized locally simply by the presence or absence of action potentials of selected neurons. A more global realization of discrete representations could involve the synchronized activity of large groups of neurons. The set of neurons which at a given time synchronize their action potentials could then be identified with a discrete representation.

Of course, despite these parallels an important difference is obvious: once the problem to be solved has been specified, a computer scientist can design an algorithm to solve this problem with the only constraints being the expressive power of the used formal programming language. Evolution always had to re-build and adapt systems to the environmental demands where only available resources could be utilized. Thus, it could well be that evolution has produced solutions which are only just good

enough and not optimal, as it would have been possible if the system is designed from scratch instead of being the result of modifying an already existing system.

However, during the last 15 years or so a systematic approach to modeling sensory systems has been developed. The general method is the following: First, one simply guesses the problem solved by the investigated sensory system. Second, one formalizes a mathematical criterion according to which the performance of the sensory system is evaluated. Third, one derives the optimal solution to this problem in terms of a mathematically formulated algorithm, which could in principle be implemented on a computer. Then, one tries to solve the same problem with a mathematical model, which is subject to the constraints of the modeled sensory system and compares the performance of this solution with the optimal one. As a last step one needs to compare the resulting predictions of the constrained model with experimental data like measured receptive field properties. If the model predictions match the measured data, then one can hope to have gained some insight into what and how well the investigated system is computing, because this was formulated in the first place.

This approach is clearly a top–down approach in the sense that the sensory system's function is formulated first and the neuronal machinery used for realizing this function is derived afterwards. Marr's analysis levels seem to support this approach. Furthermore, it is similar to the ideal case in computer science, which starts from a specification of a problem and then proceeds via well-defined refinement steps towards the executable code. Therefore, one may also refer to it as 'analysis by design' in the sense that a solution to a problem the sensory systems is guessed to solve is designed. We already noted, however, that evolution may not have found optimal, but only good enough solutions. Furthermore, when modeling biological visual systems the primary goal is still an analysis of the investigated system instead of designing an artificial vision system. Hence, one may question the validity of this 'analysis by design' in favor of a bottom–up approach starting from models of neuronal circuits which are then analyzed in terms of the computations they perform. Although in this article, we focus on top–down approaches only, we view both approaches as being distinct and almost equally important.

1.3. A taxonomy for computational principles

At a very general level, one agrees that sensory systems are computing representations of the 'outside world' which serve as the basis for an animals behavior. Concrete approaches, however, can be distinguished along multiple different lines. For example, consider the systematic top–down approach to modeling sensory systems sketched in Section 1.2. It involves a criterion for evaluating the sensory system's performance, which is usually formulated in terms of an objective function. Fig. 2(a) and (b) shows the information flow between an agent's sensory system, the read-out and action selection and the environment. In both cases the agent perceives its environment via its sensory system. Adapting the sensory system as dictated by an objective function, however, involves an error signal. We use the source of this error signal to derive a useful taxonomy according to which modeling approaches might be grouped.

Almost all proposed approaches use *unsupervised or self-supervised* organization principles, mainly because supervised principles have been viewed to be less plausible, since they require comparing the actual outputs of the sensory system with desired outputs in order to compute the error signal. Indeed, every biologically plausible organization principle needs to predict adaptation rules, which can

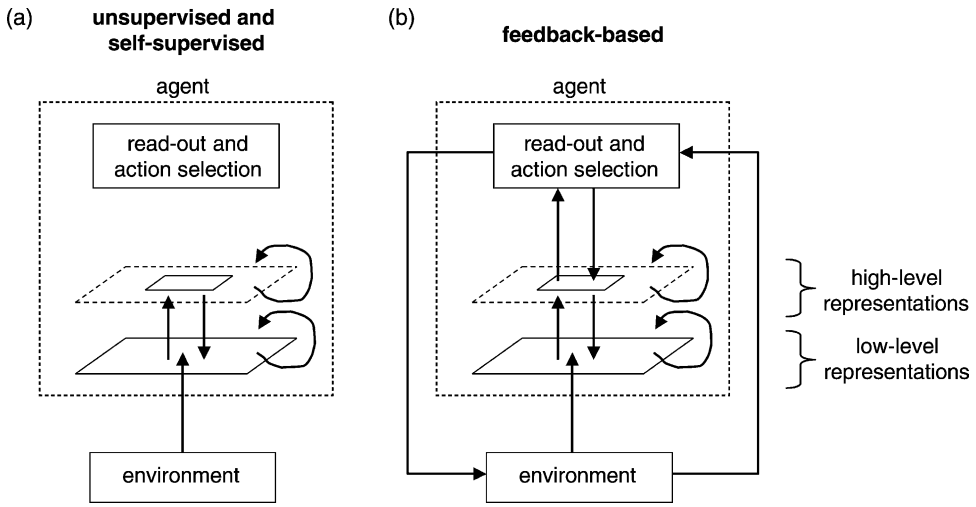

Fig. 2. Suggested taxonomy for organization principles for sensory representations. (a) Unsupervised and self-supervised principles do not rely on any feedback from the environment. All used information about the environment is already present in the feedforward input. (b) Feedback-based principles utilize feedback from the environment which is not present in the raw feedforward input like reinforcements.

be computed based on locally available signals. In unsupervised and self-supervised organization these signals are computed within the sensory system itself (see solid arrows in Fig. 2(a)). Neurons being part of high-level representations may send their output signals back via feedback projections in order to improve low-level representations, but the computed error signals do not incorporate other information about the environment than already present in the feedforward input.

In contrast, *feedback-based* organization principles utilize feedback from the environment in order to improve sensory representations (see Fig. 2(b)) which would allow for a task-specific adaptation and reorganization of a sensory system. Note, however, that feedback-based principles do not necessarily need to be supervised in the sense of relying on desired outputs of the sensory system directly provided by the environment, because those could be guessed internally, e.g. based on reinforcements received from the environment. To the best of our knowledge, no biologically plausible feedback-based organization principle for the visual system has been proposed so far, although the results of Ullman and co-workers (reviewed in Section 2.2) suggest that incorporating task-specific feedback from the environment is likely to play a role in shaping high-level representations. Although recent evidence from human psychophysics suggest that such a mechanisms may also operate in human perceptual learning (Seitz & Watanabe, 2003) of low-level visual features, it is not clear to what extent the organization of the visual system during the developmental time-scale depends on this type of feedback from the environment. Every approach we review in this article can be characterized as being either unsupervised or feedback-based. However, we decided to group the approaches reviewed here along the time-scale during which a change of receptive field properties occurs instead along this unsupervised/feedback-based taxonomy, because we think of this as another usefull taxonomy.

2. The developmental time-scale

Most principled approaches to modeling the early visual system have been made in deriving receptive field properties. Since in this article we selected the biological perspective of adaptation at multiple time-scales as a way to group these and other approaches, we do not even attempt to provide a comprehensive review of these works, but focus on some selected and representative works only.

As already sketched in Section 1.2, the general method followed by many authors is to formulate an objective function and then to optimize the free model parameters which describe properties of the neuronal system like receptive field profiles. A very prominent approach is efficient encoding. The basic idea is to recode the raw sensory inputs into neuronal firing patterns in a way that preserves all the information present in the input. The recoding, however, should also be efficient in the sense that the average 'neuronal description' of the sensory input is as short, i.e. 'cheap', as possible, which can turn out to be useful for a biological system, if the energetic costs of producing actions potentials are considered. Furthermore, short description lengths of sensory inputs allows for faster reactions times. Minimizing the average length of the 'neuronal description' of sensory inputs could be achieved by assigning short descriptions to frequently occurring inputs and longer descriptions to only rarely occurring inputs.

However, one may question whether it is a valid approach to characterize sensory systems as communication channels optimized for recoding their inputs so that all the information present in the input is preserved. For example, it is conceivable that sensory systems perform a preprocessing that leads to a high fidelity representation of only the behavioral relevant stimuli, which do not necessarily need to be the most frequently occurring ones. Nevertheless, applications of the efficient encoding idea to early sensory systems have been quite successful in the past indicating that at least to early processing stages this idea might be applicable.

On a developmental time-scale, the idea of efficient encoding with efficiency corresponding to short neuronal descriptions of the sensory inputs might indeed be applicable. In the context of adaptation at the environmental time-scale, however, we suggest the extraction of invariants as another goal of sensory systems (see Section 3.2) which turns out to be a side effects of optimizing a sensory systems based on the idea of efficient encoding.

In Section 2.1 we first introduce the idea of efficient encoding in the context of Rissanen's minimum description length framework (MDL) framework (Rissanen, 1989), because it provides a clear mathematical framework which also allows to express constraints on the sensory representations. This is important when modeling sensory systems, because those are always subject to biologically constraints like the non-negativity and limited range of neuronal firing rates.

2.1. Probabilistic inference and efficient encoding

We now shortly introduce the Bayesian approach to modeling the visual system and Rissanen's MDL framework. We also compare both approaches with an application of regularization theory to visual information processing.

A nearly paradigmatic assumption in visual neuroscience is that the visual system's goal is to compute internal representations of the external world, but for which specific purpose these representations are computed is still an open question. It is also an important question, because some representations may be more suited for a given purpose than other candidate representations. One idea is to view the visual brain as performing probabilistic Bayesian inference. In a rather general setting, this has been formulated in

Kersten and Schrater (1999), and Barlow recently envisioned its potential usefulness for investigating neuronal representations (Barlow, 2001).

In order to perform probabilistic inference, at least two things are needed: a model for how a given state H of the environment gives rise to the two-dimensional retinal image I, and a model for the prior probability of the environment's state. Denoting with $G(I|H)$ the probabilistic model for how the environment produces the images I and with $P(H)$ the prior probability for the environment's state, we can apply Bayes' rule $Q(H|I) = G(I|H)P(H)/P(I)$ to obtain the posterior $Q(H|I)$ for the environment's state given the observed image I. It is not clear, however, whether the visual brain indeed performs this type of probabilistic inference. If it does, it is also an open question which model $G(I|H)$ and prior $P(H)$ are used. Furthermore, how should neuronal activation patterns in the visual system be interpreted in terms of the above inference rule? Do neuronal activation patterns induced by visual stimulation somehow represent the full probabilistic information contained in the posterior $Q(H|I)$ as shown to be possible in Zemel, Dayan, and Pouget (1998), or do neuronal activation patterns represent the most likely state of the environment only? These issues are reviewed in greater detail in Pouget, Dayan, and Zemel (2000). Note that these questions directly relate to the levels (ii) and (iii) of the analysis levels proposed by David Marr.

Considering visual information processing as Bayesian inference is a mathematically elegant and indeed promising approach, but it might be applicable only to the computation of high-level representations much more downstream the processing hierarchy than the primary visual cortex, e.g. the inferior temporal cortex in the ventral stream which is though to perform object recognition. Furthermore, it is also only applicable as long as the underlying states of the environment can be formulated as being mutually exclusive. If the neuronal activation patterns in the primary visual cortex are also thought of as representing (probabilistic) information about the 'state of the environment', then these states are far from being intuitive, because in the primary visual cortex the visual world is viewed through the small pinholes of cortical neuron's classical receptive fields. Although experimentally so-called contextual effects from outside the classical receptive fields of neurons in the primary visual cortex (see also Section 4.2) have been reported and well characterized, the ever-changing visual stimulation due to eye movements makes its conceptually not straight forward to apply the perspective of Bayesian probabilistic inference to the primary visual cortex. It seems much more intuitive to consider activation patterns in the primary visual cortex as being representations based on which subsequent processing stages may perform probabilistic Bayesian inference.

The key idea behind the Bayesian perspective is to consider every observed, processed and propagated quantity as a random variable characterized by its probability density or distribution function. If one is interested in single quantities instead of the full probabilistic information contained in, e.g., the posterior, the Bayesian framework dictates to integrate over the probability density functions to obtain expectations or conditional expectations. Another method for extracting a single quantity based on the posterior probability density is to select the state of the environment, which maximizes the posterior given the observed image. This method is known as the maximum a posteriori method (MAP), but it is much more related to parameter estimation than to Bayesian inference in its original sense.

Rissanen's MDL framework (Rissanen, 1989) focuses on the efficient description and communication of observed data instead of inferring underlying causes which makes it a promising method to model neuronal representations even in early sensory systems like the primary visual cortex. Interestingly, for encoding models which utilize priors the MDL principle suggest a method formally equivalent to the MAP method (Rissanen, 1989).

Let $\{M(I;w)\}$ denote a family of (not necessarily probabilistic) models parameterized by w which describe observed images I. For a given observable image x the MDL principle dictates which model w to select in order to communicate the observed image x. The communication is accomplished by first sending the selected model w to the receiver and then the description of the image x using the already communicated model w. The MDL framework then dictates to select the model w which minimizes the overall code length (see below for its definition), i.e. the length for encoding the model w and the length of encoding the observed image x using the model w.

If we consider probabilistic models, i.e. $M(I;w)$ is the probability to observe image I given the model w, the MDL principle coincides with the Maximum Likelihood principle known from parameter estimation which states to select the w which maximizes the probability $M(I = x;w)$. If the $M(I;w)$ are probabilistic models composed of a model $G(I;w)$ and a probability density function $P(w)$ for the parameters, the MDL principle states to select the w which minimizes the description length $L = -\log G(I = x;w) - \log P(w)$ which is equivalent to selecting the w maximizing the posterior $Q(w|I = x) = G(I|w)P(w)/P(I)$. Within the MDL framework the negative logarithms of probabilities are used as approximations to code lengths, because a coding scheme which assigns to w a code of length -log $\times$ $P(w)$ asymptotically reaches the limit $H[W]$ for the average code length (Cover & Thomas, 1991). Here, $H[\cdot]$ denotes the entropy. In other words, in this case the MDL principle suggest a model selection formally equivalent to the MAP method from parameter estimation in a Bayesian framework. Note that now one can use the prior $P(w)$ to express properties and constraints of the representation system instead of assumptions about the environment's state. As an example, the $P(w)$ may express the desire to have so-called sparse representations (see Section 2.2). However,

the specific way in which the MDL principle should be applied to understand neuronal representations in terms of population codes is not clear yet, and the approach reviewed in Section 4.2 may turn out to be just a starting.

Now we can define the efficiency of a representation as the average length of the image description. In other words, parameters which change at the developmental time-scale like receptive fields should be adjusted so that when selecting a representation w of an image x by minimizing $L = -\log G(I = x; w) - \log P(w)$ the expected length $E[L(x)]$ averaged over the stimulus ensemble is minimized.

Finally, we also briefly like to point out the formal similarity of this objective function with a regularization theory approach to visual information processing (Poggio, Torre, & Koch, 1985): Regularization theory uses regularizers in objective functions in order to constrain the space of possible solutions to so-called ill-posed problems. Note that the prior in MAP parameter selection and its negative logarithm in the MDL framework are formally equivalent to regularizers. If the goal of the visual system is to infer the states of the environment, then this is indeed an ill-posed problem, because the information present in the 2D activation patterns of the two retinae about the 3D world is necessarily incomplete. Hence, using the regularization theory approach again allows us to formulate objective functions which need to be optimized for particular parameters of the sensory system similar to the MAP method, but now the regularizers again may correspond to prior knowledge about the environment instead of constraints for the sensory system as in the MDL framework.

These considerations reveal a great flexibility for formulating objective functions, but they do not prevent one from explicitly stating the principles based on which the objective function has been formulated. They may differ for different brain areas. For example, the MDL principle might be applicable to the primary visual cortex, whereas the Bayesian perspective might be applicable to circuits in the prefrontal cortex reading out high-level representations in the inferior temporal cortex.

2.2. Deriving receptive fields: low level representations

Let us now consider receptive field derivations for low-level representations. Although not stated in this way, the example (Olshausen & Field, 1996) we consider first intended for deriving receptive fields in the primary visual cortex can be viewed as an instance of the MDL principle.

The primary visual cortex is composed of recurrent microcircuits, and a mathematical model accounting for this recurrency is certainly desirable, but as a mathematically tractable first step one may model the neuronal responses simply as continuous valued firing rates a_i which are used as coefficients in a linear basis function model. In Olshausen and Field (1996) the following model

$$G(I(\mathbf{r})|\mathbf{a}) = N\left(\mu = \sum_i a_i \phi(\mathbf{r}),\ \sigma^2 = \sigma\right)(I(\mathbf{r}))$$

was used, where $I(\mathbf{r})$ is the intensity value of the 2D image at the position $\mathbf{r} = (r_1, r_2)$, the firing rates a_i are coefficients, the $\phi_i(\mathbf{r})$ are fixed basis functions, and $N(\mu, \sigma^2)$ denotes a 1D Gaussian density with mean μ and variance σ^2 to account for additive noise. Defining a prior

$$P(\mathbf{a}) = \prod_i \frac{1}{Z_S} \exp(-S(a_i))$$

for the coefficients one can utilize the following stochastic approximation algorithm in order to optimize the model parameters: For a given small image patch $I(\mathbf{r})$ selected randomly from a database of natural images one finds the coefficients $\mathbf{a}$ which minimize $L = -\log G(I(\mathbf{r})|\mathbf{a}) - \log P(\mathbf{a})$ via gradient descent. Then one adapts the parameters of the basis functions ϕ_i in order to minimize the expected length $E[L]$ averaged over the stimulus ensemble. Sparse coding can be achieved by using so-called sparse priors, i.e. via $S(a_i) = |a_i|$ or $S(a_i) = \log(1 + a_i^2)$. These priors enforce representations where the neurons are active only rarely which might be desirable from an energetic point of view, because firing action potentials is metabolically costly. Furthermore, the read-out of a sparse representation by subsequent processing stages might be easier compared to a non-sparse but dense representation.

Now, one may ask which basis functions should be chosen. Olshausen and Field modeled the basis functions $\phi_i(\mathbf{r})$ as long vectors $\mathbf{w}_i$ with $12 \times 12 = 144$ entries corresponding to the pixels in 12×12 image patches taken randomly from natural images. After the stochastic approximation algorithm converged, the 12×12 receptive fields (see Fig. 3, taken from Olshausen and Field (1996)) turned out to have shapes like Gabor functions: they are localized, oriented and bandpass which is a property of experimentally measurable receptive fields in primary visual cortex.

The most striking aspect of this work is that the resulting receptive fields share properties with experimentally measured receptive fields which is not achieved when changing the sparse to a Gaussian prior. This points to an important role for this constraint. Recent work by van Vreeswijk (2000b), however, suggest that receptive fields similar to Gabor functions may emerge from an optimization procedure which maximizes the mutual information (see also Section 3.1) between patches taken from natural images and the neuronal responses even without enforcing sparseness. The striking difference between both approaches is that van Vreeswijk modeled the neuronal responses as Poisson processes instead of continuous firing rates. A deeper theoretical understanding of this difference still needs to be achieved. The approach to model neuronal responses as Poisson processes instead of continuous firing

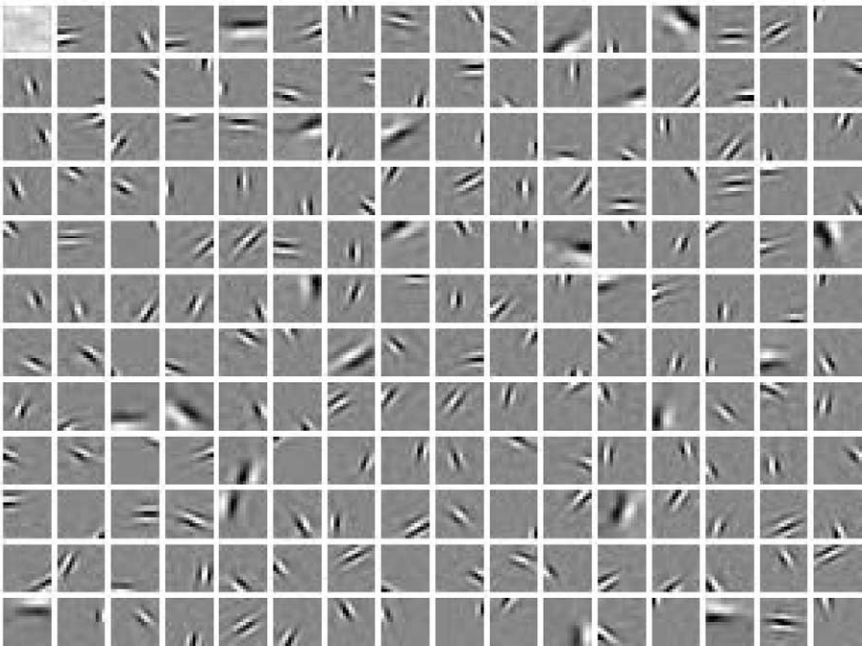

Fig. 3. Learned basis functions from natural images with sparseness prior. The image shows the result of optimizing linear receptive fields in order to minimize the average description length of 12×12 natural image patches (see text for details).

rates is a promising direction for future investigations, because this model accounts for both the noisiness and the discrete nature of neuronal spike responses. Furthermore, in van Vreeswijk (2001a) van Vreeswijk has shown that sparseness of neuronal representations modeled via renewal processes can be the natural outcome of a mutual information maximization approach instead of a constraint introduced ad hoc into an objective function.

Of course, clarifying the role of sparseness for neuronal representations is certainly also an experimental issue. Simple extracellular single cell recordings of neurons in the primary visual cortex when stimulated with natural image sequences can reveal the sparsity of individual neuron's responses ('life-time sparseness'). Another idea behind sparse coding, however, is that for a given image only a few neurons in a whole population are active ('population sparseness') which could be measured only via simultaneous recording of multiple neurons during stimulation with different images taken from the natural environment of the animal. These experimental methods are still being developed, but results may become available in the near future.

Efficient and sparse coding are certainly not the only principled approaches to deriving receptive fields in the primary visual cortex. Another principle which has been proposed is temporal coherence which is based on the assumption that objects in the environment change on a much slower time-scale than the sensory signals themself. This idea is appealing, because despite the resulting learning rules being unsupervised, they capture an essential property of an animals' environment without resorting to a feedback-based paradigm. One can recover information about the objects in the environment by extracting the slowly changing features of the raw sensory signal, which leads to the algorithm of slow feature analysis (Wiskott & Sejnowski, 2002). In summary, given sensory signals $\mathbf{x}(t) = (x_1(t),\ x_2(t), \ldots x_N(t))$ and a set $\mathcal{F}$ of real-valued basis functions, the goal is to find a slowly varying function $g(\boldsymbol{x}(t))(g_1(\boldsymbol{x})(t),\ g_2(\boldsymbol{x}(t)), \ldots,\ g_M(\boldsymbol{x}(t)))$ which can be achieved via minimizing the time-averaged squared temporal derivative $(d/dt\ g_j)^2$ as the objective function (see (Wiskott & Sejnowski, 2002) and (Berkes & Wiskott, 2002) for further details like constraints). Interestingly, the resulting receptive fields resemble properties of complex cells in the primary visual cortex like optimal stimuli similar to Gabor functions and phase invariance (Berkes & Wiskott, 2002). Although the emergence of these properties depends on the particular subset of natural image sequences selected to optimize the objective function, these results suggest temporal coherence as a plausible objective even for the early visual system. Furthermore, spatial coherence as an objective has also been used successfully in a principled approach which lead to model neurons discovering depth in random dot stereograms of curved surfaces (Becker & Hinton, 1992) indicating that intuitive principles like temporal and spatial coherence are promising candidates to be placed beside principles relying solely on issues related to efficient encoding.

2.3. Deriving receptive fields: high-level representation

So far we have considered the unsupervised derivation of receptive fields for low-level vision only, but one may speculate that simple principles may also explain receptive

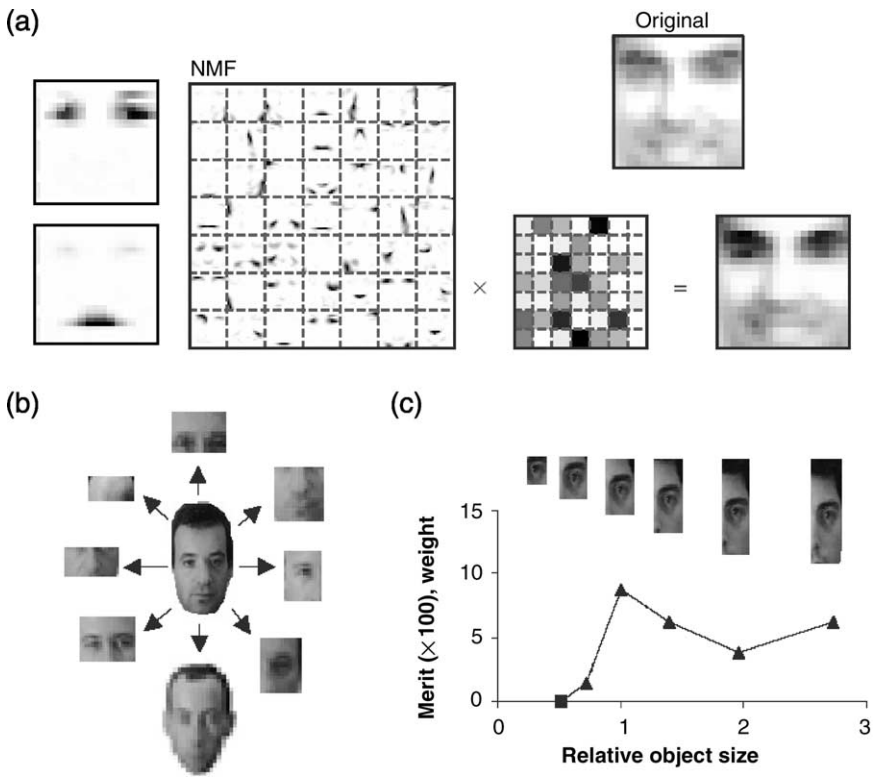

Fig. 4. Derived receptive fields for high-level representations. (a) Receptive fields resembling parts of objects produced by the NMF algorithm. Shown are all basis images and two enlarged ones. (b) Receptive fields produced by selecting the image fragments informative for a classification task (left) set up to distinguish faces from cars and the dependence of the corresponding mutual information objective function on the size of the extracted features (right) with intermediate sizes being optimal.

fields in higher visual areas like the inferior temporal cortex which is though to perform object recognition. So far, however, no explicit probabilistic model has been published that, when optimized according to the idea of efficient coding, produces model neurons that are selective to objects or parts of objects. It may come as a surprise, but the simple non-negative matrix factorization algorithm (NMF) (Lee & Seung, 1999) produces model neurons selective for parts of more complex objects. Physiological evidence supporting the idea that complex objects are represented in terms of their parts is rare and this issue is controversial. Nevertheless, this idea is plausible.

The basic idea of the NMF algorithm is to produce a representation of an ensemble of images via linearly combined basis images. The image ensemble is described as an $n \times m$ matrix V, where each of the m columns contains n non-negative image pixels. Then, the algorithm finds an approximate factorization $V \approx WH$, or

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu}$$

of the image ensemble, where the r columns of W are the basis images, and the each of the m columns in H serves as the representation of the corresponding image column in V. The factorization of V, however, is subject to the constraint of having only non-negative entries in H and W which leads to the receptive fields shown in Fig. 4(a) (taken from Lee & Seung, 1999). Visual inspection of the two enlarged basis images suggest that the algorithm indeed performs a decomposition into parts characteristic for the images in V. In this case, the parts correspond to eyes and a mouth. Relaxing the constraint of non-negativity or imposing constraints corresponding a simple quantization or to the well-known PCA algorithm leads to receptive fields, which do not correspond to object parts when inspected visually. At first, the constraint of non-negativity seems to be plausible when related to neuronal firing rates, because those cannot be negative, but receptive fields in the brain often come in opponent pairs, where the positive and negative ranges of a variable could be represented by the positive firing rates of two neurons. Nevertheless, the NMF algorithm may turn out to be an important starting point for learning of how to decompose images of objects.

Note also, that the resulting parts produced by the NMF algorithm reflect properties of the image ensemble used during training. If objects of face images are decomposed into parts, it may not be a surprise to find the parts being noses, eyes and mouths. It might be interesting to test whether the NMF algorithms also decomposes an image ensemble of face images and images of artifacts into parts which are compatible with the intuitive notion of artifact's parts. Furthermore, one may speculate that a description of object sets in terms of their parts may turn out to be efficient

in the sense of the MDL principle, which is another open question.

In another recent work related to high-level representations it has been investigated which image fragments extracted from object images turn out to be most relevant when used for solving a classification task. Using a database of face images and car images, Ullman and co-workers utilized a computational search procedure in order to determine those fragments (see Fig. 4(b) taken from (Ullman, Vidal-Naquet, & Sali, 2002)) which turn out to be most relevant to solve the classification task of distinguishing between faces from non-faces and cars from non-cars. In an iterative way they selected those image fragments which, when matched to sample images, maximize the mutual information

$$MI[C,F] = H[C] - H[C|F]$$

between the fragments F occurring in an image and the class C. Here

$$H[C] = E[-\log p(c)]_{p(c)}$$

is the class entropy and

$$H[C|F] = E[E[-\log p(c|F=f)]_{p(c|F=f)}]_{p(f)}$$

is the so-called conditional entropy with $E[\cdot]$ being the expectation operation. For a given test image of unknown class, i.e. whether a face (or a car) is present in the image or not, the image fragments in the fragment database extracted from training images are matched to the test image. Based on the degree of how well the individual fragments could be matched to the test image the conditional entropy $H[C|F]$ can be computed. The authors investigated how the mutual information $MI[C,F]$ is affected when the size of the extracted image fragments is changed and found that intermediate fragment sizes are optimal (see Fig. 4(c) taken from (Ullman et al., 2002)).

At least two factors may explain the superiority of intermediate sized fragments. First, a very large fragment may indicate the presence or absence of a particular face in an image. Since the fragments are extracted from raw training images, using large fragments leads to poor generalization which in turn leads to a lower mutual information (specificity). Second, very small fragments are much more likely to occur in different face images as well, but they also occur in non-face images and decrease the information about the image class, because small fragments also occur more often in car images (relative frequency). The authors also investigated how the mutual information changes for fixed fragment sizes, but varying resolution and found intermediate resolutions for large fragments being superior to large high resolution fragments. They conclude that intermediate complexity fragments (intermediate sized fragments with high resolution and large fragments with intermediate resolution) are optimal for stimulus classification.

Although the procedure used to extract and match the image fragments is far from being biologically plausible, their results suggest that fragments of intermediate complexity might indeed be extracted by the visual system and serve as building blocks for high-level representations. Interestingly, this result points into the same direction as the NMF algorithm, which extracts object parts as features. Note, however, that the NMF algorithm is an unsupervised algorithm whereas the latter one requires top–down feedback carrying information about the class membership. It is an interesting direction for further investigations to combine both algorithms in order to extract the optimal part-based representation for solving a classification task.

3. The environmental time-scale

So far we have considered the formation of receptive fields which is likely to happen at a more developmental time-scale even for high-level representation when compared to the environmental time-scale we consider now. Adaptation of neuronal properties to a changing environment lies at the heart of the Infomax principle whose basic idea we introduce next. In Section 3.2 we apply the Infomax principle to explain the experimentally well characterized, but in terms of its function less well understood phenomenon of contrast adaptation of simple cells in the primary visual cortex.

3.1. The Infomax principle

The Infomax principle proposes to use the mutual information

$$MI[X,Y] = H[Y] - H[Y|X]$$

between the ensemble of stimuli x and the neuronal representation Y as an objective function. Fig. 5(a) shows a 'black-box' model of a sensory system with X being the input and Y the neuronal representation of X. The stimulus X is a random variable with P_X being its probability density function. The output Y depends on the input x in a probabilistic way described by $P_{Y|X}$. Of course, the maximal mutual information between X and Y is obtained if the input X is simply copied to the output Y. However, this might not be possible for a particular sensory system due to constraints. As an example, let us consider the most prominent biological constraint: the limited dynamic range of neuronal transfer functions.

Fig. 5(b) shows the sigmoidal transfer function $g(x) = [1 + \exp(-\beta(x - \Delta x))^{-1}$. Assuming for $P_{Y|X}$ the deterministic transformation $g(x)$ with additive Gaussian noise of variance σ_n^2 it can be shown that the mutual information between X and Y as a function of the transfer function's parameters has its maximum at the maximum of the output entropy $H[Y]$. Fig. 5(c) shows two Gaussian densities for the input X with the same variance but different means. These two input densities lead to different output densities P_Y

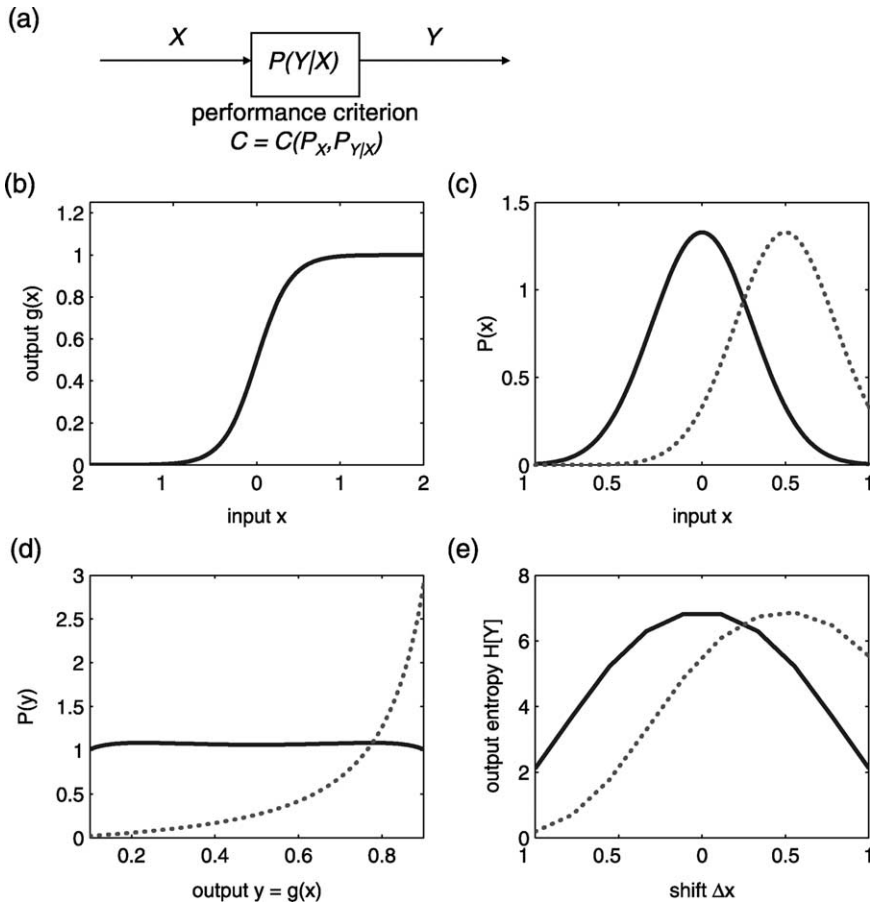

Fig. 5. (a) The Infomax performance criterion is the mutual information between the stimuli X and the neuronal representation Y and involves the statistics P_X of the environment and the probabilistic 'black box' description $P_{Y|X}$ of the representation system itself. (b) The neuronal transfer function $g(x)$ for $\beta = 5$ and $\Delta x = 0$. (c) Two Gaussian densities for the input $X$with $\sigma = 0.3$ and $\mu_1 = 0$ (solid line), $\mu_2 = 0.5$ (dotted line). (d) Densities for the output $y = g(x)$ resulting from transforming the corresponding input densities. (e) Output entropies as a function of the shift Δx of the transfer function for the two input densities.

when transformed with $g(x)$ as shown in Fig. 5(d). The probability density for the outputs caused by the Gaussian input distribution with mean $\mu_1 = 0$ is much more uniform than the one produced by the Gaussian with $\mu_2 = 0.5$ which can be understood by considering the transfer function in Fig. 5(b) which has its linear range matched to the region of high input density. For most inputs produced by the Gaussian with $\mu_2 = 0.5$ this transfer function works in its satration regime with high output firing rates (see Fig. 5(d), dotted line) occuring much more frequently than low output firing rates. Plotting the output entropy for both input densities as a function of the transfer function shift Δx reveals that the maximum is always attained at the corresponding means of the Gaussian where most of the 'probabilistic mass' is centered.

Considering a particular neuronal system with, e.g. a sigmoidal transfer function the Infomax principle predicts that the transfer functions should change if the input distribution changes. Here we have considered only the shift of the transfer function, but neuronal transfer functions may also change their shape, slope, etc.

3.2. Contrast adaptation

Let us consider now an application of the Infomax principle to contrast adaptation of simple cells in the primary visual cortex which links all three analysis levels proposed by David Marr: the computational theory, the algorithm level and the underlying neuronal mechanism.

Fig. 6(a) shows the experimental protocol used in contrast adaptation experiments. First, the cell is adapted to a high (or low) contrast adapting stimulus which in this case is a slowly moving grating. Then, the contrast response function is measured using short presentations of a test stimulus with varying contrast. Fig. 6(b) shows experimental results for intracellular recordings of a simple cell in cat primary visual cortex taken from Carandini and Ferster (1997). From left to right Fig. 6(b) shows how the contrast response functions for the stimulus-dependent first harmonic of the membrane potential (F1 component), the mean membrane potential (DC component) and the first harmonic of the firing rate adapt to the contrast of the stimulus shown during the adaptation phase. The F1 component of

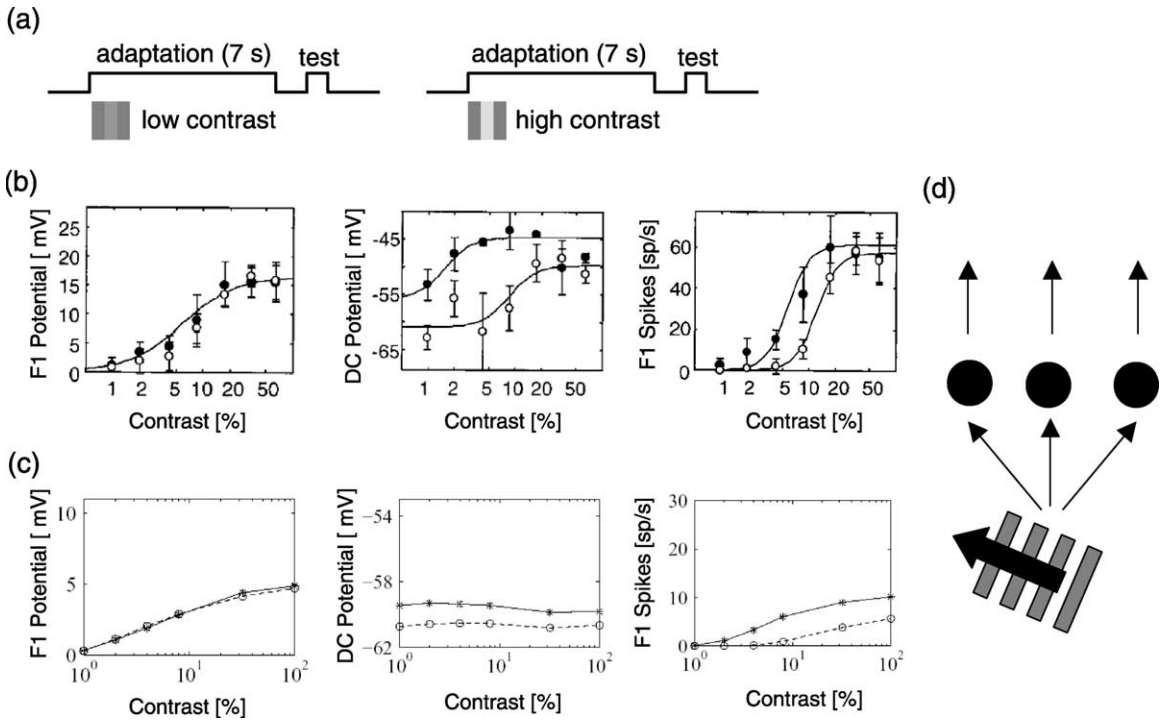

Fig. 6. Contrast adaptation of simple cells in the primary visual cortex. (a) The experimental protocol to probe contrast adaptation. (b) Experimental results for contrast adaptation in cat primary visual cortex. Shown are contrast response functions for the F1 component of the membrane potential (left), the DC component of the membrane potential (middle) and the F1 component of the firing rate after adaptation to a low contrast (solid circles) and high contrast (empty circles) stimulus. (c) Simulation results of a modeled orientation column with an online adaptation rule dictating how to change the release probability of the afferent synapses with the solid (dotted) lines corresponding to contrast response functions after adaptation to a low (high) contrast stimulus. (d) Illustration of the orientation column model.

the membrane potential does not show adaptation (left), but its DC component does (middle). Furthermore, the contrast response function of the firing rate (right) shifts towards higher contrast values following a prolonged presentation of high contrast stimuli, and shifts in the opposite direction if the preceding adapting stimulus had a low contrast.

During the last decade several ideas emerged to explain the underlying mechanism, but none of them was fully consistent with the available experimental data. Which adaptation mechanism could account for an adaptation of the mean membrane potential and the F1 component of the firing rate, but without predicting an adaptation of the membrane potential's F1 component? On the one hand, a group of studies suggested that plasticity of excitatory synaptic weights is responsible for contrast adaptation. On the other hand, an adaptation like the rescaling of synaptic weights would also predict an adaptation of the membrane potential's F1 component. In order to resolve this contradiction, we proposed a model, which involves changing the dynamic properties of the afferent geniculo-cortical, synapses from the LGN via modulating the transmitter release probability as the key mechanism (Adorjan, Piepenbrock, & Obermayer, 1999b). Thus, we modeled the afferent synapses not as static, but as dynamic synaptic weights with their value being dependent both on the recent history of stimulation and the transmitter release probability. In other words, changing the transmitter release probability changes the dynamic properties of the synaptic transmission.

Fig. 6(d) shows a sketch of an orientation column model we set up in order to explain contrast adaptation. In this model, the dynamic synapses are located at the feedforward connections. Fig. 6(c) shows simulation results of the model with high (solid lines, i.e. low contrast) and low (dotted lines, i.e. high contrast) synaptic release probability. The simulation results agree in a qualitative way well with the experimental contrast response functions measured after adaptation to stimuli with high and low contrast. This suggests the mechanism of changing the synaptic release probability as a candidate mechanism underlying contrast adaptation. Note that this mechanism also accounts for the indeed puzzling experimental observation that although the F1 component of the firing rate and the mean membrane potential adapts to the contrast of the stimulus, the F1 component of the membrane potential does not show adaptation.

Now, one may ask how the synapses adapt their transmitter release probability. How can they 'know' whether the adapting stimulus had a high or a low contrast? In Adorjan et al. (1999a) we derived an online adaptation rule for the synaptic release probability based on the Infomax objective function: Let $MI[X,Y] = H[Y] - H[Y|X]$ denote the mutual information between the input firing rates X and the output firing rates Y of a simple cell in

primary visual cortex. The input firing rates X are transformed into output firing rates Y via a transfer function $y = g(x; p)$ similar to the one used for our illustration in Section 3.1. Similar to Δx in Section 3.1, the transmitter release probability p is a parameter of this transfer function. Now, the goal is to maximize $H[Y]$ with respect to p, since we assumed the transfer function to be deterministic which implies that the mutual information as a function of the parameter p has its maximum where the output entropy $H[Y]$ is maximal. Noting that

$$\begin{aligned} H[Y] &= -E[\log p(y = g(x;p))]_{p(x)} \\ &= -E\big[\log p(x)/|\frac{\partial}{\partial x} g(x;p)|\big]_{p(x)} \\ &= E\big[\log|\frac{\partial}{\partial x} g(x;p)|\big]_{p(x)} - E[\log p(x)]_{p(x)} \end{aligned}$$

we can utilize a stochastic approximation procedure with the gradient of $H[Y]$ with respect to the release probability p being sampled in an online manner leading to an adaptation rule

$$\tau_{\text{adapt}} \frac{\partial}{\partial t} p = F(p, x)$$

which dictates how to change the transmitter release probability p as a function of the actual presynaptic firing rate x (See (Adorjan, Piepenbrock & Obermayer, 1999b) for a definition of F). We set the time constant to $\tau_{\text{adapt}} = 7$ s in order to account for the experimentally observed time-scale of contrast adaptation. The stochastic gradient approximation used in our approach is only one possible optimization method, but since it is an online learning rule, it is possible to implement it biologically. It is conceivable, however, that natural systems employ different optimization methods in order to optimize the quantity $H[Y]$.

The results in Fig. 6(c) have been generated by simulating the experimental protocol for contrast adaptation, but with the above adaptation rule governing the adaptation of the release probability p. In the context of the Infomax principle, the shift of the firing rate function towards lower contrast values can be understood as a dynamic readjustment of the firing rate function's dynamic range of those contrast values which occurred in the recent past. Compare this with Fig. 5(b) and (c) which illustrate that shifting the linear part of a sigmoidal transfer function towards the mean value of the Gaussian input distribution maximizes the output entropy $H[Y]$ by equalizing the output firing rates.

These results suggest that contrast adaptation might indeed be understood as a consequence of an ongoing optimization of neuronal response functions to changing contrast statistics in order to ensure an accurate representation of all possible contrast values. Note, however, that for a read-out to estimate the absolute value of the stimulus contrast represented by the output firing rate of neurons in the primary visual cortex, it needs to adapt as well. Therefore, another interpretation of contrast adaptation applicable to a read-out which does not adapt is a low-level preprocessing in order to ensure contrast invariance. In other words, a non-adapting read-out 'seeing' the visual world via the firing rates of adapting cells in the primary visual cortex could focus on pattern vision independent of the average contrast values in the environment.

With this exercise in modeling contrast adaptation we combined all three analysis levels: as the functional role of contrast adaptation in primary visual cortex we proposed the ongoing adaptation of the neuronal representation in order to ensure the accurate representation of absolute contrast values and/or the computation of contrast invariance. At the algorithmic level we assumed a stochastic sampling of stimulus contrasts in order to adapt the neuronal system's parameters as dictated by the Infomax objective function. At the level of the neuronal realization we assumed that contrast adaptation is mediated by adapting the transmitter release probability of afferent synapses.

4. The perceptual time-scale

Fast neuronal dynamics happen on the time-scale of a single stimulus presentation (perceptual time-scale) which prevents to consider them as an adaptation to changing statistics of the input (environmental time-scale). Let us now consider two examples which assign a functional role to these fast neuronal dynamics.

In Section 4.1 we propose an extension of the Infomax principle which takes into account a possibly time-varying model to encode the signals and assumes a continuous read-out, and in Section 4.2 we review an efficient encoding approach for explaining contextual effects in the primary visual cortex which suggests the view of fast neuronal dynamics as a search process for an efficient description of the sensory input.

4.1. Rapid adaptation to internal states

Recently, Bialek and coworkers (Brenner, Bialek, & de Ruyter van Steveninck, 2000) provided direct experimental evidence for the Infomax principle by showing that the input/output relation of the motion-sensitive H1 neuron in the fly visual system indeed adapts to the changing statistics of a selected stimulus ensemble, namely simulated visual velocity signals. They also showed that the time-scale of this adaptation is close to the physical limit imposed by statistical sampling. Therefore, it seems as if even extremely rapid adaptation could be interpreted as an optimization with respect to changing input statistics. But does this also apply to the mammalian visual system with its experimentally observed transient dynamics to flashed but otherwise static stimuli?

We argued that the functional interpretation offered by the Infomax principle is only applicable as long as

the time-scale of adaptation is slow compared to the time-scale at which the statistics of the stimulus ensemble changes. Only in this case an organism could estimate changes of stimulus statistics and adapt its sensory system. For the fly visual system this interpretation may indeed be applicable, because the statistics of the input to the H1 neuron, i.e. the variance of the velocity signal, may change abruptly corresponding to a transition from cruising flight to acrobatic or chasing behavior. In the mammalian visual system, however, even static environments are scanned with saccadic eye movements, but the neuronal responses in primary visual cortex are time-dependent on the time-scale of a typical fixation period ($\approx$ 300 ms). Interpreting these fast dynamics as an adaptation in the sense of the Infomax principle is not straight forward, because the adaptation happens during a single stimulus presentation, and can therefore not be viewed as an adaptation to changing statistics of the input, simply because statistics cannot be sampled based on a single observation.

We have shown (Adorjan, Schwabe, Wenning, & Obermayer, 2002) that these dynamics can still be understood with information-theoretic concepts by having formulated an objective function and suggested that rapid adaptation mechanisms which underly the fast and transient dynamics in sensory systems could be the result of optimizing this objective function. In particular, we extended the Infomax principle by having proposed the maximization of the mutual information

$$MI[S, R_t] = H[S] - E[H[S|R_t = r_t]]_{P(r_t;\beta_t)}$$

between the stimuli S and the neuronal representation R_t at time t after stimulus onset as the objective function for the rapid adaptation mechanisms operating during the presentation of a single stimulus. In the above equation, $H[\cdot]$ and $E[\cdot]$ denote the entropy and expectation of a random variable, and the expectation is taken with respect to $P(r_t;\beta_t) = \int \mathrm{d}s P(s)\cdot P(r_t|s;\beta_t)$ with s being the stimulus variable and $P(s)$ its assumed distribution. The time-dependent parameter β_t characterizes the state of rapid adaptation mechanisms and is assumed here to be independent of the stimulus.

The first term in the above equation reflects the prior uncertainty about the stimulus before having observed any neuronal response. The second term is the so-called noise entropy and corresponds to the average reduction of uncertainty about the actual stimulus after having observed the neuronal representation r_t at time t. The representations, however, depend on the parameterization of the sensory system which may be time-dependent as indicated by the β_t. Thus, our extension of the Infomax principle is twofold: First, we assume a continuous read-out of the neuronal representation. Second, we allow the parameterization of the sensory system to change on the time-scale of a single stimulus presentation which might become necessary when assuming a continuous read-out.

Let us consider an example in the context of encoding visual information by a local orientation hypercolumn in the primary visual cortex which is composed of recurrently connected orientation columns. Fig. 7(a) shows how a static stimulus s induces a spatial-temporal response pattern $N_1, N_2, \ldots, N_t$ in the encoding population which is relayed to a hypothesized downstream population, i.e. the read-out of the encoded information. Each N_i corresponds to the number of spikes observed in a short time-interval beginning at the ith time step, and the representation R_t denotes the accumulated spike count up to time step t. Now, even if the encoding population does not receive any feedback from the read-out, it is a valid question to ask whether the parameter β_t characterizing the instantaneous input/output relation of the encoding population should be adapted or not.

We investigated this question by setting up an effective description of a monocular patch in a local cortical orientation hypercolumn in the primary visual cortex with the real-valued inputs representing the intensities of local features, e.g. oriented edges. Given the effective network transfer function

$$\lambda_{t,i}(\mathbf{s}) := \tau \frac{\exp(\beta_t s_i)}{\sum_{j=1}^{L} \exp(\beta_t s_j)}$$

with β_t being the only free, but possibly time-dependent parameter, each $\lambda_{t,i}(\mathbf{s})$ denotes the expected number of spikes the ith neuron fires in the tth time-interval of duration τ in response to the stimulus $\mathbf{s}$. Fig. 7(b) illustrates this model. This model is a so-called divisive normalization model, which is not intended to account for all possible biophysical details of the microcircuits in primary visual cortex, but mimics the experimentally observed phenomenon of recurrent competition between orientation columns. Presenting a stimulus, which activates multiple orientation columns like, crossed bars, the orientation columns with the higher activity can be viewed as suppressing orientation columns with lower activation leading to a competition for neuronal activation. The parameter β parameterizes the level of this competition with high values corresponding to strong and low values corresponding to low competition. The outputs are modeled via Poisson processes with their time-varying rates given by the $\lambda_{t,i}$. Assuming further a non-negative and sparse distribution $P(\mathbf{s})$ for the stimuli s we performed a numerical greedy optimization to obtain the optimal parameterization for the free network parameters β_t.

We found that only a dynamic parameterization performs well as quantified by our objective function. Fig. 7(c) illustrates this result. During the initial phase of the response higher, and at later times lower values of β_t are preferred. Let us compare the dynamic parameterization with static strategies. Fig. 7(c) (dashed lines) shows static parameterizations for high and low values of β_t, and Fig. 7(d) shows

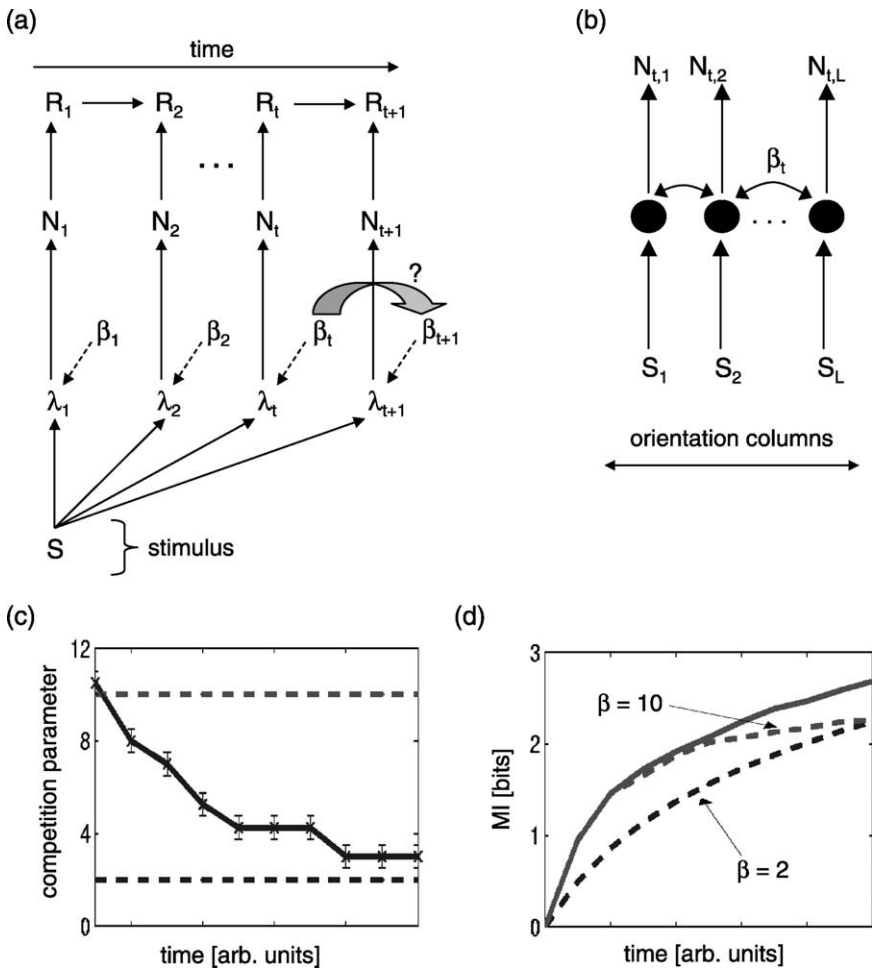

Fig. 7. The idea of dynamic coding. (a) A static sensory stimulus S is applied and induces noisy neuronal responses models as a Poisson process with time-varying mean firing rate λ_t. The instantaneous spike responses N_t are observed by a spike-counting read-out and summarized into a summed spike-count R_t. (b) Illustration of the hypercolumn model L inputs and outputs and time-dependent recurrent competition parameterized by β_t. (c) Predicted optimal dynamics of the competition β_t (solid line) and static sub-optimal static competition (dotted lines). (d) Resulting mutual information between the input S and the accumulated spike-count R_t as a function of the window used for counting the spikes for different dynamics of the competition parameter β_t.

the corresponding mutual information. For the low value the increase is slow, whereas for the high value it is faster. On the other hand, the mutual information obtained with the high value saturates and will be outperformed by the low value. In other words, a spike-counting read-out mechanism obtains on average more information about the stimulus if the sensory system is dynamic.

Intuitively, for our particular model of a network transfer function with adjustable competition this can be understood as follows: Initially, the level of output noise is high, because only a few spikes are available for representation. Strong competition, however, biases the representation towards the most salient input, which is a sensible strategy, because our assumed input distribution implies that typically only a single input line is strongly active and needs to be represented. Later, when more spikes are available for representation and averaging over time allows the signal to be separated from the noise, reduced competition allows the signaling of possibly multiple active inputs and their faithful representation. Thus, the adaptation strategy with the dynamic parameterization first transmits information about the salient features of the input, then about the less salient details.

We also applied this idea to a much more detailed biophysical model of an orientation hypercolumn in the primary visual cortex and found that the mechanisms of spike-frequency adaptation of cortical pyramidal neurons (Adorjan et al., 2002; Schwabe, Adorjan, & Obermayer, 2001) or fast synaptic depression (Adorján et al., 2000) at the recurrent excitatory synapses may serve as the underlying substrate to realize the dynamic competition suggested by our abstract model as a promising strategy in terms of the extended Infomax principle.

4.2. Contextual effects

The idea of efficient encoding has also been applied using a hierarchical probabilistic model intended to explain contextual effects in the primary visual cortex. These so-called contextual effects refer to the phenomenon that

although stimulation outside the classical receptive field (the surround) do not cause the firing of neurons, they modulate the responses to stimulation within the neuron's classical receptive field (the center). These contextual effects have been found to be dependent on the contrast and orientation of the stimuli within and outside the classical receptive field. The phenomenon reported most frequently by different groups is that if the stimulus orientation of the center and surround are equal, then the responses are suppressed compared to when the center stimulus is shown alone. Fig. 8(b) (solid line) shows the response of a complex cell in layer 2/3 of cat primary visual cortex as a function of the length of a bar used for stimulation. As the bar length increases, the center and surround are stimulated with the same orientation and the response of the center becomes suppressed. Although the detailed neuronal circuitry and the functional role of these effects are still under investigation, the reduced suppression when inactivating layer 6 (Fig. 8(b), dotted line) which is part of the feedback loop connecting the primary and secondary visual cortex suggests that top–down feedback may contribute to these contextual effects, because blocking layer 6 may also impair the re-entry of feedback signals.

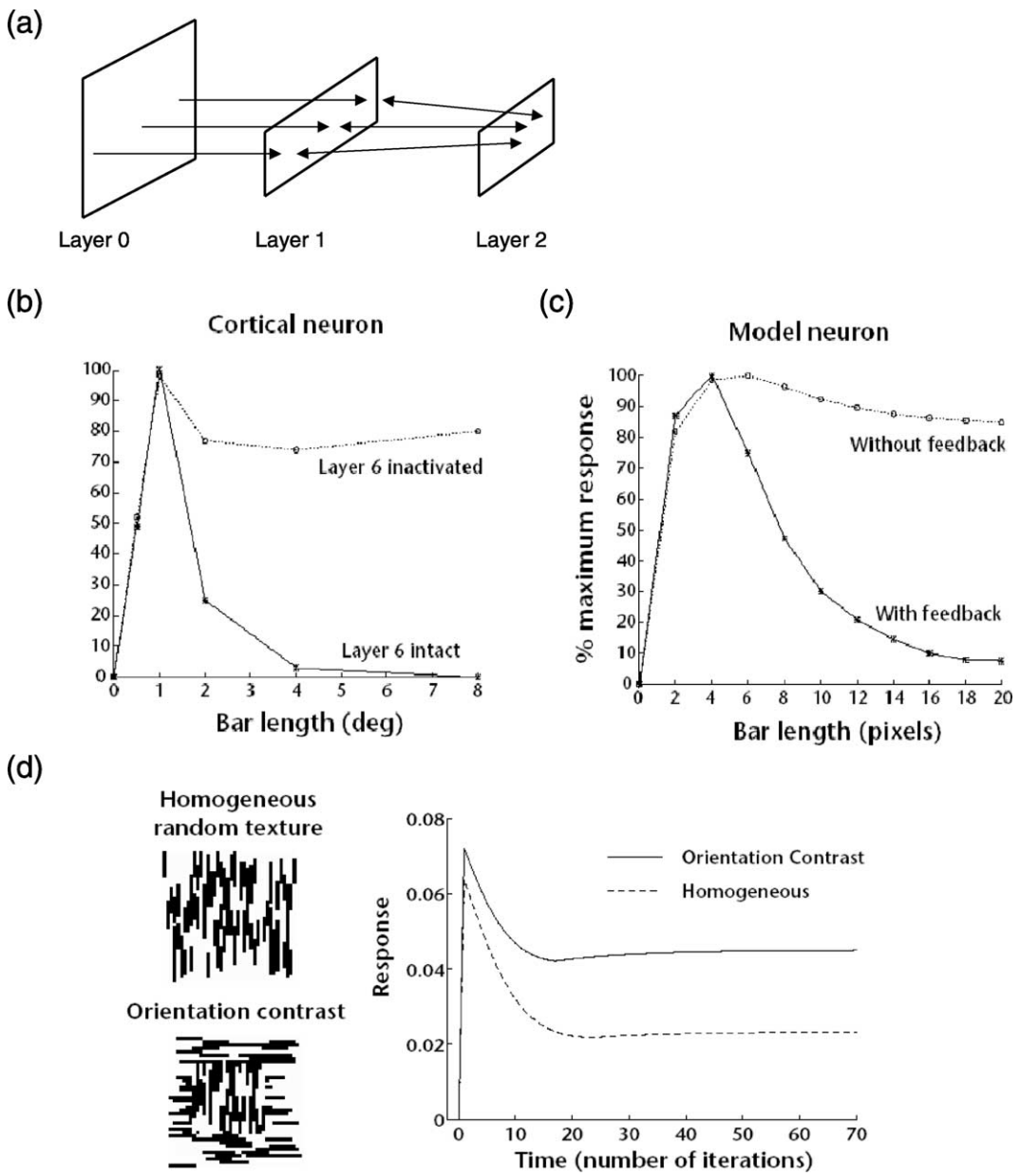

Fig. 8. Contextual effects predicted via efficient encoding with a hierarchical model. (a) Model architecture. (b) End-stopping complex cell responses in layer 2/3 of cat visual cortex when inactivating layer 6 (dotted line) and for the control case (solid line). (c) Model predictions for the case of active (solid line) and inactive (dotted line) feedback from the next higher level. (d) Predicted responses for the case of stimulating center and surround with different (solid line) and the same (dotted line) orientations. Figs. b–d taken from (Rao and Ballard, 1999).

Rao and Ballard (1999) set up a hierarchical probabilistic model and applied the idea of efficient encoding in order to derive the receptive fields of and connectivity between two layers of neurons corresponding to the primary and secondary visual cortex. Their model architecture is shown in Fig. 8(a). The raw visual input to layer 1 (the modeled primary visual cortex) is provided by layer 0 (the raw retinal image). The outputs of layer 1 serve as the inputs to layer 2 (the modeled secondary visual cortex), which projects back to layer 1. They propose that the firing rates $\mathbf{a}_n$ at each level n in the processing hierarchy serve as the representation of the feedforward input $\mathbf{I}_n$ to this layer. The representations $\mathbf{a}_n$ in turn serve as the feedforward input $\mathbf{I}_{n+1}$ to the next level $n+1$.

In contrast to the sparse coding model reviewed in Section 2.2, their hierarchical model $G(\mathbf{I}_n|\mathbf{a}_n)$ has the form

$$G(\mathbf{I}_n|\mathbf{a}_n) = N(\mu = f(\mathbf{U}_n\mathbf{a}_n)\ \sigma^2 = \sigma_n^2)(\mathbf{I}_n)$$

where the feedforward input $\mathbf{I}_n$ is a long column vector, the entries in $\mathbf{U}_n$ serve as the weights for the feedback projections from layer n to layer $n-1$, σ_n^2 is a constant variance for an assumed additive noise, and $f(\cdot)$ is a sigmoidal neuronal transfer function. The effect of feedback from layer $n+1$ to layer n is described by a model

$$P(\mathbf{a}_n|\mathbf{a}_{n+1}) = N(\mu = f(\mathbf{U}_{n+1}\mathbf{a}_{n+1}),\ \sigma^2 = \sigma_n^2)(\mathbf{a}_n)$$

which in this form can be viewed as a prior for the activations $\mathbf{a}_n$ propagated back from level $n+1$ to layer n. Utilizing the feedback projections as a way to communicate information back to the lower representation level might be beneficial, because neurons at higher levels have larger receptive fields and therefore could help to disambiguate the representation at lower levels leading to the experimentally observed contextual effects.

Rao and Ballard set up the objective function

$$E = -\log G(\mathbf{I}_n|\mathbf{a}_n) - \log P(\mathbf{a}_n|\mathbf{a}_{n+1}) + g(\mathbf{a}) + h(\mathbf{U})$$

with $h(\mathbf{a})$ and $h(\mathbf{U})$ expressing constraints for the synaptic weights $\mathbf{U}$ and the activations $\mathbf{a}$ like the sparseness constraint utilized in Section 2.2. Randomly selecting patches of natural images, then performing a search for the best representation $\mathbf{a}$ for nearly fixed synaptic weights $\mathbf{U}$ and only slowly adapting these weights during this random sampling they found receptive fields in terms of the columns of $\mathbf{U}^T$ which at the level 1 look like Gabor functions as long as $g(\mathbf{a})$ enforces sparseness. The columns of $\mathbf{U}^T$ can be interpreted as receptive fields, because in the gradient

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{a} \propto \frac{\mathrm{d}}{\mathrm{d}\mathbf{a}}E = \frac{k_1}{\sigma_n^2}\mathbf{U}^T\frac{\partial f^T}{\partial \mathbf{a}}(\mathbf{I} - f(\mathbf{U}\mathbf{a})) + \frac{k_1}{\sigma_n^2}(\mathbf{a}^{\mathrm{td}} - \mathbf{a}) - \frac{k_1}{2}g'(\mathbf{a})$$

used for searching representations $\mathbf{a}$ with k_1 being a constant and g' the derivative of g with respect to $\mathbf{a}$ the term $\mathbf{U}^T$ acts as a feedforward weight matrix. With this interpretation, the ith row of $\mathbf{U}^T$ represents the synaptic weights of a single neuron with activity a_i. They also found that when inactivating the feedback term during the gradient descent search for a good representation $\mathbf{a}$, their model produces responses similar to the one observed experimentally when inactivating layer 6 as shown in Fig. 8(c).

For our consideration of neuronal dynamics at the perceptual time-scale, however, a different observation is more relevant. Fig. 8(d) shows the responses a_i of a selected representation neuron in the first layer to visual stimulation of center and surround with the same (dotted line) and different (constant line) orientations as a function of the time after stimulus onset. Note that here the fast neuronal dynamics of $\mathbf{a}$ at the perceptual time-scale of a single stimulus presentation are considered as the signature of a dynamic search process intended to minimize the proposed objective function for a given input image with respect to the neuronal activations $\mathbf{a}$. Based on these simulation results, one could predict the time course of contextual effects which, to the best of our knowledge, have not been explored experimentally.

Although it needs to be investigated in greater detail whether transient dynamics in primary visual cortex could indeed be viewed as some sort of an online optimization of an objective function for a given input image, it is at least another possible interpretation of transient dynamics at the perceptual time-scale.

5. Summary

In summary, we have reviewed principled approaches to modeling the visual system which consist essentially of formulating an objective function and then optimizing the free parameters of simple models with respect to this objective function. However, when trying to say more about what the visual system is computing than just ‘some sort of representation’, one has to be precise and, in the ideal case, to use a coherent mathematical framework. With this review we tried to give an overview of concrete principled approaches with a focus on ideas we believe are under-represented in the literature like the MDL principle and rapid adaptation to internal states.

As a guideline we utilized the general idea of adaptation at multiple time-scales, because we believe that this idea may qualify as a general paradigm for deriving functional interpretations of changes of receptive field properties. Our own work reviewed in Section 4.1 exemplifies that this method is also applicable to fast neuronal dynamics for which almost no explicit functional interpretations are present in the literature. However, in order to test whether the view we suggest is only a ‘nice theoretical idea’, or actually realized in biological systems, further experiments are needed to pinpoint for concrete phenomena the conditions—may they be external or internal—which are changing and inducing changes of receptive field properties.

A recapitulation reveals that many proposed principles might be reduced to optimizing a regularized objective

function which often might be cast easily in terms of efficient coding using the MDL model selection principle with the logarithm of the prior corresponding to the regularizer. (Hence, our detailed review of the MDL principle, which is well suited for explicitly including biological constraints). As an example, consider the way we presented the idea of sparse coding which has been presented originally in terms of such a regularized objective function (Olshausen & Field, 1996). We simply considered sparseness as being a desirable property as expressed via a sparseness prior. We believe that more constrained models need to be proposed with the model and constraints being argued for from a biological perspective. Then, comparing model predictions with experimental measurements would allow to sort out those models and constraints hopefully corresponding to the investigated sensory system.

In this review article we have investigated mainly firing rate models. Neurons, however, are often considered to be noisy computing devices, and this noise (beside other constraints) needs to be taken into account when modeling neuronal systems. This problem is much more severe than it appears in the first place, because neuronal noise may have different characteristics than the additive noise usually assumed in probabilistic models of sensory data. Furthermore, neurons spike. Thus, (i) modeling neuronal responses with *greater biological detail* than firing rate models, and (ii) *explicitly including other biological constraints* seems to become imperative for principled approaches.

Acknowledgements

This work was supported by the Wellcome Trust (061113/Z/00) and the German Science Foundation (SFB 618 and DFG 120-3).

References

Adorjan, P., Levitt, J. B., Lund, J. S., & Obermayer, K. (1999a). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. *Visual Neuroscience*, *16*, 303–318.

Adorjan, P., Piepenbrock, C., & Obermayer, K. (1999b). Contrast adaptation and infomax in visual cortical neurons. *Reviews in the Neurosciences*, *10*, 181–200.

Adorján, P., Schwabe, L., Piepenbrock, C., & Obermayer, K. (2000). Recurrent current competition: strengthen or weaken. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), (*Vol. 12*) (pp. 89–95). *Neural Information Processing Systems NIPS*, MIT Press.

Adorjan, P., Schwabe, L., Wenning, G., Obermayer, K (2002). Rapid adaptation to internal states as a coding strategy in visual cortex? *Neuroreport*, 13(3), 337–42.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network*, *7*(2), 251–259.

Becker, S., & Hinton, G. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, *355*, 161–163.

Berkes, P., & Wiskott, L. (2002). *Applying slow feature analysis to image sequences yields a rich repertoire of complex cell Properties. In Proceedings of the ICANN'02*, Springer, pp. 81–86.

Brenner, N., Bialek, W., de Ruyter, R., & de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, *23*(3), 695–702.

Carandini, M., & Ferster, D. (1997). A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science*, *276*, 949–952.

Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York: Wiley.

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Reviews of Neuroscience*, 23.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, I: parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121–134.

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

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of Royal Society of London B*, *198*, 1–59.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligance Research*, *4*, 237–286.

Kandel, E. R., Schwarz, J. H., & Jesse, T. M. (1996). *Principles of neural science*. New York: McGraw-Hill.

Kersten, D., & Schrater, P. W. (1999). *Perception theory: Conceptual pattern inference theory: A probabilistic approach to human vision*. New York: Wiley.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 759–760.

Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, *387*, 73–76.

Marr, D. (1982). *Vision*. New York: Freeman.

Minsky, M. (1967). *Computation: Finite and infinite machines*. Engelwood Cliffs, NJ: Prentice Hall.

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.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Review in Neurosciences*, *12*, 125–132.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. Singapore: World Scientific Publishing.

Schwabe, L., Adorjan, P., & Obermayer, K. (2001). Spike-frequency adaptation as a mechanism for dynamic coding. *Neurocomputing*, *38–40*, 351–358.

Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature*, *422*(6927), 36.

Siegelmann, H. T., & Sontag, E. D. (1991). Turing computability with neural nets. *Applied Mathematics Letters*, *4*, 77–80.

Sompolinsky, H., & Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology*, *7*, 514–522.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5.

van Vreeswijk, C. (2000a). Using renewal neurons to transmit information. *European Biophysics Journal*, *29*, 245.

van Vreeswijk, C (2000b). Whence Sparseness. In: Neural Information Processing Systems NIPS, MIT Press (vol. 13), pp. 180–186.

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Computation*, *14*(4), 715–770.

Zemel, R., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*.