Helsinki University of Technology
Dissertations in Computer and Information Science
Espoo 2003 Report D3

COMPUTATIONAL MODELS RELATING PROPERTIES OF VISUAL NEURONS TO NATURAL STIMULUS STATISTICS

Jarmo Hurri

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Computer Science and Engineering for public examination and debate in Auditorium T2 at Helsinki University of Technology (Espoo, Finland) on the 5th of December, 2003, at 12 o'clock noon.

Helsinki University of Technology
Department of Computer Science and Engineering
Laboratory of Computer and Information Science
P.O.Box 5400
FIN-02015 HUT
Finland

Distribution:

Helsinki University of Technology
Laboratory of Computer and Information Science
P.O.Box 5400
FIN-02015 HUT
Finland

Tel. +358-9-451 3272
Fax +358-9-451 3277
http://www.cis.hut.fi/

Available in pdf format at http://lib.hut.fi/Diss/2003/isbn951226823X/

ISBN 951-22-6822-1 (printed version)
ISBN 951-22-6823-X (electronic version)
ISSN 1459-7020

Otamedia Oy
Espoo 2003

Hurri, J. (2003): **Computational Models Relating Properties of Visual Neurons to Natural Stimulus Statistics.** Doctoral thesis, Helsinki University of Technology, Dissertations in Computer and Information Science, Report D3, Espoo, Finland.

Keywords: computational neuroscience, cortical coding, cortical topography, primary visual cortex, simple cells, complex cells, temporal coherence, bubble coding, burst firing, independent component analysis, sparse coding.

ABSTRACT

The topic of this thesis is mathematical modeling of computations taking place in the visual system, the largest sensory system in the primate brain. While a great deal is known about how certain visual neurons respond to stimuli, a very profound question is *why* they respond as they do. Here this question is approached by formulating models of computation which might underlie the observed response properties. The main motivation is to improve our understanding of how the brain functions. A better understanding of the computational underpinnings of the visual system may also yield advances in medical technology or computer vision, such as development of visual prostheses, or design of computer vision algorithms.

In this thesis several models of computation are examined. An underlying assumption in this work is that the statistical properties of visual stimuli are related to the structure of the visual system. The relationship has formed through the mechanisms of evolution and development. A model of computation specifies this relationship between the visual system and stimulus statistics. Such a model also contains free parameters which correspond to properties of visual neurons. The experimental evaluation of a model consists of estimation of these parameters from a large amount of natural visual data, and comparison of the resulting parameter values against neurophysiological knowledge of the properties of the neurons, or results obtained with other models.

The main contribution of this thesis is the introduction of new models of computation in the primary visual cortex. The results obtained with these models suggest that one defining feature of the computations performed by a class of neurons called simple cells, is that the output of a neuron consists of periods of intense neuronal activity. It also seems that the activity levels of nearby simple cells are positively correlated over short time intervals. In addition, the probability of the occurrence of such regions of intense activity in the joint space of time and cortical area seems to be small. Another contribution of the thesis is the examination of the relationship between two previous computational models, namely independent component analysis and local spatial frequency analysis. This examination suggests that results obtained with independent component analysis share some important properties with wavelets, in the way their localization in space and frequency depends on their average spatial frequency.

Acknowledgements

I am very aware that during the time I have worked towards this thesis, I have been very fortunate: I have had good financial support, excellent supervision, professional collaborators, extremely nice co-workers, and first-class facilities.

The work presented in this thesis was funded by Helsinki Graduate School of Computer Science and Engineering, and the Laboratory of Computer Science and Engineering at Helsinki University of Technology. This funding made it possible for me to prepare this thesis on the topic that was of greatest interest to me. I am also grateful to the Finnish Foundation for the Promotion of Technology for additional financial support.

The contents of this thesis are a product of close co-operation with Dr. Aapo Hyvärinen. Not only has this collaboration been close, but from my point of view it has also been seamless. In addition to having had the chance to work with one of the top researchers in this field, I have also enjoyed greatly the fact that Aapo is a very easy person to work with. I can only say that in terms of supervision and collaboration, there is nothing more I could have asked for.

The opportunity to do the work reported in this thesis was given to me by Academy Professor Erkki Oja. I would like to thank him for making it possible for me to pursue a doctoral degree. Professor Oja was one of the persons who initially directed me towards the analysis of image data, a research topic which proved to be very fertile. We also collaborated in the preparation of one of the articles in this thesis, and in finalizing the thesis. Professor Oja is also the head of the Neural Networks Research Centre, which in my opinion has been a practically perfect place to prepare this thesis. Here I have had the chance to enjoy academic freedom, the company of some of the nicest and brightest researchers, and access to excellent facilities, especially computational resources.

Mr. Jaakko Väyrynen has been a coauthor in some of the articles that form an integral part of this thesis. Jaakko has provided valuable insights during all of the stages of our joint work. All his work in the experiments has been of very high quality, and he has always been able to complete our experiments admirably even with our tight schedules.

Entry into the analysis of natural image sequences was greatly facilitated by Professor Hans van Hateren, who provided us with access to his database of natural image sequences. In addition, Professor van Hateren also provided us with the program used to measure the properties of spatial filters, thus saving us time, and ensuring the comparability of our results.

During the preliminary inspection of this thesis, the reviewers Dr. Jukka Heikkonen and Docent Pentti Laurinen made a very important contribution in improving the overall quality and coherence of the thesis.

All of my co-workers and friends have contributed in making this period of my

life very enjoyable. In particular, I would like to thank the following persons for their assistance and support, as well as interesting discussions, both on and off the topic: Professor Olli Simula, Dr. Esa Alhoniemi, Dr. Patrik Hoyer, Mr. Jaakko Särelä, Mr. Mika Inki, Mr. Ville Könönen, and Docent Ilpo Kojo.

Finally, I thank my wife Jenni for sharing also the professional part of my life with me.

Contents

Notations and abbreviations

Constants and variables

lower- or uppercase letter	scalar, or random variable
i, k, ℓ, n	general-purpose indices
t, τ	time indices
Δt	delay in time
K, N, T	general-purpose upper limits for indices
x_k	kth component of observed data
w_k	kth weight in a linear model
y_k	kth (processing unit) output or latent variable
s_k	kth latent variable (in Publication 7)
$y_k(t)$	kth output signal or latent signal
$s_k(t)$	kth latent signal (in Publication 7)
$v_k(t)$	kth variance signal or latent signal
$z_k(t)$, $u_k(t)$	latent signals
(x, y)	spatial image coordinates
(x, y, t)	spatiotemporal image sequence coordinates
boldface lowercase letter	column vector
$\mathbf{x} = [x_1 \ \cdots \ x_N]^T$	observed data
$\mathbf{x}(t)$	observed signals
$\mathbf{w}_k$	vector of weights corresponding to kth linear unit
$\mathbf{y}$	outputs or latent sources
$\mathbf{y}(t)$	output signals or latent signals
$\mathbf{v}(t)$	latent signals
boldface uppercase letter	matrix
$\mathbf{W}$	matrix relating observations to outputs
$\mathbf{A}$, $\mathbf{M}$	weight matrices in generative models

Functions

$f(\cdot)$, $g(\cdot)$, $G(\cdot)$	scalar-valued nonlinear functions
$\mathrm{E}\{\cdot\}$	expected value (over a set of samples)
$\mathrm{E}_t\{\cdot\}$	expected value (over time)
$\mathrm{cov}\{\cdot,\cdot\}$	covariance
$\kappa_4(\cdot)$	kurtosis (a fourth-order cumulant)
$\mathrm{var}_\lambda(\cdot)$	local variance with decay parameter λ
$p(\cdot)$	probability density function
$I(x,y)$	(spatial) image
$I(x,y,t)$	(spatiotemporal) image sequence
$h(\cdot,\cdot)$	spatial neighborhood function
$\phi(t)$	temporal filter
$\mathbf{abs}(\cdot)$	function mapping vector components to absolute values

Abbreviations

CRF	classical receptive field
DC	direct current (constant part of a signal)
ICA	independent component analysis
IT	inferior temporal
K	koniocellular
LGN	the lateral geniculate nucleus
M	magnocellular
MST	medial superior temporal
MT	medial temporal
P	parvocellular
PCA	principal component analysis
V1	the primary visual cortex
V2–V5	visual cortical areas 2–5

Chapter 1

Introduction

1.1 Motivation and overview

At the back of our heads, there is a brain area called the *primary visual cortex*. As the name suggests, this brain area is involved in the sense of vision, in transforming light entering our eyes into visual experiences. The brain area has been labeled "primary" because it is generally thought that the majority of visual processing that takes place in the cortex is initiated in this area (Wurtz and Kandel 2000a).

Visual processing in the primary visual cortex is performed by a network of *neurons*, basic cells of the nervous system (Kandel 2000). Information is transmitted out of a neuron in a train of nerve impulses, or *spikes*. Neurons in the network can be connected to each other and to neurons outside the primary visual cortex. Neurons in the primary visual cortex can be divided into different classes based on the way they respond to visual stimuli. Probably the two most important classes bear the names *simple cells* and *complex cells*. These names were given by Hubel and Wiesel (1962, 1968), who first described these neuron classes, and the names reflect the fact that these researchers found it simpler to describe the response properties of the cells in the first class.

There exists a large amount of research results on how the spike trains emitted by simple and complex cells change in response to various visual stimuli. For example, it is known that many simple cells respond vigorously – that is, emit a large number of spikes in a given unit of time – when a line or an edge with a certain orientation is shown at a particular location of the visual field (e.g., Palmer 1999). Simple and complex cells may well be the most thoroughly studied types of neurons in the whole brain. Research has also shed light on the properties of many other neurons, located in different brain areas, which seem to take part in visual perception. To name just a few examples of recent research, there have been suggestions of neurons that signal the orientation of surfaces (Tsutsui et al. 2002), and neurons representing the quantity of a small number of visual objects (Nieder et al. 2002).

All of the previous examples were cases where an intuitively important visual quality – a small line or an edge, a surface, or the number of objects – was associated with the responses of a class of neurons. This approach is very intuitive, since it relates neural responses to something that we find easy to understand. Despite these advances in relating neural activity to visual qualities, there are several key open issues regarding the way in which visual information is represented in the brain.

First, some lines of evidence point to more complicated representations in the brain. For example, already in the 1960s results obtained in a branch of vision science called *visual psychophysics* – the quantitative study of the relations between visual sensations and stimuli – suggested that some processing in the visual system might utilize the representation of the visual world in terms of *spatial frequencies*, instead of basic spatial image elements like lines and edges (Blakemore and Campbell 1969). This suggestion received support later in neurophysiological measurements (Albrecht et al. 1980; DeValois et al. 1982). These measurements indicated that the response properties of simple cells are more complicated than what would be expected of line and edge detectors, and perhaps more suitable for a frequency-based representation.

Let us assume for a while that the visual system indeed utilizes a frequency representation. This prompts the obvious question: *why*? We certainly do not perceive ordinary visual scenes in terms of frequencies. To put it in another way, those physical properties of objects that are important to us seem to be related to everyday concepts like edges and surfaces, and not some esoteric properties like frequencies, which are alien to everyday thinking. Thus, it is difficult to find a match for the frequency representation in the way we intuitively perceive the visual world. Perhaps, then, the frequency representation is used by the visual system because it is suitable for the type of information processing tasks, or computations, that the brain performs.

Second, the same question *why* is just as important, but less obvious, even if we assume that edges, lines, and bars do constitute a low-level representation of the visual world in the brain. Currently we do not know how such a representation enables an animal to do its tasks. In fact, we do not know how any other information processing system, for that matter, could utilize such a representation to perform a wide variety of visual tasks; the design of general-purpose computer vision systems has proven notoriously difficult. While it may seem intuitive to first identify the lines and edges in an image, and then proceed to the identification of surfaces and objects, how this could really be accomplished in a complex visual scene remains a mystery. As Mumford (1994) has stated: "Introspection turns out often to be a very poor guide to the complexity of a problem." Therefore, in this case as well, we lack knowledge of the possible computational advantages of such a representation. Furthermore, cells in the primary visual cortex can not be described only by stating that they respond vigorously to basic image elements. The cells have other properties: for example, they respond differently to different types of motion (DeAngelis et al. 1993a), and the principles that seem to determine their locations on the cortex are complicated (Blasdel 1992). Again, we may ask the question: *why*? In order to provide an answer, we need quantitative theories that are able to predict the properties of visual neurons.

At this point a word of caution regarding the question *why* is in order. While it certainly looks like a good question to be posed, it is also a risky research topic, because the answer may be *very* complicated. Our biological characteristics are a result of millions of years of evolution. The way in which these characteristics are present in a population depends upon the traits of preceding populations, is influenced by random components, and is constrained by laws of physics and chemistry (Stearns and Hoekstra 2000). But, on the other hand, the potential payoff in terms of a significant increase in our understanding of the brain balances this risk.

An underlying assumption in this work is that a central element in the answer to the question *why* is provided by the properties of stimuli that we typically see.

In particular, we assume that the *statistical properties of visual stimuli* are related to the structure of the visual system (Attneave 1954; Barlow 1961; Simoncelli 2003; Olshausen 2003). Mechanisms that have enabled the formation of this relationship are evolution, and development of the visual system during the early stages of the life of an animal, including the fetal period. The models of computation examined in this thesis specify this relationship between simplified models of visual neurons and stimulus statistics. Our models of visual neurons are simplified models of real neurons – these simplified models can be used efficiently to perform various calculations, for example, in our estimation algorithms.

The models of computation discussed in this thesis can be thought of as ideas of what kind of a representation of visual stimuli the brain employs. They do not try to answer the question *why* with a complete answer: they do not state how the results of the computation affect the reproductive success, or the behavior, of an animal. Instead, the models try to offer a description of what seem to be the important, defining properties of the representation. Previous research on such models has produced intriguing connections between the structure of the visual system and the statistics of natural stimuli. For example, it has been demonstrated that a high proportion of a set of basic simple-cell models tends to have very low activity in a natural scene (Olshausen and Field 1996). This points to a representation in which activity on the primary visual cortex is *sparse*. To be more precise, the activity seems to be *maximally* sparse. Note again that this model of computation does not relate the properties of visual neurons to the way in which we observe the visual world, nor to the tasks of an animal, but to hypothetical information-processing principles.

To specify the relationship between stimuli and properties of visual neurons, a computational model contains a number of *parameters* whose exact values are not specified in the model. These parameters correspond to properties of visual neurons. The exact values of the parameters can be estimated by applying the relationship between the simplified models of visual neurons and natural stimulus statistics, as specified by the computational model. The resulting parameter values are then compared against neurophysiological measurements, or results obtained with other computational models. For example, we have compared the orientation selectivity of basic simple-cell models, specified completely after the parameters were estimated, to the orientation selectivity of simple cells. Such comparisons are an integral part of the assessment of the neuroscientific contribution of a model.

On the whole, the assessment of the neuroscientific contribution of a model consists of two parts. One is the analysis of the degree to which the model and real neurons support the same set of input-output mappings – a property which we will here call *implementational equivalence*. The second is the evaluation of the *predictive power* of the model where, ideally, we would like the implications of a model to agree with all previous physiological data, and also to provide an array of new hypotheses, which could then be verified or falsified by new neuroscientific measurements. Our models, as well as other existing similar computational models, are not completely implementationally equivalent to real neurons, nor do they possess perfect predictive power:

- Neurons are much more complicated than the simplified neuron models (see Section 2.4), which are mathematical abstractions that we use to represent the response properties of neurons; also, neurons do not perform exact mathematical calculations like our neuron models. On the other hand, our simplified neuron models seem to capture many important response properties of real

neurons, thereby providing a partial match to the observed computational (response) properties of real neurons.

- The evaluation of the predictive power of a model can be divided roughly into three stages: i) qualitative evaluation based on visual inspection of estimated parameter values and physiological measurements, ii) quantitative comparison against earlier physiological measurements, and iii) quantitative comparison against new physiological measurements inspired by the model. In this thesis, the evaluation of models of computation takes place at the level of the first two stages – that is, we evaluate the models qualitatively, and quantitatively against earlier physiological measurements. No new physiological measurements are reported in this thesis. Furthermore, the comparison of our predictions against earlier physiological measurements does not result in a perfect match.

The neurophysiological contribution of the models will be discussed further in Section 4.6, where we summarize the results obtained with our models.

The main scientific contribution of this thesis is the introduction of new models of computation in the primary visual cortex. The results obtained with these models suggest that one defining feature of the computations performed by simple cells is that the output of a neuron consists of limited periods of intense neuronal activity. We call this principle *temporal coherence of activity levels.* It also seems that the activity levels of nearby simple cells are positively correlated over short time intervals, resulting in a principle we call *spatiotemporal activity level dependencies.* Combining these principles, and the principles of sparseness and *spatial activity level dependencies* described in previous research, points to a model of cortical computation in which activation is limited to a few cortical patches at any given time, but when a patch is activated, it tends to remain active for a while. This is what we have called *bubble coding,* because one can visualize such activity as a set of bubbles in a three-dimensional space, defined by time and the two-dimensional cortical surface. An additional contribution of this thesis is the examination of the relationship between two previous computational models, independent component analysis (ICA) and local spatial frequency analysis. This analysis suggest that the results obtained with ICA are similar to wavelet representations, in that the spatial localization of the basis vectors obtained with ICA increases with their mean spatial frequency, while their localization in frequency decreases.

This thesis consists of an introductory part and 7 original articles. The rest of the introductory part is organized as follows. The publications of this thesis, along with the contributions of the current author, are described in the next section. Then an overview of the primary visual cortex is provided in Chapter 2. Chapter 3 contains a review of some important previous computational models, and a description of the contribution made in the first publication of this thesis. The main contribution of this thesis is described in Chapter 4, and a short summary is given in Chapter 5.

1.2 Publications of this thesis

The ordering of the publications of this thesis follows a logical progression from the analysis of previous computational models (Publication 1) to new models (Publications 2–7), so that the model describing the computational properties of individual neurons (Publication 2) is extended to multiple neurons (Publications 3 and 4),

and the two approaches are presented in a common generative model framework (Publication 5). The case of spatiotemporal receptive fields is then studied (Publication 6), and, finally, a model unifying a number of computational principles is presented (Publication 7).

Before their publication, all of the articles have gone through a review process in which at least two independent reviewers have assessed the manuscript.

Publication 1 Jarmo Hurri, Aapo Hyvärinen, and Erkki Oja. Wavelets and natural image statistics. In M. Frydrych, J. Parkkinen, and A. Visa (Eds.), *Proceedings of the 10th Scandinavian Conference on Image Analysis*, pages 13–18, 1997.

In this paper we analyzed the connections between wavelets and results obtained by applying independent component analysis (ICA) to image data, by using concepts from time-frequency analysis.

The current author suggested the application of concepts from time-frequency analysis to analyze ICA results, designed and performed the experiments and wrote the paper, with Dr. Hyvärinen and Prof. Oja taking part in the editing. The independent component analysis algorithm applied in the paper was originally developed by Dr. Hyvärinen and Prof. Oja.

Publication 2 Jarmo Hurri and Aapo Hyvärinen. Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Computation*, volume 15, number 3, pages 663–691, 2003.

This paper introduced temporal coherence of activity levels as a possible computational principle behind simple-cell receptive field structure. A gradient-based algorithm was developed for solving the resulting constrained optimization problem. The results (linear receptive field models) estimated from natural image sequences were compared quantitatively against corresponding results obtained with earlier models (independent component analysis / sparse coding). Several control experiments were also done to verify the novelty and validity of the results.

The concept of temporal coherence of activity levels was developed jointly by the authors of the paper. The current author also developed the algorithm, performed the experiments and wrote the paper, with Dr. Hyvärinen taking part in the editing.

Publication 3 Jarmo Hurri and Aapo Hyvärinen. A novel temporal generative model of natural video as an internal model in early vision. In A. E. C. Pece (Ed.), *Proceedings of the First International Workshop on Generative-Model-Based Vision*, pages 33–38, 2002.

This paper extended the concept of temporal coherence of activity levels by including inter-cell dependencies over time. These dependencies seem to capture important properties of the connectivity and organization of simple cells. Both intra- and inter-cell dependencies were captured by the first layer of the formulated two-layer multivariate autoregressive generative model. An algorithm based on the method of moments and least mean squares estimation was developed to solve the estimation problem.

The formulation of activity level dependencies as an autoregressive generative model was originally suggested by Dr. Hyvärinen. The model presented in this paper was developed jointly by the authors of the paper. The current author also developed the algorithm, performed the experiments and wrote the paper, with Dr. Hyvärinen taking part in the editing.

Publication 4 Jarmo Hurri and Aapo Hyvärinen. Temporal and spatiotemporal coherence in simple-cell responses: A generative model of natural image sequences. *Network: Computation in Neural Systems*, volume 14, number 3, pages 527–551, 2003.

This paper extended the work started in Publication 3 in several ways. First, we introduced a new and faster algorithm for solving the problem. Second, the differences between the model and standard independent component analysis were assessed. Third, the effect of the approximations made in the estimation was analyzed. An intuitive explanation of the results was also presented.

The new algorithm presented in this paper was developed jointly by the authors of the paper. The current author also performed the experiments and wrote the paper, with Dr. Hyvärinen taking part in the editing.

Publication 5 Jarmo Hurri and Aapo Hyvärinen. Temporal coherence, natural image sequences, and the visual cortex. In S. Becker, S. Thrun, and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems*, volume 15, pages 141–148, 2003.

In this paper temporal activity level coherence and inter-cell temporal activity level dependencies were presented in a unified generative model framework. This framework was originally suggested by Dr. Hyvärinen. The current author performed the experiments and wrote the paper, with Dr. Hyvärinen taking part in the editing.

Publication 6 Jarmo Hurri, Jaakko Väyrynen and Aapo Hyvärinen. Spatiotemporal linear simple-cell models based on temporal coherence and independent component analysis. *Proceedings of the Eighth Neural Computation and Psychology Workshop*, in press.

In this paper the concept of temporal coherence of activity levels was studied in the case of spatiotemporal receptive fields. The results were compared against physiological measurements and results obtained with independent component analysis / sparse coding.

The current author suggested the topic of the research described in the paper. The experiments were designed mainly by the current author, in co-operation with other authors. The experiments were done jointly by Mr. Väyrynen and the current author, with the current author as the supervisor. The current author also wrote the paper, with the other authors taking part in the editing.

Publication 7 Aapo Hyvärinen, Jarmo Hurri, and Jaakko Väyrynen. Bubbles: A unifying framework for low-level statistical properties of natural image sequences. *Journal of the Optical Society of America A*, volume 20, number 7, pages 1237–1252, 2003.

In this paper we defined a model which unifies three types of suggested coding principles in the visual cortex: sparseness, temporal activity level dependencies, and inter-cell activity level dependencies. The resulting coding principle, bubble coding, describes computation in the cortex as patches of activity which are contiguous over both time and cortical surface, and also occur rarely both in time and on the surface.

The generative model and the estimation method presented in this paper were developed by Dr. Hyvärinen, and the experiments were performed by Mr. Väyrynen. The current author participated in the design of the experiments, contributed part of the software used in the experiments and took part in the editing of the paper.

Chapter 2

The primary visual cortex: overview and basic neuron models

2.1 The primate visual system

This section is a very short introduction to the visual system of primates. While in general in this work we refer to measurements made from primates, we will in some cases also refer to studies concerning the visual system of the cat. There are two reasons for this. First, some groundbreaking measurements in the field of visual neuroscience were made from cats. Second, there are some important research results concerning cats which, to my knowledge, have not been reproduced in primates. For a review of the visual system of the cat, the reader is referred to (Sherman and Spear 1982).

The different brain areas related to perception are traditionally divided into *unimodal* and *multimodal* areas. The unimodal areas are thought to process information of one particular sensory type, such as visual information, or auditory information. The multimodal areas combine sensory information from a number of different modalities. For simplicity, here we limit our discussion to unimodal visual processing.

Unimodal visual processing in the primate brain is performed mainly in the *retina*, *thalamus*, and a number of *cortical areas* residing at different locations of the cortex.[1] The retina (e.g., Tessier-Lavigne 2000; Masland 2001) is located in the eye, and is the organ responsible for converting light intensities and spectra into neural activity. It also presumably takes part in other information processing tasks – such as contrast enhancement, motion detection, and control of pupil size and eye velocity – and has over ten different types of output (*ganglion*) cells. In the thalamus, which is located in the middle of the brain between the two *cerebral hemispheres*, the *lateral geniculate nucleus* (LGN) (e.g., Wurtz and Kandel 2000a; Sherman and Guillery 2002), is generally seen as a relay between the retina and visual cortical areas. Approximately 90% of outgoing fibers (*axons*) from the retina

[1] The *superior colliculus* and the *pulvinar* have been excluded from these considerations. The role of the superior colliculus in visual processing seems to be related to the control of voluntary eye movement; the role of the pulvinar is unclear (Bullier 2002).

terminate in the LGN. The neurons in the LGN are driven by the input from the retina, but they are also modulated by a large number of inputs from other brain regions; in fact, only 5–10% of the inputs to the relay neurons in the LGN are from the retina, while approximately 60% are from cortical areas and the *brainstem* (the rest are local connections coming from inside the LGN). For a short review of some computational models of the retina and the LGN, see Section 3.4.

Further unimodal visual processing is performed in different visual cortical areas, including

- the primary visual cortex (also known as *area V1*, or *striate cortex*)
- *area V2* and *area V4*
- *area V5* (also known as *medial temporal cortex*[2] or *area MT*),
- *inferior temporal cortex*[3] (also known as *area IT*), and
- *medial superior temporal cortex*[4] (also known as *area MST*).

The identification of these different areas is sometimes fairly complicated. In cases where visual areas are located adjacent to each other on the cortex, the identification is based on factors such as differences in the response properties of cells to visual stimuli, size and density of cell bodies, density of *myelin* (a substance which forms a sheath around axons), and changes in the way in which the response properties of nearby neurons are related to each other (Bullier 2002). The primary visual cortex is the largest visual area, and will be described in more detail below. The other visual cortical areas are generally thought to have various tasks in perception of form, color, motion, depth etc. (e.g., Wurtz and Kandel 2000b; Lennie 2000).

While a great deal is known about the properties of different visual cortical areas, the way in which a visual perception is formed in the brain is not clear. One of the most influential theories has been the classic feedforward processing paradigm (e.g., Lamme and Roelfsema 2000; Bullier 2002). In this paradigm, the visual system is considered as a hierarchy of processing layers, where the upper layers process increasingly more complex visual information. Neurons at low layers of the hierarchy, such as at the level of V1, are considered to respond to simple image features like short edges and lines (see also Section 3.2). At the higher layers, such as V4, the outputs of lower layers would then be combined to form representations of more complicated visual elements, such as shapes of objects (Pasupathy and Connor 2002). As useful and influential as the feedforward paradigm may be, it seems to be an insufficient model of the visual system (Lamme and Roelfsema 2000; Bullier 2002). Visual areas residing at different layers in the hierarchical model are in reality heavily interconnected, so feedback connections are abundant. The activities of neurons at the lower layers have been observed to be affected by phenomena that are supposed to take place at higher layers – one example is the perceptual completion of partially observed objects (Sugita 1999). While the feedforward model has definitely not been abandoned, a different type of model has gained considerable evidence: a model in which perception is a result of the dynamic interactions of different cortical areas (Lamme and Roelfsema 2000; Bullier 2001;

[2] medial: being or occurring in the middle; temporal: of or relating to the temples or the sides of the skull behind the orbits

[3] inferior: lower

[4] superior: upper

Tong 2003). We will return to the role of these interactions below when we discuss the primary visual cortex in more detail.

Parallel processing and specialization are common properties in the brain. For example, different parts of the visual field can be processed in parallel by different neural circuits, although these circuits typically have at least some interconnections. Regarding specialization, certain types of neurons can be sensitive, for example, to stimulus color, while the response of other neurons depends only on light intensity changes. Parallelism and specialization can be seen clearly in the case of different *visual pathways*: qualitatively different types of visual information are processed in at least partially physiologically separate channels in parallel. In the earliest parts of primate visual systems – the retina and the lateral geniculate nucleus – at least three different pathways have been identified (e.g., Wurtz and Kandel 2000a; Hendry and Reid 2000; Bullier 2002): *magno-* (M), *parvo-* (P), and *koniocellular* (K). Cells in the magno- and parvocellular pathway differ from each other in their response latencies, and the way they respond to changes in color, luminance, spatial frequency, temporal frequency: M cells respond faster than P cells, are insensitive to color, less sensitive to spatial frequency than P cells, and more sensitive to luminance and temporal frequency. Functional properties of cells in the koniocellular pathway seem to be more heterogenous and not very well known (Hendry and Reid 2000).

In the study of cortical visual areas, the two most prominent pathway candidates that have been identified are the *ventral*[5] and *dorsal*[6] pathways (e.g., Kandel and Wurtz 2000). Both pathways pass through V1, and supposedly also through V2; thereafter the ventral pathway extends via V4 to the inferior temporal cortex, while the dorsal pathway leads to the *posterior parietal*[7] cortex through area MT. The ventral pathway has been labeled as the "what" pathway, and seems to be mostly concerned with color and form. The dorsal pathway, on the other hand, has been named either as the "where" (Mishkin et al. 1983) or "how" (Goodale and Milner 1992) pathway, depending on whether it is thought to exhibit specialization in depth and motion, or in information about how to act on the objects (as opposed to representing the objects). It is possible that the ventral pathway is driven by the parvocellular pathway, and the dorsal pathway by the magnocellular pathway, but this seems to be a matter of some controversy (Hendry and Reid 2000). It seems that considerable convergence of the magnocellular, parvocellular and koniocellular pathways takes place in the primate V1 (Sawatari and Callaway 1996; Callaway 1998; Vidyasagar et al. 2002). Thus, in general it is not feasible to associate V1 neurons specifically with any of these prominent pathways.

Another general property of the brain is that the response properties of nearby neurons tend to have certain similarities. For example, *retinotopy* refers to the preservation of the spatial arrangement of inputs from the retina (Kandel and Wurtz 2000). That is, the spatial arrangement of neurons in a cortical area is similar to the spatial arrangement of their inputs in the retina. In the next section we will see some other such *topographic* properties of neurons in the primary visual cortex.

[5] ventral: abdominal (in this case, lower)

[6] dorsal: located near or on the back (in this case, upper)

[7] posterior: situated behind; parietal: of, relating to, or forming the upper posterior wall of the head

2.2 The primary visual cortex

This thesis focuses on modeling the response properties of neurons in the primary visual cortex. In this section we discuss shortly some aspects of current neurophysiological knowledge on the response properties of neurons in this visual area. However, it must immediately be noted that the responses of neurons in the primary visual cortex are affected by other parts of the visual system, especially those that lie in the pathway from the eye to this area of the brain. Therefore, when we talk about response properties of V1 cells, what we actually mean is the mapping from visual input to V1 responses; cells in the retina and the lateral geniculate nucleus obviously influence this response, and other parts of the brain may also do so.

In the term "primary visual cortex", the word "primary" refers to the idea that in a hierarchy of cortical areas, this part of the cortex receives/processes visual information first. What this really means is a complicated matter, since visual stimulus information can reach other cortical areas via routes that bypass the primary visual cortex (e.g., Schoenfeld et al. 2002), the visual areas are interconnected, and there are other visual areas which respond to visual stimuli almost as quickly as V1 (Vanni et al. 2001; Bullier 2002). But in general it seems that in the cortex, the processing of the majority of visual information at least begins in V1.

Originally, three major classes of cells were identified in the primary visual cortex: *simple cells*, *complex cells* and *end-stopped cells*, also known as *hypercomplex cells* (e.g., Hubel and Wiesel 1968; Palmer 1999). More recent measurements suggest that end-stopping is not a property of a particular class of neurons, but can be exhibited by simple or complex cells to different degrees (DeAngelis et al. 1994). In what follows, we will limit our discussion to simple and complex cells.

Simple and complex cells share some important response properties: they typically respond strongly to stimuli located at a certain position in the visual field and having a particular frequency (scale) and orientation (e.g., Hubel and Wiesel 1968; Palmer 1999; Wurtz and Kandel 2000a). Typical stimuli that were initially used to elicit strong responses from these cells included lines, edges, and bars, with different orientations (Hubel and Wiesel 1968); modern research describes stimuli that elicit strong responses in terms of *sinusoidal gratings* (see Figure 2.1). Most simple and complex cells are insensitive to variations in the color of the stimulus, that is, the neurons are *achromatic* (Hubel and Wiesel 1968; Lennie 2000); in what follows, we will limit our discussion to achromatic response properties. Also, in both cell classes, some cells are *directionally selective*, that is, the most vigorous response is elicited from these cells when the stimulus is moved in a certain direction with a certain speed (e.g., Hubel and Wiesel 1968; DeAngelis et al. 1993a).

But there are some fundamental differences between simple and complex cells. Originally the classification was based on whether a cell responded to both onset and offset of light at a certain position in the visual field (complex cells), or just one of these (simple cells) (Hubel and Wiesel 1962; Hubel and Wiesel 1968). A more modern description is that simple and complex cells differ in their sensitivity to the *phase/position* of the stimulus. This difference in the response properties of the cells is perhaps best illustrated by considering their responses to sinusoidal gratings, which are standard tools in the study of visual cells. A sinusoidal grating is completely specified by four parameters: amplitude, orientation, frequency, and phase. Figure 2.1 illustrates two such gratings, having the same amplitude, orientation, and frequency, but differing in their phase. When a simple cell is excited with these gratings, its response varies greatly with the phase of the grating, while

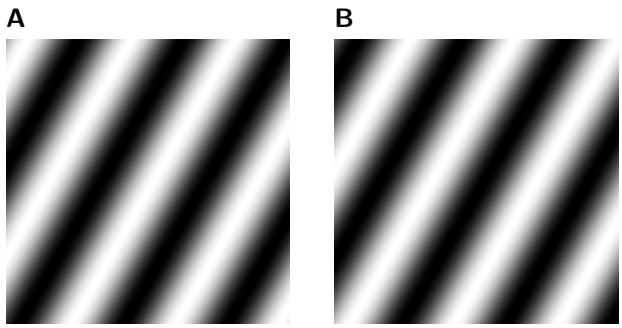


Figure 2.1: Two sinusoidal gratings. The grating in (B) has the same amplitude, frequency, and orientation as the grating in (A), but different phase.

a complex cell is insensitive to phase changes.

One theory is that complex cells exhibit this phase/position invariance because they combine the outputs of multiple similarly oriented simple cells with similar preferred spatial frequencies and directional selectivities, but different phases/positions (Hubel and Wiesel 1962). Recent neurophysiological measurements made from cats support this theory (Alonso and Martinez 1998; Martinez and Alonso 2001). We will return to this theory below in Section 2.4. There are also other theories hypothesizing, for example, that physiologically different pathways lead to simple and complex cells (see, e.g., the short review Callaway 2001; see also Chance et al. 1999).

Much of the research on the primary visual cortex concerns the way in which simple and complex cells are organized in this inherently three-dimensional cortical area. It seems that the organization is such that several properties of nearby cells tend to be continuous, that is, change slowly (e.g., Callaway 1998; Blasdel and Campbell 2001). The most important of these properties seems to be

- the preferred orientation of the cell
- *ocular dominance*: which eye dominates in determining the response of the cell
- retinotopy (see Section 2.1, page 17).

The resulting ordering with respect to these properties seems to follow primarily *columnar* organization: if one moves directly into the cortex from the surface of the cortex, the properties stay the same. Gradual changes in the properties take place when one moves along the surface of the cortex. This means that when one examines how preferred orientation, ocular dominance, and retinal position change in the cortex, the examination can be restricted to two dimensions, that is, to changes along the surface of the cortex. It seems that different properties dominate in the ordering at different scales: preferred orientation seems to be the dominant property at the smallest scale, and retinotopy at the largest (Blasdel and Campbell 2001). The resulting topography is very complex (Blasdel 1992). Still, in general, cells that code for similar orientation, input coming from the same eye, and similar retinal locations, tend to be close to each other. Another property of the cells that may have an effect on the topographic organization is preferred spatial frequency (Tootell et al. 1988; DeAngelis et al. 1999).

In the classic feedforward processing paradigm (see Section 2.1), the role of V1 has been the identification of elementary image features (e.g., Palmer 1999; Wurtz and Kandel 2000a). According to this traditional textbook view, simple and complex cells detect elementary image features – like lines, bars, and edges – and pass these on to other visual cortical areas for further processing. However, the primary visual cortex is connected reciprocally to other visual areas (e.g., areas V2 and V5) – that is, there is a dense network of feedback connections to V1 from "higher" cortical visual areas (Bullier 2002). Also, as will be discussed in the next section, the way in which a V1 neuron responds is affected by the context in which the stimulus is perceived. It seems likely that V1 is not only a servant of the higher visual areas, but an active component in visual perception (Bullier 2001; Rees et al. 2002; Tong 2003). This complicates the description of response properties of V1 neurons, as we will see next.

2.3 Classical receptive fields and contextual effects

The output of a neuron in the primary visual cortex is a continuous train of action potentials (spikes), all-or-none type of impulses with a duration of approximately 1 ms (Kandel 2000). With no visual input into the retina, the visual neurons still spike occasionally. The rate at which this spiking occurs is called the *spontaneous firing rate*. In the primary visual cortex, the spontaneous activity of different neurons is correlated, and forms similar activity patterns as when actual visual input activates the cells (Tsodyks et al. 1999). The spontaneous firing rate decreases when moving up the visual pathway, that is, it is high in the retina but low in the cortex (e.g., Bair et al. 2002). What is particularly important from the point of view of this work is that the spontaneous firing rate of simple cells is typically low (Heeger 1992; DeAngelis et al. 1993a).

For a given neuron, the part of the visual field where a change in light intensity can raise the firing rate above the spontaneous firing rate is called the *classical (spatial) receptive field* of the neuron (e.g, Freeman et al. 2001; however, see also Sugita 1999 for an experiment in which simple cells exhibited excitation even with no change inside the classical receptive field). For example, the simple cells of monkeys can be excited from visual areas ranging typically between $\frac{1}{4}^\circ$ and $\frac{3}{4}^\circ$ (Hubel and Wiesel 1968). Complex-cell receptive fields are generally thought to be larger than simple-cell receptive fields (Hubel and Wiesel 1968), although some more recent measurements indicate that in cats, the receptive fields of these two cell types have similar sizes (Freeman et al. 2001). *Binocular* neurons – that is, neurons which can be excited from either eye – actually have two receptive fields, one for each eye (Bullier 2002); however, in this work we focus on modeling *monocular* neurons. The same term "receptive field" is also used to refer to the part of the retina where a change in light intensity can excite the neuron.

If there is no change in incoming light inside the classical receptive field of a visual neuron, the activity of the neuron remains very low (Bullier 2002). That is why achromatic neuronal response properties in fact describe responses to *light intensity changes*. Below we will see that the description of a classical receptive field is often accompanied by a map showing the effect of light intensity changes at different positions of the visual field. In fact, in neuroscience the term "receptive field" is tightly associated with this map. When this kind of terminology is used, the (classical) *spatiotemporal* receptive field refers to a combination of the classical spatial receptive field, and a map showing the temporal profiles of the effect of

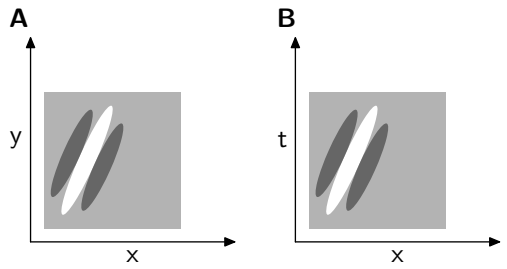


Figure 2.2: A simplified illustration of classical spatial and spatiotemporal receptive fields. (A) A prototype of an orientation-selective spatial receptive field that would respond strongly to a suitably oriented line or bar. The intensity of each point (x, y) in the receptive field describes how a light change at time τ at that position would affect the firing rate of the cell at time $\tau + \Delta\tau$, where $\Delta\tau$ denotes the time it takes for the cell to respond to visual stimuli. The medium gray background indicates the area where light intensity changes have no effect. White color inside the medium gray background indicates area where changes to brighter tend to increase the firing rate, and changes to darker tend to decrease the firing rate; in the darker gray areas, the effects of light changes are the opposite. (B) The cross-section of a prototype of a directionally selective spatiotemporal receptive field that would respond strongly to a vertical line or bar moving into a suitable direction with an appropriate speed. The actual receptive field is three-dimensional, with coordinates (x, y, t) – for simplicity, we show only one x–t cross-section of the receptive field at a fixed y-coordinate $y = y_0$, and assume that the x–t profile of the receptive field is similar for all y. The intensity of each point (x, y_0, t) in the receptive field describes how a light change at time $\tau - t$, at spatial position (x, y_0), would affect the firing rate of the cell at time $\tau + \Delta\tau$, where $\Delta\tau$ denotes the time it takes for the cell to respond to visual stimuli. The grayscale coding is the same as in (A).

light intensity changes at different positions. Simplified examples of spatial and spatiotemporal receptive fields are shown in Figure 2.2.

While excitation of achromatic visual neurons above the spontaneous firing rate is possible only if there are light intensity changes within the classical receptive field, there are many other factors which influence the induced response.

- Spatially, the strength of the response can be affected from regions outside the classical receptive field (e.g., Zipser et al. 1996; Freeman et al. 2001). These regions are often described as belonging to the *nonclassical* receptive field.

- Temporally, adaptation to input can modify the response strength of the neuron (Marlin et al. 1988), or even the structure of the receptive field (Marlin et al. 1991; see also Stanley 2002).

- Factors other than visual stimuli – such as the behavioral state of the animal – also affect the responses of visual neurons. For example, attention or arousal can change the response properties of neurons in the primary visual cortex (Wörgötter and Eysel 2000).

Collectively these phenomena are called *contextual effects* (Albright and Stoner 2002). Most of these effects influence the response properties on a longer timescale

than what is relevant for the models discussed in this thesis. In what follows, we will focus on the relatively fast modulatory spatial contextual effects, which are inherently related to the nonclassical receptive field.

There exists evidence that classical response properties (as defined by classical receptive fields) and spatial contextual effects represent two different stages of processing of stimulus information in the visual system: spatial contextual modulation seems to take place some 100–250 ms after the initial response of the neuron (Zipser et al. 1996). This points to a first stage of a fast feedforward sweep – described by the classical receptive fields of neurons – followed by recurrent processing which mediates spatial contextual effects (Lamme and Roelfsema 2000). When an understanding of the functioning of the visual system is developed, it is important to understand both the properties of the feedforward stage, and the complicated, recurrent interactions taking place at later stages. In the case of physiological studies of the primary visual cortex, studies of the classical receptive fields of simple and complex cells, as well as research on how complex cells utilize simple-cell outputs in the feedforward stage, represent the former effort.

In this work we too, like those neuroscientists who have studied classical receptive fields, apply the reductionist approach to the problem of understanding vision: we focus on computational modeling of the feedforward stage and, as a consequence, on the classical receptive fields of neurons. We think that a fundamentally important approach to building computational models of the brain is to break the computations performed the whole system into smaller constituents, insofar as this is possible. Also, from the practical point of view of building computational models, we have to dissect the visual system into parts which lend themselves to such modeling and further analysis. Some models of classical receptive fields are suitable for these purposes, as we shall see in the next section. Furthermore, such descriptions of classical receptive fields can be used to predict a number of neuronal response properties: preferred orientation, direction, spatial frequency etc. Thus, as we shall see next, they provide one answer to a central problem in the description of neurons: how well can we capture the response properties of a visual neuron with a convenient mathematical model?

2.4 Basic neuron models

Capturing response properties

There are many different ways to characterize the responses of visual neurons to light intensity changes inside their receptive fields (e.g., Palmer 1999). For example, one can measure the (spatial) *orientation selectivity* of the neurons by examining how the orientation of a bar or a sinusoidal grating (see Figure 2.1, page 19) influences the response. Similarly, *frequency selectivity* indicates how the response depends on the frequency of the sinusoidal grating, and *phase selectivity* measures dependence on the phase of the grating. When the temporal domain is included, one can for example measure the *direction selectivity* of the neuron, that is, how the response depends on the direction of movement of a stimulus.

All of these selectivities are important characterizations of visual neurons, but in order to obtain a full description of neural response properties, a model giving the output of the neuron as a function of the input is desired. The development of such models involves model design and evaluation, and for models with free parameters, also parameter estimation from physiological measurements.

One of the open issues in neuroscience is the question of the type of information carried by the signal at different timescales (e.g., deCharms and Zador 2000). That is, what properties of the neural representation of stimuli, motor commands, or other concepts presented in the brain can be "read out" by considering average firing rates over a certain time interval. For example, in the middle temporal (MT) visual area, the temporal precision of spike trains seems to be of order 2–3 ms, as measured by the standard deviation of spike times in repeated experiments with the same stimulus (Buračas et al. 1998). Here the mean firing rate seems to convey information about the mean direction of motion, and measuring the rate over shorter time intervals seems to provide more detailed information about the fine temporal structure of the motion. This suggests that in this case, a shorter time window means better temporal precision. However, it is an open issue whether qualitatively different type of information than temporal structure – such as the spatial structure of a stimulus – can be passed at different timescales (deCharms and Zador 2000). In general, the question of exact timing is also related to the possibility of synchronized activity among a set of neurons.

In the computational models studied in this thesis, the output of a neuron is the mean firing rate taken over nonoverlapping 40 ms time windows. This temporal precision is set by the 25 Hz sampling rate of the natural image sequence data we are using. Ideally, a faster sampling rate would be desirable – however, previous research has shown that the 25 Hz sampling rate is adequate, for example, for quantitative analysis of resulting spatiotemporal receptive field models. In fact, one of the advantages of our data set is that it has also been used in other research (van Hateren and Ruderman 1998; Olshausen 2000), which improves the comparability of the results.

After the mean firing rate has been chosen as a descriptor of cell output, and we want to use a model giving the output of the neuron as a function of visual stimulus, the question then becomes: what kind of a model is sufficiently powerful? On one hand, we want the model to be as simple as possible, so that its parameters can be estimated successfully from measurement data, or from natural stimulus data (by using the rules specified by the computational model). On the other hand, the model should be expressive enough to account for the observed properties of the actual neurons: for example, we might want the model to account for the observed orientation selectivity of simple cells. Below, we will consider some basic neuron models used in computational neuroscience, and discuss their applicability in modeling cells in the early visual system, especially neurons in the primary visual cortex.

Linear models

Linear models are the ubiquitous workhorses of science and engineering. They are also the simplest successful neuron models of the visual system, despite the fact that their output can be negative, while the firing rate of a neuron is always nonnegative. For our purposes the best way to present them is in a vector-matrix formulation. In this formulation, spatial or spatiotemporal visual input – that is, an image or an image sequence – is *vectorized* before it is used as input in the model. Vectorization means that we transform two- or three-dimensional data into a one-dimensional form which lends itself easily to certain computations. For images, vectorization can be done by concatenating the columns of the two-dimensional image into a vector – for an image of size $N \times N$ pixels this yields a vector of length N^2. For

a sequence of length T of such images, we concatenate all the T vectors obtained by applying the previous procedure to each image in the sequence – this yields a vector of length TN^2.

For static input – that is, images – we use the following notation to denote a linear model. Let $\mathbf{x}$ denote a vectorized image patch. For such an image patch input, the output of a neuron with index k, denoted by y_k, $k = 1, ..., K$, is given by

$$y_k = \mathbf{w}_k^T \mathbf{x} = \sum_i w_{k,i} x_i, \tag{2.1}$$

where vector $\mathbf{w}_k$ is the vectorized *spatial filter* implementing a *linear transformation* from $\mathbf{x}$ to y_k. Collecting the outputs of a set of K neurons into one output vector $\mathbf{y} = [y_1 \cdots y_K]^T$, the input-output relationship can be expressed as

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{2.2}$$

where $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_K]^T$. Temporal input, such as image sequences, can be transformed linearly by using spatial, temporal, or spatiotemporal linear models. Let us denote vectorized visual input at time t by vector $\mathbf{x}(t)$. Then in the linear model the output of a neuron with index k at time t, denoted by $y_k(t)$, is given by

$$y_k(t) = \mathbf{w}_k^T \mathbf{x}(t), \tag{2.3}$$

or, denoting $\mathbf{y}(t) = [y_1(t) \cdots y_K(t)]^T$,

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t). \tag{2.4}$$

If we are using a spatial linear model, then $\mathbf{x}(t)$ is just a vectorized image from the image sequence at time t. If the model is spatiotemporal, then $\mathbf{x}(t)$ is a vectorized subsequence of images from the sequence at times $t, t-\Delta t, ..., t-(T-1)\Delta t$, and two temporally close input vectors $\mathbf{x}(t)$ and $\mathbf{x}(t-\Delta t)$ correspond to partly overlapping image sequences. The purely temporal model can be considered, for example, as a spatiotemporal model where each image consists of only a single pixel.

In the previous section, we discussed the map describing the effect of intensity changes at different positions of the receptive field of the cell. The linear model is a basic mathematical description of this map. The average image intensity (mean of $\mathbf{x}(t)$) is typically assumed to be $\mathbf{0}$, so that average intensity yields a zero firing rate – this is then defined to be the spontaneous firing rate (see Section 2.3). Thus, in the model, $\mathbf{x}(t)$ is equivalent to changes from average image intensity, and $\mathbf{y}(t)$ is equivalent to deviations from the spontaneous firing rate, and the model specifies a relationship between these two. The filter $\mathbf{w}$ is simply called the receptive field of the cell. Simplified graphical examples of what the contents of $\mathbf{w}$ could be were shown in Figure 2.2 (page 21). In those examples, an element of $\mathbf{w}$ could only have one of three values: 0 (medium gray region), $a > 0$ (white region), or $b < 0$ (dark gray region).

Linear receptive-field models can be estimated from visual neurons by employing a method called *reverse correlation* (Dayan and Abbott 2001). In this method, a linear filter is estimated so that the mean square error between the estimated $y_k(t)$ in equation (2.3), and change in the actual firing rate is minimized, where the mean is taken over a large set of visual stimuli. The name "reverse correlation" comes from the fact that the general solution to this problem involves the computation of the time-correlation of stimulus and firing rate. However, the solution is simplified

when temporally uncorrelated (white noise) sequences are used as visual stimuli – in this case, the optimal $\mathbf{w}$ is obtained by computing an average stimulus over those stimuli which elicited a spike.

As was mentioned above, the linear model maps changes in light intensity to changes in the firing rate. If a neuron has a relatively low spontaneous firing rate, this may be problematic: the firing rates predicted by the linear model may then tend to be negative, thus providing a poor match with how the neuron responds. This is less of a problem for ganglion and LGN cells, since their spontaneous firing rates are relatively high. Linear models have turned out to account for several properties of retinal ganglion cells (Enroth-Cugell et al. 1983; Enroth-Cugell and Robson 1984), and have proved to be useful models in the description of cells in the lateral geniculate nucleus (Cai et al. 1997).

As for modeling simple and complex cells, neurons in the primary visual cortex have relatively low spontaneous firing rates (Heeger 1992; DeAngelis et al. 1993a). Nevertheless, in the case of simple cells, linear models have turned out to be good predictors of orientation selectivity (orientation preference and tuning), and also account reasonably well for spatial and temporal frequency selectivity, and direction of preferred motion (DeAngelis et al. 1993b; Lampl et al. 2001). There are a number of simple-cell response properties that linear models are not able to predict. The most important may be the magnitudes of the responses to moving stimuli, which also determine the degree of directional selectivity (Tolhurst and Dean 1991; DeAngelis et al. 1993b; Lampl et al. 2001). Some other discrepancies between linear predictions and actual cell responses will be described in the next section, along with more advanced models. For complex cells, linear models are clearly inadequate (e.g., Movshon et al. 1978; Szulborski and Palmer 1990). For example, the original characteristic feature of complex cells was considered to be their positive response to both light increases and decreases at the same position in the visual field – this is something a linear model can not predict.

As for the use of linear neuron models in computational models of visual neurons, linear models have been used widely in computational models of retinal neurons (e.g., Atick and Redlich 1990), neurons in the lateral geniculate nucleus (e.g., Dong and Atick 1995b), and simple cells (e.g., Olshausen and Field 1996; Olshausen and Field 1997; Bell and Sejnowski 1997; van Hateren and van der Schaaf 1998; van Hateren and Ruderman 1998; Olshausen 2000). Despite their limitations as simple-cell models, the results obtained with linear models have created great interest in the neuroscience community. For example, computational models utilizing sparseness (Olshausen and Field 1996; Olshausen and Field 1997) and independent component analysis (Bell and Sejnowski 1997; Hyvärinen et al. 2001) (see Section 3.4) have prompted comparisons against neurophysiological data (van Hateren and van der Schaaf 1998; van Hateren and Ruderman 1998) and new neurophysiological measurements (Vinje and Gallant 2000; Weliky et al. 2003).

The starting point in the new computational models introduced in this thesis is the application of linear models, as was also the case with sparse coding and independent component analysis. However, we will be paying attention to the interpretation of the results when some basic nonlinearities are taken into account. In what follows we will consider some nonlinear models for simple and complex cells.

Basic nonlinear models for simple and complex cells

The response properties of simple cells include a number of nonlinear characteristics, and complex cells require nonlinear models in the first place. Here we will shortly discuss three different types of models to handle nonlinearities in V1 neurons: the *Wiener model*, the *divisive normalization model*, and the *energy model*.

In the Wiener model (e.g., Mathews and Sicuranza 2000), a linear stage is followed by a static nonlinearity f:

$$y_k(t) = f\left(\mathbf{w}_k^T \mathbf{x}(t)\right). \tag{2.5}$$

In neuron models, f is typically nonnegative so that $y_k(t)$ will fulfill this basic requirement of the firing rate. A special case of the Wiener model is *half-wave rectification* (e.g., Heeger 1992), defined by

$$f(\alpha) = \max\{0, \alpha\}. \tag{2.6}$$

This is the nonlinearity we will occasionally apply in this thesis. Half-wave rectification offers one way to interpret the purely linear model (2.3) in a more physiologically plausible way: the linear model combines the outputs of two half-wave rectified (nonnegative) cells with *reversed polarities* into a single output $y_k(t)$ – one cell corresponds to linear filter $\mathbf{w}$ and the other to filter $-\mathbf{w}$ (see Publication 2).

It was mentioned in the previous section that linear models of simple cells fail to predict the degree of directional selectivity of simple cells. A Wiener model which performs better in this prediction is one consisting of a cascade of half-squaring and an *expansive exponent* (Albrecht and Geisler 1991; DeAngelis et al. 1993b):

$$f(\alpha) = (\max\{0, \alpha\})^n, \tag{2.7}$$

where $n > 1$. When the exponent n is estimated from neurophysiological measurements, its value turns out to be different for different neurons, with a mean of approximately 2.3–2.5 (Albrecht and Geisler 1991; DeAngelis et al. 1993b).

One of the most accurate currently known simple-cell models, in terms of predictive power, is the divisive normalization model (e.g., Heeger 1992; Carandini et al. 1997; Carandini et al. 1999). Let $\mathbf{w}_1, ...\mathbf{w}_K$ denote a set of filters, and σ a scalar parameter. In the divisive normalization model, the output of the cell corresponding to filter $\mathbf{w}_k$ is given by

$$y_k(t) = \frac{f\left(\mathbf{w}_k^T \mathbf{x}(t)\right)}{\sum_{i=1}^{K} f\left(\mathbf{w}_i^T \mathbf{x}(t)\right) + \sigma^2}, \tag{2.8}$$

where f is again a static nonlinearity, such as half-wave rectification. The divisive normalization model can account for a number of observed simple-cell nonlinearities, including *response saturation*: at high contrast values, a change in input contrast yields a smaller change in cell response than what is predicted by the linear model (Carandini et al. 1999).

As was already mentioned above, for complex cells the linear model is completely inadequate. A basic nonlinear complex-cell model is the *energy model* (Adelson and Bergen 1985; see also Watson and Ahumada 1985). In this model, the output of a complex cell is computed as a sum of squares of the responses of linear filters:

$$y_k(t) = \sum_i \left(\mathbf{w}_i^T \mathbf{x}(t)\right)^2. \tag{2.9}$$

This model bears obvious resemblance with the idea that a complex cell combines the responses of a number of simple cells to achieve phase invariance, as was discussed in Section 2.2. Each of the linear transformations $\mathbf{w}_i^T \mathbf{x}(t)$ in equation (2.9) would then correspond to the outputs of two half-rectified simple cells, with receptive fields $\mathbf{w}_i$ and $-\mathbf{w}_i$, as discussed above, and the different filters $\mathbf{w}_i$ would have similar orientation, frequency, and scale, but different phase/position. The energy model has received some support from neurophysiological measurements (Emerson et al. 1992; Gaska et al. 1994).

Chapter 3

Some previous computational principles for early vision

3.1 Approaches to understanding visual processing in the brain

Different approaches can be taken in order to study the nature of computations in the visual system. Perhaps the most straightforward one is to try to relate neural responses to visual qualities that are familiar to us in everyday life – we will examine an example of this approach in more detail in Section 3.2. In the history of vision science, another approach resulted from the examination of the capabilities of the visual system in terms of their response to different spatial frequencies – we will return to this approach in Section 3.3.

A majority of the studies discussed here, including the major contribution of this thesis (Chapter 4), and many of the previously suggested computational principles described in this chapter (Sections 3.4–3.6), belong to an approach in which it is assumed that the visual system has adapted to the statistics of visual stimuli that the animal receives in its environment (e.g., Attneave 1954; Barlow 1961; Field 1987; Field 1994; Simoncelli and Olshausen 2001; Olshausen 2003; Simoncelli 2003). The approach has been applied especially to the parts of the visual system which are thought to be involved in the very first stages of visual processing – the retina, the LGN, and the primary visual cortex – which are often called collectively the early vision.

In this approach, it is assumed that the properties of visual stimuli have influenced the structure and functionality of the visual system either genetically or during development – although it must be remembered that developmental mechanisms themselves are under the pressure of evolutionary selection – so that some properties of the system have become optimal for statistics of the stimuli. Research on visually deprived cats (Sherman and Spear 1982), cats reared in strobe illuminated environments (Humphrey and Saul 1998), and cats reared in environments where a single orientation dominates the visual input (Blakemore and Cooper 1970; Sengpiel et al. 1999), clearly shows that the properties of visual input during the early stages of the life of an animal play a role in determining the response properties of visual cells. It is therefore clear that properties of visual stimuli have a major influence on the response characteristics of visual neurons.

But there are several possible pitfalls in this hypothesis which relates typical natural stimuli to the way in which the visual system functions. First, the concept of a "typical" stimulus is very hard to define (Simoncelli 2003). Customarily in this field a set of visual stimuli, such as images or image sequences, is collected, and this set is considered to be representative of typical stimuli; however, there is currently no way to quantify the degree to which the data set covers the space of natural stimuli. Second, it is also dangerous to assume that evolution or development yields behavior or structures that are absolutely optimal. Evolution is nondeterministic and greedy, and builds upon earlier solutions – in many cases this yields nonoptimal solutions. Even though at least some behaviors – such as deciding the composition of the diet, or sampling uncertain food locations – of some species seem to be optimal or near-optimal (e.g., Krebs and Davies 1993), in some cases the results of evolutionary selection are not optimal. In the domain of vision, visual illusions (e.g., Palmer 1999; Eagleman 2001) represent one type of nonoptimality. In addition, even if the outward behavior of an animal would have been driven by evolution and development to be optimal or near-optimal, there can be many different internal mechanisms which yield similar fitness values. In terms of evolutionary theory, this is a consequence of *neutral evolution* – genetic variation that is not correlated with reproductive success – which has been demonstrated experimentally with replicated bacteria populations (e.g., Stearns and Hoekstra 2000). It is therefore even more difficult to draw any conclusions about the optimality of internal physiological mechanisms.

Furthermore, the computational models described so far in literature practically ignore the tasks of the organism and computational constraints set by the limitations of neurons (Simoncelli and Olshausen 2001). The tasks of the organism, in turn, depend on factors such as predator-prey relationships, the diet of an animal, mating behavior etc. It seems plausible that in higher animals, such as mammals, the representation of sensory stimuli is least dependent on the external functions of the organism (such as motor commands) on the primary sensory cortices, and the dependency increases in neural networks closer to the motor cortex or similar areas (deCharms and Zador 2000). This seems reasonable because low-level representations are used for many different behavioral purposes. Therefore, examining the relationship between behavior and neural activity on the primary visual cortex can be very difficult. However, the relationship between neuronal activity and cognitive tasks, such as perception, memory and learning, is also of great importance, and might be easier to investigate. But current computational models have not been able to establish this connection.

In spite of all these reservations, the study of the relationship between sensory systems and the statistics of stimuli has gained considerable interest (e.g., Field 1987; Olshausen 2003; Simoncelli 2003). Within this research field, two different types of computational model families can be identified. The first family of models focuses on optimal properties of the outputs of neurons. These models typically optimize the output of linear or nonlinear models using some optimality principle, such as statistical independence, or minimal change over time. The second family consists of *generative models* of natural stimuli (Hinton and Ghahramani 1997; Olshausen 2003). In these models, natural stimuli are generated by underlying, hidden (latent) variables, according to the rules specified by the model. When the model is estimated from natural stimulus data, the parameters of the model and the latent variables may then represent the properties and outputs of neurons. The idea of describing natural stimuli by a generative model, and interpreting the hidden

variables of this model as a neural representation, may at first seem counterintuitive, because the stimuli are not generated by the neural network. However, a generative model can express explicitly information about the *regularities in the stimuli* as *properties of hidden variables*. If these regularities can be used to make inferences about the underlying real world, the visual system might benefit from such an internal representation of its stimuli (Knill and Richards 1996).

What is most important for research on the relationship between sensory systems and the statistics of stimuli is that this research has also spurred new neurophysiological measurements in two sensory systems – the visual system (van Hateren 1992; Dan et al. 1996; Vinje and Gallant 2000; Nirenberg et al. 2001; Weliky et al. 2003; see also the review Reinagel 2001) and the auditory system (Chechik et al. 2001) – and also new psychophysical measurements (Párrage et al. 2000). In computational modeling, the emphasis has been in the utilization of redundancy – which is discussed in more detailed below – but other models have been developed as well.

What follows below is a short review of some of the central computational principles and models related to early parts of the visual system.

3.2 Line and edge detection with filters

Line and edge detection are basic image processing tasks, used for example in the initial stages of image segmentation in machine vision (e.g., Gonzalez and Woods 1992). A common way to detect these basic image elements is to identify certain properties in the responses of one or more filters. To illustrate this, let us consider the following simple example. Let $\mathbf{x}_1, ..., \mathbf{x}_N$ denote a set of 9-dimensional vectorized (see Section 2.4) small image patches, originally of size 3×3 pixels, taken from a larger image. Let us remove the mean intensity from each of these patches, and normalize them to unit Euclidean norm; these steps are useful when we want to discard the mean intensity and local average contrast. Let us denote the resulting vectors by $\mathbf{x}_{n,*}$, $n = 1, ..., N$, for which, because of the previous two steps, we have $\sum_i x_{n,*,i} = 0$ and $\|\mathbf{x}_{n,*}\| = 1$. Now let $\mathbf{w}$ denote the vectorized form of the spatial filter shown in Figure 3.1A. Note that $\sum_i w_i = 0$, and that the multiplier $1/(3\sqrt{2})$ in front of the spatial filter ensures that $\|\mathbf{w}\| = 1$; these properties simplify things a bit below. Let y_n denote the result obtained when $\mathbf{x}_{n,*}$ is filtered with $\mathbf{w}$:

$$y_n = \mathbf{w}^T \mathbf{x}_{n,*} = \sum_i w_i x_{n,*,i}. \tag{3.1}$$

Equation (3.1) is the *inner product* between vectors $\mathbf{w}$ and $\mathbf{x}_{n,*}$. Since each $\mathbf{x}_{n,*}$ has unit Euclidean norm, the maximal y_n is obtained when $\mathbf{x}_{n,*} = \mathbf{w}$, that is, when $\mathbf{x}_{n,*}$ corresponds to a thin oblique white line. The minimal y_n is obtained when $\mathbf{x}_{n,*} = -\mathbf{w}$, that is, when $\mathbf{x}_{n,*}$ corresponds to a thin oblique black line. Thus, each y_n can be considered to be a measure of the similarity between $\mathbf{w}$ and $\mathbf{x}_{n,*}$: the larger y_n, the more similar the image patch is with the template $\mathbf{w}$. By applying a (nonlinear) scheme like the one described here – removal of mean intensity, normalization, and linear transformation – at every position of a larger image, the resulting values y_n could be used to detect image positions containing thin oblique lines. Similarly, the filter in Figure 3.1B could be used to detect vertical edges. Other properties than maxima or minima can also be considered. For example, one operational definition of an edge is that it is a maximum of

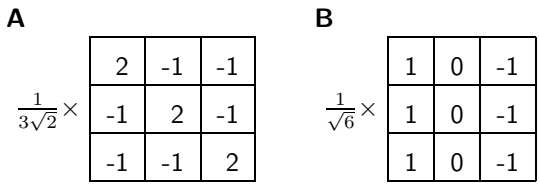


Figure 3.1: Simple spatial filters for the detection of lines and edges. (A) A filter which can be used to detect thin oblique lines. (B) A filter which can be used to detect vertical edges. The multipliers in front of the filters scale the filters so that they have unit norm. See text for details.

the derivative of intensity – such maxima can be detected as zero-crossings of filters computing some approximation of the second derivative (Gonzalez and Woods 1992; Palmer 1999). Also, spatiotemporal filters similar to the ones in Figure 3.1 can be devised to detect moving lines and edges.

The idea of simple cells as detectors of lines, edges, and bars (thicker lines) is without a doubt the most influential suggestion of the computations performed by these cells. As was mentioned in Section 2.2, in the original studies by Hubel and Wiesel (1962, 1968), simple cells were observed to respond most vigorously to one of these image elements with a certain orientation. This observation, together with the importance of the detection of lines and edges in engineering-oriented approaches, and the construction of filters such as those presented in Figure 3.1, led to the idea that the task of simple cells was to identify low-level image features (see also Marr and Hildreth 1980). Because of its fundamental importance, this theory of the computational role of simple cells is the one that is presented, for example, in some neuroscientific textbooks.

However, there are some reasons that have led researcher to study other hypotheses about the tasks of these cells. The first two reasons are related to important psychophysical and neurophysiological observations:

1. Psychophysical observations led to a suggestion that images might not be processed in the early parts of the visual system as a collection of local primitive spatial elements, such as short edges and lines, but spatial frequency elements (sinusoidal grating patches) (Blakemore and Campbell 1969; Palmer 1999). To be more precise, it was suggested that there are neurons in the visual system that are selective especially to the spatial frequency of the stimulus.

2. More detailed measurements of simple-cell responses to different stimuli, and of the receptive fields of simple cells, suggested that the spatial structure of these receptive fields is somewhat more complicated than that of a basic edge or line detector, and provided support for the frequency element processing model (Albrecht et al. 1980; DeValois et al. 1982).

We will return to these observations and models in the next section.

In addition, while the edge and line detection hypothesis is important and intuitive, it seems that it is too simplistic to offer a true explanation of the role of these cells. Engineering efforts utilizing such simple detectors have shown that the identification of objects is a very difficult task – in many cases other methods have to be used, and the selection of the method depends largely on the particular characteristics of the problem (Gonzalez and Woods 1992). From the point of view of

visual neuroscience, many questions related to the role of simple cells remain unanswered. If simple cells detect elementary image features, how are these combined to form a perception? What is the role of feedback connections? Why do simple cells, when stimulated with moving lines and edges, exhibit various degrees of direction selectivity? What is the contribution of simple cells in tasks like texture analysis, or determination of the structure of objects from shading (Palmer 1999)?

In the following sections we will see some alternatives to the edge and line detection hypothesis. We start by considering a model in which images are considered as combinations of different frequencies, instead of combinations of simple localized image elements.

3.3 Local spatial frequency analysis

In the theory of the Fourier transform there is a deep mathematical result, which concerns the localization of the representation of signals in time (ordinary signal representation) and frequency (Fourier representation). This theorem, known as the *time-bandwidth product theorem,* or the *uncertainty principle*, (e.g., Cohen 1995), sets a limit on the joint localization of the signal in the time domain and in the frequency domain. Informally, if a signal is highly localized in time, it must contain a wide spectrum of different frequencies, whereas if the signal contains a very limited range of frequencies, it must have a long duration. Although in its basic form the uncertainty principle applies to functions with one argument – for example, a signal and its Fourier transform – similar limitations apply to multivariate functions, such as images, and their Fourier representations (Daugman 1985).

The uncertainty principle has important implications for linear systems. A linear system can be specified completely by its response to an impulse in the input. Linear filtering (*convolution*) can be represented as the product of the Fourier transforms of this *impulse response* and the input (e.g., Ifeachor and Jervis 2002). Because the impulse response is also a signal (in the one dimensional case) or an image (in the two-dimensional case, where it is often called the *point spread function*), the uncertainty principle applies to the impulse response as well. Suppose that one wants to build a linear spatial filter to sift a very limited band of frequencies from a certain location in an image. From the point of view of multiplication in the frequency domain, this means that the filter would have to be highly localized in frequency. The uncertainty principle would then limit the degree to which the filter could be localized spatially.

What makes this issue important for computational modeling of the visual system is the theory of *psychophysical spatial frequency channels* and its possible neurophysiological implementation in the primary visual cortex (e.g., Blakemore and Campbell 1969; Palmer 1999). The psychophysical spatial frequency channel theory posits that some functioning of the visual system can be decomposed into operations on a set of distinct spatial frequency channels, each channel being selective to a range of frequencies and orientations. Some impressive experimental results support the theory (Palmer 1999). For example, decreased sensitivity – or fatiguing, or adaptation – caused by prolonged exposure to visual stimuli seems to take place selectively in different spatial frequency channels. Because simple and complex cells are selective to frequency and orientation, and since their receptive fields are also spatially localized, it has been suggested that they might be contributing to spatial frequency channels by performing *local spatial frequency analysis* (Marcelja 1980; Daugman 1980; Pollen and Ronner 1983). When linear models are used to model

simple cells, localization and frequency selectivity are limited by the uncertainty principle.

Two particularly important linear models are closely related to local spatial frequency analysis: *wavelets* and *Gabor functions* are linear operators localized in both spatial and frequency domains, with the uncertainty principle as a limitation (for a related theory called *scale-space analysis*, see, e.g., Lindeberg and Mardia 1994). A wavelet is a member of a *wavelet family*, a set of functions generated from a 'model function' called a *mother wavelet* by elementary operations such as scaling, translation, and rotation (e.g., Cohen and Kovačević 1996; Hess-Nielsen and Wickerhausen 1996). Many different mother wavelets exist. Some applications in which wavelets have proven to be useful are feature detection, compression, noise removal, computer vision and graphics, and time-frequency description of signals (e.g., Kovačević and Daubechies 1996). A Gabor function is formed by modulating a sinusoidal (see Figure 2.1 on page 19) by a Gaussian (bell-curve-like) window, that is, it is the product of a sinusoidal and a Gaussian (e.g., Field 1987). The localization of Gabor functions in space and frequency reach the lower bound of the uncertainty principle (Daugman 1985). Certain subsets of Gabor functions can be constructed to have similar spatial and frequency properties as wavelet families (e.g., Lee 1996). In a classic paper, Jones and Palmer (1987) showed that Gabor functions provide a good fit to spatial simple-cell receptive fields in the cat primary visual cortex (however, see Palmer et al. 1991 for a critique of the Gabor model of simple cells). Gabor functions are used extensively as computational models of simple cells, and in the estimation of descriptive parameters from computationally obtained receptive-field models.

The local spatial frequency analysis theory can be considered as the second paradigm about the function of the cells in the primary visual cortex, the first being the line and edge detector paradigm put forward by Hubel and Wiesel. The local spatial frequency analysis theory represents a major step in the development of computational theories of vision, in that it departs from analyzing visually obvious features in stimuli.

3.4 Utilizing redundancies in sensory data

Sensory data are highly *redundant*: for example, knowing the brightness of a pixel in a digital image enables us to make a pretty good guess about the brightness of a nearby pixel (Kersten 1987; see also Eckert and Buchsbaum 1993; Dong and Atick 1995a for descriptions of redundancies in natural image sequences). The utilization of redundancies has been associated with the processing of sensory data since the classic papers by Attneave (1954) and Barlow (1961). The foundations of much of the research focusing on redundancy can be found in *information theory* which (Cover and Thomas 1991) "answers two fundamental questions in communication theory: what is the ultimate data compression, and what is the ultimate transmission rate of communication." Originally the central idea in the application of information theory to the visual system was that information about the environment might be presented in the brain as compactly as possible, so that redundancy would have been stripped away (Attneave 1954; Barlow 1961). One of the motivations of this view was the thought that it was uneconomical, or perhaps even impossible, for the brain to handle all visual input data (Attneave 1954; Barlow 2001). In the case of no noise, another way to state this idea of maximizing the capacity of available computational resources is that each cell should use its output

channel with maximum capacity, and there should be no redundancies between the outputs of different cells (Simoncelli 2003). In this short review we will discuss only some models in which redundancy plays a central role, namely *decorrelation* and *independent component analysis / sparse coding* (why the last two of these methods are related will be discussed below).

There have been hypotheses about both spatial and temporal decorrelation by visual neural networks. In the spatial case, decorrelation removes linear correlations – that is, correlations of the form $\mathrm{E}\{y_k y_\ell\}$ – between the outputs of different neurons, whereas in the temporal case it removes temporal linear correlations of the form $\mathrm{E}_t\{y_k(t)y_k(t-\Delta t)\}$ in the output of an individual cell. When combined with low-pass filtering to attenuate high frequency noise, decorrelation has proved to be a prominent model of the first layers of mammalian vision: in the case of spatial decorrelation, of retinal ganglion cells (Atick and Redlich 1992; see also Atick and Redlich 1990; Nirenberg et al. 2001), and in the case of temporal decorrelation, of the lateral geniculate nucleus (Dong and Atick 1995b; Dan et al. 1996).

Independent component analysis (ICA) and sparse coding have probably been the most promising statistical models in linking simple-cell receptive-field structure to natural stimulus statistics. One of the motivations for the development of these statistical methods was the observation that decorrelation alone is not sufficient to lead to the emergence of oriented and localized filters (Field 1987). Independent component analysis (e.g., Jutten and Herault 1991; Comon 1994; Bell and Sejnowski 1995; Cardoso 1998; Lee 1998; Girolami 1999; Hyvärinen et al. 2001; Cichocki and Amari 2002) is a method in which traditionally (several extensions of the basic model exist) a special form of generative model is assumed for the observed data $\mathbf{x}$: the assumption is that the data have been generated linearly from a set of statistically independent source (latent) variables $\mathbf{s}$, that is, $\mathbf{x} = \mathbf{A}\mathbf{s}$ (and $\mathbf{A}$ is invertible). Both the mixing coefficients – that is, the *mixing matrix* $\mathbf{A}$ – and the underlying sources $\mathbf{s}$ are unknown. If the generative model holds, the only assumption that is needed to recover both the mixing matrix and the sources, up to permutation and scaling, is that at most one of the latent variables has a normal distribution (Hyvärinen et al. 2001). One intuitive view of why this is possible is based on the central limit theorem (Papoulis 1991), which states that the linear combination of independent random variables approaches a normal distribution as the number of variables grows. Intuitively, then, if two non-Gaussian sources are mixed linearly together, the mixtures are "more Gaussian" than the sources – therefore, in the space of all random variables that can be obtained linearly from the sources, the *sources themselves* are "maximally non-Gaussian". Quantifications of "non-Gaussianity" can then be used as objective functions to find the inverse of the mixing matrix, that is, the inverse which gives the sources from the observations. This is one approach to independent component analysis, for an extensive review of different approaches see (Hyvärinen et al. 2001).

In natural image and image sequence data, the ICA model does not hold. One consequence of this is that the "independent components" obtained in linear ICA are not fully independent; we will discuss the nature of these dependencies in more detail in Section 3.5. However, since the model does not hold, we have to take a closer look at the objective function in order to understand what an ICA algorithm is really doing. As was mentioned above, independent analysis is related to sparse coding; in fact, in many cases an objective function used in ICA can also be interpreted as a measure of sparseness.

In sparse coding, a data set is transformed into another data set so that in this

new data, the occurrence of values that are close to zero – that is, have very small magnitudes – is maximized. To put it another way, the occurrence of values that are "significantly" or substantially different from zero is minimized. What "significant" or substantial really mean depends on the measure of sparseness, and has to be defined mathematically. Sparse coding, or learning a sparse code, means that the transformation from the original data $\mathbf{x}$ to new data $\mathbf{y}$ is optimized so that the new data set is as sparse as possible. There are two kinds of sparseness (Willmore and Tolhurst 2001): *population sparseness* and *lifetime sparseness*. To illustrate these, consider a random vector $\mathbf{y}$. This random vector $\mathbf{y}$ has high population sparseness if, on the average, the number of components differing substantially from zero is small in the samples drawn from $\mathbf{y}$. On the other hand, high lifetime sparseness is a property of a *single* random variable – in our example, a random variable y_k (a component of $\mathbf{y}$) would have high lifetime sparseness if, on the average, samples drawn from y_k would seldom attain values with substantial magnitudes. In computational models, additional constraints, such as the uncorrelatedness of the components of $\mathbf{y}$, can be used to learn a code that exhibits both lifetime and population sparseness. Further constraints, such as unit variance constraint on each of the y_k's, are needed to avoid degenerate solutions.

One way to explain the connection between sparse coding and ICA is to point out that a Gaussian random variable is neither very sparse nor very dense (dense being here the opposite of sparse), so measures of sparseness can also be used to measure "non-Gaussianity". This is why some statistical measures, such as *kurtosis*

$$\kappa_4 (y_k) = \mathrm{E}\left\{y_k^4\right\} - 3\left(\mathrm{E}\left\{y_k^2\right\}\right)^2 \tag{3.2}$$

can be used in both ICA and sparse coding (e.g., Field 1994; Hyvärinen et al. 2001). An additional link between the two methods can be observed by noting that some sparse coding algorithms can be interpreted as ICA algorithms; most notably the algorithm introduced in (Olshausen and Field 1996) can be recast as estimation of a linear generative model, in which the underlying components – in addition to being sparse – are also statistically independent (Simoncelli and Olshausen 2001).

One further note about the role of redundancy is in order before discussing the application of ICA and sparse coding to modeling the primary visual cortex. As was mentioned above, the original work done in this field emphasized the role of redundancy *reduction*. However, in some cases the redundancy is not reduced, it is *transformed*. For example in ICA / sparse coding, the variance of the output of a filter is typically fixed. In the family of random variables with a fixed mean and variance, the Gaussian random variable has maximal *differential entropy* (Cover and Thomas 1991), a quantity which can be used to compare the uncertainties of continuous random variables. Therefore, when maximizing non-Gaussianity with a fixed variance constraint, the redundancy in a single component *increases*, while inter-component dependencies are decreased. Or, as Barlow (2001) has stated in his recent review: "There is therefore no hidden redundancy; it is all manifest in the nonoptimal frequencies of activity in the elements." So there has been a shift in the paradigm towards the discovery and recognition of redundancy. Pinpointing redundancies may, for example, help find important regularities in the data. One demonstration in which the importance of redundancy was shown clearly was the work reported in (Becker and Hinton 1992), where surface depth was discovered from random dot stereograms: this was done by extracting the redundant disparity information from multiple stereograms of the same scene.

Returning to the role of ICA and sparse coding in modeling properties of the

primary visual cortex, there have been a number of studies linking these models to spatial (Olshausen and Field 1996; Bell and Sejnowski 1997; see also Publications 1 and 2) and spatiotemporal (van Hateren and Ruderman 1998; Olshausen 2000; see also Publication 6) simple-cell receptive fields, and also some reports that describe how the theories could be related to chromatic and binocular processing (Hoyer and Hyvärinen 2000), the structure of complex cells (Hyvärinen and Hoyer 2001; Szatmáry and Lőrincz 2002), and end-stopping and contour coding (Hoyer and Hyvärinen 2002). The obtained results have shown good agreement with physiological measurements in both the case of spatial (van Hateren and van der Schaaf 1998) and spatiotemporal (van Hateren and Ruderman 1998) linear models (however, see Ringach 2002 for critique of results obtained with ICA / sparse coding). The theories have also spurred new neurophysiological measurements (Vinje and Gallant 2000; Weliky et al. 2003), which have indicated that neural responses of simple cells are relatively sparse when the cells are stimulated from within the classical receptive field. Furthermore, the responses become even sparser if a naturalistic visual stimulus also overlaps the nonclassical receptive field (Vinje and Gallant 2000). This is a promising observation which will hopefully lead to new theoretical developments.

There are also interesting connections between the results obtained with ICA, and the local spatial frequency theory, which was discussed in Section 3.3. It was shown in Publication 1 that the basis vectors (columns of matrix $\mathbf{A}$) obtained with ICA are localized in both space and frequency, and that there is a relationship between the spatial frequencies of the basis vectors, and the degree of their localization: spatial localization increases, and frequency localization decreases, with spatial frequency. This is also typical of wavelets (e.g., Hess-Nielsen and Wickerhausen 1996).

3.5 Modeling dependencies between linear filters

In the previous section we saw that in independent component analysis (ICA), and in some sparse coding algorithms, it is assumed that the outputs of the linear filters are statistically independent of each other. However, further research on the topic has suggested that important dependencies exist between simple-cell-like filters. These dependencies have intriguing connections to the topographic layout of simple cells in the cortex, and to the way in which complex cells presumably pool the outputs of groups of simple cells.

Although the results of applying basic ICA on natural visual stimuli are difficult to interpret because the linear generative model does not hold, the results still suggest that the outputs of simple-cell-like filters could be *approximately* statistically independent when computed over a wide range of natural scenes. Further studies have shed light on the remaining dependencies between the outputs in the case of static image input (Zetzsche and Krieger 1999; Hyvärinen and Hoyer 2000; Wainwright and Simoncelli 2000; Schwartz and Simoncelli 2001; Hyvärinen and Hoyer 2001; Hyvärinen et al. 2001; Welling et al. 2003; Karklin and Lewicki 2003). In these studies, it has been discovered that in natural image data, simple-cell-like filters exhibit correlations in the *energies* (or variances) of their outputs. That is, for some pairs of filters, the large magnitude of the output of one of the filters implies that, on the average, the output of the other filter will also have a large magnitude. For example, such dependencies are strong in filters with similar frequency and orientation, but slightly different positions (Schwartz and Simoncelli 2001).

In some earlier work, Hyvärinen et al. have obtained interesting results by modeling these dependencies. First, when a model was specified so that groups of filters with strong dependencies would be pooled together, and the actual filters (free parameters) in the model were subsequently learned from natural image data, the resulting filters in each group possessed a number of similarities (Hyvärinen and Hoyer 2000). In particular, filters belonging to the same group had similar orientation and frequency, but differed in their phase and, to some extent, in their spatial position. A similar kind of pooling of a group of simple cells with similar orientation and frequency, but different phase and location, is thought to take place at the level of complex cells in the primary visual cortex (see Section 2.2). In a second model, the filter locations were specified in a two-dimensional lattice so that filters with strong dependencies would tend to be located close to each other in the lattice (Hyvärinen and Hoyer 2001; Hyvärinen et al. 2001; for another model introducing a similar idea see Welling et al. 2003). When the filters were learned from natural image data, the result was that filters which were close to each other in the lattice had similar spatial location and/or orientation and/or frequency, but differed in their phase. A similar type of topographical ordering with respect to location, orientation, and frequency has also been observed in the locations of simple cells in the primary visual cortex (see Section 2.2).

As a final note on the subject of inter-filter dependencies, the work of Schwartz and Simoncelli (2001) suggests that the kind of energy dependencies described here can be removed by a slightly modified version of the divisive normalization cell model (see equation (2.8) on page 26)

$$y_k(t) = \frac{\left(\mathbf{w}_k \mathbf{x}(t)\right)^2}{\sum_{\ell=1}^{K} \beta_{k,\ell} \left(\mathbf{w}_\ell \mathbf{x}(t)\right)^2 + \sigma^2}, \tag{3.3}$$

where the parameters $\beta_{k,\ell}$ and σ are estimated from natural image data. This kind of a model seems to be able to remove the dependencies between the energies of the filters, and also accounts for a number of nonlinearities observed at the simple-cell level – see (Schwartz and Simoncelli 2001) for details.

3.6 Temporal coherence

The term "temporal coherence" refers to a coding principle in which, when processing temporal input, the representation of this input in the computational system changes as little as possible over time. In computational visual neuroscience, temporal coherence has traditionally been associated with the invariance properties of complex cells. In this section we will review shortly some of the most important research concerning temporal coherence. The work presented in some of the articles of this thesis can also be considered to contribute into this research area – these contributions are presented separately in detail below in Chapter 4.

Földiák was one of the first authors to suggest the usefulness of temporal coherence in computational neuroscience (Földiák 1991; see also Hinton 1989). He developed a two-layer network which was able to learn to identify a fixed feature, such as a line with a fixed orientation, even if the way the feature was expressed in the data changed, for example, if the line was translated. Földiák used temporal coherence as a tool to learn translation invariances: artificially generated input data were temporally coherent (consecutive input frames contained translated versions of a line with the same orientation), and by using competition and short-term

memory, the output was also taught to be temporally coherent. This associated translated versions of a feature with each other.

Since Földiák several researchers have studied temporal coherence. For example, Stone (1996) used a multi-layer nonlinear network to discover surface depth from stereograms; learning was achieved by using a temporal sequence of slightly different stereograms, and by maximizing short-term temporal smoothness of output while preserving long-term variability in output. Let $\mathbf{w}$ denote the parameters in the model, f be the mapping from input to output, and var_λ a measure of local variance with decay parameter λ (the smaller λ, the faster the decay, and the more temporally localized the measure of variance). Learning was accomplished by maximizing the following objective function

$$O(\mathbf{w}) = \log \frac{\mathrm{var}_{\lambda_L} \{f(\mathbf{w}, \mathbf{x})\}}{\mathrm{var}_{\lambda_S} \{f(\mathbf{w}, \mathbf{x})\}}, \tag{3.4}$$

where $\lambda_S \ll \lambda_L$, that is, the numerator measures long-term variance, while the denominator measures short-term variance. What is common in the studies reported in (Földiák 1991; Stone 1996) is that the input data sets were generated so that there was an underlying, coherent parameter in the data, and the objective was to find that parameter by using coherence. Thus, the main result was the demonstration of the usefulness of temporal coherence using simulated data.

Recently researchers have started to apply temporal coherence to natural visual stimuli (Kayser et al. 2001; Berkes and Wiskott 2002; Hashimoto 2003; see also Kohonen et al. 1997). In (Kayser et al. 2001), a network with a number of "slowly varying subspaces" was learned from natural image sequence data. The outputs of this network were complex-cell-like energy detectors (see Section 2.4) which pooled the energies of a number of filters – the filters that formed a subspace. Learning of multiple temporally coherent outputs was achieved by using an objective function containing a sum of terms similar to equation (3.4). As a result, the outputs of the network exhibited orientation selectivity and translation invariance, thereby resembling the response properties of complex cells. In (Berkes and Wiskott 2002), the authors applied slow feature analysis (Wiskott and Sejnowski 2002) to simulated image sequence data. In slow feature analysis, the average change in the output is minimized, with the constraint that the output signal must have unit variance – this is equivalent to maximizing linear temporal correlation at the output (see Publication 2). The simulated image sequences were obtained by moving a window in natural images using different transformations: translations, rotations, and zooming (translation towards or away from camera). The class of functions used to compute the output from the input was the set of second-degree polynomials. Also in this case, the resulting network exhibited various complex-cell-like response properties, most notably phase invariance and orientation selectivity, but also other characteristics such as end-inhibition.

Chapter 4

New computational models utilizing stimulus dynamics

4.1 Introduction

In this chapter we focus on the main contribution of this thesis: the research concerning temporal coherence of activity levels, spatiotemporal activity level dependencies, and bubble coding (Publications 2–7). A short overview of the underlying principles is provided here – more detailed descriptions of the models and results can be found in the publications.

Whereas previous research has focused mostly on static properties of the outputs of simple-cell-like filters (see Sections 3.2–3.5 and Publication 1), or temporal properties of complex-cell-like invariant feature detectors (see Section 3.6), the starting point of the work described in this chapter is the nature of temporal properties of simple-cell-like linear filters. We begin by discussing the natural image sequences used in the experiments, and then proceed to consider the temporal response properties of simple-cell-like filters when the input consists of image sequences.

4.2 Natural stimulus data

A key feature in this work is the use of large quantities of natural image sequence data in the experimental evaluation of our models. That is, we do *not* generate simplified image sequence data ourselves from simple object worlds or static images. This, in our opinion, is of major importance, since it exposes the model to the complex phenomena that are present in real dynamic visual stimuli.

The natural image sequences used as data in the experiments of Publications 2–7 were a subset of those used in (van Hateren and Ruderman 1998). The original data set consists of 216 monochrome video clips of 192 seconds each, taken from television broadcasts. More than half of the videos feature wildlife, the rest show various topics such as sports and movies. Sampling frequency in the image sequences is 25 frames per second, and each frame has been block-averaged to a resolution of 128×128 pixels. Unfortunately the reproduction of example image sequences in this thesis is not possible because of copyright issues.

For our experiments this data set was pruned to remove the effect of human-made objects and artifacts (see Publication 2), which left us with 129 videos. The

motivation behind the pruning of the videos was to make the data as "natural" as possible. The pruning does have a real effect on the results: for example, if pruning is not done, the resulting spatial ICA filters computed from data sampled from the videos show more horizontally or vertically oriented, elongated receptive fields.

In most of our experiments utilizing the image sequence data we assume that a neuron receives its input from a small, stationary window, containing a spatially small proportion of the whole image sequence. While the existence of spatially limited classical receptive fields justifies the assumption of a restricted input area, the assumption of a stationary window is not as well-founded. This is because in real visual systems, there are dynamic components which cause the part of the visual field feeding into the receptive field to change in time. One obvious component is observer movement, which is actually included to some degree in our data, since the camera is moving and zooming. However, another important factor is eye movement, which is not included in the data, or our models.

Eye movement is a very complicated phenomenon, because the trajectory of the eyes depends on the task of an animal, as well as the visual input. These dependencies would have to be modeled, if eye movements were to be fully incorporated into our stimulus data or into our models. Since the task of an animal and analysis of the contents of a scene are concepts which are beyond the early levels of the visual system, we have decided to not include eye movements into our experiments.

4.3 Temporal coherence of activity levels

Simplified intuitive illustration

Objects can undergo a number of transformations in image sequences: translation, rotation, occlusion, and, for objects that are not rigid, deformation. A transformation in the three-dimensional space can induce a different transformation in an image sequence. For example, a translation towards the camera induces a change of object size in the image sequence. It seems that most typical transformations of objects in the three-dimensional world result in something similar to *local translations* of lines and edges in image sequences. This is obvious in the case of three-dimensional translations, and is illustrated in Figure 4.1A for two other types of transformations: rotation and bending. The phenomenon illustrated here is also related to the *aperture problem*: when the movement of a long straight line is observed inside a small window, the observer always perceives motion perpendicular to the direction of the line, regardless of the direction of the actual motion (e.g., Palmer 1999).

What happens at the output of a simple-cell-like filter in the case of such a local translation? This is illustrated in Figure 4.1B. When the filter is suitably oriented, it tends to respond strongly at consecutive time points, but the sign of the response may change. In other words, the variance of the output is large. We call this kind of temporal coherence *temporal coherence of activity levels*, and measure such coherence with *temporal response strength correlation*.

Temporal response strength correlation

In Publication 2 we showed that not only do simple-cell-like filters exhibit temporal coherence of activity levels, but they are *optimal* with respect to a measure of such coherence as follows. We use the basic linear cell model, as described by equation (2.4). Temporal response strength correlation, the objective function, is

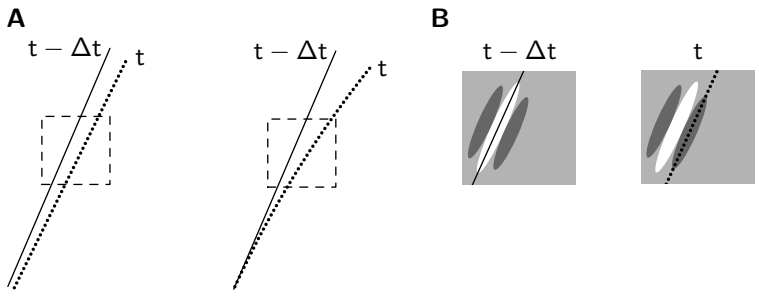


Figure 4.1: A simplified illustration of temporal coherence of activity levels in the outputs of simple-cell-like filters. (A) Transformations of objects in the three-dimensional world induce local translations of edges and lines in local regions in image sequences: rotation (left) and bending (right). The solid line shows the position/shape of a line in the image sequence at time $t - \Delta t$, and the dotted line shows its new position/shape at time t. The dashed square indicates the sampling window. (B) Temporal coherence of activity levels: in the case of a local translation of an edge or a line, the response of a simple-cell-like filter with a suitable position and orientation tends to be strong at consecutive time points, but the sign may change. The figure shows a translating line superimposed on an oriented and localized receptive field at two different time instants (time $t - \Delta t$, solid line, left; time t, dotted line, right).

defined by

$$f(\mathbf{W}) = \sum_{k=1}^{K} \mathrm{E}_t \left\{ g(y_k(t)) g(y_k(t - \Delta t)) \right\}, \tag{4.1}$$

where the nonlinearity g is strictly convex, even (rectifying), and differentiable. The symbol Δt denotes a delay in time, which varied between 40 ms and 960 ms in our experiments (see Publication 2). The nonlinearity g measures the strength (amplitude) of the response of the filter, and emphasizes large responses over small ones. Examples of choices for this nonlinearity are $g_1(x) = x^2$ and $g_2(x) = \ln \cosh x$. A set of filters which has a large temporal response strength correlation is such that the same filters *often respond strongly at consecutive time points*, outputting large (either positive or negative) values, thereby expressing temporal coherence of the activity of populations of neurons. In addition to the objective function, some constraints are also needed to limit the dynamic range of the outputs, and to rule out the noninteresting solution where all the filters (rows of $\mathbf{W}$) are identical – see Publication 2 for details.

Note that, in comparison with a "traditional" measure of temporal coherence, the objective function (4.1) is a measure of short-term *nonlinear* temporal correlation, whereas in slow feature analysis (Wiskott and Sejnowski 2002), the objective function measures short-term *linear* temporal correlation (see Section 3.6).

Figure 4.2A shows the results (rows of $\mathbf{W}$) after optimizing equation (4.1) for a large set of natural image sequence samples. In this case, the spatial filters are of size 16×16 pixels (see Publication 5 for more information on the experimental setup which yielded these results). As can be seen, the resulting filters are localized, oriented, and have different scales, which are the defining qualitative features of simple-cell receptive fields (see Section 2.2). In Publication 2, results obtained by

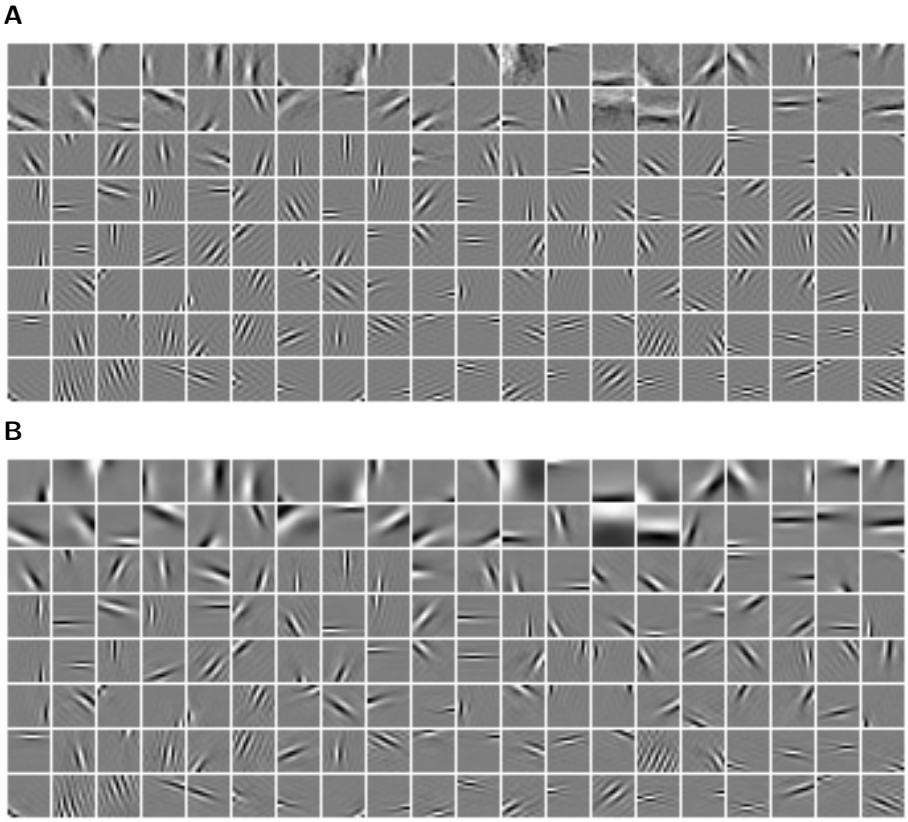


Figure 4.2: Filters and basis vectors maximizing temporal response strength correlation. (A) A set of filters (rows of matrix $\mathbf{W}$), estimated from natural image sequences by optimizing temporal response strength correlation (equation (4.1), here nonlinearity $g(x) = \ln\cosh x$ and $\Delta t = 40\,\mathrm{ms}$). The filters have been ordered according to $\mathrm{E}_t\left\{g(y_k(t))g(y_k(t-\Delta t))\right\}$, that is, according to their "contribution" into the final objective value (filters with largest values top left). (B) The corresponding set of basis vectors (columns of matrix $\mathbf{A}$) – see text for details.

maximizing temporal response strength correlation were compared quantitatively against results obtained with independent component analysis / sparse coding (see Section 3.4); the results indicate that the two methods produce quantitatively similar results with respect to several important parameters. A similar comparison between ICA results and physiological measurements from simple cells had been reported earlier in (van Hateren and van der Schaaf 1998).

Several control experiments, described in detail in Publication 2, were performed to ensure the validity and novelty of the results. Also, while the linear model applied here in fact models the output of two nonnegative cells (see Section 2.4), preliminary results suggest that a similar principle may apply even in the case of a half-rectified (nonlinear) model (see Section 2.4 and Publication 2).

Generative-model interpretation and some related physiological observations

As was mentioned in Section 3.1, within the research linking natural stimulus statistics to the properties of the visual system, two different lines of research can be identified: one where optimal properties of the outputs of neurons are employed, and another in which a generative model of natural stimulus data is postulated. The interpretation of equation (4.1) given above is in the spirit of the optimal neural output approach. However, another view of equation (4.1) is to consider it as a way to estimate a generative model, as shown in Publication 5. This interpretation is based on the concept of sources with *nonstationary variances* (Matsuoka et al. 1995; Hyvärinen 2001; Pham and Cardoso 2001), and the use of objective function (4.1) in the estimation of linear generative models where sources have nonstationary variances, is analogous to the application of measures of sparseness in the estimation of linear generative models with non-Gaussian sources (see Section 3.4).

The linear generative model for $\mathbf{x}(t)$, the counterpart of equation (2.4), is similar to the one in (Hyvärinen and Hoyer 2001; Olshausen and Field 1996):

$$\mathbf{x}(t) = \mathbf{A}\mathbf{y}(t). \tag{4.2}$$

Here $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_K]$ denotes a matrix which relates the image patch or short image sequence (in the case of spatiotemporal simple-cell models) $\mathbf{x}(t)$ to the activities of the simple cells, so that each column $\mathbf{a}_k$, $k = 1, ..., K$, gives the two- or three-dimensional feature that is coded by the corresponding simple cell. Within the generative-model community, it has become customary to regard columns of $\mathbf{A}$ as descriptors of features coded by the neurons (Olshausen and Field 1996; Hyvärinen and Hoyer 2001; Hyvärinen et al. 2001); the set of basis vectors corresponding to the filters in Figure 4.2A is shown in Figure 4.2B. The basis vectors are otherwise similar to the filters, except that high frequencies have been attenuated – see (Hyvärinen and Hoyer 2001) for additional discussion.

The nonstationarity of the variances of sources $\mathbf{y}(t)$ means that their variances change over time. In practice, it is typically also assumed that the variance is correlated at nearby time points, that is, that the variance changes smoothly. An example of a signal with nonstationary, smoothly changing variance is shown in Figure 4.3. It can be shown that optimization of a cumulant-based criterion, which is similar to equation (4.1), can separate independent sources with nonstationary variances (Hyvärinen 2001). Thus, the maximization of the objective function can also be interpreted as estimation of a generative model in which the activity levels

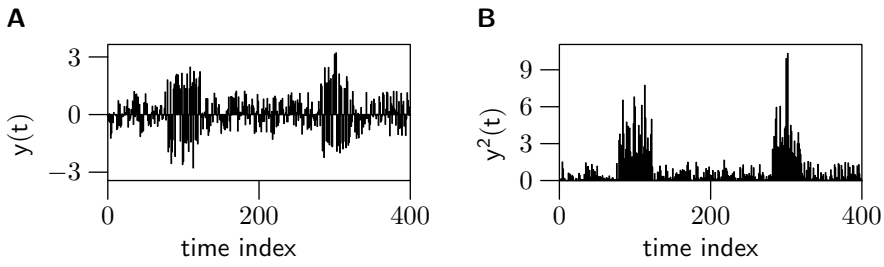


Figure 4.3: Illustration of nonstationarity of variance. (A) A temporally uncorrelated signal $y(t)$ with nonstationary, smoothly changing variance. (B) Plot of $y^2(t)$.

of the sources vary over time, and are temporally correlated over time. The sources would then correspond to simple-cell outputs.

In fact, in a recent review of research on neural responses to natural stimuli, Reinagel (2001) described these responses as follows: "The experiments summarized above show that spiking neurons respond to many natural visual stimuli with discrete high-frequency firing events separated by periods of low firing or complete silence." This verbal description of the outputs of real neurons is strikingly similar to what a signal with nonstationary variance looks like.

Another interesting neurophysiological observation related to the model presented here is the importance of *spike bursts* in *synaptic transmission*. It seems that synapses – sites at which neural activity is transmitted from one neuron to another – transmit information about individual (isolated) spikes with very low probability, but respond much more readily to bursts of spikes, even if the bursts are very short (see Lisman 1997 for a review). In other words, the synapses seem to act as "filters that transmit bursts, but filter out single spikes" (Lisman 1997). An objective function such as equation (4.1) rewards burst-like high-activity periods, and may therefore be related to increased transmission reliability in synapses.

The case of spatiotemporal filters

The discussion in the previous sections was limited to spatial filters, but a similar phenomenon seems to apply also in the spatiotemporal case. Figure 4.4 shows an intuitive illustration of how directionally selective spatiotemporal filters could also exhibit temporal response strength correlation.

Figure 4.5 shows a subset of spatiotemporal basis vectors maximizing temporal response strength correlation. Quantitative analysis of the spatiotemporal results can be found in Publication 6. The most important difference between these results and neurophysiological measurements is that our results are not localized temporally. A similar discrepancy was also found when we extracted a corresponding set of spatiotemporal filters by using ICA. The ICA results become more temporally localized if a *deflationary* algorithm – in which the filters are extracted one by one (Hyvärinen et al. 2001) – is used and dimensionality reduction is applied to the data. This observation is in concordance with the results obtained by van Hateren and Ruderman (1998). An analogous change in the algorithm and preprocessing methods improves slightly the temporal localization of results obtained by maximizing temporal response strength correlation, but not to the same degree as in the

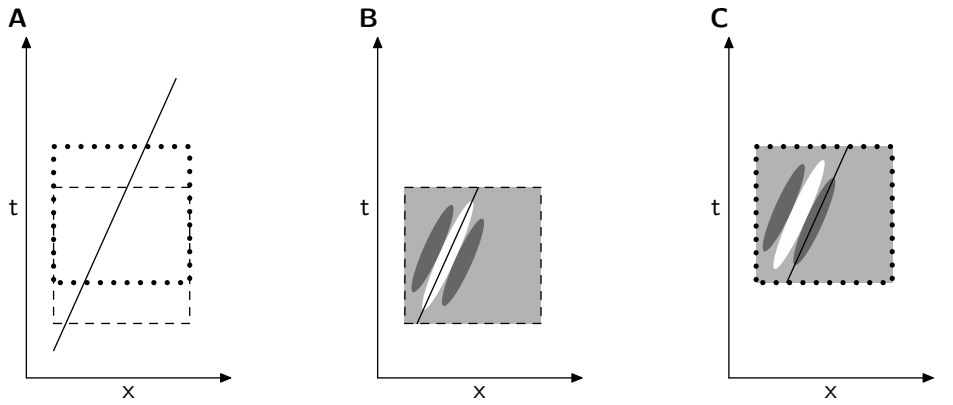


Figure 4.4: An illustration of how the outputs of simple-cell-like *spatiotemporal* filters could exhibit large temporal response strength correlation. A phenomenon analog to translation in the spatial case can be observed in the spatiotemporal domain. Let x and y denote the horizontal and vertical spatial coordinates, respectively, and let t denote the temporal coordinate. (A) The spatiotemporal trace (solid line) of a horizontally moving vertical line is shown here in the x–t coordinate system. The plot is similar for all y-coordinates because the moving line is vertical. Two different overlapping spatiotemporal input windows, separated by a small time difference, are also marked, one with dashed line, and the other with dotted line. (B) A simple-cell-like spatiotemporal filter, with position and orientation that match the initial position of the line and its direction of movement, responds strongly to the moving line. Here the spatiotemporal filter has been superimposed over the dashed temporal window – white color indicates large positive values in the filter, dark color indicates large negative values, and middle gray indicates zero values. (C) When the same spatiotemporal filter, at the same spatial position, is applied to the same input a moment later (dotted spatiotemporal input window), the response is still strong, but the sign changes. Therefore the temporal response strength correlation of the outputs of the simple-cell-like spatiotemporal filter would be large for this kind of input.

case of ICA. So far we have been unable to pinpoint the reason for this difference. Further research is needed to clarify the issue.

To be exact, some of the results in Publication 2 also describe spatiotemporal filters. This is because in some experiments, a temporal filter was used in preprocessing. A cascade consisting of a temporal filter and a spatial filter is in fact a *space-time separable* spatiotemporal filter; the spatiotemporal filter illustrated in Figure 4.4, on the other hand, can not be represented as such a cascade, and is called *space-time inseparable* (e.g., Adelson and Bergen 1985; DeAngelis et al. 1995). The difference between the spatiotemporal results in Publication 2 and Publication 6 is, then, that in Publication 2 the spatiotemporal filters were *forced* to be space-time separable.

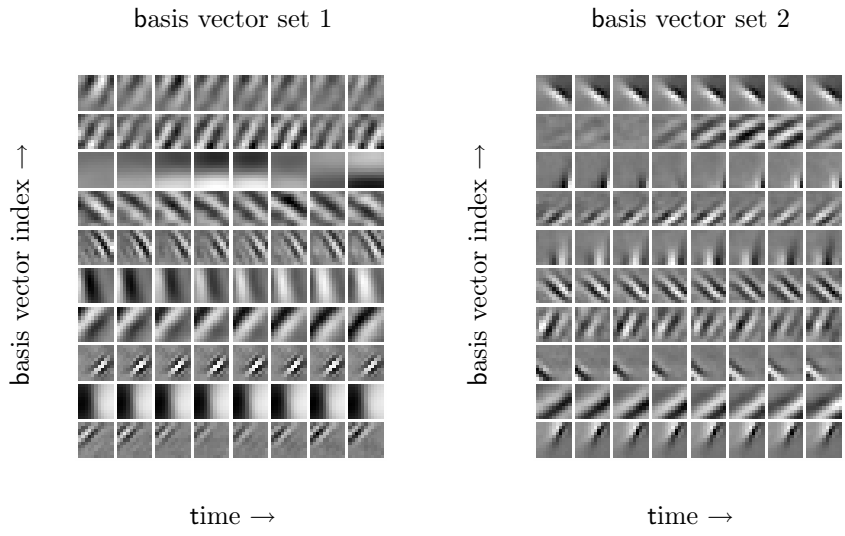


Figure 4.5: Spatiotemporal basis vectors maximizing temporal response strength correlation. A subset of 20 spatiotemporal basis vectors obtained by maximizing temporal coherence of activity levels in natural image sequences (10 basis vectors in the image on the left and 10 on the right). Each of the 20 rows corresponds to one spatiotemporal basis vector, and the frames in the row correspond to spatial basis vectors at consecutive time instants.

4.4 Spatiotemporal activity level dependencies

Simplified intuitive illustration

In the previous section we discussed maximization of nonlinear time-correlation of output activity levels, and saw that it provides an alternative to sparse coding and independent component analysis as a computational principle underlying simple-cell receptive-field structure. This time-correlation can be considered as temporal activity level dependency: the activity level of a filter at time t is not independent of the activity at time $t - \Delta t$. It seems that this temporal dependency is not the only type of activity level dependency in a set of simple-cell-like filters. Figure 4.6 illustrates how *two different simple-cell-like filters* with similar profiles – having the same orientation and scale but slightly different positions – respond at consecutive time instants when the input is a translating line. It can be seen that the outputs of *both filters* tend to be highly active at *both time instants,* but again, the signs of the outputs vary. This means that in addition to temporal activity dependencies (the activity of a filter is large at time $t - \Delta t$ and time t), there are two other kinds of activity level dependencies.

Spatial (static) dependencies Both filters are highly active at a single time instant. This kind of dependency is an example of the activity level dependencies modeled in previous research on static images (Zetzsche and Krieger 1999; Hyvärinen and Hoyer 2000; Wainwright and Simoncelli 2000; Hyvärinen et al. 2001; Schwartz and Simoncelli 2001; Welling et al. 2003).

Spatiotemporal dependencies The activity levels of different filters are also re-

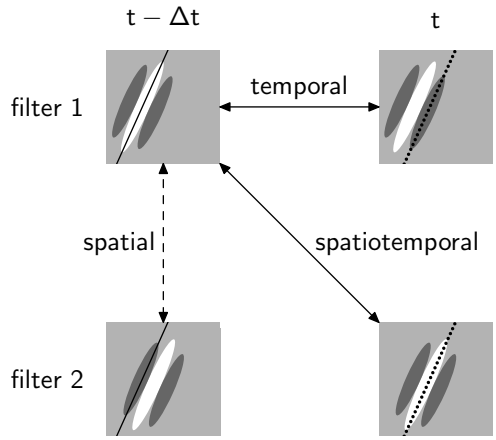


Figure 4.6: A simplified illustration of static and short-time temporal and spatiotemporal activity level dependencies in the outputs of simple-cell-like filters. For a translating edge or line, the responses of two similar filters with slightly different positions (filter 1, top row; filter 2, bottom row) are large at nearby time instants (time $t - \Delta t$, solid line, left column; time t, dotted line, right column). Each subfigure shows the translating line superimposed on a spatial filter. The magnitudes of the responses of *both* filters are large at *both* time instants. This introduces three types of activity level dependencies: temporal (in the output of a single filter at nearby time instants), spatial (between two different filters at a single time instant) and spatiotemporal (between two different filters at nearby time instants). The model introduced in this section includes temporal and spatiotemporal activity level dependencies (marked with solid arrowheaded lines).

lated over time. For example, the activity of filter 1 at time $t - \Delta t$ is related to the activity of filter 2 at time t.

In what follows, we describe a model – originally introduced in Publication 3 – which incorporates both temporal and spatiotemporal activity level dependencies; the model and the related estimation algorithm were developed and analyzed further in Publication 4.

A two-layer generative model with activity level dependencies

The generative model of natural image sequences employing activity level dependencies has two layers, as illustrated in Figure 4.7. The first layer, which captures the temporal and spatiotemporal activity level dependencies, is a multivariate autoregressive model between the activity levels (amplitudes) of filter responses at time t and time $t - \Delta t$. The signs of the responses are generated by a latent signal between the first and the second layer. The second layer is linear, and maps responses to the image space. The mathematical details of the model, including additional constraints, are described in Publication 4. The algorithms for the simultaneous estimation of both layers employ the method of moments and the method of least mean squares – see Publications 3 and 4 for details. In the estimation algorithms, both of the layers of the model are estimated simultaneously; this is a significant improvement on most multi-layer statistical models of early vision (Hyvärinen and

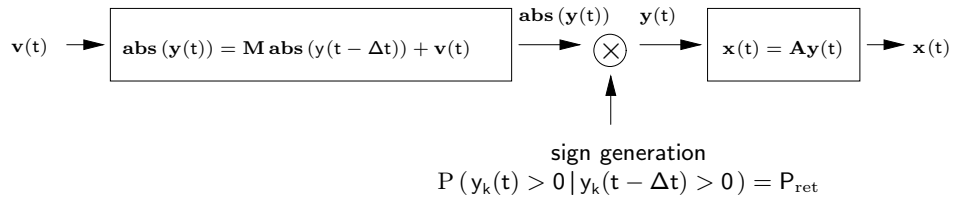


Figure 4.7: The two layers of the generative model with temporal and spatiotemporal activity level dependencies. Let $\mathbf{abs}\,(\mathbf{y}(t)) = [|y_1(t)| \cdots |y_K(t)|]^T$ denote the activity levels (amplitudes) of filter responses. In the first layer, the driving noise signal $\mathbf{v}(t)$ generates the activities of the filters $\mathbf{abs}\,(\mathbf{y}(t))$ via a multivariate autoregressive model. Matrix $\mathbf{M}$ captures the temporal and spatiotemporal activity level dependencies in the model. The signs of the responses are generated between the first and the second layer to yield signed responses $\mathbf{y}(t)$. The probability that a latent signal $y_k(t)$ retains its sign is P_{ret}. In the second layer, the natural image sequence $\mathbf{x}(t)$ is generated linearly from filter responses. In addition to the relations shown here, the generation of $\mathbf{v}(t)$ is affected by $\mathbf{M}\,\mathbf{abs}\,(y(t - \Delta t))$ to ensure nonnegativity of $\mathbf{abs}\,(\mathbf{y}(t))$. See Publication 4 for details.

Hoyer 2000; Wainwright and Simoncelli 2000; Hyvärinen and Hoyer 2001), because no a priori fixing of *either* of these layers is needed. Two other models where both layers of a two-layer model are estimated simultaneously have recently been introduced in (Hashimoto 2003; Welling et al. 2003).

The basis vectors and their spatiotemporal dependencies – that is, matrices $\mathbf{A}$ and $\mathbf{M}$ – can be visualized simultaneously by using an interpretation of $\mathbf{M}$ as a similarity matrix (see Publication 4). Figure 4.8 illustrates the basis vectors laid out at spatial coordinates derived from $\mathbf{M}$ in in this way. The resulting basis vectors are again oriented, localized and multiscale, as in the previous section. We can also see that local topography emerges in the results: those basis vectors which are close to each other seem to be mostly coding for similarly oriented features at nearby spatial positions. As was discussed above in Section 2.2, this kind of grouping is characteristic of pooling of simple-cell outputs at the complex-cell level, and also similar to the topographic relationships in the primary visual cortex. It should also be noted that the qualitative properties of the results in Figure 4.8 do not change even if we adopt the idea in which a linear model in fact represents two neurons with reversed polarities (see Section 2.4). For additional analysis of the results, see Publication 4.

4.5 Bubble coding

In order to motivate the development of the bubble coding model, let us summarize the key research results on modeling the properties of the neural representation at the simple-cell level. Results obtained using sparse coding / independent component analysis suggest that, on the average, at a single time instant relatively few simple cells are active on the cortex (see Section 3.4). In this thesis, we have described an alternative model, which suggests that simple cells tend to be highly active at consecutive time instants – that is, their outputs are burst-like (see Section 4.3). On the other hand, previous research on static dependencies between

Figure 4.8: Grouping similar to complex-cell pooling of simple-cell outputs, and to the topographic properties of neurons in the primary visual cortex, emerges from spatiotemporal activity level dependencies. Here we have plotted the basis vectors (columns of $\mathbf{A}$) at two-dimensional coordinates, obtained by applying multidimensional scaling to the similarity values defined by $\mathbf{M}$ (see Publication 4 for details). As can be seen, nearby basis vectors seem to be mostly coding for similarly oriented features with similar frequencies at nearby spatial positions. In addition, some global topographic organization also emerges: those basis vectors which code for horizontal features are on the left in the figure, while those that code for vertical features are on the right. Some short distances were magnified in order to be able to show the basis vectors in a reasonable scale.

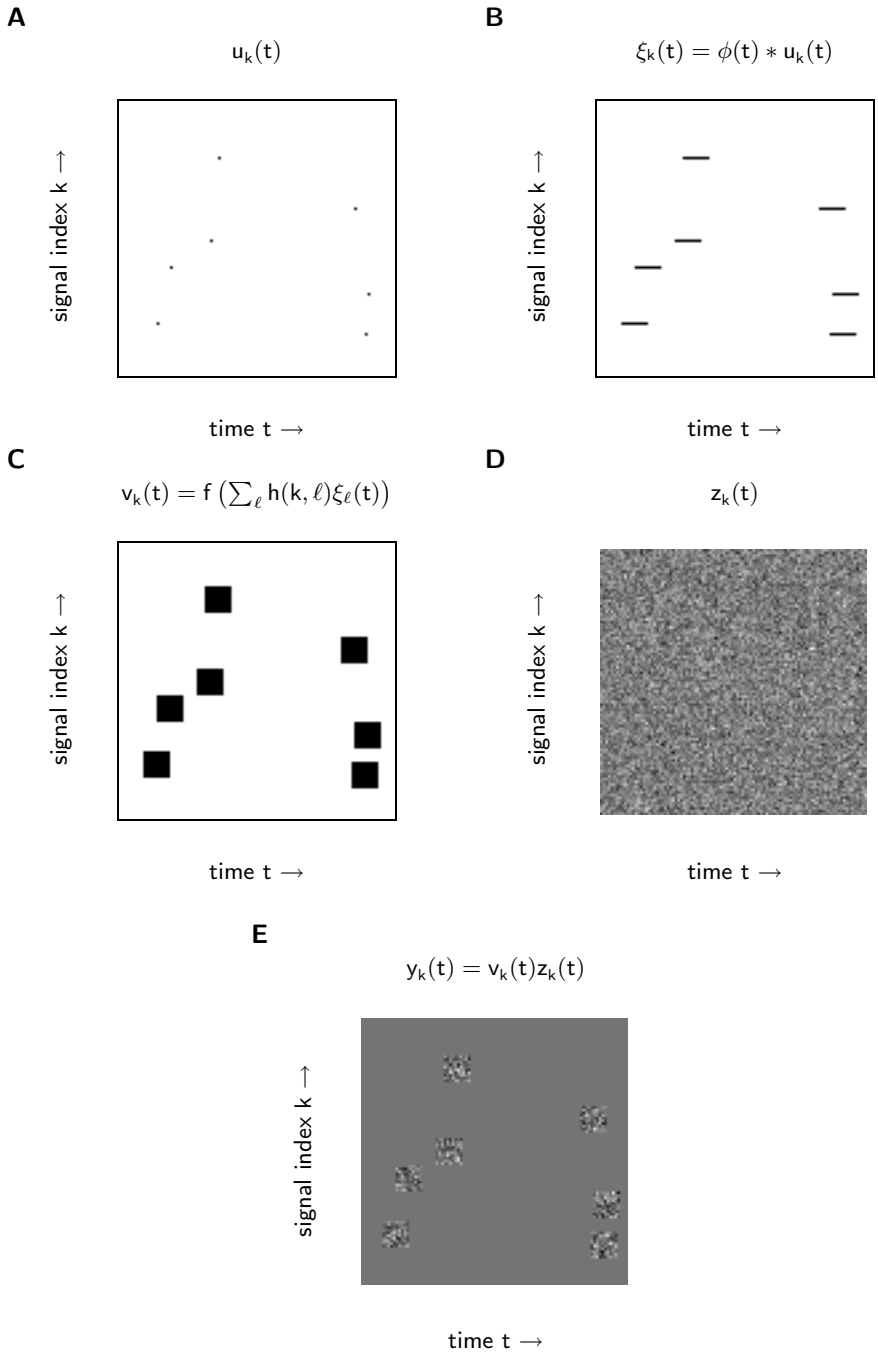
simple-cell-like filters, and the relationship between these dependencies and cortical topography, suggest that the active cells tend to be located in cortical patches, that is, close to each other on the cortex (see Section 3.5). Again, as we saw in the previous section, an alternative model also leads to topographic properties resembling cortical topography, a model which utilizes time-correlations between the outputs of different filters.

These different principles – sparseness, temporal coherence of activity levels, spatial activity level dependencies, and spatiotemporal activity level dependencies – are not conflicting. That is, none of the principles excludes the existence of another. Perhaps, then, each of these models offers just a limited view to a more complete model of cortical coding at the simple-cell level. In fact, the following description of simple-cell activation is in accordance with all of the principles: when an animal is viewing a natural scene, a relatively small number of patches of cortical area are highly active in the primary visual cortex, and the activity in these areas tends to be sustained for a while. That is, activity is sparse, and contiguous both in space and time. This is the bubble coding model, proposed in Publication 7.

In order to use consistent notation in this introductory part of the thesis, we denote latent response signals again with $\mathbf{y}(t)$ (in Publication 7, these responses are denoted by $\mathbf{s}(t)$). In the bubble coding model, the generative mapping from latent responses to natural image sequence data is linear, like in the previous sections: $\mathbf{x}(t) = \mathbf{A}\mathbf{y}(t)$. The main idea in the bubble coding model is the way in which the responses are generated so that they have bubble-like activity. This is accomplished by introducing a bubble-like variance signal for $\mathbf{y}(t)$, as illustrated by an example in Figure 4.9. The spatiotemporal locations of the variance bubbles are determined by a sparse process $\mathbf{u}(t)$ (Figure 4.9A). A temporal filter ϕ and spatial pooling function h, both of which are fixed a priori in the model, spread the variance bubbles temporally and spatially (Figures 4.9B and C). The resulting variance bubbles can also overlap each other, in which case the variance in the overlapping area is obtained as a sum of the variances in each bubble; in Figure 4.9, however, the variance bubbles are nonoverlapping. It is also possible that at this point a fixed static nonlinearity f is applied to rescale the magnitudes of the variance bubbles. These steps yield

---

Figure 4.9: (*facing page*) Illustration of the generation of response signals $y_k(t)$ in the bubble coding model for one-dimensional topography. (A) The starting point is the set of sparse signals $u_k(t)$. (B) Each sparse signal $u_k(t)$ is filtered with a temporal low-pass filter $\phi(t)$, yielding signals $\phi(t)*u_k(t)$. In this example, the filter $\phi(t)$ simply spreads the impulses uniformly over an interval. (C) In the next step, a neighborhood function $h(k,\ell)$ is applied to spread the bubbles spatially. A static nonlinearity f may also be applied at this point to rescale the magnitudes of the variance bubbles. This yields variance bubble signals $v_k(t) = f\big(\sum_\ell h(k,\ell)\,[\phi(t) * u_\ell(t)]\big)$. In this example, the neighborhood function h is simply 1 close-by and 0 elsewhere, and the static nonlinearity f is just the identity mapping $f(\alpha) = \alpha$. (D) Gaussian temporally uncorrelated (white noise) signals $z_k(t)$. (E) Responses are defined as products of the Gaussian white noise signals and the spatiotemporally spread bubble signals: $y_k(t) = z_k(t)v_k(t)$. Note that in subfigures (A)–(C), white denotes value zero and black denotes value 1, while in subfigures (D) and (E), medium gray denotes zero, and black and white denote negative and positive values, respectively.


A
$u_k(t)$
signal index k →
time t →
B
$\xi_k(t) = \phi(t) * u_k(t)$
signal index k →
time t →
C
$v_k(t) = f\left(\sum_\ell h(k,\ell)\xi_\ell(t)\right)$
signal index k →
time t →
D
$z_k(t)$
signal index k →
time t →
E
$y_k(t) = v_k(t)z_k(t)$
signal index k →
time t →

the variance signals

$$v_k(t) = f\left(\sum_\ell h(k,\ell)\,[\phi(t) * u_\ell(t)]\right). \tag{4.3}$$

The burst-like oscillating nature of the responses inside the bubbles is introduced through a Gaussian temporally uncorrelated (white noise) process $\mathbf{z}(t)$ (Figure 4.9D). Finally, the responses are generated from the variance bubbles and the noise signals by multiplying the two together (Figure 4.9E):

$$y_k(t) = v_k(t) z_k(t). \tag{4.4}$$

Note that all three different types of activity level dependencies – temporal, spatial, and spatiotemporal (see Figure 4.6 on page 47) – are present in the bubble-coding model.

In order to estimate the bubble coding model, an approximative maximum likelihood scheme is used (details can be found in Publication 7). Note that because the pooling function h is fixed, it enforces the spatial pooling, while in the two-layer model described in the previous section, this pooling was learned from the data. The temporal smoothing (low-pass) filter ϕ is also fixed in the model. In the experiments with image sequence data, the choice of the length of this temporal filter was directed by the experiment in which the effect of different lengths was examined; this was done by generating signals having similar temporal dynamics as the presumably underlying source signals in natural image sequences (see Section 3.C. in Publication 7). The size of the spatial pooling function h was chosen arbitrarily to be 1 inside a 3×3 window around each unit in the lattice.

Figure 4.10 shows the resulting spatial basis vectors, obtained when the bubble coding model was estimated from natural image sequence data. The basis consists of simple-cell-like linear receptive-field models, similar to those obtained with the models introduced in the previous sections, or by maximization of sparseness. The orientation and the location of the feature coded by the vectors change smoothly when moving on the topographic grid. Low-frequency basis vectors are spatially segregated from the other vectors, so there also seems to be some ordering based on preferred spatial frequency. Such an organization with respect to orientation, location, and spatial frequency is similar to the topographic ordering of simple cells in the primary visual cortex, as was discussed in Section 2.2. An animated example of a spatiotemporal basis estimated using this method can be found at `http://www.cis.hut.fi/jarmo/animations/bubbleanimation.gif` . Note that also in this case, as in the case of spatiotemporal filters maximizing temporal response strength correlation, the spatiotemporal results do not exhibit a great degree of temporal localization.

4.6 Discussion of neuroscientific contribution

In this chapter, we have shortly described three sets of results obtained with new computational models utilizing stimulus dynamics, namely the results obtained by

- maximizing the temporal coherence of activity levels (Section 4.3)
- estimating the two-layer model utilizing temporal and spatiotemporal activity level dependencies (Section 4.4)

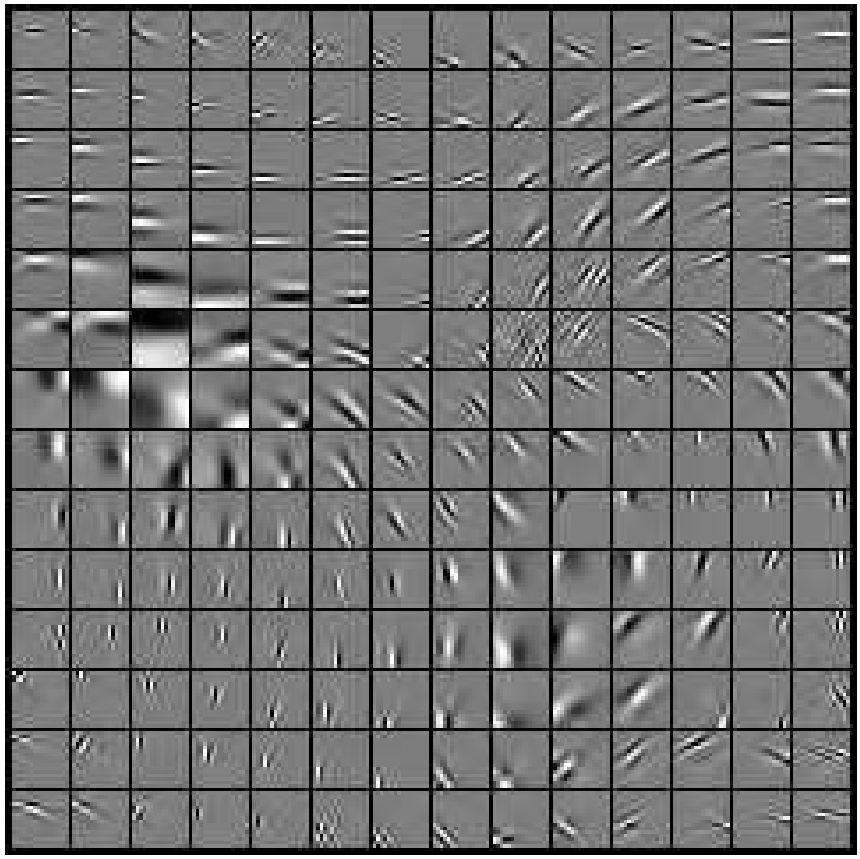


Figure 4.10: A set of spatial basis vectors, estimated from natural image using the bubble coding estimation method, and laid out at spatial coordinates defined by the lattice in the bubble coding model. The topographic organization of the basis vectors exhibits ordering with respect to orientation, location, and spatial frequency of the vectors.

- estimating the bubble coding model (Section 4.5).

Here we will discuss the neuroscientific contribution of these models.

In Chapter 1, the assessment of the neuroscientific contribution of a model was divided into two parts: evaluation of the implementational equivalence of the mathematical model and its biological counterpart, and evaluation of the predictive power of the model. In what follows, we will discuss both of these aspects in detail.

In the development of computational models of the brain, one must strive for implementational equivalence and predictive power. But it must also be remembered that the models are abstractions of the real world, and do not provide a perfect match with modeled phenomena. Therefore, in order to be able to position a model realistically inside a research field, not only should we compare the model against the ideal, but also against other comparable research. In current research on computational models of the primary visual cortex, the best established research is offered by the work on ICA/sparse coding, including (Olshausen and Field 1996; Bell and Sejnowski 1997; van Hateren and van der Schaaf 1998; van Hateren and Ruderman 1998; Hyvärinen and Hoyer 2000; Hyvärinen et al. 2001; Schwartz and Simoncelli 2001). Thus, in the following discussion, we will contrast the results presented in this chapter with results obtained with ICA/sparse coding.

Implementational equivalence

In this work, we have used linear and half-wave rectified simple-cell models (see Section 2.4). The results obtained by maximizing temporal response strength correlation, including the spatiotemporal results, were based on the linear model (see Publications 2 and 6). An experiment in which half-wave rectification was used was reported in the spatial case (see Figure 12 in Publication 2). However, in that experiment the same set of constraints was used as in the experiments with linear models – unit variance and uncorrelatedness constraints on the signed outputs – so the results are more difficult to interpret. Also, the filters that were obtained in the experiment were somewhat different from results obtained with a purely linear simple-cell model: the filters were not as well defined, and seemed to span a smaller range of scales (frequencies).

The results obtained by estimating the two-layer model can be readily interpreted from the point of view of either the linear or the half-wave-rectified cell model. For the half-wave-rectified cell model, the results suggest that simple-cell-like *filter pairs* have temporal and spatiotemporal activity level dependencies (see Section 7.2. in Publication 4). For the bubble-coding model, interpretation in the case of a half-wave-rectified cell model is complicated, because it implies that within a bubble, half of the cells have zero activity.

In comparison, the vast majority of research applying ICA/sparse coding to learn receptive-field models has been based on linear cell models (Olshausen and Field 1996; Bell and Sejnowski 1997; van Hateren and van der Schaaf 1998; van Hateren and Ruderman 1998; Hyvärinen and Hoyer 2000; Hyvärinen et al. 2001). However, in recent years, nonnegative sparse coding models have been introduced and applied in learning receptive-field models (Hoyer and Hyvärinen 2002; Hoyer 2003). It seems that with respect to implementational equivalence, the research reported in this thesis lies somewhere between traditional ICA/sparse coding research, and these new nonnegative models.

Predictive power

In the discussion of the predictive power of the models, we will employ the rough division of the level of evaluation, introduced in Chapter 1:

1. qualitative comparison of results against previous neurophysiological measurements
2. quantitative comparison of results against previous neurophysiological measurements
3. quantitative comparison of results against new neurophysiological measurements inspired by the model.

The results obtained with models presented in this thesis have so far not led to new neurophysiological experiments, while ICA/sparse coding research has already prompted new measurements from the brain (Vinje and Gallant 2000; Weliky et al. 2003). However, it must be noted that initial theoretical results concerning ICA/sparse coding have existed for considerably longer than our results. It is our hope that the work presented here will eventually lead to new hypotheses and their empirical evaluation. In what follows we will limit our discussion to comparison against previous neurophysiological measurements.

In the case of spatial filters, the results obtained with all three models exhibit the qualitative properties of spatial simple-cell receptive fields – localization, orientation, and different scales – as can be seen in Figures 4.2 (page 42), 4.8 (page 49), and 4.10 (page 53). These results link the models strongly to these fundamental qualitative properties of spatial receptive fields of simple cells. Some important quantitative properties – related to selectivity to spatial frequency and orientation – of our results are similar to those of filters obtained with ICA (see Figure 4 in Publication 2). This comparison against ICA results also provides a link to neurophysiological measurements, because ICA results have been compared against physiological measurements in (van Hateren and van der Schaaf 1998), although the preprocessing methods in our study were a bit different (in particular, we had no significant dimensionality reduction). Measurements made from ICA results differ from physiological measurements in the distribution of peak spatial frequencies, but show reasonable agreement with respect to other measurements such as preferred orientation, orientation bandwidth, and spatial frequency bandwidth (van Hateren and van der Schaaf 1998). It must be noted, however, that in comparisons of spatial selectivity, a common unit (visual angle) can not be established for computational results and neurophysiological measurements (van Hateren and van der Schaaf 1998). This reservation also applies to other comparisons of spatial selectivity mentioned below. Overall, it seems that in the case of spatial filters, the predictive power of our models and that of ICA/sparse coding models are approximately similar.

For spatiotemporal filters, the spatial properties of the filters seem to be in reasonable agreement with physiological measurements (see Publication 6). In particular, the distribution of peak spatial frequencies agrees better with physiological measurements than in the case of spatial filters, as was also noted by van Hateren and Ruderman (1998). However, temporal properties of our results seem to differ from measurements made from simple cells. In particular, our results seem to exhibit a smaller degree of temporal localization than the spatiotemporal receptive fields of simple cells, both in the case of maximization of temporal coherence

(see Figure 4.5) as well as bubble coding (see Section 4.5). Thus, with respect to temporal properties of spatiotemporal filters, the predictions of ICA/sparse coding seem to match neurophysiological observations somewhat better (van Hateren and Ruderman 1998).

Two of the models presented in this thesis – the two-layer generative model and the bubble coding model – also make predictions related to the topographical ordering of simple cells in the primary visual cortex, and the way in which complex cells presumably pool the outputs of a number of simple cells. These predictions are qualitative: filters nearby each other in the resulting topography tend to have similar location and/or orientation and/or frequency, and filters presumably pooled by higher-order units share similar characteristics (see Figure 4.8 on page 49 and Figure 4.10 on page 53). The topography observed in the primary visual cortex is very complicated, with different organizational principles governing at different scales (Blasdel 1992; Blasdel and Campbell 2001). Some ICA/sparse coding models make similar predictions as our models (Hyvärinen and Hoyer 2000; Hyvärinen et al. 2001; Hyvärinen and Hoyer 2001; Welling et al. 2003) – some quantitative comparison against physiological measurements can also be found in (Hyvärinen and Hoyer 2001). Overall, it seems that ICA/sparse coding models and our models make similar predictions in this area; however, the predictions made by ICA/sparse coding models have been analyzed somewhat more quantitatively than ours.

Chapter 5

Summary

Neurophysiological measurements of the brain have revealed a wide variety of response properties of neurons in the primary visual cortex, describing how the neurons respond to different visual stimuli. What remains largely unknown, however, is *why* the neurons have these observed properties.

In this thesis, this question has been approached by formulating models of computation for some cells and cell groups in the primary visual cortex. These computational models relate properties of a class of visual neurons, called simple cells, to natural stimulus statistics. An underlying assumption in the development of these models has been that stimulus statistics have influenced the response properties of the neurons through evolution and development.

The main contribution of this thesis is the introduction of three new models of computation. The first of these models characterizes computation on the simple-cell level as bursts of activity. The second model relates the activity levels of different simple cells, located close to each other on the cortex, at nearby time instants to each other. The third model combines these two characterizations, along with other characterizations described in previous research, yielding a unifying model with sparse bubble-like activity regions in the three-dimensional space formed by time and the two-dimensional cortical surface. An additional contribution is the examination of the relationship between two previous models of computation, namely independent component analysis (ICA) and local spatial frequency analysis.

In this thesis, the computational models have been evaluated experimentally by estimating the free parameters of the models, and comparing the resulting parameter values against our knowledge of the properties of visual neurons, or results obtained with other models of computation. The experimental results concerning the new computational models link these models to several observed properties of simple cells, including spatial localization, orientation selectivity, spatial frequency selectivity, directional selectivity, and topographic organization in the cortex. The experimental examination of the relationship between ICA and local spatial frequency analysis suggest that results obtained with ICA share some properties with wavelets: spatial localization tends to increase with mean spatial frequency, while frequency localization tends to decrease. Ideally, the evaluation of a model should also include generation of new hypotheses, and their validation or falsification by new neurophysiological measurements. The models presented here have not yet entered that stage of evaluation.

The computational models examined in this thesis can be considered as poten-

tial partial answers to the question *why*: the models do not link the properties of visual neurons to the tasks of the animal – instead, they specify hypothetical characteristics of the neural representation at the simple-cell level. Some suggestions as to why these characteristics might be useful have been provided in this thesis, but in future research, the question *why* has to, in turn, be applied to these new hypothetical coding principles.

Bibliography

Adelson, E. H. and J. R. Bergen (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A 2*(2), 284–299.

Albrecht, D. G., R. L. DeValois, and L. G. Thorell (1980). Visual cortical neurons: are bars or gratings the optimal stimuli? *Science 207*(4426), 88–90.

Albrecht, D. G. and W. S. Geisler (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience 7*(6), 531–546.

Albright, T. D. and G. R. Stoner (2002). Contextual influences on visual processing. *Annual Review of Neuroscience 25*, 339–379.

Alonso, J.-M. and L. M. Martinez (1998). Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature Neuroscience 1*(5), 395–403.

Atick, J. J. and A. N. Redlich (1990). Towards a theory of early visual processing. *Neural Computation 2*(3), 308–320.

Atick, J. J. and A. N. Redlich (1992). What does the retina know about natural scenes? *Neural Computation 4*(2), 196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Reviews 61*(3), 183–193.

Bair, W., J. R. Cavanaugh, M. Smith, and J. A. Movshon (2002). The timing of response onset and offset in macaque visual neurons. *The Journal of Neuroscience 22*(8), 3189–3205.

Barlow, H. B. (1961). The coding of sensory messages. In W. H. Thorpe and O. L. Zangwill (Eds.), *Current Problems in Animal Behaviour*, Chapter XIII, pp. 331–360. Cambridge University Press.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems 12*(3), 241–253.

Becker, S. and G. E. Hinton (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature 355*(6356), 161–163.

Bell, A. and T. J. Sejnowski (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation 7*(6), 1129–1159.

Bell, A. and T. J. Sejnowski (1997). The independent components of natural scenes are edge filters. *Vision Research 37*(23), 3327–3338.

Berkes, P. and L. Wiskott (2002). Applying slow feature analysis to image sequences yields a rich repertoire of complex cell properties. In J. R. Dorronsoro (Ed.), *Artificial Neural Networks – ICANN 2002*, Volume 2415 of *Lecture notes in computer science*, pp. 81–86. Springer.

Blakemore, C. and F. W. Campbell (1969). On the existence of neurons in the human visual system selectively responsive to the orientation and size of retinal images. *Journal of Physiology 203*(1), 237–260.

Blakemore, C. and G. F. Cooper (1970). Development of the brain depends on the visual environment. *Nature 228*(270), 477–478.

Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *The Journal of Neuroscience 12*(8), 3139–3161.

Blasdel, G. G. and D. Campbell (2001). Functional retinotopy of monkey visual cortex. *The Journal of Neuroscience 21*(20), 8286–8301.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews 36*(2–3), 96–107.

Bullier, J. (2002). Neural basis of vision. In H. Pashler, S. Yantis, D. Medin, R. Gallistel, and J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology* (3rd ed.), Volume 1, pp. 1–40. John Wiley & Sons.

Buračas, G. T., A. M. Zador, M. R. DeWeese, and T. D. Albright (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron 20*(5), 959–969.

Cai, D., G. C. DeAngelis, and R. D. Freeman (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology 78*(2), 1045–1061.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience 21*, 47–74.

Callaway, E. M. (2001). Neural mechanisms for the generation of visual complex cells. *Neuron 32*(3), 378–380.

Carandini, M., D. J. Heeger, and J. A. Movshon (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience 17*(21), 8621–8644.

Carandini, M., D. J. Heeger, and J. A. Movshon (1999). Linearity and gain control in V1 simple cells. In E. G. Jones and P. S. Ulinski (Eds.), *Models of cortical function*, Volume 13 of *Cerebral Cortex*, pp. 401–443. Plenum Press.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE 86*(10), 2009–2025.

Chance, F. S., S. B. Nelson, and L. F. Abbott (1999). Complex cells as cortically amplified simple cells. *Nature Neuroscience 2*(3), 277–282.

Chechik, G., A. Globerson, N. Tishby, M. J. Anderson, E. D. Young, and I. Nelken (2001). Group redundancy measures reveal redundancy reduction in the auditory pathway. In T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems*, Volume 14, pp. 173–180. The MIT Press.

Cichocki, A. and S.-i. Amari (2002). *Adaptive Blind Signal and Image Processing*. John Wiley & Sons.

Cohen, A. and J. Kovačević (1996). Wavelets: the mathematical background. *Proceedings of the IEEE 84*(4), 514–522.

Cohen, L. (1995). *Time-Frequency Analysis*. Prentice Hall Signal Processing Series. Prentice Hall.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing 36*(3), 287–314.

Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory*. John Wiley & Sons.

Dan, Y., J. J. Atick, and R. C. Reid (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *The Journal of Neuroscience 16*(10), 3351–3362.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research 20*(10), 847–856.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A 2*(7), 1160–1169.

Dayan, P. and L. F. Abbott (2001). *Theoretical Neuroscience*. The MIT Press.

DeAngelis, G. C., R. D. Freeman, and I. Ohzawa (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology 71*(1), 347–374.

DeAngelis, G. C., G. M. Ghose, I. Ohzawa, and R. D. Freeman (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *The Journal of Neuroscience 19*(9), 4046–4064.

DeAngelis, G. C., I. Ohzawa, and R. D. Freeman (1993a). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology 69*(4), 1091–1117.

DeAngelis, G. C., I. Ohzawa, and R. D. Freeman (1993b). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology 69*(4), 1118–1135.

DeAngelis, G. C., I. Ohzawa, and R. D. Freeman (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences 18*(10), 451–458.

deCharms, R. C. and A. M. Zador (2000). Neural representation and the cortical code. *Annual Review of Neuroscience 23*, 613–647.

DeValois, R. L., D. G. Albrecht, and L. G. Thorell (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research 22*(5), 545–559.

Dong, D. W. and J. J. Atick (1995a). Statistics of natural time-varying images. *Network: Computation in Neural Systems 6*(3), 345–358.

Dong, D. W. and J. J. Atick (1995b). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems 6*(2), 159–178.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience 2*(12), 920–926.

Eckert, M. P. and G. Buchsbaum (1993). Efficient coding of natural time varying images in the early visual system. *Philosophical Transactions of the Royal Society B 339*(1290), 385–395.

Emerson, R. C., J. R. Bergen, and E. H. Adelson (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research 32*(2), 203–218.

Enroth-Cugell, C. and J. G. Robson (1984). Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description. *Investigative Opthalmology & Visual Science 25*(3), 250–267.

Enroth-Cugell, C., J. G. Robson, D. E. Schweitzer-Tong, and A. B. Watson (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *Journal of Physiology 341*(1), 279–307.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A 4*(12), 2379–2394.

Field, D. (1994). What is the goal of sensory coding? *Neural Computation 6*(4), 559–601.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation 3*(2), 194–200.

Freeman, R. D., I. Ohzawa, and G. A. Walker (2001). Beyond the classical receptive field in the visual cortex. *Progress in Brain Research 134*, 157–170.

Gaska, J. P., L. D. Jacobson, H.-W. Chen, and D. A. Pollen (1994). Space-time spectra of complex cell filters in the macaque monkey: A comparison of results obtained with pseudowhite noise and grating stimuli. *Visual Neuroscience 11*(4), 805–821.

Girolami, M. (1999). *Self-Organising Neural Networks: Independent Component Analysis and Blind Signal Separation*. Springer.

Gonzalez, R. C. and R. E. Woods (1992). *Digital Image Processing*. Addison-Wesley.

Goodale, M. A. and A. D. Milner (1992). Separate visual pathways for perception and action. *Trends in Neurosciences 15*(1), 20–25.

Hashimoto, W. (2003). Quadratic forms in natural images. *Network: Computation in Neural Systems 14*(4), 765–788.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience 9*(2), 181–197.

Hendry, S. H. C. and R. C. Reid (2000). The koniocellular pathway in primate vision. *Annual Review of Neuroscience 23*, 127–153.

Hess-Nielsen, N. and M. V. Wickerhausen (1996). Wavelets and time-frequency analysis. *Proceedings of the IEEE 84*(4), 523–540.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence 40*(1–3), 185–234.

Hinton, G. E. and Z. Ghahramani (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society B 352*(1358), 1177–1190.

Hoyer, P. O. (2003). Modeling receptive fields with non-negative sparse coding. In E. De Schutter (Ed.), *Computational Neuroscience: Trends in Research 2003*, pp. 547–552. Elsevier.

Hoyer, P. O. and A. Hyvärinen (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems 11*(3), 191–210.

Hoyer, P. O. and A. Hyvärinen (2002). A multi-layer sparse coding network learns contour coding from natural images. *Vision Research 42*(12), 1593–1605.

Hubel, D. H. and T. N. Wiesel (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology 160*(1), 106–154.

Hubel, D. H. and T. N. Wiesel (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology 195*(1), 215–243.

Humphrey, A. L. and A. B. Saul (1998). Strobe rearing reduces direction selectivity in area 17 by altering spatiotemporal receptive-field structure. *Journal of Neurophysiology 80*(6), 2991–3004.

Hyvärinen, A. (2001). Blind source separation by nonstationarity of variance: A cumulant-based approach. *IEEE Transactions on Neural Networks 12*(6), 1471–1474.

Hyvärinen, A. and P. O. Hoyer (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation 12*(7), 1705–1720.

Hyvärinen, A. and P. O. Hoyer (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research 41*(18), 2413–2423.

Hyvärinen, A., P. O. Hoyer, and M. Inki (2001). Topographic independent component analysis. *Neural Computation 13*(7), 1525–1558.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent Component Analysis*. John Wiley & Sons.

Ifeachor, E. and B. W. Jervis (2002). *Digital Signal Processing – A Practical Approach* (2nd ed.). Addison-Wesley.

Jones, J. P. and L. A. Palmer (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology 58*(6), 1233–1258.

Jutten, C. and J. Herault (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing 24*(1), 1–10.

Kandel, E. R. (2000). Nerve cells and behavior. See Kandel, Schwartz, and Jessell (2000), Chapter 2, pp. 19–35.

Kandel, E. R., J. H. Schwartz, and T. M. Jessell (Eds.) (2000). *Principles of Neural Science* (4th ed.). McGraw-Hill.

Kandel, E. R. and R. H. Wurtz (2000). Constructing the visual image. See Kandel, Schwartz, and Jessell (2000), Chapter 25, pp. 492–506.

Karklin, Y. and M. S. Lewicki (2003). Higher-order structure of natural images. *Network: Computation in Neural Systems 14*(3), 483–499.

Kayser, C., W. Einhäuser, O. Dümmer, P. König, and K. Körding (2001). Extracting slow subspaces from natural videos leads to complex cells. In G. Dorffner, H. Bischof, and K. Hornik (Eds.), *Artificial Neural Networks – ICANN 2001*, Volume 2130 of *Lecture notes in computer science*, pp. 1075–1080. Springer.

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A 4*(12), 2395–2400.

Knill, D. C. and W. Richards (Eds.) (1996). *Perception as Bayesian inference*. Cambridge University Press.

Kohonen, T., S. Kaski, and H. Lappalainen (1997). Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural Computation 9*(6), 1321–1344.

Kovačević, J. and I. Daubechies (Eds.) (1996). *Proceedings of the IEEE*. Volume 84, number 4, special issue on wavelets.

Krebs, J. R. and N. B. Davies (1993). *An Introduction to Behavioural Ecology* (3rd ed.). Blackwell Science.

Lamme, V. A. F. and P. R. Roelfsema (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences 23*(11), 571–579.

Lampl, I., J. S. Anderson, D. C. Gillespie, and D. Ferster (2001). Prediction of orientation selectivity from receptive field architecture in simple cells of cat visual cortex. *Neuron 30*(1), 263–274.

Lee, T. S. (1996). Image representations using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*(10), 959–971.

Lee, T.-W. (1998). *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers.

Lennie, P. (2000). Color vision. See Kandel, Schwartz, and Jessell (2000), Chapter 29, pp. 572–589.

Lindeberg, T. and K. V. Mardia (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics 21*(1/2), 224–270.

Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neurosciences 20*(1), 38–43.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America 70*(11), 1297–1300.

Marlin, S. G., R. M. Douglas, and M. S. Cynader (1991). Position-specific adaptation in simple cell receptive fields of the cat striate cortex. *Journal of Neurophysiology 66*(5), 1769–1784.

Marlin, S. G., S. J. Hasan, and M. S. Cynader (1988). Direction-selective adaptation in simple and complex cells in cat striate cortex. *Journal of Neurophysiology 59*(4), 1314–1330.

Marr, D. and E. Hildreth (1980). Theory of edge detection. *Proceedings of the Royal Society of London B 207*(1167), 187–217.

Martinez, L. M. and J.-M. Alonso (2001). Construction of complex receptive fields in cat primary visual cortex. *Neuron 32*(3), 515–525.

Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience 4*(9), 877–886.

Mathews, V. J. and G. L. Sicuranza (2000). *Polynomial Signal Processing*. John Wiley & Sons.

Matsuoka, K., M. Ohya, and M. Kawamoto (1995). A neural net for blind separation of nonstationary signals. *Neural Networks 8*(3), 411–419.

Mishkin, M., L. G. Ungerleider, and K. A. Macko (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences 6*, 414–417.

Movshon, J. A., I. D. Thompson, and D. J. Tolhurst (1978). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology 283*(1), 79–99.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch and J. L. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain*, Chapter 7, pp. 125–152. The MIT Press.

Nieder, A., D. J. Freedman, and E. K. Miller (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science 297*(5587), 1708–1711.

Nirenberg, S., S. M. Carcieri, A. L. Jacobs, and P. E. Latham (2001). Retinal ganglion cells act largely as independent encoders. *Nature 411*(6838), 698–701.

Olshausen, B. A. (2000). Sparse coding of time-varying natural images. In P. Pajunen and J. Karhunen (Eds.), *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 603–608.

Olshausen, B. A. (2003). Principles of image representation in visual cortex. In L. Chalupa and J. Werner (Eds.), *The Visual Neurosciences*. The MIT Press.

Olshausen, B. A. and D. Field (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature 381*(6583), 607–609.

Olshausen, B. A. and D. Field (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research 37*(23), 3311–3325.

Palmer, L. A., J. P. Jones, and R. A. Stepnoski (1991). Striate receptive fields as linear filters: characterization in two dimensions of space. In A. G. Leventhal (Ed.), *The Neural Basis of Visual Function*, Volume 4 of *Vision and Visual Dysfunction*, pp. 246–265. CRC Press.

Palmer, S. E. (1999). *Vision Science – Photons to Phenomenology*. The MIT Press.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes* (3rd ed.). Electrical & Electronic Engineering Series. McGraw-Hill.

Párrage, C. A., T. Troscianko, and D. J. Tolhurst (2000). The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology 10*(1), 35–38.

Pasupathy, A. and C. E. Connor (2002). Population coding of shape in area V4. *Nature Neuroscience 5*(12), 1332–1338.

Pham, D.-T. and J.-F. Cardoso (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing 49*(9), 1837–1848.

Pollen, D. A. and S. F. Ronner (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man and Cybernetics 13*(5), 907–916.

Rees, G., G. Kreiman, and K. Koch (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience 3*(4), 261–270.

Reinagel, P. (2001). How do visual neurons respond in the real world? *Current Opinion in Neurobiology 11*(4), 437–442.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology 88*(1), 455–463.

Sawatari, A. and E. M. Callaway (1996). Convergence of magno- and parvocellular pathways in layer 4B of macaque primary visual cortex. *Nature 380*(6573), 442–446.

Schoenfeld, M. A., H.-J. Heize, and M. G. Woldorff (2002). Unmasking motion-processing activity in human brain area V5/MT+ mediated by pathways that bypass primary visual cortex. *NeuroImage 17*(2), 769–779.

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

Sengpiel, F., P. Stawinski, and T. Bonhoeffer (1999). Influence of experience on orientation maps in cat visual cortex. *Nature Neuroscience 2*(8), 727–732.

Sherman, S. M. and R. W. Guillery (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society B 357*(1428), 1695–1708.

Sherman, S. M. and P. D. Spear (1982). Organization of visual pathways in normal and visually deprived cats. *Physiological Reviews 62*(2), 738–855.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology 13*(2), 144–149.

Simoncelli, E. P. and B. A. Olshausen (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience 24*, 1193–1216.

Stanley, G. B. (2002). Adaptive spatiotemporal receptive field estimation in the visual pathway. *Neural Computation 14*(12), 2925–2946.

Stearns, S. C. and R. F. Hoekstra (2000). *Evolution: an introduction*. Oxford University Press.

Stone, J. (1996). Learning visual parameters using spatiotemporal smoothness constraints. *Neural Computation 8*(7), 1463–1492.

Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature 401*(6750), 269–272.

Szatmáry, B. and A. Lőrincz (2002). Independent component analysis of temporal sequences subject to constraints by lateral geniculate nucleus inputs yields all the three major cell types of the primary visual cortex. *Journal of Computational Neuroscience 11*(3), 241–248.

Szulborski, R. G. and L. A. Palmer (1990). The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. *Vision Research 30*(2), 249–254.

Tessier-Lavigne, M. (2000). Visual processing by the retina. See Kandel, Schwartz, and Jessell (2000), Chapter 26, pp. 507–522.

Tolhurst, D. J. and A. F. Dean (1991). Evaluation of a linear model of directional selectivity in simple cells of the cat's striate cortex. *Visual Neuroscience 6*(5), 421–428.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience 4*(3), 219–229.

Tootell, R. B. H., M. S. Silverman, S. L. Hamilton, E. Switkes, and R. L. DeValois (1988). Functional anatomy of macaque striate cortex. V. Spatial frequency. *The Journal of Neuroscience 8*(5), 1610–1624.

Tsodyks, M., T. Kenet, A. Grinvald, and A. Arieli (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science 286*(5446), 1943–1946.

Tsutsui, K.-I., H. Sakata, T. Naganuma, and M. Taira (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science 298*(5592), 409–412.

van Hateren, J. H. (1992). Real and optimal neural images in early vision. *Nature 360*(6399), 68–70.

van Hateren, J. H. and D. L. Ruderman (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London B 265*(1412), 2315–2320.

van Hateren, J. H. and A. van der Schaaf (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B 265*(1394), 359–366.

Vanni, S., T. Tanskanen, M. Seppä, K. Uutela, and R. Hari (2001). Coinciding early activation of the human primary visual cortex and anteromedial cuneus. *Proceedings of the National Academy of Sciences of the USA 98*(5), 2776–2780.

Vidyasagar, T. R., J. J. Kulikowski, D. M. Lipnicki, and B. Dreher (2002). Convergence of parvocellular and magnocellular information channels in the primary visual cortex of the macaque. *European Journal of Neuroscience 16*(5), 945–956.

Vinje, W. E. and J. L. Gallant (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science 287*(5456), 1273–1276.

Wainwright, M. J. and E. P. Simoncelli (2000). Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen, and K.-R. Müller (Eds.), *Advances in Neural Information Processing Systems*, Volume 12, pp. 855–861. The MIT Press.

Watson, A. B. and A. J. Ahumada (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A 2*(2), 322–342.

Weliky, M., J. Fiser, R. H. Hunt, and D. N. Wagner (2003). Coding of natural scenes in primary visual cortex. *Neuron 37*(4), 703–718.

Welling, M., S. Osindero, and G. E. Hinton (2003). Learning sparse topographic representations with products of Student-t distributions. In S. Becker, S. Thrun, and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems*, Volume 15, pp. 1359–1366. The MIT Press.

Willmore, B. and D. J. Tolhurst (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems 12*(3), 255–270.

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

Wörgötter, F. and U. T. Eysel (2000). Context, state and the receptive fields of striatal cortex cells. *Trends in Neurosciences 23*(10), 497–503.

Wurtz, R. H. and E. R. Kandel (2000a). Central visual pathways. See Kandel, Schwartz, and Jessell (2000), Chapter 27, pp. 523–547.

Wurtz, R. H. and E. R. Kandel (2000b). Perception of motion, depth, and form. See Kandel, Schwartz, and Jessell (2000), Chapter 28, pp. 548–571.

Zetzsche, C. and G. Krieger (1999). Nonlinear neurons and high-order statistics: New approaches to human vision and electronic image processing. In B. E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging IV*, Volume 3644 of *Proceedings of SPIE*, pp. 2–33. SPIE—The International Society for Optical Engineering.

Zipser, K., V. A. F. Lamme, and P. H. Schiller (1996). Contextual modulation in primary visual cortex. *The Journal of Neuroscience 16*(22), 7376–7389.