Technische Universität Berlin

Bernstein Center for Computational Neuroscience Berlin

Master's Thesis

Self-organization Of V1 Complex-Cells Based On Slow Feature Analysis And Retinal Waves

Sven Dähne

June 11, 2010

Supervisors:

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum
Germany

Niko Wilbert
Institute for Theoretical Biology
Humboldt-Universität zu Berlin
Germany

Abstract

The developing visual system of many mammalian species is partially structured and organized even before the onset of vision. Spontaneous neural activity, which spreads in waves across the retina, has been suggested to play a major role in these prenatal structuring processes. Recently, it has been shown that when employing an efficient coding strategy, such as sparse coding, these retinal activity patterns lead to basis functions that resemble optimal stimuli of simple cells in primary visual cortex (V1).

Here I present the results of applying a coding strategy that optimizes for temporal slowness, namely Slow Feature Analysis (SFA), to a biologically plausible model of retinal waves. Previously, SFA has been successfully applied in modeling parts of the visual system, most notably in reproducing a rich set of complex-cell features by training SFA with natural image sequences. In this work, I was able to obtain units that share a number of properties with cortical complex-cells by training with simulated retinal waves.

The results support the idea that retinal waves share relevant temporal and spatial properties with natural visual input. Hence, retinal waves seem suitable training stimuli to learn invariances and thereby shape the developing early visual system so that it is best prepared for coding input from the natural world.

Acknowledgments

I would like to thank my supervisor Prof. Laurenz Wiskott for giving me the opportunity to work on this interesting project. The numerous discussions with him and Niko Wilbert have provided me with great insights and intuitions, which were most helpful in the process of completing this thesis.

I am particularly grateful for the constant support of Niko Wilbert. His remarkable Python-skills have been a great help on many occasions. The opportunity to use and extend his simulation framework has saved a lot of time and is much appreciated. Furthermore, I thank Niko for patiently reading the first version of this thesis and providing most helpful comments and suggestions.

I would also like to thank Christian Hinze and Henning Sprekeler for valuable discussions and useful hints in several stages of the project. The idea for this project was conceived by Henning Sprekeler.

Most importantly, I would like to thank my parents. Without their constant support during my entire duration of study, none of this would have been possible.

Contents

1. Introduction **6**

1.1. The Early Visual System 8

1.1.1. From The Retina To V1 9

1.1.2. Retinal Waves 11

1.2. Normative Computational Models 14

1.2.1. Sparse Coding, Information Maximization, and ICA 15

1.2.2. Temporal Stability, Slowness, and SFA 17

1.3. Thesis Hypothesis 18

2. Methods **20**

2.1. Slow Feature Analysis 20

2.2. Input Data 23

3. Results **27**

3.1. Receptive Fields 27

3.2. Sinusoidal Test Stimuli 28

4. Control Experiments **32**

4.1. Phase Invariance 32

4.2. Orientation Selectivity 34

4.3. Gabor-patch Quadrature Filter Pair Model 36

4.4. Discrete vs Continuous Training Data 40

5. Discussion **43**

A. Gabor Quadrature Filter Pair Model **47**

B. Affirmation **50**

1. Introduction

The brain is an organ of remarkable complexity. In humans, it contains approximately 10^{10} nerve cells (neurons) each making up to 10^4 connections to other nerve cells. Thereby, a network of neurons is formed with a total number of connections in the range of 10^{14} to 10^{15}. This network defines who we are. It gives rise to all our thoughts, all our feelings, all our memories, and all our senses. It allows us to form perceptions and, via those perceptions, experience the world that is around us.

The brain however, does not "see" or "hear". It merely receives electrical signals (action potentials) from nerve cells that are connected to our sensory organs. It is these electrical signals that constitute the input to the brain, and form the basis of all neuronal computations. By means of neurons exchanging action potentials, the brain interprets the signals that come from our sensory organs and only thereby enables us to hear, to smell, to taste, and most importantly to see. The computations performed by the brain when interpreting signals are a result of the physical properties of individual neurons and of the connectivity patterns between them. While the properties of individual neurons are rather fixed, the connectivity patterns are subject to change and thereby allow adaption and learning.

Perhaps for us humans, the dominant sense among all other senses is the sense of vision. Loosing it causes much more severe impairment than for example loosing the ability to taste. Hence, it does not come as a surprise that visual areas take up the most space among the sensory areas[1] on the surface of the cortex[2]. A more comprehensive introduction into the brain areas involved in visual processing will be given later in this chapter (see section 1.1). For now it suffices to acknowledge the fact that there *are* regions in the brain that are exclusively devoted to the processing of visual information.

The regions involved in the earlier stages of processing have been studied extensively, one such region being the "primary visual cortex" (V1). The conjunction of neuroscientific

[1] Neurons that play a similar functional role, such as being involved in visual perception or in motor control, usually occupy spatially connected regions, or areas, of the brain. Consequently, those regions that contain neurons involved in sensory processing are referred to as sensory areas.

[2] Cerebral cortex, or just *cortex*, is the out-most layer of the brain. Among other, it is where the higher cognitive functions such as memory, attention, language, and perceptual awareness are believed to be localized.

results and the study of the statistical properties of natural images (i.e. the input to the visual system) have led to the idea, that neurons found in V1 are actually very well adapted to the statistical regularities present in natural images (Field [1994]). In fact, some of their response properties can be regarded near-optimal with respect to certain efficiency criteria (Olshausen and Field [1996], Bell and Sejnowski [1997], Einhauser et al. [2002], Berkes and Wiskott [2005]). So it seems that the design of early visual processing areas, i.e. the connectivity patterns between neurons in these areas, has evolved to cope best with images provided by the natural environment. This notion agrees very much with evolutionary theory, so it does not come as too big a surprise. What is interesting though, is the question of how these well-designed connectivity patterns emerge as the visual system develops in the newly born (or even unborn) infant.

Two possible answers to that question come to mind:

1. The connectivity patterns are stored in the organism's genetic code and are expressed as the organism matures.

2. The connectivity patterns are acquired once the organism is confronted with its natural input, such that it can adapt the "wiring" of the visual system to the regularities in natural images.

Proposition (1) is unlikely to be true for two major reasons. First of all, the sheer number of connections to be stored would exceed the capacity of the genetic code. There are simply too many. Second of all, a number of studies point to the fact, that the visual system needs visual input to fully develop its characteristic properties (Chapman and Stryker [1993], Chapman et al. [1996]). Proposition (2) is able to accommodate the arguments brought up against proposition (1). However, it can also not be entirely true, because there are animals that can see right after birth (Albert et al. [2008]). Furthermore, there are studies that indicate that some species have a partially functioning visual cortex before eye-opening (Wiesel and Hubel [1963], Horton and Hocking [1996]).

Essentially, propositions (1) and (2) represent the old debate of nature versus nurture, or innate versus learned. Perhaps closer to the truth would be a theory about the development of the early visual system that can be regarded as half-way between the two points of view mentioned. Such a theory has been brought forward by Albert, Schnabel and Field in a publication from 2008, entitled "Innate visual learning through spontaneous activity patterns" (Albert et al. [2008]). This theory proposes an innate learning approach, based on internally generated retinal input. This internally generated input drives the development *before* the onset of vision and thereby prepares the system for

further refinement that occurs *after* the onset of vision. In other words, the visual system is trained to be best adapted to its input, but the input is first internally generated, almost like a simulation of what is to be expected once the animal's eyes open. Next to the internally generated input, the second important ingredient to the theory is the learning objective, i.e. the goal that governs what regularities of the input the system has to adapt to. Albert et al. have chosen an objective that ensures a good representation of the input while keeping the level of activity (the number of action potentials) of individual model neurons low, thus minimizing the energy consumption of the system. This objective can also be interpreted as maximizing the statistical independence of the outputs of the individual model neurons. In their simulations, such an objective together with the mentioned internally generated input was able to reproduce important properties of the early visual system. See section 1.2.1 for a more detailed description of the relevant concepts.

In my thesis, I want to show that the emergence of specific properties of the visual system can be explained using an entirely different learning objective, namely the objective of "temporal stability" or "slowness". However, before I can go on and make my hypothesis more explicit, I have to make a digression and explain a few more terms and key concepts concerning the neurobiology of the visual system and a particular class of recent computational models. The necessary basics of the visual system are provided in section 1.1. Readers who are familiar with terms such as "V1", "simple cells", and "complex cells" as well as their properties may choose to skip this section. A brief introduction into normative computational models is given in section 1.2. Again, readers familiar with terms such as "sparse coding", "temporal coherence", "slow feature analysis", and normative modeling in general may choose to skip this section entirely.

Once these concepts have been introduced, I formulate my thesis hypothesis in a more detailed fashion and give a brief outline of how I intend to test it in section 1.3.

1.1. The Early Visual System

In this section, I first give a short introduction into the major parts of the early visual system. The second part of this section introduces retinal waves, which are the pre-onset-of-vision internally generated stimuli that were mentioned earlier. In both cases, I confine my descriptions to the parts that are important for the context of this thesis.

1.1.1. From The Retina To V1

The neural visual system begins in the back of the eye with the retina. After the light has passed through the lens and entered the eyeball, images form on the back of the eye were the retina is located. The retina consists of several layers of cell types through which the light passes before it reaches the light sensitive receptor cells in the last retinal layer. The layers and their spatial arrangement are schematically depicted in figure 1.1 A. There are two types of light sensitive cells which are called rods and cones due to their specific shapes. The rods and cones convert the physical stimulus (light intensity) into changes of their membrane potential, which are passed on to the next layer of cells. The horizontal cell layer integrates membrane potential changes from several neighboring receptor cells. A similar function is performed by the layer of amacrine cells. Ultimately, the membrane potential variations reach the ganglion cell layer in which action potentials are generated. A single retinal ganglion cell encodes the light intensity detected by all receptor cells that feed input into it. The ganglion cell thereby signals information about the light that is received by a spatially confined patch of the retina. This area of the visual field to which a ganglion cell is responsive is called the *receptive field* (RV) of the cell. As visually responsive cells are better stimulated with images that contain structure (e.g. oriented bars or gratings) the structured images that stimulate a cell are often also referred to as the cell's receptive field.

Together, the fibers of the ganglion cells along which the action potentials travel constitute the optic nerve, which leaves the eyeball and projects to the back of the brain. The path from the retina to the primary visual cortex is called optic tract and is indicated in figure 1.1 B. At this point, the nerve bundles coming from either eye contain information about both sides of the visual field, the right and the left side. This changes after the next station on the optic tract, which is called the optic chiasm. The optic chiasm is a point of branching and conversion for the fibers that represent the two sides of the visual field. Coming from the right eye, those fibers that represent the right side of the visual field cross over to the left side of the brain whereas the fibers that represent the left visual field remain on the right side of the brain. They are joined by fibers that come from the left eye and also represent the left visual field. After the optic chiasm, the information from one side of the visual field now travels on the opposite side of the brain.

The next station on the optic tract is the lateral geniculate nucleus (LGN). The LGN is considered the first processing center for visual information and is located in the thalamus. Cells in the LGN have similar receptive fields compared to retinal ganglion cells, see figure 1.2 A. The receptive fields are round with subregions that excite the cells when stimulated with light and subregions that inhibit the cell when stimulated with light. One subregion

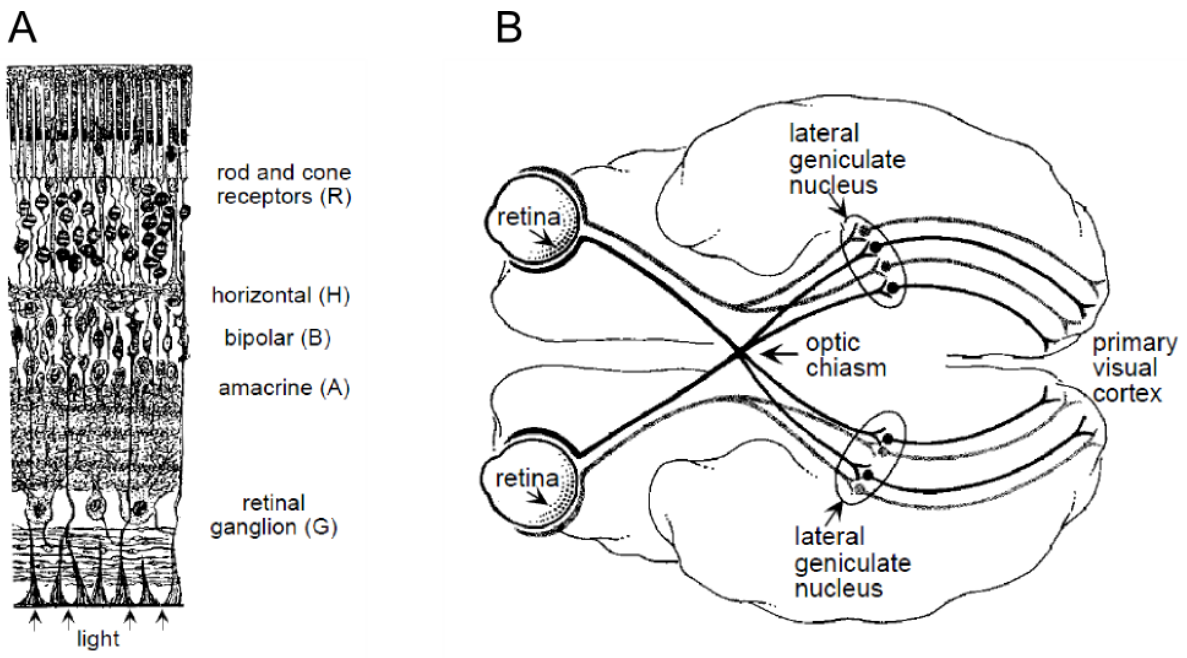

Figure 1.1.: The early visual system. **A** Schematic of the different cell type layers that constitute the retina. Note that the light enters from below in the diagram. **B** Visual pathway. The visual information travels from the retina along the visual pathway to the primary visual cortex. From there it is dispatched to further processing areas (not shown). Figures are adapted from Dayan and Abbott [2001]

usually encloses another, which has lead o the terms ON-center receptive field if the excitatory subregion is enclosed by an inhibitory one, and OFF-center in the inverse case. Consequently, the optimal stimulus for an ON-center cell would be a bright dot, while an OFF-center cell would be best stimulated by bright ring with a dark center. Among others, one functional role of the LGN is presumed to be the temporal decorrelation of visual input (Dong and Atick [1995b]).

Primary visual cortex (V1) is the final stage in the early visual system that will be considered in this introduction. The task of most V1 neurons seems to be the extraction and representation of low-level image features. Extracting features from images means, that there are cells that are active only if a certain feature (such as straight lines having a specific orientation) is present in the image. Examples of receptive fields of these cells are shown in figure 1.2 B. Orientation is one feature to extract, spatial frequency would be another. Spatial frequency is a bit more technical and harder to grasp than orientation, but intuitively speaking it can be understood as the degree of details present in an image. In order to understand the content of an image, feature extraction is fundamental. After identifying structures such as edges in images, these can be grouped and understood to form a shape, which in turn can be interpreted to belong to an actual object. Our entire perception is constructed from these basic image features.

V1 contains these "feature detector" cells and they are usually classified into two classes: *simple cells* and *complex cells*. The stimuli that are preferred by simple and complex cells are images that contain parallel bright and dark lines (figure 1.2 C). In these stimuli the transition from dark to light follows a sinusoidal function, and therefore they are called sinusoidal gratings, or planar waves. These gratings are parameterized by their orientation, spatial frequency[3], and phase[4]. Both, simple and complex cells, generally show a selectivity for orientation and spatial frequency. However, only simple cells are also sensitive to the phase of the grating. In other words, while simple cells can encode the position of an oriented bar within their receptive field, complex cells can only encode the orientation and width of such a bar stimulus. This *invariance* with respect to phase can be regarded as a first level of abstraction, which is an important step towards the reliable extraction of higher order features including object identity for example. The

[3]Here the spatial frequency is the frequency of the cosine that marks the transition from light to dark in the planar wave. Alternatively it can also be understood as the width of the light and dark bars. Higher spatial frequency means narrower bars.

[4]The phase of the grating describes a shift in the onset of the light/dark oscillation relative to the center of the image. This shift is given in radians or alternatively in degrees. A zero degree phase shift means that the center of the bright bar is in the middle of the image. A 90 degree phase shift means a shift of half an oscillation period to the right, i.e. now the dark bar is in the center of the image.

described phase selectivity in the response of simple cells and phase invariance in the response of complex cells is illustrated in figure 1.2 E and F, respectively. The plots show the response as a function of phase varying over time.

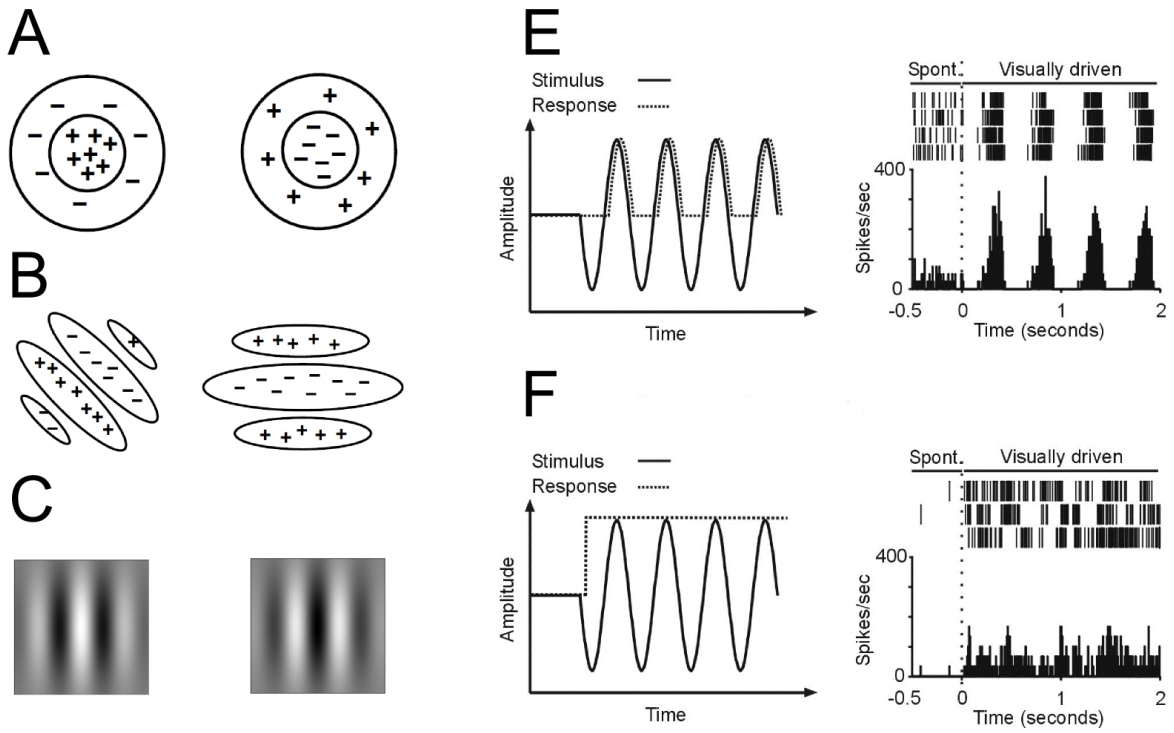

Figure 1.2.: **A** Typical receptive fields of LGN cells. Plus signs denote regions that excite the cell when illuminated, minus signs denote regions that inhibit the cell when stimulated with light. **B** Receptive fields of V1 simple cells. **C** Sinusoidal gratings used to stimulate visually responsive V1 cells. Note the similarity to the receptive fields in **B**. The right sinusoid has a phase shift of 90 degrees relative to the left one. **E** and **F** Response of simple (**E**) and complex cells (**F**) to changes in phase over time. The right plots show experimental data and the left plots show an idealization of the response properties. Figures **A** and **B** taken from Hyvärinen et al. [2009]. Figures **E** and **F** are taken from the Scholarpedia article on receptive fields (www.scholarpedia.org/article/Receptive_field).

1.1.2. Retinal Waves

In the introductory paragraphs I have mentioned internally generated input to the developing early visual system, prior to the onset of vision. Where does this internally generated input come from?

The immature and yet light-insensitive retina of many species generates spontaneous bursting activity. This activity occurs in spatiotemporal patterns spreading in waves

across the retina bringing the spontaneous bursts of neighboring cells into synchrony. Because of their wave-like shape these spontaneous activity patterns are called *retinal waves*. Figure 1.3 depicts the traveling dynamics of one such wave in the top row of plots.

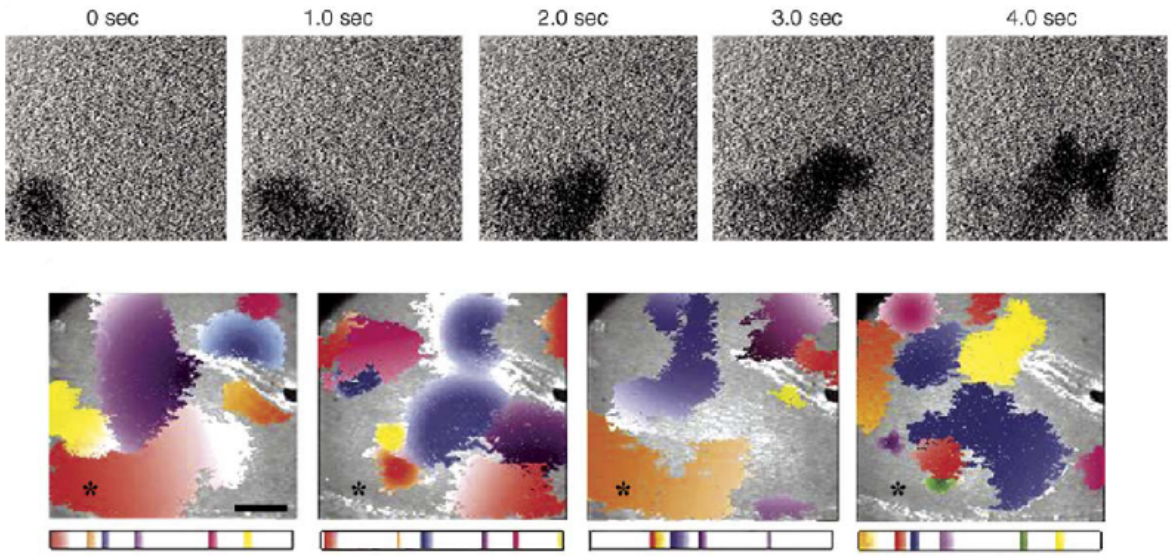

Figure 1.3.: Retinal waves. Top row: Calcium imaging plots of a ferret retina. The five plots show the wave-like bursting activity (dark regions) over five consecutive seconds. Bottom row: Retinal wave domains. Each plot summarizes the wave activity of one minute in the same retina patch as in the top row plots. Colored regions are wave domains, i.e. regions within which a single wave emerged, traveled, and disappeared. Wave domains are not overlapping and not fixed. Figures are taken from Firth et al. [2005].

The facts and figures of the following three paragraphs, which describe the most important properties of retinal waves and how they emerge, are taken from the following review articles: Wong [1999], Thompson [1997], Firth et al. [2005], Blankenship and Feller [2009].

Retinal waves travel within certain spatial domains. This means that they emerge, travel some distance, and then fade away all within spatially confined regions of the retina (see figure 1.3, bottom row of plots). The boundaries of these domains, however, are not fixed. They change over time, invading all areas of the retina. Hence, there are no special pace-maker regions from which waves might be sent out. The direction of wave travel is not predictable, with the exception that waves tend not to spread into domains that have been recently active, thus establishing bounds of the active domains.

Measured parameters such as speed, size, burst frequency, and burst durations vary

over species and also over time within a species. In ferrets for example, the burst duration measured at a fixed location on the retina is 1 to 4 seconds in the first week after birth but goes up to 2 to 19 seconds in the third week after birth (ferrets open their eyes during postnatal week 5 and 6 White et al. [2001]). These values are comparable with burst durations measured in rats and mice. Wave speed has been reported to be in the range of 0.1 to 0.3 mm/s in ferrets.

The underlying mechanism that give rise to the correlated bursting patterns are rather complicated and still subject of study. In the developing retina of mice, three different mechanisms have been identified that generate waves at subsequent developmental stages. The driving force behind wave initiation seem to be retinal amacrine cells (see figure 1.1 A), which, in the absence of synaptic input, depolarize regularly (approximately every 15 seconds, Zheng et al. [2006]). They, in turn, excite retinal ganglion cells. In the earliest developmental stages, the amacrine and ganglion cell activity is postulated to be propagated horizontally via gap junctions (i.e. direct electrical transmission) between ganglion cells or amacrine cells. In a later stage, the activity is mediated horizontally via neurotransmitters (acetylcholine) between amacrine cells. Shortly before eye opening the acetylcholine mediated mechanism is yet again replaced by another. The transmission is henceforth taken over by bipolar interneurons (also indicated in figure 1.1 A) and the neurotransmitter glutamate. The variations in the generation mechanism are most likely the reason for the subtle changes in some wave properties (e.g. slight increase in propagation speed) as the visual system develops.

Theoretical as well as experimental studies (Willshaw and Von Der Malsburg [1976], Lowel and Singer [1992], respectively) suggest that spatially correlated input is required for the proper development of ordered connections from the retina to the LGN and then to visual cortex. Examples of these ordered connections include retinotopic maps, ocular dominance columns, and orientation maps in V1. Torborg et al. [2004] report that retinal waves drive the establishment of orderly connections from the retina to the LGN. Chemically abolishing retinal waves results in severe developmental impairment of cortical ocular dominance columns (Wiesel and Hubel [1963], Horton and Hocking [1996]) and orientation selectivity (Chapman and Stryker [1993], Chapman et al. [1996], Weliky and Katz [1997]). These and other studies indicate that retinal waves indeed seem to be a necessary condition for the emergence of many important properties of the early visual system (Thompson [1994, 1997]).

As described above, the underlying mechanisms that give rise to retinal waves have been studied quite thoroughly. In result, many of the functionally relevant components and neuronal circuits have been identified. Also, a number of computational models

have been proposed in order to better understand the emergence of retinal waves (Feller et al. [1997], Nenadic et al. [2003], Godfrey and Swindale [2007], Hennig et al. [2009]). A representative of these models, the one by Godfrey and Swindale [2007], is explained in greater depth in section 2.2.

Please note, that I have only given a simplified textbook account of the presented topics. For example, the early visual system was introduced as a purely feed-forward processing stream, where input images from the retina get passed on via the LGN to V1 and then to higher processing areas. What was not mentioned, for example, is the large portion of feedback connections from higher processing areas back to V1 and to the LGN. I also could not go into the amount of processing that already takes place on the level of the retina or the LGN. See, for example, Olshausen and Field [2005] and Carandini et al. [2005] for comprehensive reviews on these and other matters. Finally, the distinction between simple and complex is not as clear cut as presented above. The introduced dichotomy is based to a large degree on a specific measure for the phase selectivity of a cell, proposed by Skottun et al. [1991]. There is, however, a debate about the validity of the conclusions that there truly are two distinct classes of cells in V1 (Dean and Tolhurst [1983], Mechler [2002]). Thus, the given review of the early visual system was by no means complete. Instead, it was intended to serve as an accessible introduction to the fundamentals of visual neuroscience that are relevant for the context of this thesis.

1.2. Normative Computational Models

In order to better understand the results of probing any kind of real-life system (e.g. the visual system or individual neurons therein) in experiments, it has proven useful to construct a simplified model of the real-life system.

What is a model, or more specifically, a *computational model*? A computational model is a formal, and abstract description of how one beliefs a real-life process functions. First a conceptual idea is formulated about what properties of the system bear relevance to the experimental question. Thereafter, the conceptual model is first expressed in mathematical form and then implemented on a computer to simulate the processes that are assumed to take place in the real-life system. A good model should be able to reproduce the experimental results, i.e. it should deliver the same output as the real-life system when presented with the same input. Furthermore, the model should allow for predictions about how the real-life system will respond to input that it was not yet presented with or how it will behave in a different experimental paradigm. Finally, in order to be informative the model should be simple. With a complicated enough structure,

any model can reproduce any data and would therefore not be very instructive. This principle is known as "Occam's razor". In essence, a good model should be simple, yet have sufficient descriptive as well as predictive power.

An example for a model is the model of retinal waves proposed in Godfrey and Swindale [2007]. The equations of the retinal wave model describe the rate of change of the membrane potential and the change of the firing threshold of retinal amacrine cells as a function of various variables, including the activity of neighboring amacrine cells. Numerically integrating these differential equations on a computer (i.e. "running the model") yields time courses of amacrine cell activity which when properly visualized closely resemble retinal waves as described above. The retinal wave model is a classical example of a descriptive and mechanistic model. It makes reference to relevant biological components of the corresponding real-life system (the retina) and explicitly describes their interactions that lead to the emergence of experimentally observed properties (the activity waves).

There is, however, a different kind of modeling approach that has proven rather successful in visual neuroscience. Instead of focusing on the question *how* certain properties of the real-life system emerge, this approach specifically addresses the question *why* the system has the observed properties. Horace Barlow has motivated this approach in a very intuitive manner in his article from 1961 entitled "Possible Principles Underlying the Transformations of Sensory Messages" (Barlow [1961]). He writes

> A wing would be a most mystifying structure if one did not know that birds flew. One might observe that it could be extended a considerable distance, that it had a smooth covering of feathers with conspicuous markings, that it was operated by powerful muscles, and that strength and lightness were prominent features of this construction. These are important facts, but by themselves they do not tell us that birds fly. Yet without knowing this, and without understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding.

The idea behind this approach is to formulate a hypothesis about the problem that the system might be solving or the goal it is trying to reach, for example to fly. This goal is then translated into a mathematical function, which is called the *objective function*. The parameters of the objective function may or may not have real-life counterparts, as these models don't necessarily have to make reference to biological components. In the final modeling step, the parameters are tuned to optimize the objective function using standard optimization tools. After the optimal parameters have been found, they

describe how the real-life system *should* behave if it were indeed optimized with respect to the hypothesized objective. Hence the name *normative* models.

To make the concept more plastic, I now present two groups of normative models, which are both highly relevant to the topic of my thesis.

1.2.1. Sparse Coding, Information Maximization, and ICA

In 1996, Bruno Olshausen and David Field published a seminal paper entitled "Emergence of simple-cell receptive field properties by learning a sparse code for natural images" (Olshausen and Field [1996]). The model therein is a normative model of simple cells, as found in V1.

Olshausen and Field employed the generally accepted assumption, that the computation performed by simple cells can be described as the dot product between an input image and the receptive field image (RF) of the simple cell. Upon doing so, the authors chose the pixel values of the receptive field images to be the parameters of their objective function. The objective function itself was a combination of two aspects. The first aspect was how well the input image could be described by the population of the model simple cells (i.e. the reconstruction error after projection onto the RF image basis). This is important and intuitive, because, after all, the cells in our visual cortex should be able to form a representation of the images that enter our visual system. The second aspect of the objective function was the idea that the activity of the individual cells in the population should be as statistically independent as possible when exposed to natural images. However, instead of enforcing statistical independence directly, a different aspect of the activity of cells was chosen to be maximized, namely *sparseness*. A sparse coding of an image means that it is represented with only few active cells (or RF basis function). The reasoning behind the idea of statistically independent cells, is the conjecture that natural images are the result of a mixture of a number of independent causes (e.g. objects, animals, people). Also, a pixel by pixel representation is not very efficient, because neighboring pixels exhibit large correlations in natural images (Simoncelli and Olshausen [2001]). Having a representation with statistically independent activities requires a reduction of such redundancies, and would thus make the representation more efficient. A sparse representation is also desirable with respect to metabolic considerations. The production of action potentials consumes a considerable amount of energy because next to voltage gated ion channels, there are channels that involve second messenger cascades, which in turn rely on endergonic chemical processes. Therefore, a sparse code is more energy efficient than a non-sparse code.

After having optimized the objective function by adjusting the receptive field images of the model simple cells, it turned out, that the obtained model receptive fields were remarkably similar to those found for actual simple cells in cat and monkey primary visual cortex. The obtained receptive fields resembled localized, oriented filters of various spatial frequencies, similar to Gabor functions. This finding led to the conclusion that the specific shape of V1 simple cell receptive fields may have been the result of evolutionary processes that have optimized the early visual system to form a faithful, yet efficient representation of natural images.

Bell and Sejnowski [1997] reported similar results using a different objective function. Their optimization was based on information maximization (Bell and Sejnowski [1995]) and also led to Gabor-like receptive fields and maximally independent filter outputs (i.e. simple cell activities).

Both objectives, sparse coding as well as information maximization, are examples of a general class of algorithms that is called *independent component analysis* (ICA). ICA algorithms are designed to disentangle (possibly non-linear) mixtures of independent signals. Their successful application to model the emergence of simple cells supports the notion, that the early visual system might have evolved to solve a similar task.

There are, however, other normative models that have also shed light on functions that the early visual system may be optimized to perform. One such other class of normative models is the topic of the next section.

1.2.2. Temporal Stability, Slowness, and SFA

The class of models presented in this section is also based on the assumption that cortical neurons are (near-) optimally adapted to the statistics of their natural input (i.e. to the regularities present therein). However, rather than optimizing the statistical independence of a sensory representation, its *temporal stability*, or *slowness*, is to be maximized.

The slowness approach is mainly motivated by the observation that behaviorally relevant aspects of the world tend to change their properties on a slower time scale than their representation in the earliest sensory organs does. In other words, changes that occur in a visual scene containing objects, animals, or people generally happen on a slower time scale than variations of the low-level features such luminance values in small areas of the visual field (e.g. what would be detected by a retinal ganglion cell).

Application of the slowness objective in several flavors has been successfully applied to model the emergence of V1 complex cells in a number of studies. Földiák [1991] has used a modified Hebbian learning rule (trace rule) together with moving bar stimuli to

train an artificial neural network. The network learned invariance to shifts of the bar images. Körding, Kayser, Einhäuser, and König used more natural stimuli in their study (Einhauser et al. [2002], Kording et al. [2004]). They mounted a camera onto a cat's head and thereby recorded the visual input from the cat's perspective. A three layered neural network was trained with the captured "cat-cam" movies, resulting in units in the top layer that were not only translation invariant but also showed selectivity for spatial frequency and receptive field aspect ratio comparable to physiological findings. Finally, Berkes and Wiskott [2005] presented a model for complex cells which was able to reproduce a rich set of physiological properties found in cortical complex cells, including direction selectivity, orientation as well as frequency tuning, end-inhibition, and side-inhibition.

Next to modeling complex cells found in visual cortex, the slowness hypothesis has been successfully applied to model other types of cells found in the brain as well. Franzius et al. [2007] have reported a model for the self-organized emergence of so-called place cells, head-direction cells, and spatial-view cells. All of these cell types are found in the hippocampal formation of the rodent brain. Their model consisted of a hierarchical neural network of slow feature analysis units (see below) and it was trained with simulated, quasi-natural image sequences. Wyss et al. [2006] also reported a model for hippocampal place cells based on temporal slowness.

The mentioned studies are all based on slowness, yet use different means to implement and optimize the objective function. Wiskott and Sejnowski [2002] have proposed an algorithm that is able to extract slowly varying components from high dimensional input signals. This algorithm is called Slow Feature Analysis (SFA) and will be described in a more detailed fashion in section 2.1.

1.3. Thesis Hypothesis

My thesis work is mainly based on two conceptions about the early visual system, which I have introduced in the previous sections. Firstly, there is the idea that neurons in primary visual cortex have evolved to be optimally adapted to the regularities present in their natural inputs (Attneave [1957], Barlow [1961], Field [1994], section 1.2 and references therein). Secondly, the idea that retinal waves serve as a suitable training stimulus for the development of the visual system prior to the onset of vision (see Albert et al. [2008], Sprekeler and Wiskott [2010] and section 1.1.2).

In their control experiments, Berkes and Wiskott [2005] were able to obtain SFA units that share relevant features with complex cells even if the units were trained with image

sequences derived from colored noise images with a $\frac{1}{f^2}$ power spectrum instead of natural image sequences. These results suggest that spatial second-order statistics are sufficient to learn SFA units that resemble complex cells. However, some properties of the obtained SFA units, such as the shape of their optimal stimuli, varied strongly with respect to the spatial transformations that were applied in order to generate the image sequences. This empirical finding was supported by analytical considerations reported in Sprekeler and Wiskott [2010]. In fact, if one assumes that the spatial statistics of the input data do not vary with respect to the transformations that are applied to generate them, it turns out that the equations that determine the solution to the SFA optimization problem are independent of the input statistics. Hence, the observed *independence* of the SFA unit properties on higher order spatial correlations and the observed *dependence* on the transformation that were used to generate the image sequences. These results led to the conjecture that complex cell properties can be learned from retinal waves, because the moving waves could resemble a prenatal analog of translation that is present in natural image sequences.

Albert et al. [2008] have proposed that there exists an innate learning strategy which structures the visual system with the same objective before and after the onset vision. Prior to visual experience the mechanism acts on internally generated input (retinal waves), whereas after eye-opening the connectivity patters are refined by training with natural visual input. In their simulations, Albert et al. have used sparse coding as a learning objective and obtained simple cell receptive fields, similar to those derived from natural images by Olshausen and Field [1996] and Bell and Sejnowski [1997].

Thesis Hypothesis Applying the objective of temporal slowness to retinal waves leads to similar results as have been obtained when applying it to (quasi-)natural input sequences (Einhauser et al. [2002], Berkes and Wiskott [2005], Sprekeler and Wiskott [2010]). Hence, the slowness approach (Földiák [1991]), as manifested by slow feature analysis (Wiskott and Sejnowski [2002]), is compatible with the retinal-wave-based, innate learning mechanism proposed by Albert et al. and others (Albert et al. [2008], Wong [1999]).

Why is this question an important one? The question of why the early visual system is structured the way it is has not yet been answered unequivocally. Theoretical arguments suggest that V1 neurons have evolved to optimally represent natural input. However, the question "Optimal with respect to what?" is still unanswered as there are several possible candidates[5]. Showing that temporal slowness can also yield meaningful results

[5]Of course, the individual objectives not necessarily exclude each other.

when applied to retinal waves would supply further support to the argument that the brain indeed employs it as coding strategy (possibly amongst others).

Furthermore, the validity of my hypothesis is not obvious. As the slowness approach, and SFA in particular, rely on the temporal structure for learning a sensory representation, the claim would imply that retinal waves share relevant features in their spatiotemporal structure with that of natural visual input. While the spatiotemporal statistics of natural movies have been analyzed to some extent (Dong and Atick [1995a,b]), to the best of my knowledge, similar analyses have not yet been done for retinal waves at all.

All my work was done in simulation and analytically. Firstly, a biological plausible model of retinal waves (Godfrey and Swindale [2007]) was implemented, which generates image sequences of a simulated patch of retina. Along with this model, various other training stimuli were implemented for testing purposes. Then, a network of slow feature analysis nodes was constructed, similar to that of Berkes and Wiskott [2005]. The network was then trained with the retinal wave image sequences. Subsequently, the obtained model neurons were analyzed in a similar fashion as is done in physiological experiments. Finally, the results are interpreted and compared to experimental findings.

2. Methods

This section describes slow feature analysis (SFA) introduced by Wiskott and Sejnowski [2002]. SFA is an algorithm that implements the slowness principle that was motivated in the introduction. Furthermore, the training stimuli for SFA and the way they are generated is explained.

2.1. Slow Feature Analysis

The goal of SFA is to find instantaneous input-output functions that extract slowly varying scalar output signals from a high-dimensional input signal. To ensure that the extracted output signals are informative, they are required to be uncorrelated and to have unit variance. The learning objective can be mathematically formalized as follows:

Optimization problem: Given a multidimensional input signal $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$, $t \in [t_0, t_1]$, find a set of real-valued functions $g_1(\mathbf{x}), \ldots, g_K(\mathbf{x})$ from a function space F, such that for the output signals $y_j(t) := g_j(\mathbf{x}(t))$ the expression

$$\Delta(y_j) \;:=\; \left\langle \dot{y}_j^2 \right\rangle_t \text{ is minimal} \tag{2.1}$$

under the constraints

$$\begin{aligned} \langle y_j \rangle_t &= 0 \text{ (zero mean)}, && (2.2)\\ \left\langle y_j^2 \right\rangle_t &= 1 \text{ (unit variance)}, && (2.3)\\ \forall i < j, \langle y_i y_j \rangle_t &= 0 \text{ (decorrelation and order)}, && (2.4) \end{aligned}$$

with $\langle \cdot \rangle_t$ and $\dot{y}$ indicating time-averaging and the time derivative of y, respectively.

The expression to be minimized (2.1) is a measure of the temporal slowness of the signal $y_j(t)$, with small Δ-values indicating a slowly varying signal. The trivial solution is a set of functions that is constant for all t. Constraints (2.2) and (2.3) avoid this trivial solution and constraint (2.4) ensures that different functions code for different aspects of

the input signal.

The following subsections outline how the SFA algorithm finds the set of solutions that optimize the objective function (i.e. the set of functions $g_j(\mathbf{x}(t))$ from the function space F).

Linear Function Space

Consider the case in which the function space F is the space of all linear functions, i.e. all functions of the form $g_j(\mathbf{x}) = \mathbf{w}_j^T\mathbf{x} = \sum_i w_{j,i}x_i$. In the first step of the algorithm the training data is "sphered". Sphering means that the data is first centered in the coordinate system by subtracting the mean and then whitened by linearly transforming the centered data to have a unit covariance matrix, i.e. $\left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t = \mathbf{I}$, where $\mathbf{z}$ denotes the sphered data. Whitening can be achieved using principal component analysis (PCA). After such preprocessing of the training data, weight vectors $\mathbf{w}_j$ have to be found such that

$$\Delta(y_j) \;:=\; \left\langle \dot{y}_j^2 \right\rangle_t = \mathbf{w}_j^T \left\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \right\rangle_t \mathbf{w}_j \tag{2.5}$$

is minimal. The constraints of the optimization problem take on the following form:

$$\langle y_j \rangle_t \;=\; \mathbf{w}_j^T \underbrace{\langle \mathbf{z} \rangle_t}_{=0} = 0, \tag{2.6}$$

$$\left\langle y_j^2 \right\rangle_t \;=\; \mathbf{w}_j^T \underbrace{\left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t}_{=\mathbf{I}} \mathbf{w}_j = \mathbf{w}_j^T\mathbf{w}_j = 1, \tag{2.7}$$

$$\forall i < j, \langle y_i y_j \rangle_t \;=\; \mathbf{w}_i^T \underbrace{\left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t}_{=\mathbf{I}} \mathbf{w}_j = \mathbf{w}_i^T\mathbf{w}_j = 0. \tag{2.8}$$

It follows that after the mentioned preprocessing of the training data the constraints are fulfilled if and only if the weight vectors $\mathbf{w}_j$ form an orthonormal set of vectors.

The set of solutions to the optimization problem posed in equation (2.5) is given by the set of eigenvectors of the matrix $\left\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \right\rangle_t$, which is the second moment matrix of the time derivative of the sphered training data (Wiskott and Sejnowski [2002]). The resulting eigenvectors are ordered according to slowness by ordering them according to the corresponding eigenvalues, starting with the smallest. Thus, the eigenvector that corresponds to the smallest eigenvalue of $\left\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \right\rangle_t$ constitutes the slowest output signal.

Note that before the weight vectors $\mathbf{w}_j$ are applied to new data $\tilde{\mathbf{x}}$, this data has to be mapped into the coordinate system of $\mathbf{z}$. This is done by applying the same affine transformation that was applied to the training data $\mathbf{x}$.

In summary, if the function space F is confined to linear functions only, the solution to the optimization problem posed in (2.5) can be found efficiently by solving a simple eigenvalue problem.

Non-linear Function Space

In some applications it can be useful to consider a non-linear function space F within to search for the solutions to the SFA problem.

If F has a finite number of basis functions $h_1, \ldots, h_M$, then every function $g \in \mathrm{F}$ can be expressed as a linear combination of these basis functions. For example, the space of all polynomials of degree n is spanned by all monomials up to the order of n. We can now map the input $\mathbf{x}$ into the non-linear function space via

$$\mathbf{h}(\mathbf{x}) := (h_1(\mathbf{x}), \ldots, h_M(\mathbf{x}))^T. \tag{2.9}$$

We define $\mathbf{h}$ to be the *expanded input* and thereby generate every function $g \in \mathrm{F}$ as

$$g(\mathbf{x}) = \sum_{k=1}^{M} w_k h_k(\mathbf{x}) = \mathbf{w}^T \mathbf{h}(\mathbf{x}). \tag{2.10}$$

This leaves us with a problem formulation in terms of weights belonging to the basis functions of the specific function space F. Thereby, the problem is cast into a linear form and can be solved as outlined above, by simply replacing $\mathbf{x}$ by $\mathbf{h}(\mathbf{x})$.

SFA Algorithm

Given the considerations presented in the previous section, the SFA algorithm can be summarized in the following steps:

Nonlinear expansion: Map the input data into the selected function space F.

Sphering: Subtract the mean and transform the expanded data to have a covariance matrix which is the unit matrix $\mathbf{I}$.

Slow feature extraction: Compute the second moment matrix of the temporal derivative of the sphered training data and solve the eigen-decomposition using standard linear algebra tools.

The result of SFA is a set of weight vectors $\mathbf{w}_j$ that constitute the slowly varying output components $g(\mathbf{x})_j = \mathbf{w}_j^T \mathbf{z}$, with $\mathbf{z}$ denoting the expanded and sphered input signal. These components are also referred to as *SFA units*.

In the present application, the function space F was chosen to be the space of all polynomials up to degree two. Therefore, SFA yields weight vectors which linearly combine all the monomials up to degree two of all input dimensions. In this case, it is possible to express the solution in terms of a *quadratic form*:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{z} = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{f}^T\mathbf{x} + c \tag{2.11}$$

The coefficients of $\mathbf{H}$ and $\mathbf{f}$ are determined by $\mathbf{w}$. Expressing $g(\mathbf{x})$ in terms of a quadratic form in the input space rather than a linear function in the expanded space can be instructive when interpreting its response properties to input stimuli.

2.2. Input Data

In order to train the SFA units, image sequences were used that were derived from a biologically plausible model of retinal waves proposed in Godfrey and Swindale [2007]. The model assumes that spontaneous activity of retinal amacrine cells drives the wave activity. The wave initiation and propagation mechanism is based on the assumptions that (1) amacrine cells are spontaneously active, (2) have local excitatory connections, and (3) that the rate with which spontaneous activity occurs is inversely related the magnitude of excitatory input during depolarization, i.e. the larger the input was that caused an amacrine cell to become active, the longer it will take for the next spontaneous activation to occur.

In the model, the activity of a cell is governed by two variables: one variable representing the cell's membrane potential and one variable representing the firing threshold. Once the potential reaches the threshold, the cell becomes active for a fixed amount of time. The membrane potential is modeled as a leaky integrator. In the presence of input, the membrane potential approaches the level of input (thereby possibly exceeding the activation threshold) and in the absence of input it will decay to zero. The evolution of the threshold variable on the other hand depends only partially on the amount of excitatory input. In the absence of input, the threshold decays linearly to zero with a fixed rate. Thereby, the cell can become active spontaneously. Once the cell becomes active, the threshold rises linearly with a fixed rate as long as the cell stays active. If there is input to the cell while it is active, the increase in threshold is facilitated, leading to a much higher threshold after activation. Hence, once a cell has been part of an activity wave (i.e. its neighbors were active congruently) it is harder to activate. The result of this activity dependent refractoriness is that waves of activation remain confined to certain

domains and cannot recruit the entire retina. The borders of these domains, however, are changing as the system evolves in time and thereby produce non-repeating and random (yet spatially coherent) activity patterns.

In the simulations, the amacrine cells were arranged in a regular grid consisting of 128 by 128 cells. The activity of each cell was represented by an either black or white pixel in a correspondingly sized image. Each cell received input from other cells that were within a six cell radius, where the synaptic weights between cells were inversely proportional to the distance of cell's respective positions. In order to avoid inhomogeneities at the borders of the simulated patch of retina[1], circular boundary conditions on the connectivity between cells were imposed. This means, that cells at the image borders also received input from cells at the opposite image border. The retinal wave model was parameterized such that it produced waves that are similar in size and velocity to those observed in mice during the first two weeks after birth. The parameters were taken from Godfrey and Swindale [2007]. However, the parameter that controls the frequency of spontaneous depolarizations was changed in order to have waves occur more often. With the original parameter value, a very large portion of the simulated retinal images contained no activity at all and where therefore useless for the SFA algorithm.

All simulations were started with random membrane potentials and threshold levels for each cell. After an initial warm-up phase of 30 minutes simulated time, the produced retinal activity patterns were recorded for another 30 min simulated time, which resulted in 18000 images (10 per second). The temporal dynamics of the simulated waves is a property of the wave model and should therefore reflect the dynamics of real retinal waves. Figure 2.1 A shows four example frames of the full simulated retina. In order to illustrate the dynamics of the waves, the images in the example sequence are each 20 simulated time steps apart, which corresponds to 2 seconds. So the four depicted images show 6 seconds of simulated activity. The average amplitude spectrum of the retinal wave images is shown in figure 2.1 B. The spectrum shows the same amplitude fall-off in all directions, indicating that on average all orientations are equally present in the training images. Taking the mean over orientations and plotting the result in log-log coordinates yields the amplitude spectrum plot in figure 2.1 C. See the discussion section of this thesis for further remarks regarding the properties of the Fourier spectrum of retinal waves.

The obtained image sequence was then tiled into overlapping receptive fields of size 16 by 16 pixels, with an overlap of 5 pixels. This resulted in 289 image sequences, each being 18000 images long and having a dimension of $16 \cdot 16 = 256$. The dimensionality

[1]Inhomogeneity effects would otherwise occur, because the cells at the image borders otherwise receive less input than those that are further away from border.

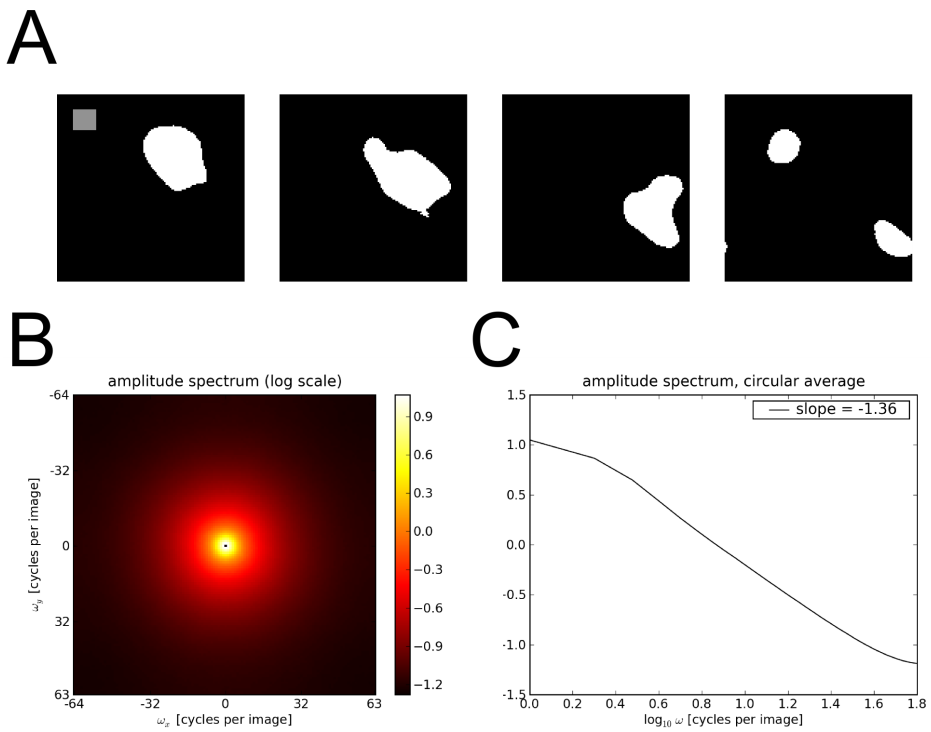

Figure 2.1.: Simulated retinal waves. **A** Four example frames of simulated amacrine cell activity with a time distance of 20 frames, corresponding to 2 seconds simulated time. Receptive field size is illustrated by the gray square in the top left corner of the first frame. **B** Average amplitude spectrum of all frames of this simulation run. **C** Same spectrum as in **B**, averaged over orientations and plotted in log-log coordinates.

was reduced to 50 by applying principal component analysis (PCA) to all receptive field image sequences. Dimensionality reduction was a necessary step because the number of dimensions M after quadratic expansion of an N dimensional input signal is given by $M = N + (N^2 - N)/2 + N$, summing the purely quadratic monomials (x_i^2), the mixed monomials ($x_i x_j$), and the linear monomials (x_i), respectively. For 256 input dimensions, the number of entries of the second moment matrix of the quadratically expanded signal is in the order of 10^9, while for 50 input dimensions the number of entries to compute is in the order of 10^6. Yet, even after PCA, the number of parameters to estimate from the data is still very high and requires a correspondingly large amount of data. For this reason, the receptive field image sequences were concatenated to yield a single image sequence of length $289 \cdot 18000 = 5202000$ images. This is a permissible course of action, because the statistical properties of the image sequences of each individual receptive field are expected to be equal, which is due to the uniform distribution of wave initiation points over the simulated retina and the circular boundary conditions on the connectivity. This single retinal wave image sequence then served as a training input to SFA.

Other models exist that also produce retinal waves. Feller et al. [1997] presented a two-layer model (amacrine and ganglion cell layers), which reproduced observed spatiotemporal patterns of retinal waves. In this model also, spontaneous activity and horizontal coupling in the amacrine cell layer are essential to the emergence of spatially coherent activity patterns. Recently, Hennig et al. [2009] proposed a theoretical account for early-stage retinal waves that is based on slow after-hyperpolarization currents occurring after depolarization, i.e. after an amacrine cell became active. The strength of these currents depends on the intensity of the depolarizing input. This mechanism is very similar in spirit to the activity depend refractoriness contained in the model by Godfrey and Swindale (Godfrey and Swindale [2007]). Other retinal wave models have been proposed in Nenadic et al. [2003], Butts et al. [1999].

Here, the Godfrey and Swindale model was used, because it is straight forward to implement and runs sufficiently fast. Furthermore, the authors provided parameter settings to simulate waves of different species, which would allow for a comparison of SFA training results. However, this option has not yet been explored.

3. Results

After training with the retinal wave image sequence, the first 50 SFA units were analyzed. Since the SFA output signals are ordered with respect to slowness, those units represent the 50 most slowly varying components that can be extracted from the training input, given the restrictions of the chosen function space. The trained SFA units are characterized by showing those input images to which the units respond with the largest output and by visualizing the units' responses to sinusoidal gratings.

3.1. Receptive Fields

In physiological experiments it is common practice to characterize visually responsive neurons via their spatial receptive fields. In the introduction (section 1.1) I have mentioned, that the term receptive field can be somewhat ambiguous. It is used to refer to the location in visual space to which a neuron is responsive if a stimulus is presented there. At the same time, the term receptive field is used to describe the stimulus that makes the neuron respond most strongly. In order to be consistent with the neuroscience literature, the latter convention is adopted here: Optimal stimuli of cortical cells and SFA units are referred to as their receptive fields.

Once the structure of a cortical cell's receptive field has been estimated in neurophysiological experiments, it allows inferences about the preferred orientation, frequency, and (in case of simple cells) the preferred phase of the cell. Simple cell receptive fields can be mapped by computing the spike-triggered average of random dot input stimuli and are well described by 2D Gabor functions (Daugman [1985], Jones and Palmer [1987b,a]). Complex cells, on the other hand, require more elaborate schemes for receptive field mapping, due to their largely non-linear input-output relation. Using methods such as spike-triggered covariance or second order interaction maps has revealed many insights about the spatial structure of complex cell receptive fields (Touryan et al. [2005], Sasaki and Ohzawa [2007], Livingstone and Conway [2003]). For example, just like the ones of simple cells, the receptive fields of complex cells also posses subregions with opposite polarity (similar to ON and OFF regions), from which the frequency tuning of the cell

can be predicted.

Since the trained SFA units are quadratic forms of the pixel intensities, there is an explicit formulation of the input-output relation. Given a fixed-norm constraint on the input images, it is therefore possible to compute the optimal excitatory and inhibitory stimuli (Berkes and Wiskott [2007]). The polarity of the weight vector $\mathbf{w}$ that constitutes an SFA unit is arbitrary, and therefore the sign of the output signal is also arbitrary. Here, the the units were sign-corrected such that the stimulus that causes the largest-magnitude response, yields a positive response, i.e. is the maximally excitatory stimulus. Figure 3.1 A shows the maximally excitatory images for the first 25 SFA units (i.e. the 25 slowest), whereas figure 3.1 B shows the maximally inhibitory images for the same units. Most of the optimal stimuli (excitatory as well as inhibitory) show spatially segregated and elongated ON and OFF regions, which is in close correspondence with experimental data. See figure 3.1 C for a comparison with receptive field structures obtained from adult cats reported in Touryan et al. [2005].

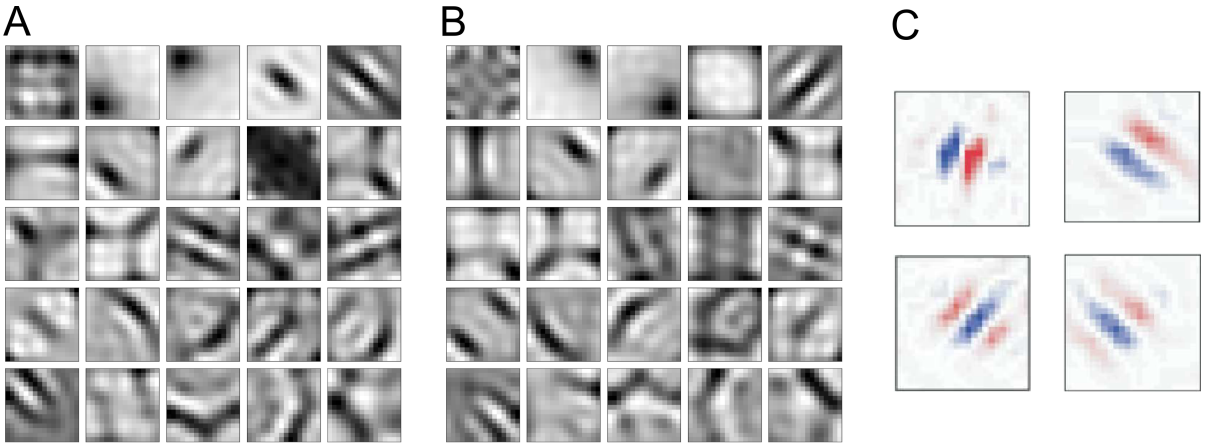

Figure 3.1.: Optimal stimuli. **A** Maximally excitatory stimuli, plotted for the first 25 SFA units. **B** Maximally inhibitory stimuli for the same selection of units as in **A**. **C** For comparison, receptive field structures of complex cells of cats estimated by Touryan et al. [2005].

3.2. Sinusoidal Test Stimuli

Further response properties of the SFA units are visualized by showing their responses to sinusoidal gratings. When V1 cells are probed with sinusoidal gratings in neurophysiological experiments, the used gratings are usually parameterized along three dimensions: orientation, spatial frequency, and phase. In order to compute the response of the SFA

units to gratings, the orientation and phase were confined to the range from 0 to 2π, whereas the maximal possible spatial frequency was eight cycles per receptive field, due to the 16 by 16 pixel size of the receptive field.

For better comparison of the response properties between individual SFA units, the outputs of each unit were normalized in the following manner: First of all, the response of the unit to a gray input image was subtracted. Secondly, the entire output was sign corrected such that the maximal positive output is larger in magnitude than the maximal negative output. Finally, the output was normalized to have unit variance over the explored range of input stimulus parameters.

Figure 3.2 depicts the responses of the first 25 SFA unit as a function of orientation and phase of the input sinusoidal grating. I refer to the plots as *orientation/phase response functions*. The spatial frequency of the gratings was set to the value that maximizes the response of the respective SFA unit. Most of the orientation/phase response functions exhibit horizontal stripe patterns, indicating that the response varies stronger along the orientation axis compared to the phase axis. This rather qualitative observation was quantified using the *response modulation index* (Skottun et al. [1991]) (also referred to as F1/F0 ratio) and the *orientation selectivity index* (OSI) (Chapman and Stryker [1993]), which measure the response variation with respect to phase and orientation, respectively. The F1/F0 ratio is a spectral measure of the phase response curve[1]. Specifically, it is the ratio of the first harmonic of this curve to its DC component, hence the name F1/F0 ratio. Cortical cells with F1/F0 ratio smaller than one are classified as complex cells, whereas cells with a ratio larger than one are classified as simple cells. The OSI is a spectral measure of the orientation response curve[2], and is given by $\frac{F2}{(F0+F2)} * 100$, where F2 is the amplitude of the second harmonic and F0 is the DC component of the orientation response curve. Chapman and Stryker [1993] have reported average OSI values of approximately 40 in adult ferrets and approximately 48 in adult cats.

Figure 3.2 B and C show the respective histograms of the these two measures, computed for the 50 SFA units of this particular simulation run. The majority of SFA units have a response modulation index smaller than one. Hence, they would be classified as complex cells in a physiological experiment. The distribution of orientation selectivity values shows a wide spread over the possible range of values, which is consistent with experimental findings (Chapman and Stryker [1993]). However, the comparatively high population mean (68.7) indicates a rather specific orientation tuning in the majority of the SFA

[1]The phase response curve corresponds to a horizontal cut through the orientation/phase response function at the orientation that maximizes the cell's output.

[2]The orientation response curve corresponds to a vertical cut through the orientation/phase response function at the phase that maximizes the response

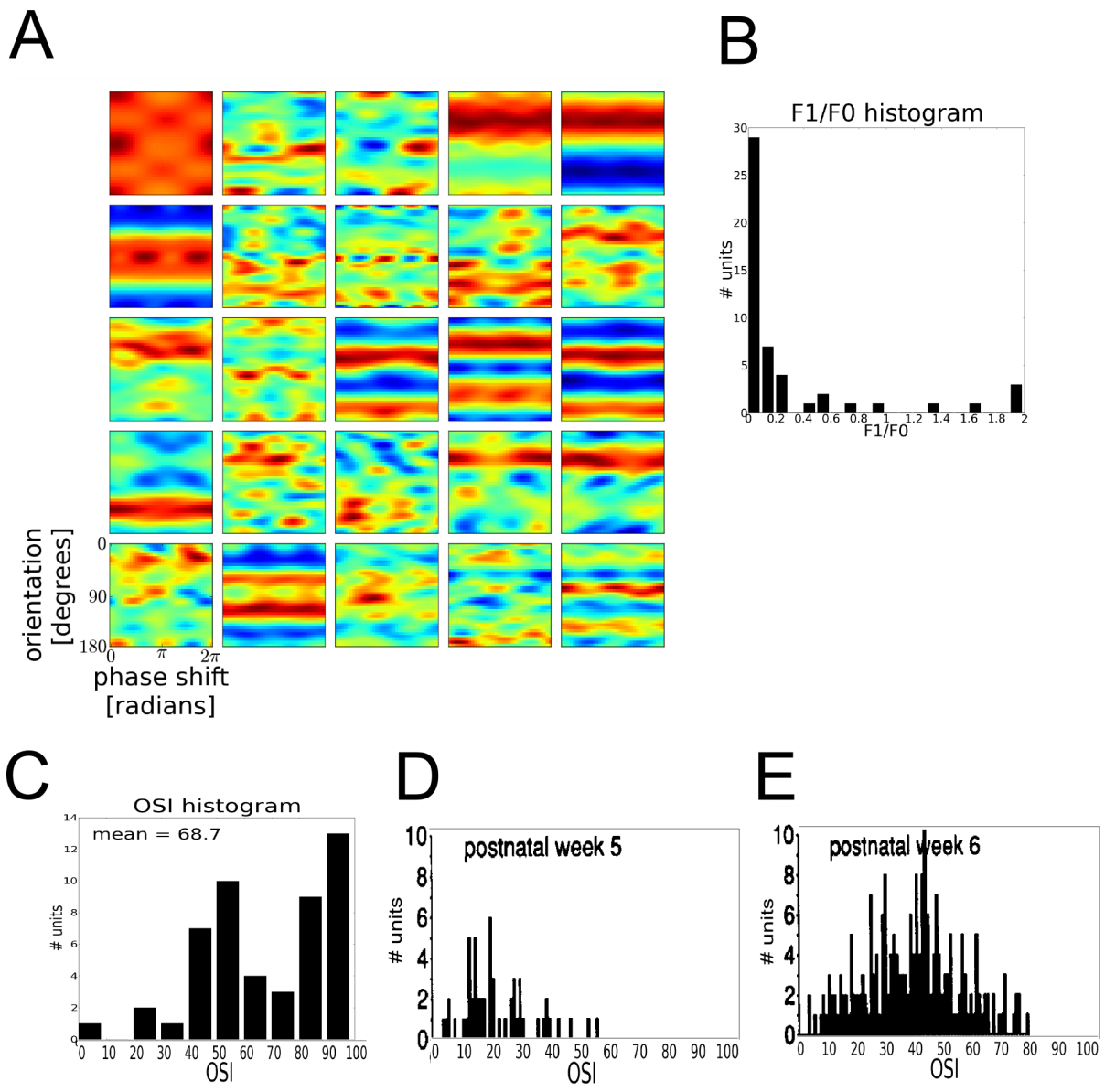

Figure 3.2.: **A** Response of the SFA units to sinusoidal gratings depicted as a function of orientation (y-axis) and phase (x-axis) of the grating. The spatial frequency was set to the unit's preferred value. **B** Histogram of response modulation (F1/F0) values, indicating the susceptibility to the phase parameter of the input grating. A large portion of units from this simulation run (45 out of 50) have an F1/F0 value smaller than one and would thus be classified as complex-cells in a physiological experiment. **C** Histogram of orientation selectivity (OSI) values, indicating the susceptibility to the orientation parameter of the input grating. The OSI values are rather high compared to experimental findings, displayed in **D** and **E**. These plots show OSI histograms obtained from ferrets (Chapman and Stryker [1993]). Eye opening occurs between postnatal week 5 and 6. See section 4.2 for possible explanations regarding the very high OSI values of the SFA units.

units. Some units even reach OSI values above 90, which is much higher than the values reported for adult cats or ferrets. OSI histograms for ferrets shortly before eye-opening and shortly afterwards are shown in figure 3.2 D and E, respectively. How exactly SFA achieves the phase invariance and the large orientation selectivity is an interesting issue and is considered in more detail in section 4.1.

Figure 3.3 A shows the response of the SFA units as a function of spatial frequency of the input gratings (averaged over phase). These plots can also be interpreted as the response of the units in the frequency plane. Ringach et al. [2002] have investigated the response of V1 cells in macaque monkeys in the same manner. Their results are shown in figure 3.3 B for comparison. Similar to Ringach et al.'s findings, almost all SFA units exhibit active inhibition to stimuli that are not orientated in the preferred direction. This inhibition takes place for orientations orthogonal to the unit's preferred orientation but in some units also for non-orthogonal directions.

The orthogonal and non-orthogonal suppression is also visible in typical orientation tuning polar plots, depicted in figure 3.3 C for the same SFA units. These plots show the orientation tuning function at the unit's preferred frequency and phase. Unlike traditional plots of this kind, here also the negative response (inhibition) of the SFA units is shown. Excitatory activity is plotted in solid red lines, while inhibition is plotted in dashed blue lines. The majority of SFA units show a clear orientation preference. There are units that prefer only a single orientation as well as units that also respond strongly to a second direction, which is in some cases orthogonal to the first one and in other cases not. This phenomenon manifests itself in the polar plots as so-called secondary response lobes, which are also not untypical for cells found in mammalian V1. Figure 3.3 D shows experimental data obtained by Devalois et al. [1982]. Here only the excitatory response is plotted, but the different types of responses (not tuned for orientation, single orientation preference, secondary response lobes) are well exemplified. The inhibitory response of the SFA units seems to follow similar patterns as the excitatory response. There is inhibition in a single direction only but also secondary inhibition response lobes as well.

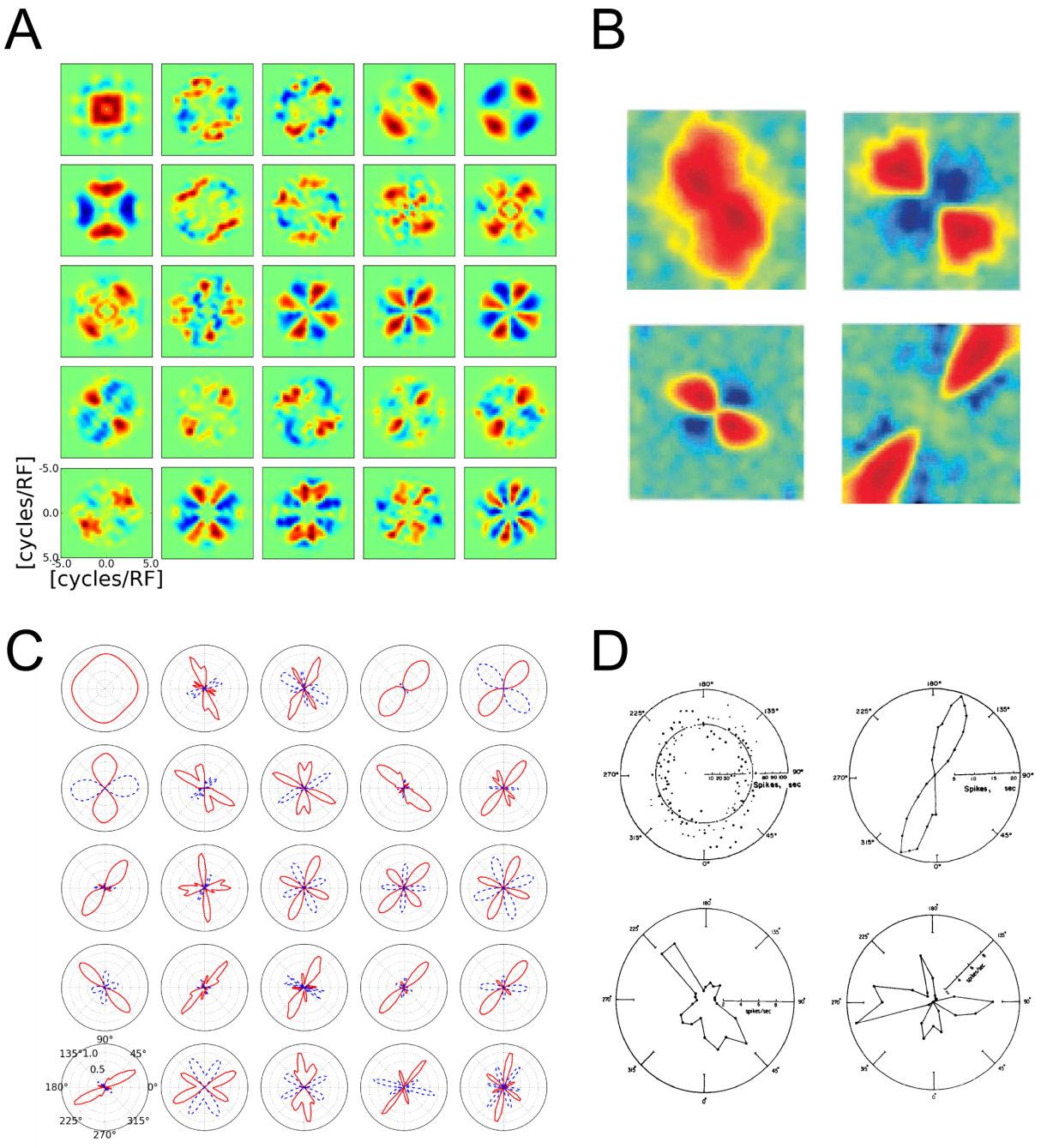

Figure 3.3.: **A** Response of the SFA units to sinusoidal gratings depicted as a function of the gratings' spatial frequency. **B** Same information as in **A**, plotted for V1 neurons in macaque monkeys (Ringach et al. [2002], fig.2). **C** SFA unit response as a function of orientation only. Solid red lines indicate positive response, dashed blue lines indicate negative (inhibition) response. **D** Orientation tuning functions of V1 cells of macaque monkey (Devalois et al. [1982]).

4. Control Experiments

4.1. Phase Invariance

Most of the SFA units show a large invariance in their response with respect to phase (or position) of an input grating. In the classical complex cell model (Adelson and Bergen [1985]) this phase invariance is a direct result of pooling the squared outputs of two linear Gabor filters that have the same preferred frequency and orientation but are in 90° phase shift relative to each other. Such a pair of filters is called a *quadrature filter pair* (QFP). Recall that the SFA units are essentially a quadratic form of the pixel intensities, in which the contribution of the quadratic and the linear terms can be separated (see section 2). The constant term is of no interest in the analysis, because it cannot convey any information about a changing stimulus. If $\mathbf{x}(t)$ is a sinusoidal grating with a phase that changes over time, then the output of the linear term ($\mathbf{f}^T\mathbf{x}(t)$) will be an oscillation which follows the phase shift. In other words, the linear term alone cannot achieve phase invariance.

Consider the quadratic form equation from section 2.1:

$$g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{f}^T\mathbf{x} + c. \tag{4.1}$$

The equation shows how the quadratic term $\mathbf{H}$, the linear term $\mathbf{f}$, and the constant term c contribute to the computation of a unit's response $g(\mathbf{x})$. In figure 4.1, $g(\mathbf{x}) - c$ is shown as a function of phase of the input grating. This plot is called the phase response curve. Spatial frequency and orientation of the input gratings have been set to the units' preferred values. The chosen SFA units include the one that exhibits the largest phase invariance (smallest F1/F0 ratio) and the one that exhibits the smallest phase invariance (largest F1/F0 ratio), top row and bottom row respectively. The first column in the figure displays the phase response of the quadratic and the linear term separately, as well as the sum of the two (which corresponds to $g(\mathbf{x}) - c$). For the most phase invariant unit, it can be seen that the phase invariant response is mostly carried by the contribution of the quadratic term. The contribution of the linear term on the other hand is comparatively

marginal in this case. For the second unit (middle row in the same figure) the response of the linear term is larger and the phase dependent fluctuations of the quadratic term seem to compensate those of the linear term. However, for the unit with the lowest phase invariance, both terms show an almost aligned phase dependence and contribute almost equally to the overall output, which leads to a very high F1/F0 ratio for $g(\mathbf{x})$, i.e. small phase invariance.

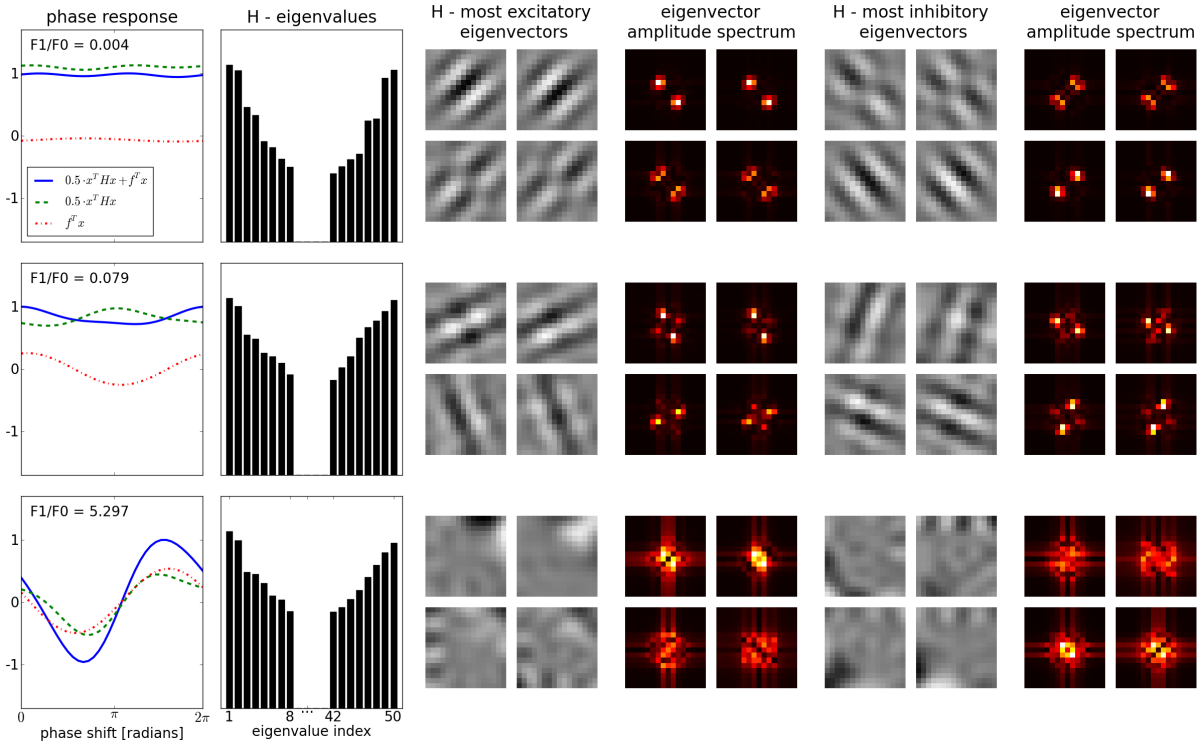

Figure 4.1.: Emergence of phase invariance. The two top rows correspond to the most phase invariant SFA units, the bottom row to a unit whose response is phase dependent. Phase dependence of the linear and the quadratic term as well as their sum is depicted in the first column. The second column shows the two ends of the eigenvalue spectrum of $\mathbf{H}$. Eigenvalues are sorted from highest to lowest, then their absolute values are plotted. Columns three and five show the eigenvectors that belong to the four largest, respectively smallest, eigenvalues. Columns four and six show the corresponding Fourier amplitude spectra.

In order to gain a better understanding of the computation performed by the quadratic term, it is instructing to consider the eigenvalue decomposition of matrix $\mathbf{H}$, which is given by the sum over the outer products of its eigenvectors $\mathbf{v}_i$, weighted by the corresponding eigenvalues λ_i:

$$\mathbf{H} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T. \tag{4.2}$$

In this formulation, the computation of the quadratic term becomes

$$\frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} = \frac{1}{2}\mathbf{x}^T\left(\sum_i \lambda_i \mathbf{v}_i\mathbf{v}_i^T\right)\mathbf{x} = \frac{1}{2}\sum_i \lambda_i\left(\mathbf{x}^T\mathbf{v}_i\right)\left(\mathbf{v}_i^T\mathbf{x}\right) = \frac{1}{2}\sum_i \lambda_i\left(\mathbf{x}^T\mathbf{v}_i\right)^2. \quad (4.3)$$

Hence, the quadratic term can be regarded as a weighted sum of squared filter outputs to an input image, where the weights are given by the eigenvalues λ_i and the corresponding filters are given by the eigenvectors $\mathbf{v}_i$. Figure 4.1 shows the largest-magnitude eigenvalues and the corresponding eigenvectors for the selected SFA units, plotted as images.

It is noteworthy that for the very phase invariant units, the largest-magnitude eigenvectors seem to come in pairs of two. This means, that not only the corresponding eigenvectors would elicit the largest-magnitude response, but also any linear combination of the two with appropriately normalized weights (the squared weights of the linear combination have to sum to one). For those SFA units, interpolating between their two most excitatory eigenvectors is essentially equivalent to a phase shift. They constitute a quadrature filter pair. Hence, the output of the quadratic term remains rather constant when stimulated with a sinusoidal grating with changing phase but constant (preferred) frequency and orientation. As can be seen in the figure, the two eigenvectors corresponding to the second largest pair of eigenvalues also form a quadrature filter pair. Depending on their preferred orientation and spatial frequency, their contribution can either lead to a broadening of orientation tuning (top row SFA unit) or emergence of a second preferred orientation, i.e. secondary response lobes (middle row SFA unit). The same considerations apply for the maximally-inhibiting eigenvalue/eigenvector pairs, although the pairing of eigenvalues is less obvious.

In case of the phase sensitive unit, the paired occurrence of eigenvalue and, more importantly, the sinusoidal structure of the eigenvectors is a lot less pronounced. Hence the larger susceptibility to the phase of the input grating.

The observed emergence of quadrature filter pairs is in line with the analytical derivation of complex cell properties from the slowness principle, presented in Sprekeler and Wiskott [2006, 2010]. Sprekeler and Wiskott predict the formation of quadrature filter pairs to ensure slowly varying output signals on training input that is derived from applying translation to a static input image. At least locally, this seems a reasonable first-order approximation of retinal waves, because on the receptive field level, the waves primarily appear as passing white edges. This is of course not true in those cases where a wave emerges or decays within a receptive field, changes its size and shape, or, for example, the wave itself is smaller than the receptive field. These cases occur frequently and thereby provide an explanation for the fact that not all of the first 50 SFA units can be

characterized as being a weighted superposition of quadrature filter pairs. However, test simulations in which SFA units were trained with purely translation-based input stimuli show that all of the first 50 units indeed have this quadrature filter pair property.

4.2. Orientation Selectivity

The histogram of OSI values in figure 3.2 shows that a large number of SFA units seem to posses an extremely high, even almost maximal, orientation selectivity. OSI values of 90 and higher are, to the best of my knowledge, not reached by actual cortical neurons. Values of up to 80 (and equivalent values in other scales) are possible, yet very rare (Chapman and Stryker [1993], Chapman et al. [1996]).

Mechanistically, the unphysiologically high OSI values can be understood as a direct result of the rather large active inhibition of non-preferred orientations. In some of the units, the inhibitory response to the non-preferred orientations is as large as the excitatory response to preferred orientations. This leads to a very small mean response when the average is taken over all orientations. Recall the definition of the OSI, which is given by $100 * F2/(F0 + F2)$, where $F2$ is the amplitude of the second harmonic and $F0$ is the mean. If $F0$ now becomes small due to the fact that excitation and inhibition cancel each other out in the average, it becomes evident how the unphysiologically high OSI values arise. The reason is the balanced, in some units even symmetrical, strength of excitation and inhibition.

It is not clear how much rotation is present in the retinal wave image sequences that is detectable within a the receptive field size. The presence of rotation is predicted to lead to harmonic oscillations in the orientation tuning curve (Wiskott [2003], Sprekeler and Wiskott [2010]). When visualizing such an orientation tuning in polar plots, the harmonic oscillation is manifested by inhibitory lobes with amplitudes comparable to the excitatory lobes. The presence of secondary (and possibly more) excitatory and inhibitory lobes in the polar plots represent harmonic oscillations of higher frequency. Such orientation tuning patterns are found in some of the obtained SFA units, indicating that there might be enough rotation present in the retinal wave input stimuli to have caused them. This, and the considerations of the previous paragraph can explain the very high OSI values in light of the theoretical foundations of SFA.

Interestingly, complete absence of rotation in the training input, is predicted to lead to a rather erratic orientation tuning (Sprekeler and Wiskott [2010]). In the extreme case of infinitely large receptive fields, the theory predicts that the translation in the input leads to the emergence of infinitely large quadrature filter pairs, which corresponds to infinitely

sharp peaks in Fourier space, known as Dirac δ peaks. No rotation in the training input means there is no incentive for SFA to yield slowly changing output with respect to rotation. Thus, the predicted orientation tuning curve would be a weighted sum of such Dirac δ peaks. However, due to the spatially finite receptive fields, the localization in Fourier space becomes less sharp, leading to a smoothing of the orientation tuning curve. Hence, if trained with stimuli absent of rotation, a smooth, i.e. somewhat slowly varying, orientation tuning curve can still be expected, with the smoothness not being caused by the slowness objective but by finiteness of the receptive field.

In order to get a feeling for the dependence of orientation selectivity on the transformations present in the training stimuli, simulations were run in which SFA units were trained with input sequences that contained either only translation, only rotation, or a combination of both applied to pink-noise images. The type of noise image used, exhibits the same kind of second order correlation structure as found in natural images (Field [1994], Simoncelli and Olshausen [2001]). This control experiment is similar in nature to those performed by Berkes and Wiskott (Berkes and Wiskott [2005]), in which they also applied SFA to image sequences derived from pink-noise image templates. The image sequences were obtained by placing a 16 by 16 pixel window in the center of the image template and then, for each frame, applying a small amount of transformation (translation, rotation or a mixture of both, depending on the condition) to the window and record the image content momentarily enclosed by it. The resulting properties of the trained SFA units are shown in figure 4.2.

When translation is the only transformation present in the training input, then the optimal stimuli resemble Gabor-patches, the units show almost perfect phase-invariance, and the orientation tuning is very sharp, see figure 4.2 A. Figure 4.2 B depicts the resulting properties of SFA units trained with a mixture of rotation and translation of pink noise images. Here translation was still the dominant transformation, which means that the image templates were faster translated than rotated to create the image sequence. The majority of SFA units are just as phase invariant as in the translation-only case. However, there is a difference with respect to their orientation tuning curves, which have become much smoother, in some units even sinusoidal. This reflects the rotation component in the training data, because these SFA units now react with slowly changing output to rotation in the input. In the case of input stimuli containing only rotation, depicted in figure 4.2 C, the units exhibit a strong phase-dependence but their orientation tuning curves seldom fall below 50% activity. As expected, most of these SFA units are characterized by a strong orientation invariance, which leads to a very unspecific orientation tuning.

Qualitative assessment indicates that the results for retinal waves (see figure 3.2 and

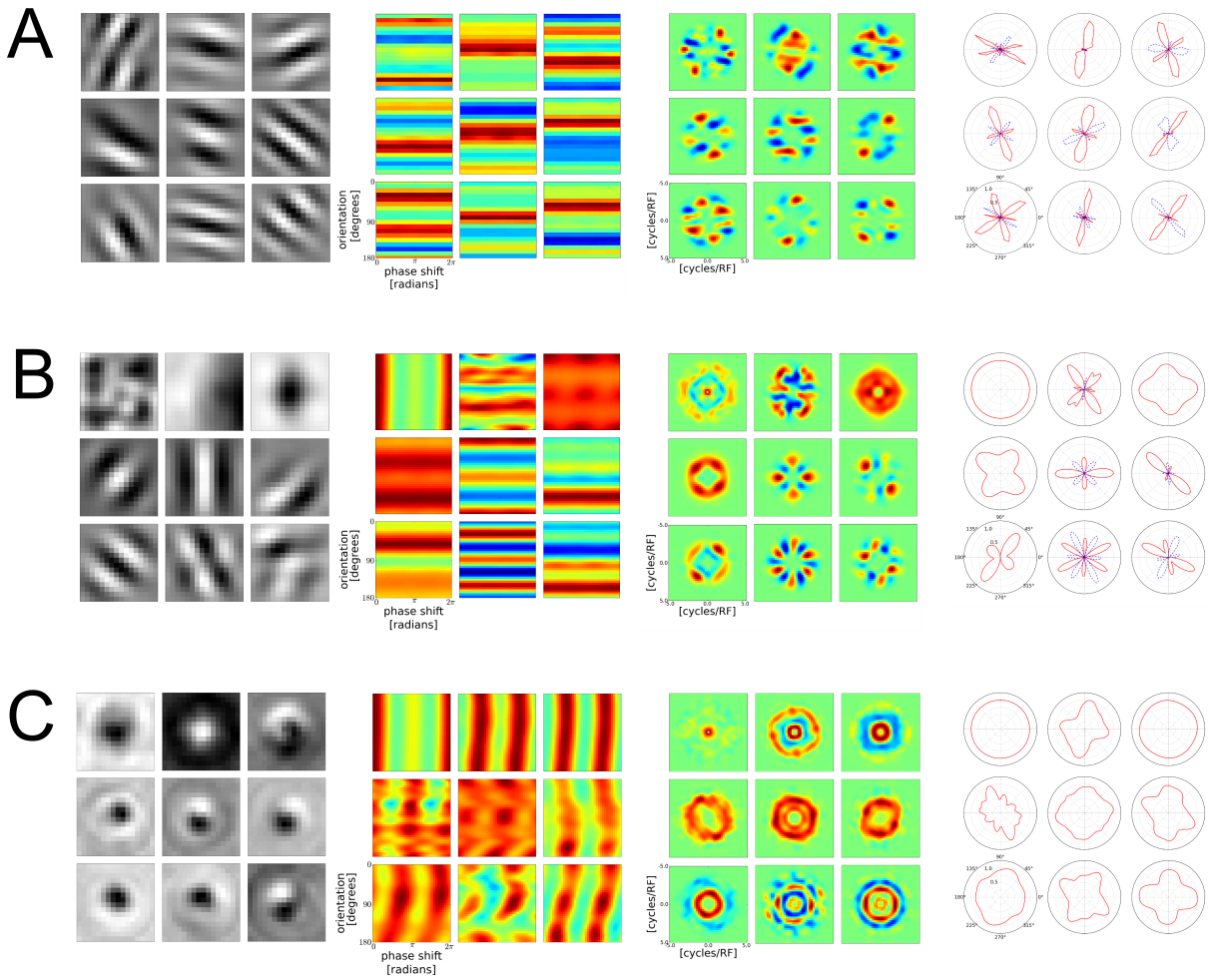

Figure 4.2.: Translation versus rotation as dominant training input feature. Every plot group visualizes a different aspect of the trained SFA units and within each plot group, nine units are shown (every third, starting from the first, up to SFA unit number 25). Columns from left to right: optimal excitatory stimuli, orientation/phase response, response in Fourier space, orientation tuning plots. Rows: **A** SFA units trained with pink noise images that were subject to translation only. **B** Units trained with a mixture of rotation and translation. **C** Units trained with input that contained rotation only.

3.3) fall in between row A and row B of figure 4.2. This is in line with the observation that retinal wave image sequences locally contain more (or faster) translation than rotation.

4.3. Gabor-patch Quadrature Filter Pair Model

With the experiments described above, two extremes were explored concerning the transformation content of the training input. When SFA is trained with input sequences containing only translation, the output of the resulting units can be described as linear combinations of Gabor-patch quadrature filter pairs (Gabor-QFPs). But what happens if the weights of such a linear combination are chosen randomly, as opposed to SFA finding them to optimize its objective? Such a randomly mixed Gabor-QFP model would represent a baseline for orientation selectivity against which the training result can be compared in case of no rotation present in the training data.

The tuning curve of a linear combination of Gabor-QFPs is given by a linear combination of the tuning curves of the individual Gabor-QFPs. Using this insight and assuming that the individual tuning curves are equal (except for their preferred orientation), a compact expression for the OSI of such an ensemble of Gabor-QFPs can be derived (see appendix A for the derivation):

$$\mathrm{OSI}_{\mathrm{G-QFP}} = \frac{100}{1 + e^{\frac{1}{2}(\pi\sigma_x f^*)^{-2}} \sqrt{\frac{\sum_{i,j}^{N} w_i w_j}{\sum_{i,j}^{N} w_i w_j \cos(2(\phi_i^* - \phi_j^*))}}} \tag{4.4}$$

with f^* being the preferred frequency, σ_x the standard deviation of the Gaussian window in image space, ϕ_i^* the preferred orientations of the individual Gabor-QFPs, and w_i the respective weights in the linear combination.

When constructing this model, a number of design questions arose:

What frequency preference should they have? The orientation tuning curve can be regarded as a circular section out of the response surface around the origin, when plotted in a spatial frequency coordinate system with the axes denoting the spatial frequency in x and y direction (see figure 3.3 A). The preferred frequency f^* of a Gabor-QFP is then the radius of the circle on which the orientation tuning curve is found. In formulating the expression for the OSI of Gabor-QFP model, f^* thus becomes a parameter. The value chosen for f^* was 3.5 cycles per receptive field. In the retinal wave simulations, the majority of units had a preferred spatial frequency in the range of 2.5 to 3.5 cycles per receptive field. 3.5 is approximately the maximal spatial frequency possible after PCA.

How many Gabor-QFPs to include? Two filters are needed for each QFP and in constructing the filters, SFA is limited to linear combinations of the PCA components. Therefore the number of QFPs that each SFA unit can consist of is bounded. In the limit of infinite receptive field size, PCA becomes equivalent to a Fourier basis Unser [1984]. This means, that in the limit, SFA can create QFPs by combining the two components of the Fourier basis that have the same frequency vector but different phases. Thus, the number of PCA components with most power at the preferred frequency was estimated by comparison of their Fourier transforms[1]. The resulting number was divided by two, leading to $N = 9$.

What orientation preference should they have? The orientation tuning curve of the linear combination of Gabor-QFPs is the linear combination of the tuning curves of the individual QFPs. Hence, the positioning of the individual tuning curves (which are Gaussians) together with their total number constraints the shape of the resulting linear combination. Preferred orientations were assigned to the Gabor-QFPs such that the full orientation axis is sampled in equally sized steps. If there are more than 4 Gabor-QFPs in the linear combination, then this allows in principle for non-orthogonal inhibition and non-orthogonal side lobes, as observed in physiological experiments and simulations with retinal wave input.

How to pick the weights for the linear combination? It is not obvious from what distribution to draw the weights of the linear combination. Assume Gabor-QFPs were chosen as basis functions for the function space in which SFA searches for solutions to its objective. The weights of the linear combination would then be determined by linear SFA on the projected data. The second moment matrices of the projected data and its temporal derivative would reflect the correlation structure which exists between the QFPs and which depends on parameters such as their number, their preferred orientation, their width, and so on. An attempt to find the corresponding distribution for the coefficients of the generalized eigenvectors analytical was deemed too difficult and, hence, not pursued. Instead, the weight distribution was approximated by a bootstrapping approach. The $\mathbf{H}$ eigenvalue spectra of SFA units from simulations with translation-only training data were pooled and then resampled when drawing weights for the Gabor-QFP linear combination[2]. As described above, the eigenvalues in these spectra come in pairs of two and can be

[1] Those PCA components are the most likely candidates for the construction of the QFPs.

[2] Prior to pooling, the eigenvalue spectra had to be normalized to unit variance, because their range differed. Such a step seemed admissible, as we were only interested in the relative weighting between QFPs within an individual SFA unit.

interpreted as the weights for QFPs composed of the corresponding eigenvectors.

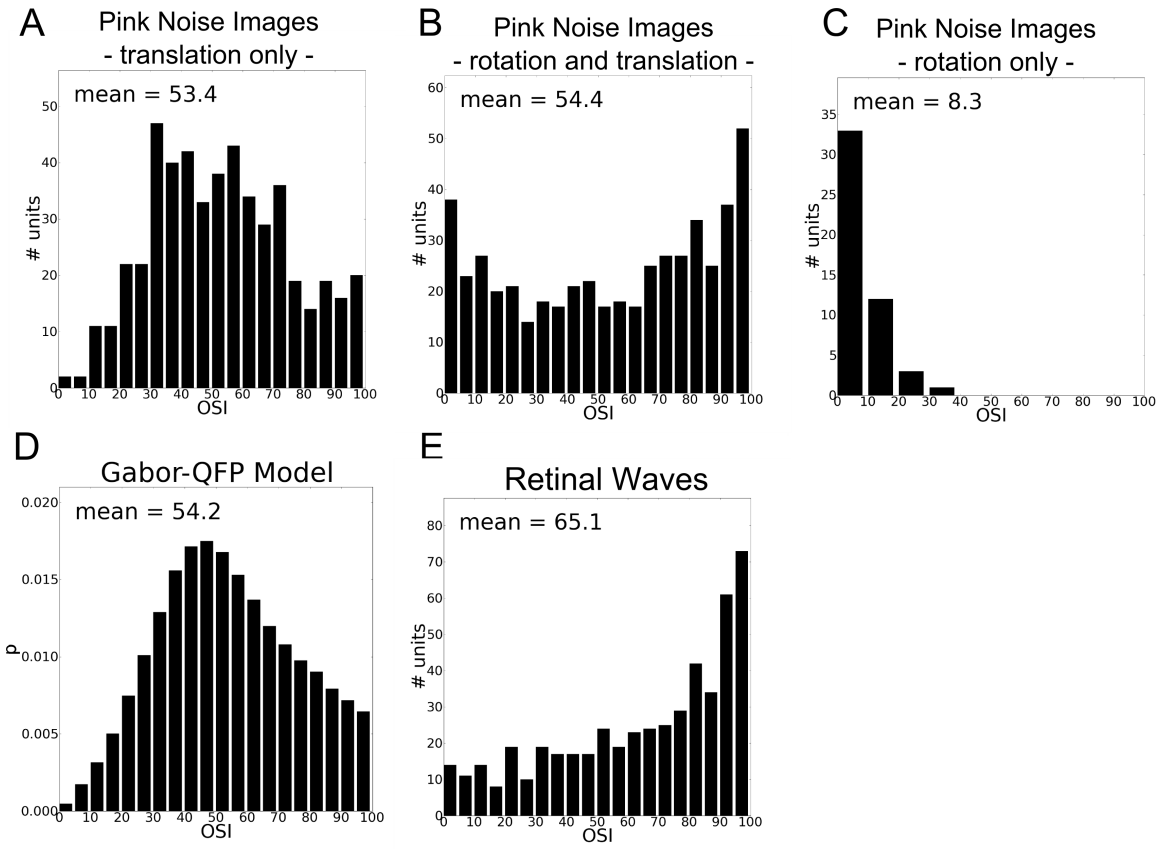

Figure 4.3.: Orientation selectivity index (OSI) histograms of SFA units trained with different types of stimuli. **A** Training with pink noise images that were translated only. **B** Training with pink noise images that were rotated and translated. **C** Training with pink noise images that were rotated only. **E** Training with retinal wave image sequences. **D** OSI distribution of the Gabor-QFP model. Histograms in **A**, **B**, and **E** were obtained by pooling OSI values from the first 50 SFA units of 10 simulation runs with identical parameters for the training input generation. For the histogram in **C** OSI values from the first 50 SFA units of a single simulation run were pooled.

Figure 4.3 compares the obtained OSI histograms of the control experiments and of the retinal wave simulations with the distribution computed for the Gabor-QFP model. The histograms were obtained by pooling OSI values from the first 50 SFA units of 10 simulation runs with identical parameters for the training input generation.

Training with inputs that contain translation only led to medium range OSI values (figure 4.3 A). The resulting SFA units exhibit excitation as well as inhibition in their orientation tuning, which leads to a lower average in activity than in the rotation-only

case. Due to the lack of rotation in the training data, there is no incentive to achieve low-frequency harmonic oscillations in the orientation tuning, which would elevate the F2 component and thereby the OSI. However, the smoothness of the tuning curve which is caused by the finiteness of the receptive field may have a similar effect, also leading to a moderate elevation of the F2 component. The OSI distribution of the Gabor-QFP model, shown in figure 4.3 D, resembles the translation-only histogram quite well, which indicates that the model indeed approximates the SFA solution for this specific case of training input.

When the training input contained a mixture of rotation and translation, the shape of the OSI histogram becomes more similar to the histogram obtained from training with retinal waves (same figure B and E, respectively). In the mixed-transformation case, the histogram shows a peak at the high and a peak at the low end of possible OSI values. The peak at the high end is due the reasons that were elaborated in section 4.2, i.e. that those units with very high OSI values are best described as a linear combination of Gabor-QFPs with a specific choice of the weights. The frequency vectors of the Gabors and weights in the linear combination are chosen by SFA in such a way that harmonic oscillations in the orientation tuning curve emerge. The Gabor-QFPs cause slowly varying output with respect to translation while the harmonic oscillations in the orientation tuning lead to slowly varying output with respect to rotation, thereby optimizing the SFA objective. The peak at the low end of possible OSI values can be attributed to the SFA units that are exclusively adjusted to the rotation in the training stimuli, see figure 4.2 B for comparison.

In fact, the histogram in figure 4.2 B can be regarded as intermediate between the histograms in A and C of the same figure. Changing the relative contribution of rotation and translation (for example by adjusting the respective speeds) should result in a morphing of OSI histograms between A, B, and C.

SFA units trained with stimuli that contain only rotation exhibit very low OSI values (see figure 4.3 A). The units respond to gratings of all orientations with almost equal strength, which is indicative of a rotation invariance (compare with figure 4.2, bottom row of plot groups). This is to be expected, because if the training input contains only rotation, the best SFA can do in order to create slowly varying output is to make the resulting units rotation invariant.

4.4. Discrete vs Continuous Training Data

In further control experiments, I have used moving sinusoidal waves and moving Gaussian blobs as training data. These can, at least locally, be regarded as crude approximations to retinal waves because they appear as a passing front of activity on the level of receptive field size.

Training Data

The main parameters of the planar sinusoidal waves are the oscillation frequency, the orientation of the grating, the traveling direction of the wave, a start point, and a traveling distance. For the simulations, the preferred frequency was drawn from a uniform distribution over the interval $[1.5, 10]$ oscillations (this corresponds to a wavelength from 5 to 35 pixels). The orientation was drawn from a uniform distribution over the interval $[0, 360]$ degrees. A single animation consists of a wave with given frequency traveling from a randomly chosen starting point for a given distance. From one frame to the next, the wave moved one pixel and when the chosen traveling distance was reached, new parameter values were drawn from their respective distributions and a new animation began. This procedure of animating a wave and resetting the parameters was repeated until a maximum number of frames was reached. The resulting image sequences contained 5000 images, each image having a size of 64x64 pixels. This image sequence was then tiled and concatenated in the same manner as the retinal wave image sequences, see section 2.2 for the description of the procedure. Next to the straight planar wave, a circular wave generation mechanism was also implemented that works in principle similar to the straight wave generator.

The moving Gaussian blob essentially consists of a 2D Gaussian distribution that is moved across the image plane. By changing the variances along the two principal axis of the distribution, different elliptical blob shapes can be obtained. In all simulations, the trajectory of the blob was a mere straight line. The image size and image sequence length were the same as for the sinusoidal planar waves.

Both input types (Gaussian blob and sinusoidal waves) were also used in a thresholded fashion. In case of the Gaussian blob for example, this means that instead of having a continuous increase of activity as the blob moves across the image, the image was thresholded to yield full activity in those pixels (or modeled retinal cells) that were above threshold and no activity in the others. Thresholded and continuous (i.e. unthresholded) image sequences were used in separate SFA training runs.

Results

The results for the thresholded input sequences are qualitatively similar to those obtained for retinal wave image sequences. Figure 4.4 A shows the results of training with the thresholded Gaussian blob image sequences. Depicted are the optimal stimuli (excitatory and inhibitory) and the first invariance computed according to the algorithm proposed in Berkes and Wiskott [2007]. Most of the optimal stimuli are show a structure reminiscent of Gabor patches. The first invariances are in most cases phase shifted versions of the respective optimal stimuli, which indicates a phase invariance. The results for the thresholded sinusoidal waves (straight and circular) are very similar and thus not shown.

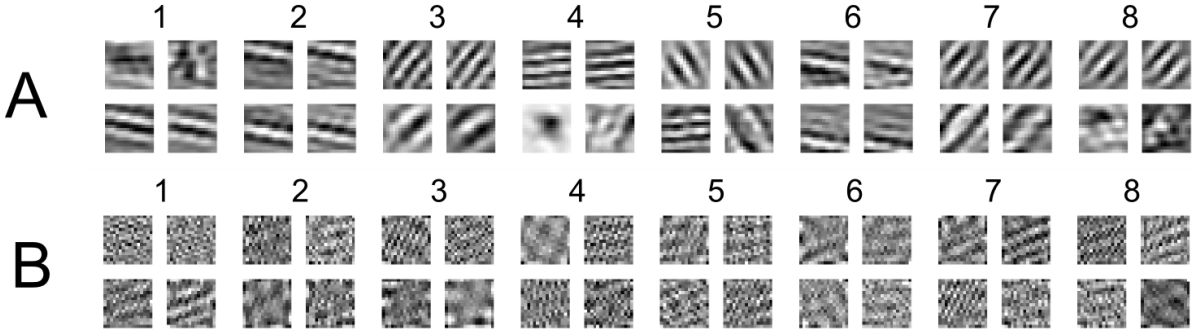

Figure 4.4.: Optimal stimuli and first invariance of SFA units trained with thresholded Gaussian blob image sequences (**A**) and with continuous (i.e. unthresholded) sinusoidal planar waves (**B**). Numbers indicate SFA unit index. For each SFA unit, the most excitatory (inhibitory) stimulus is shown in the upper left (lower left) and the corresponding first invariance to the right of it.

Interestingly, the results for the continuous Gaussian blobs and sinusoidal waves are rather different. Figure 4.4 B shows the results of training with unthresholded sinusoidal waves. The optimal stimuli seem contaminated with high frequency components (almost like "salt and pepper noise"). Yet underneath the noise, sinusoidal structure alluding to Gabor-like optimal stimuli is visible in some of the units. The results for the continuous Gaussian blobs are similar.

It seems that a too strong fall off of power in the Fourier spectrum leads to units with optimal stimuli whose high frequency components are very prominent, i.e. the optimal stimuli appear to be very noisy. When high frequencies are relatively absent in the input, there is no incentive to suppress them during the optimization for slowness. Hence, they are not punished in the process of finding the optimal stimulus for a unit.

5. Discussion

In this thesis, I presented the results of applying slow feature analysis to image sequences derived from a model of retinal waves. The resulting SFA units share a number of properties with complex cells, which are found in adult mammalian primary visual cortex. The defining feature of cortical complex cells is that they respond very well to sinusoidal gratings and show little variation in their response when the animal is presented with a grating that moves. This most important feature could be reproduced with the SFA model, i.e. the response of the SFA units is largely invariant with respect to the phase of a sinusoidal input grating. Secondly, the optimal stimuli of a large portion of the SFA units shows structure similar to that found in experimentally observed receptive fields. All optimal stimuli have ON and OFF regions. Some optimal stimuli resemble Gabor patches, while others do not. Thirdly, similar to cortical simple- and complex cells, the SFA units respond stronger to some orientations of the input grating than to others, i.e. they exhibit an orientation tuning. However, some specifics of their orientation tuning are not in accordance with physiological observations (see the next paragraph for further discussion of this issue). Finally, the learned SFA units exhibit frequency tuning, which is also found to be the case in cortical simple and complex cells.

The results of this study shed some light on the strength and weaknesses of SFA with respect to biological aspects. The phase invariance of cortical complex cells could be well reproduced, which seems to render SFA a suitable model for the emergence of such. In addition to the reproduced properties already mentioned, the modeling study of Berkes and Wiskott (Berkes and Wiskott [2005]) showed that SFA can also account for more complex cell features including direction selectivity as well as end- and side-inhibition. In my study, I did not test for these properties, but visual inspection of the optimal stimuli indicates that some of the obtained SFA units could also exhibit end- and side-inhibition. The orientation selectivity exhibited by the SFA units obtained in the present study is well above experimental findings when measured with the orientation selectivity index OSI (Chapman and Stryker [1993]). The reasons for such high OSI values have been investigated in section 4.2. The fact that orientation selectivity is overestimated, however, does not imply that the slowness hypothesis has to be dismissed. An additional

objective may be necessary to capture all properties of complex cells. Perhaps a different preprocessing of the SFA units' responses needs to be done (e.g. rectification) before computing the OSI. An alternative explanation might be that the high OSI values arise as an artifact of the rather abstract implementation of the slowness principle through SFA. The SFA algorithm as applied in this study is not intended to be biologically plausible and could thus produce effects that an actual neural system, which also implements the slowness principle, would not. Neural network implementations of the slowness principle have been applied in other studies (Einhauser et al. [2002], Kording et al. [2004]) and a neural implementation of SFA has been proposed in Sprekeler et al. [2007]. When trained with retinal wave image sequences, those implementations should also reproduce the phase invariance property. It would be interesting to explore how well these implementations fare in accounting for other complex cell properties, compared to the SFA algorithm used in this study.

The retinal wave image sequences used for training the SFA units were derived from a model of retinal waves proposed by Godfrey and Swindale (Godfrey and Swindale [2007]). Their model explains the emergence of spatially coherent patterns and their propagation on the basis of spontaneous depolarization and activity dependent refractoriness of amacrine cells. Other models of retinal waves exist and have been briefly described in section 2.2. The results presented in this report, however, should not depend on the particular choice of the retinal wave model with which the input sequences are generated. Despite the (sometimes subtle) differences in the wave generation mechanisms, all of the mentioned models reproduce relevant retinal wave characteristics such as the distribution of size, speed, and inter-wave-interval, as well as the spatial coherence and the spatially limited and changing wave domains. The Godfrey and Swindale model was chosen for practical reasons including the fact that their model was the most recent when the work in this project began, their model is easily implemented, it runs fairly quickly, and the authors included parameter settings for retinal waves of several animal species. However, there is no principal reason for favoring the Godfrey and Swindale model over the others. The obtained results should not be much different when using a different retinal wave model, provided that the used model generates waves with a similar statistics.

In the introduction of this thesis I have referred to the work of Albert (Albert et al. [2008]), in which sparse coding (Olshausen and Field [1996]) was applied to retinal wave images and resulted in the emergence of basis functions having receptive fields similar to those found for cortical simple cells. The authors propose that the early visual system is structured under the same learning objective before and during visual experience. For retinal waves to be adequate training stimuli under a fixed objective, the waves must

share relevant statistical properties with input acquired after the onset of vision, i.e. natural image sequences. The average Fourier amplitude spectrum of the generated retinal wave images, that were used in this study, shows an exponential fall-off with an exponent of about -1.36. This corresponds to an amplitude to frequency relation of $\frac{c}{f^{1.36}}$, or equivalently to a spectral power to frequency relation of $\frac{c^2}{f^{2.72}}$, with c being a scaling constant and f denoting spatial frequency. For natural images, the spectral power to frequency relation is very well approximated by $\frac{c^2}{f^n}$ with n being close to 2 (see Ruderman and Bialek [1994], Dong and Atick [1995a], and Simoncelli and Olshausen [2001]). This suggests, that retinal waves have a similar second-order spatial correlation structure compared to natural images. Albert et al. [2008] point out that retinal wave images also contain higher-order correlations similar to natural images, due to their wavefront and edge-like structure. And it is these higher order correlations, the authors emphasize, that the visual system may exploit when structuring the receptive fields of cortical simple cells under the objective of sparse coding. For the slowness objective, on the other hand, the temporal correlation structure in the training data is of great importance. The temporal statistics of an image sequence are governed by the spatial statistics and the type of image transformation that lead from one image to the next. The results presented in this report, indicate that retinal waves also share relevant temporal statistical properties with natural visual input.

To the best of my knowledge, there are no studies that describe the properties of complex cells shortly after or shorty before birth. The results of this thesis, however, predict that complex cells could already be present at the time of birth, possibly in some preliminary form. It seems likely that this is the case at least in some animal species, such as horse or giraffes, because their offspring is born at a very advanced developmental level. The freshly born foals or calves are able to stand and follow their mother within the first hour after birth. It seems likely that these animals, even at this young age, can recognize and differentiate objects (e.g. other members of their species) independently of their position in the visual field[1]. Complex cells are a prime candidate for the basis of position invariant object recognition (Shams and von der Malsburg [2002]) and therefore it is presumable that the mentioned animal species are born with an, at least partially developed, complex cell system already in place.

The performed control experiments demonstrate the dependence of the SFA units on the transformations that are present in the training stimuli. From this, an experimental prediction can be derived concerning the emergence of simple and complex cells. If one could disrupt the temporal structure of retinal waves while leaving the spatial proper-

[1]Possibly using also other sensory cues such as olfactory or auditory signals.

ties intact, then, according to the theory behind SFA, the development of complex cells should be impaired. The development of simple cells, on the other hand, should not be hindered, because the theory of sparse coding is based on spatial statistics only, making no reference to the temporal properties of the training data. This prediction can be tested with the following experimental setup: First of all, the naturally occurring endogenous activity patterns have to be abolished. This can be achieved using pharmacological neurotransmitter blockers to sever the functional connection between amacrine and retinal ganglion cells as was done, for example, in a study by Chapman (Chapman et al. [1986]). Then, activity patterns in retinal ganglion cells can be artificially induced using implanted multi-electrode grids (Alteheld et al. [2004], Rodger et al. [2008]). Such an experimental setting may still be difficult to achieve, but it is not impossible. Once the natural retinal waves can be fully substituted by artificial ones, their spatiotemporal statistics are subject to manipulation. As stated above, replaying recorded natural retinal waves in a randomized order (by shuffling the frames) should impair development of complex cells but not of simple cells.

In conclusion, I find that my simulation results support the hypothesis stated in the introduction: The slowness objective, manifested by SFA, is compatible with an innate learning mechanism that learns on endogenous activity in the same manner as on actual visual input. A large portion of the SFA units obtained from training with retinal wave image sequences share relevant properties with cortical complex cells and thereby provide a theoretical account for their emergence prior to the exposure to natural visual input.

A. Gabor Quadrature Filter Pair Model

In this section, a formula for the orientation selectivity index (OSI) Chapman and Stryker [1993] of a linear combination of Gabor quadrature filter pairs (Gabor-QFPs) is derived.

The orientation tuning curve of the i$^{\text{th}}$ Gabor-QFP, here denoted by $t_i(\phi)$, is a one dimensional function of the input grating's orientation angle. The curve has two maxima, one at $\phi_i^* < \pi$ and one at $\phi_i^{**} = \phi_i^* + \pi$, and it is obtained by a circular section through the amplitude spectrum of one of the Gabor filters[1] at the preferred frequency f^*. By evaluating the amplitude spectrum in polar coordinates, i.e. as a function of frequency f and orientation ϕ, here denoted by $A(f, \phi)$, the tuning curve is given by $t_i(\phi) = A_i(f^*, \phi)$.

The Fourier amplitude spectrum of a 2d Gabor function is equal to the sum of two Gaussians in Fourier space, centered at the points (f^*, ϕ^*) and (f^*, ϕ^{**}). The width of the Gaussians in Fourier space σ_f is inversely proportional to the width of the Gaussian envelope of the Gabor function in image space σ_x, i.e. $\sigma_f = \frac{1}{2\pi\sigma_x}$. The curved section through the 2d Gaussians in Fourier space can be approximated by a straight line section, which is a 1d Gaussian of width σ_f. Thus, $t_i(\phi)$ is modeled with two Gaussians, centered at ϕ_i^* and ϕ_i^{**}, respectively.

Recall the definition of the OSI: $100 * \frac{F2}{F2+F0} = \frac{100}{1+\frac{F0}{F2}}$, with $F0$ being the mean value and $F2$ the amplitude of the second harmonic. Because the Gabor-QFP tuning curve is periodic function with a period of π radians (or 180°), it is equivalent to consider $t_i(\phi)$ only in the range from 0 to π and to compute $F0$ and $F1$ of $t_i(\phi)$ confined to this interval.

The tuning curve of a linear combination of Gabor-QFP is the linear combination of the individual tuning curves, i.e. $t(\phi) = \sum_i^N w_i t_i(\phi)$. In order to compute the OSI for $t(\phi)$, an expression for its complex Fourier spectrum, $\mathcal{F}_t(f)$ is required. Once the spectrum is obtained, the amplitudes are given by the absolute value of the complex Fourier spectrum, e.g $F1 = |\mathcal{F}_t(1)|$ with $|\mathcal{F}_t| = \sqrt{\mathcal{F}_t \overline{\mathcal{F}}_t}$. Thus, an expression for $\mathcal{F}_t$ is required.

First of all, the linearity of the Fourier transform is applied to yield an expression for $\mathcal{F}_t$ in terms of the Fourier transforms of the individual tuning curves $\mathcal{F}_{t_i}$, and the respective

[1] Both filters of a quadrature filter pair have the same Fourier amplitude spectrum. They only differ in their phase spectrum.

weights w_i

$$\mathcal{F}_t = \sum_i^N w_i \mathcal{F}_{t_i}. \tag{A.1}$$

Because all $t_i(\phi)$ are shifted versions of each other, each t_i can be expressed as the result of the convolution of a Gaussian, centered at 0°, with a shifted Dirac δ function, i.e. $t_i(\phi) = t_0(\phi) * \delta_i(\phi)$, where $*$ denotes convolution and $\delta_i(\phi) = \delta(\phi - \phi_i^*)$. This simplifies equation A.1, because we can apply the convolution theorem, which expresses the Fourier transform of a convolution of two functions as the product of the Fourier transformed functions. $\mathcal{F}_{t_i}$ therefore becomes $\mathcal{F}_{t_0} \cdot \mathcal{F}_{\delta_i}$, which decomposes $\mathcal{F}_t$ further into

$$\mathcal{F}_t = \sum_i^N w_i \mathcal{F}_{t_i} = \sum_i^N w_i \mathcal{F}_{t_0} \mathcal{F}_{\delta_i} = \mathcal{F}_{t_0} \sum_i^N w_i \mathcal{F}_{\delta_i}. \tag{A.2}$$

Now an expression for $|\mathcal{F}_t|$ in terms of $\mathcal{F}_{t_0}$ and $\mathcal{F}_{\delta_i}$ can be formulated:

$$|\mathcal{F}_t| = \sqrt{\mathcal{F}_t \overline{\mathcal{F}}_t} = \sqrt{\mathcal{F}_{t_0} \sum_i^N w_i \mathcal{F}_{\delta_i} \overline{\mathcal{F}_{t_0} \sum_j^N w_j \mathcal{F}_{\delta_j}}} = \sqrt{\mathcal{F}_{t_0} \overline{\mathcal{F}_{t_0}} \sum_{i,j}^N w_i w_j \mathcal{F}_{\delta_i} \overline{\mathcal{F}_{\delta_j}}}. \tag{A.3}$$

In order to plug in actual values, expressions for $\mathcal{F}_{t_0}$ and $\mathcal{F}_{\delta_i}$ are given next. The zero-centered tuning curve, normalized to unit amplitude, is given by $t_0(\phi) = \exp\left(-\frac{1}{2}(\frac{\phi}{\sigma_\phi})^2\right)$, with $\sigma_\phi = \frac{\sigma_f}{\pi f^*} = \frac{1}{2\pi^2 \sigma_x f^*}$. Note that σ_ϕ is given as a fraction of 1, not in radians or degrees. Likewise, σ_x is specified in units of image side length, not in pixels. The Fourier transform of $t_0(\phi)$ is also a Gaussian:

$$\mathcal{F}_{t_0}(f) = \overline{\mathcal{F}_{t_0}(f)} = \sqrt{2\pi\sigma_\phi^2} \exp\left(-\frac{1}{2}\left(2\pi\sigma_\phi f\right)^2\right). \tag{A.4}$$

The Fourier transform of $\mathcal{F}_{\delta_i}$ is a complex sinusoid in frequency space:

$$\mathcal{F}_{\delta_i}(f) = \exp\left(-\mathrm{i}2\phi_i^* f\right). \tag{A.5}$$

Equations A.4 and A.5 can now be substituted into equation A.3:

$$|\mathcal{F}_t(f)| = 2\pi\sigma_\phi^2 \exp\left(-\left(2\pi\sigma_\phi f\right)^2\right) \sum_{i,j}^N w_i w_j \exp\left(\mathrm{i}2f\left(\phi_j^* - \phi_i^*\right)\right) \tag{A.6}$$

$$= 2\pi\sigma_\phi^2 \exp\left(-\left(2\pi\sigma_\phi f\right)^2\right) \sum_{i,j}^N w_i w_j \cos\left(2f\left(\phi_j^* - \phi_i^*\right)\right). \tag{A.7}$$

The complex exponential in the sum simplifies to a cosine, because summand (i,j) is

the complex conjugate of summand (j,i), canceling out the imaginary part. All that is left now, is to evaluate the ratio of $|\mathcal{F}_t(f)|$ for $f = 0$ and $f = 1$, and to substitute this ratio in the OSI definition. By using equation A.7 and the definition of the OSI, we have

$$\mathrm{OSI}_{\mathrm{G-QFP}} = \frac{100}{1 + \frac{F0}{F1}} \tag{A.8}$$

$$= \frac{100}{1 + \frac{|\mathcal{F}_t(0)|}{|\mathcal{F}_t(1)|}} \tag{A.9}$$

$$= \frac{100}{1 + \exp\left(2\pi^2\sigma_\phi^2\right)\sqrt{\frac{\sum_{i,j}^{N} w_i w_j}{\sum_{i,j}^{N} w_i w_j \cos\left(2\left(\phi_j^* - \phi_i^*\right)\right)}}} \tag{A.10}$$

$$= \frac{100}{1 + \exp\left(\frac{1}{2(\pi\sigma_x f^*)^2}\right)\sqrt{\frac{\sum_{i,j}^{N} w_i w_j}{\sum_{i,j}^{N} w_i w_j \cos\left(2\left(\phi_j^* - \phi_i^*\right)\right)}}} \tag{A.11}$$

Equation A.11 gives the orientation selectivity index of a linear combination of Gabor quadrature filter pairs in terms of their respective weights w_i and the parameters σ_x and f^*.

There is an interesting consequence that follows from equation A.10. According to the textbook view, the orientation tuning curve of actual visual responsive neurons can be well described by a Gaussian function Dayan and Abbott [2001]. This corresponds to the tuning curve of a G-QFP model with $N = 1$. In that case, the term under the square root in equation A.10 becomes 1 and only the exponential remains. The OSI now only depends on the (fitted) standard deviation σ_ϕ. Since σ_ϕ is always positive, the exponential in the OSI equation will always be larger or equal to 1, causing the denominator to be larger or equal to 2. Thus, the OSI cannot exceed 50, thereby exhausting only half the range of possible values!

B. Affirmation

Hereby I confirm that I wrote this thesis independently and that I have not made use of any other resources or means than those indicated.

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, June 11, 2010

–
Sven Dähne

Bibliography

Edward H. Adelson and James R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):284–299, February 1985. doi: 10.1364/JOSAA.2.000284. URL http://dx.doi.org/10.1364/JOSAA.2.000284.

M. V. Albert, A. Schnabel, and D. J. Field. Innate visual learning through spontaneous activity patterns. *PLoS computational biology*, 4 (8), 2008. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000137. URL http://dx.doi.org/10.1371/journal.pcbi.1000137.

N. Alteheld, G. Roessler, M. Vobig, and P. Walter. The retina implant–new approach to a visual prosthesis. *Biomedizinische Technik. Biomedical engineering*, 49(4):99–103, April 2004. ISSN 0013-5585. doi: 10.1515/BMT.2004.020. URL http://dx.doi.org/10.1515/BMT.2004.020.

F. Attneave. Physical determinants of the judged complexity of shapes. *Journal of experimental psychology*, 53(4):221–227, April 1957. ISSN 0022-1015. URL http://view.ncbi.nlm.nih.gov/pubmed/13416488.

Horace B. Barlow. Possible principles underlying the transformation of sensory messages. *Sensory Communication*, pages 217–234, 1961.

Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995. ISSN 0899-7667. doi: 10.1162/neco.1995.7.6.1129. URL http://dx.doi.org/10.1162/neco.1995.7.6.1129.

Anthony J. Bell and Terrence J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.3911.

P. Berkes and L. Wiskott. Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis*, 5(6):579–602, July 2005. ISSN 1534-7362. doi: 10:1167/5.6.9. URL http://dx.doi.org/10:1167/5.6.9.

P. Berkes and L. Wiskott. Analysis and interpretation of quadratic models of receptive fields. *Nat Protoc*, 2(2):400–407, 2007. ISSN 1750-2799. doi: 10.1038/nprot.2007.27. URL http://dx.doi.org/10.1038/nprot.2007.27.

Aaron G. Blankenship and Marla B. Feller. Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nature Reviews Neuroscience*, 11(1):18–29, December 2009. ISSN 1471-003X. doi: 10.1038/nrn2759. URL http://dx.doi.org/10.1038/nrn2759.

Daniel A. Butts, Marla B. Feller, Carla J. Shatz, and Daniel S. Rokhsar. Retinal waves are governed by collective network properties. *J. Neurosci.*, 19(9):3580–3593, May 1999. URL http://www.jneurosci.org/cgi/content/abstract/19/9/3580.

Matteo Carandini, Jonathan B. Demb, Valerio Mante, David J. Tolhurst, Yang Dan, Bruno A. Olshausen, Jack L. Gallant, and Nicole C. Rust. Do we know what the early visual system does? *J. Neurosci.*, 25(46):10577–10597, November 2005. doi: 10.1523/JNEUROSCI.3726. URL http://dx.doi.org/10.1523/JNEUROSCI.3726.

B. Chapman and M. P. Stryker. Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.*, 13(12):5251–5262, December 1993. URL http://www.jneurosci.org/cgi/content/abstract/13/12/5251.

B. Chapman, M. D. Jacobson, H. O. Reiter, and M. P. Stryker. Ocular dominance shift in kitten visual cortex caused by imbalance in retinal electrical activity. *Nature*, 324(6093):154–156, 1986. ISSN 0028-0836. doi: 10.1038/324154a0. URL http://dx.doi.org/10.1038/324154a0.

Barbara Chapman, Michael P. Stryker, and Tobias Bonhoeffer. Development of orientation preference maps in ferret primary visual cortex. *J. Neurosci.*, 16(20):6443–6453, October 1996. ISSN 0270-6474. URL http://www.jneurosci.org/cgi/content/abstract/16/20/6443.

John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, July 1985. doi: 10.1364/JOSAA.2.001160. URL http://dx.doi.org/10.1364/JOSAA.2.001160.

Peter Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 1st edition, December 2001. ISBN 0262041995. URL http://www.worldcat.org/isbn/0262041995.

A. F. Dean and D. J. Tolhurst. On the distinctness of simple and complex cells in the visual cortex of the cat. *The Journal of Physiology*, 344(1):305–325, November 1983. URL http://jp.physoc.org/content/344/1/305.abstract.

R. Devalois, E. Williamyund, and N. Hepler. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5): 531–544, 1982. ISSN 00426989. doi: 10.1016/0042-6989(82)90112-2. URL http://dx.doi.org/10.1016/0042-6989(82)90112-2.

D. Dong and J. Atick. Statistics of natural time-varying images, 1995a. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.1878.

Dawei W. Dong and Joseph J. Atick. Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, 6(2):159–178, January 1995b. doi: 10.1088/0954-898X_6_2_003. URL http://dx.doi.org/10.1088/0954-898X_6_2_003.

Wolfgang Einhauser, Christoph Kayser, Peter Konig, and Konrad P. Kording. Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience*, pages 475–486, February 2002. ISSN 0953-816X. doi: 10.1046/j.0953-816x.2001.01885.x. URL http://dx.doi.org/10.1046/j.0953-816x.2001.01885.x.

M. B. Feller, D. A. Butts, H. L. Aaron, D. S. Rokhsar, and C. J. Shatz. Dynamic processes shape spatiotemporal properties of retinal waves. *Neuron*, 19(2):293–306, August 1997. ISSN 0896-6273. URL http://view.ncbi.nlm.nih.gov/pubmed/9292720.

David J. Field. What is the goal of sensory coding? *Neural Computation*, 6 (4):559–601, July 1994. ISSN 0899-7667. doi: 10.1162/neco.1994.6.4.559. URL http://dx.doi.org/10.1162/neco.1994.6.4.559.

S. Firth, C. Wang, and M. Feller. Retinal waves: mechanisms and function in visual system development. *Cell Calcium*, 37(5):425–432, May 2005. ISSN 01434160. doi: 10.1016/j.ceca.2005.01.010. URL http://dx.doi.org/10.1016/j.ceca.2005.01.010.

Peter Földiák. Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200, June 1991. ISSN 0899-7667. doi: 10.1162/neco.1991.3.2.194. URL http://dx.doi.org/10.1162/neco.1991.3.2.194.

Mathias Franzius, Henning Sprekeler, and Laurenz Wiskott. Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput Biol*, 3(8): e166+, August 2007. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0030166. URL http://dx.doi.org/10.1371/journal.pcbi.0030166.

Keith B. Godfrey and Nicholas V. Swindale. Retinal wave behavior through activity-dependent refractory periods. *PLoS Comput Biol*, 3 (11):e245+, November 2007. doi: 10.1371/journal.pcbi.0030245. URL http://dx.doi.org/10.1371/journal.pcbi.0030245.

Matthias H. Hennig, Christopher Adams, David Willshaw, and Evelyne Sernagor. Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(4):1077–1086, January 2009. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4880-08.2009. URL http://dx.doi.org/10.1523/JNEUROSCI.4880-08.2009.

J. C. Horton and D. R. Hocking. An adult-like pattern of ocular dominance columns in striate cortex of newborn monkeys prior to visual experience. *J. Neurosci.*, 16(5):1791–1807, March 1996. ISSN 0270-6474. URL http://www.jneurosci.org/cgi/content/abstract/16/5/1791.

Aapo Hyvärinen, Jarmo Hurri, and Patrick O. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. (Computational Imaging and Vision)*. Springer, 1 edition, June 2009. ISBN 1848824904. URL http://www.worldcat.org/isbn/1848824904.

J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1233–1258, December 1987a. ISSN 0022-3077. URL http://jn.physiology.org/cgi/content/abstract/58/6/1233.

J. P. Jones and L. A. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1187–1211, December 1987b. URL http://jn.physiology.org/cgi/content/abstract/58/6/1187.

Konrad P. Kording, Christoph Kayser, Wolfgang Einhauser, and Peter Konig. How are complex cell properties adapted to the statistics of natural stimuli? *J Neurophysiol*, 91(1):206–212, January 2004. ISSN 0022-3077. doi: 10.1152/jn.00149.2003. URL http://dx.doi.org/10.1152/jn.00149.2003.

Margaret S. Livingstone and Bevil R. Conway. Substructure of direction-selective receptive fields in macaque v1. *J Neurophysiol*, 89(5):2743–2759, May 2003. doi: 10.1152/jn.00822.2002. URL http://dx.doi.org/10.1152/jn.00822.2002.

S. Lowel and W. Singer. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255(5041):209–212, January 1992. doi: 10.1126/science.1372754. URL http://dx.doi.org/10.1126/science.1372754.

F. Mechler. On the classification of simple and complex cells. *Vision Research*, 42(8): 1017–1033, April 2002. ISSN 00426989. doi: 10.1016/S0042-6989(02)00025-1. URL http://dx.doi.org/10.1016/S0042-6989(02)00025-1.

Zoran Nenadic, Bijoy K. Ghosh, and Philip Ulinski. Propagating waves in visual cortex: A large-scale model of turtle visual cortex. *Journal of Computational Neuroscience*, 14(2):161–184, March 2003. ISSN 09295313. doi: 10.1023/A:1021954701494. URL http://dx.doi.org/10.1023/A:1021954701494.

Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996. URL http://redwood.psych.cornell.edu/papers/olshausen_field_nature_1996.pdf.

Bruno A. Olshausen and David J. Field. How close are we to understanding v1? *Neural Computation*, 17(8):1665–1699, August 2005. ISSN 0899-7667. doi: 10.1162/0899766054026639. URL http://dx.doi.org/10.1162/0899766054026639.

D. L. Ringach, C. E. Bredfeldt, R. M. Shapley, and M. J. Hawken. Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque v1. *J Neurophysiol*, 87(2):1018–1027, February 2002. URL http://jn.physiology.org/cgi/content/abstract/87/2/1018.

D. Rodger, A. Fong, W. Li, H. Ameri, A. Ahuja, C. Gutierrez, I. Lavrov, H. Zhong, P. Menon, and E. Meng. Flexible parylene-based multielectrode array technology for high-density neural stimulation and recording. *Sensors and Actuators B: Chemical*, 132(2):449–460, June 2008. ISSN 09254005. doi: 10.1016/j.snb.2007.10.069. URL http://dx.doi.org/10.1016/j.snb.2007.10.069.

Daniel L. Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, Aug 1994. doi: 10.1103/PhysRevLett.73.814. URL http://dx.doi.org/10.1103/PhysRevLett.73.814.

Kota S. Sasaki and Izumi Ohzawa. Internal spatial organization of receptive fields of complex cells in the early visual cortex. *J Neurophysiol*, 98(3):1194–1212, September 2007. doi: 10.1152/jn.00429.2007. URL http://dx.doi.org/10.1152/jn.00429.2007.

Ladan Shams and Christoph von der Malsburg. The role of complex cells in object recognition. *Vision research*, 42(22):2547–2554, October 2002. ISSN 0042-6989. URL http://view.ncbi.nlm.nih.gov/pubmed/12445848.

Eero P. Simoncelli and Bruno A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001. ISSN 0147-006X. doi: 10.1146/annurev.neuro.24.1.1193. URL http://dx.doi.org/10.1146/annurev.neuro.24.1.1193.

Bernt C. Skottun, Russell L. De Valois, David H. Grosof, Anthony J. Movshon, Duane G. Albrecht, and A. B. Bonds. Classifying simple and complex cells on the basis of response modulation. *Vision Research*, 31(7-8):1078–1086, 1991. doi: 10.1016/0042-6989(91)90033-2. URL http://dx.doi.org/10.1016/0042-6989(91)90033-2.

Henning Sprekeler and Laurenz Wiskott. Analytical derivation of complex cell properties from the slowness principle. In *Proceedings CNS 2006*, 2006.

Henning Sprekeler and Laurenz Wiskott. A theory of slow feature analysis for transformation-based input signals. *Neural Computation*, 2010.

Henning Sprekeler, Christian Michaelis, and Laurenz Wiskott. Slowness: An objective for spike-timingâdependent plasticity? *PLoS Comput Biol*, 3(6): e112+, June 2007. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0030112. URL http://dx.doi.org/10.1371/journal.pcbi.0030112.

I. Thompson. Visual development: From darkness into light. *Current Biology*, 4(5): 458–461, May 1994. ISSN 09609822. doi: 10.1016/S0960-9822(00)00103-2. URL http://dx.doi.org/10.1016/S0960-9822(00)00103-2.

Ian Thompson. Cortical development: A role for spontaneous activity? 7(5):R324–R326, May 1997. URL http://www.cell.com/current-biology/abstract/S0960-9822(06)00150-3.

Christine L. Torborg, Kristi A. Hansen, and Marla B. Feller. High frequency, synchronized bursting drives eye-specific segregation of retinogeniculate projections. *Nature Neuroscience*, 8(1):72–78, December 2004. ISSN 1097-6256. doi: 10.1038/nn1376. URL http://dx.doi.org/10.1038/nn1376.

Jon Touryan, Gidon Felsen, and Yang Dan. Spatial structure of complex cell receptive fields measured with natural images. *Neuron*, 45(5):781–791, March 2005. ISSN 08966273. doi: 10.1016/j.neuron.2005.01.029. URL http://dx.doi.org/10.1016/j.neuron.2005.01.029.

Michael Unser. On the approximation of the discrete karhunen-loeve transform for stationary processes. *Signal Processing*, 7(3):231–249, December 1984. ISSN 01651684. doi: 10.1016/0165-1684(84)90002-1. URL http://dx.doi.org/10.1016/0165-1684(84)90002-1.

M. Weliky and L. C. Katz. Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature*, 386(6626):680–685, April 1997. ISSN 0028-0836. doi: 10.1038/386680a0. URL http://dx.doi.org/10.1038/386680a0.

Leonard E. White, David M. Coppola, and David Fitzpatrick. The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature*, 411(6841):1049–1052, June 2001. ISSN 0028-0836. doi: 10.1038/35082568. URL http://dx.doi.org/10.1038/35082568.

T. N. Wiesel and D. H. Hubel. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of neurophysiology*, 26:1003–1017, November 1963. ISSN 0022-3077. URL http://view.ncbi.nlm.nih.gov/pubmed/14084161.

D. J. Willshaw and C. Von Der Malsburg. How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 194(1117):431–445, 1976. ISSN 00804649. doi: 10.2307/77138. URL http://dx.doi.org/10.2307/77138.

L. Wiskott and T. J. Sejnowski. Slow feature analysis: unsupervised learning of invariances. *Neural computation*, 14(4):715–770, April 2002. ISSN 0899-7667. doi: http://dx.doi.org/10.1162/089976602317318938. URL http://dx.doi.org/10.1162/089976602317318938.

Laurenz Wiskott. Slow feature analysis: A theoretical analysis of optimal free responses. *Neural Computation*, 15(9):2147–2177, September 2003. ISSN 0899-7667. doi: 10.1162/089976603322297331. URL http://dx.doi.org/10.1162/089976603322297331.

Rachel O. L. Wong. Retinal waves and visual system development. *Annual Review of*

Neuroscience, 22(1):29–47, 1999. ISSN 0147-006X. doi: 10.1146/annurev.neuro.22.1.29. URL http://dx.doi.org/10.1146/annurev.neuro.22.1.29.

Reto Wyss, Peter König, and Paul F. M. J. Verschure. A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol*, 4(5):e120+, April 2006. doi: 10.1371/journal.pbio.0040120. URL http://dx.doi.org/10.1371/journal.pbio.0040120.

Jijian Zheng, Seunghoon Lee, and Z. Jimmy Zhou. A transient network of intrinsically bursting starburst cells underlies the generation of retinal waves. *Nature Neuroscience*, 9(3):363–371, February 2006. ISSN 1097-6256. doi: 10.1038/nn1644. URL http://dx.doi.org/10.1038/nn1644.